CARDIFF UNIVERSITY
PRIFYSGOL CAERDYDD

Natural Environment Research Council
WRc-NSF

THE DERIVATION OF THE MOLECULAR BASIS OF ENDOCRINE DISRUPTION IN AQUATIC INVERTEBRATES

SUBMITTED FOR THE DEGREE OF PH.D.

BY

CHRISTINE SAMBLES (B.SC. M.RES)

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

DECLARATION

This work has not previously been accepted in substance for any degree and it not concurrently submitted in candidature for any degree.

Signed.....(Candidate) Date...30/9/07.....

This thesis is being submitted in partial fulfilment of the requirements for the degree of PhD.

Signed.....(Candidate) Date...30/9/07.....

This thesis is the result of my own independent work and investigation, except where otherwise stated. Other sources are acknowledged with explicit references. A bibliography is appended at the end of this thesis.

Signed.....(Candidate) Date....30/9/07.....

I hereby give consent for my thesis, if accepted, to be available for photocopying and for inter-library loans and for the title and summary to be made available to outside organisations.

Signed.....(Candidate) Date....30/9/07.....

"The one process now going on that will take millions of years to correct is the loss of genetic and species diversity by the destruction of natural habitats. This is the folly our descendants are least likely to forgive us."

(Edward O. Wilson)

"Hey farmer farmer
Put away that DDT now
Give me spots on my apples
But leave me the birds and the bees
Please!"

(From the song "Big Yellow Taxi" by Joni Mitchell)

ACKNOWLEDGEMENTS

This study was supported financially by the Natural Environment Research Council (NERC), additional financial support was provided by WRc-NSF as CASE partner. I would firstly like to thank my supervisors Dr. Peter Kille (Cardiff University), Dr. Jason Weeks (WRc-NSF) and Dr. Ian Johnson (WRc-NSF) for their help and enthusiasm. I would also like to thank Dr. Alex Ford for his *E. marinus* specimens, without which, the work in Chapter 9 would not be possible.

I would like to thank all those who contributed to the work in this thesis. The *G. pulex* cDNA libraries were created by Dr. Judith Richards and exposure RNA extractions were performed by Mr. John Morgan. I would also like to thank everyone at Cynrig Fish Culture Unit, Llanfrynach, Powys and the Environment Agency (Wales) for allowing me access to their site to collect amphipods and always making me welcome. I would like to thank the support staff in the Cardiff University Molecular Biology Support Unit (MBSU): Ezra and the Steves, and to Q-Chip for their emergency loan of a spectrophotometer at a great time of need! My thanks also go to Steffan Adams for his bioinformatics and database support and perl scripts which saved me so much time. Thanks to the University Of Edinburgh School Of Biological Sciences Sequencing Service (SBSSS) for the sequencing of the libraries and the amazing speed at which they worked and to Anne Hedley and Mark Blaxter for integrating the sequences into the Crustacea EST and NCBI dbEST databases.

Thank you to the many people who acted as a source of help and inspiration during this endeavour:

Jenny, thanks for always having time to answer my numerous questions and giving me the opportunity to learn so much from you.

Jodie, thanks for your incredible ability to motivate and encourage me even when it seemed it was all going wrong. I loved our trips out 'shrimping' and I can't thank you enough for all the driving across South Wales to stand in a freezing river whilst the

rain and snow showered down. Even the bemused sheep were sheltering whilst the wind tried to blow us downstream.

Tracey, thanks for the support and help you have given me, not just for my 'PhD years' but for the last 10 years since we first met as Cardiff Uni undergraduates, lets hope the next 10 years are less stressful for both of us!

Vicki, I can't thank you enough for the constant support, reassurance, advice and encouragement over the last 4 years, and for buying me a drink at the end of a bad day (there were many!) and for giving me a place to stay when I missed my last train home. I will never forget you staying until after midnight to finish off a problematic print run as I had been in since 6am and was about to fall over!

My thanks go to my fellow labmates Jane, Eleftheria and Vega. You all brought laughter and fun to my world of rather dull and repetitive microarrays. Also, my thanks go to the remaining people who I have also worked alongside as part of the fantastic PK/AJM team, John, Graham, Sam and Suresh.

Finally, I am forever indebted to my parents, grandmothers and to Philip and Tim for their understanding, endless patience and encouragement when it was most required.

ABSTRACT

Endocrine disrupting chemicals (EDCs) have attracted worldwide media attention due to their feminisation effects on aquatic organisms. Studies on the environmental effects of these compounds have become increasingly important due to fears of increased infertility and their influence on the dynamics of an ecological niche. The benthic invertebrate *Gammarus pulex* has been selected as a sentinel species for studying the effects of these pollutants on the amphipod endocrine system, and subsequently on reproduction.

The aims of this project were to gain a greater understanding of the endocrine system of *Gammarus pulex* at the molecular level and to identify how pollutants may effect gene expression in intersex and normal *Echinogammarus marinus*.

Two proteins found in this freshwater shrimp have been identified as possible molecular indices of endocrine disruption, the female specific yolk protein, vitellogenin, and the exoskeletal cuticle protein. Cuticle protein is the first molecular tool to determine moult stage in *G. pulex*, which has previously relied on morphological microscopy. Rapid moult staging of cDNA, and subsequent analysis of gene expression at the different moult stages, are critical steps towards understanding normal and disrupted endocrine regulation and control. Increased vitellogenin levels in female *G. pulex* have confirmed its effectiveness as a sex-specific marker.

The experimental processes and results presented have significantly enhanced not only the quantity of molecular knowledge of the freshwater amphipod *G. pulex*, but also offers great insight into the genetic profiles of different genders, developmental and moult cycle stages in *G. pulex*. It also provides important information on endocrine disruption and intersex gene expression profiles in a related species *E. marinus*. A major potential output from this study is the production of a suite of novel and established molecular markers to detect early stages of endocrine disruption in the ubiquitous benthic invertebrate, *G. pulex* and related amphipods.

OVERVIEW

CHAPTER 1

GENERAL INTRODUCTION.....1

CHAPTER 2

CHAPTER 3

DEVELOPMENT OF A *GAMMARUS PULEX* CDNA LIBRARY....94

CHAPTER 4

IDENTIFICATION OF CANDIDATE GENES.....158

CHAPTER 5

ANALYSIS OF MOLECULAR BIOMARKERS.....185

CHAPTER 6

OPTIMISATION OF *GAMMARUS PULEX* MICROARRAYS.....217

CHAPTER 7

CHANGES IN GENE EXPRESSION IN ADULT *GAMMARUS PULEX*.....268

CHAPTER 8

GENE EXPRESSION CHANGES IN DIFFERENT DEVELOPMENTAL STAGES OF *GAMMARUS PULEX*.....309

CHAPTER 9

DETERMINING DIFFERENTIAL GENE EXPRESSION IN INTERSEX AND NORMAL *ECHINOGAMMARUS MARINUS*...329

CHAPTER 10

GENERAL DISCUSSION AND FUTURE WORK.....353

REFERENCES

APPENDIX

ABBREVIATIONS

20E	20-hydroxyecdysone
AFLP	Amplified Length Fragment Polymorphism
AGH	androgenic gland hormone
AR	androgen receptor
ATP	Adenosine Triphosphate
CA	corpora allata
cDNA	Complementary Deoxyribonucleic Acid
CHH	crustacean hyperglycaemic hormone
Cp/CUT	Cuticle protein
CYP	cytochrome P450 enzymes
dATP	Deoxyadenosine Triphosphate
DBD	DNA binding domain
dCTP	Deoxycytidine Triphosphate
dGTP	Deoxyguanosine Triphosphate
DNA	Deoxyribonucleic Acid
dNTP	Deoxyribonucleoside Triphosphate
dsDNA	Double Stranded Deoxyribonucleic Acid
DTT	Dithiothreitol
dTTP	Deoxythymidine Triphosphate
EcR	ecdysteroid receptor
EcRE	ecdysteroid receptor element
EDC	endocrine disrupting chemicals
EDTA	Ethylenediaminetetraacetate
ER	oestrogen receptor
EST	Expressed Sequence Tag
FAM	Carboxyfluorescein
GAPDH	Glyceraldehyde-3-phosphate dehydrogenase
gDNA	Genomic DNA
GIH	gonad inhibiting hormone
GSP	Gene Specific Primers
HPLC	High Performance Liquid Chromatography

HRE	hormone receptor element
HSP	heat shock protein
IPTG	Isopropylthiogalactopyranoside
JH	juvenile hormone
NaOAc	Sodium Acetate
PAHs	polycyclic aromatic hydrocarbons
PCBs	polychlorinated biphenyls
PCDDs	polychlorinated dibenzo-p-dioxins
LBD	ligand binding domain
MF	methyl farnesoate
MO	mandibular organ
MOI	moult-inhibiting hormone
MOIH	mandibular organ inhibiting hormone
mRNA	Messenger Ribonucleic Acid
NLS	nuclear localisation signal
NUP	Nested Universal Primer
PCR	Polymerase Chain Reaction
PR	progesterone receptor
QPCR	Quantitative Polymerase Chain Reaction
RA	retinoic acid
RACE	Rapid Amplification of cDNA Ends
RAP	receptor associated protein
RNA	Ribonucleic Acid
RO	Reverse Osmosis
ROX	Carboxy-X-rhodamine[1]
RT-PCR	Reverse Transcription Polymerase Chain Reaction
SMART	Switching Mechanism At 5'end of RNA Transcript
ssDNA	Single Stranded Deoxyribonucleic Acid
STW	sewage treatment works
TAMRA	Carboxytetramethylrhodamine
TBT	Tributyltin
TEMED	Tetramethylethylenediamine
UPM	Universal Primer Mix
USP	ultraspiracle protein

Vg/Vtg	Vitellogenin
VIH	vitellogenesis inhibiting hormone
X-gal	5-Bromo-4-Chloro-3-Indolyl- β-D-galactoside

[1]The X prefix of the X-rhodamines, refers to the fluorophore's extra julolidine rings. These rings prevent rotation about the nitrogen atoms, resulting in a shift in the fluorophore's spectra to longer wavelengths and usually an increase in its fluorescence quantum yield. (Molecular Probes, Leiden, The Netherlands).

CHAPTER 1

GENERAL INTRODUCTION

1.1 BIOLOGY

Invertebrates constitute approximately 95% of all terrestrial and aquatic species and play an important role in the aquatic ecosystem. It is therefore important to take these organisms into account when assessing the effects of pollution in the aquatic environment.

The freshwater amphipod, *Gammarus pulex* (*G. pulex*), is a gonochoristic benthic invertebrate that feeds on the detritus found on the river bed (Gross *et al.* 2001). The external structure of this gammarid can be seen in Figure 1.1.

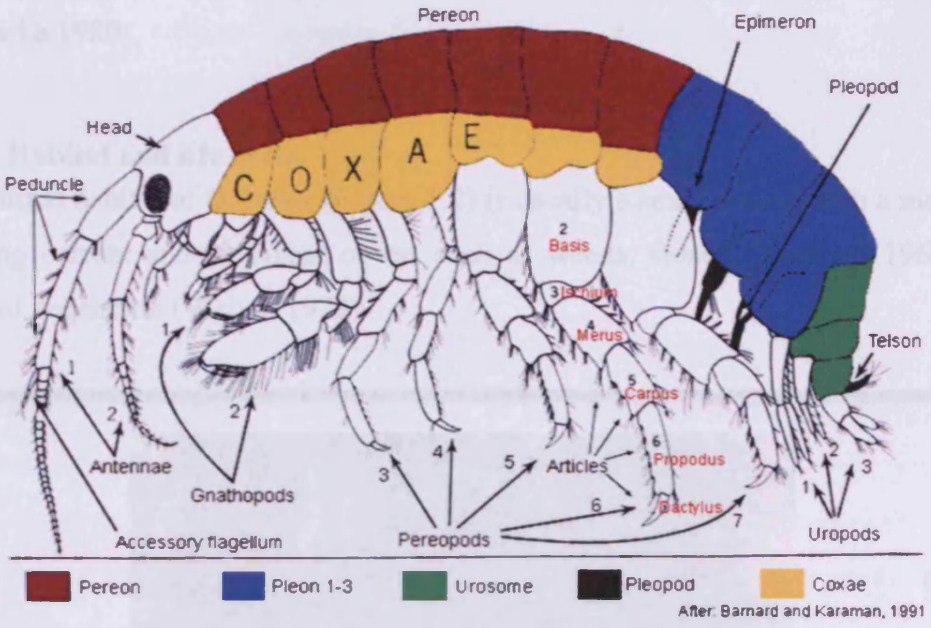


Figure 1.1: External structure of an amphipod.

The pereon consists of the thoracic segments 2-8 carrying the gnathopods and pereopods (locomotory appendages). The pleon consists of the first 3 segments of the abdomen and are attached to the pleopods (paired biramous appendages used for swimming). The urosome consists of the last 3 segments of the abdomen and contains the uropods (paired appendages used for swimming or to aid in the moulting process). The telson is the plate attached to the last urosomal segment the 6th abdominal segment). (Barnard and Karaman 1991)

On average females live between 1 and 2 years, and males between 2 and 5 years (Welton and Clarke 1980; Cold and Forbes 2004). Female *G. pulex* contain brood plates, behind which the young develop until they hatch as free swimming independent neonates. A single female can produce between 2 and 5 clutches of eggs during her lifetime, with an average of 16 eggs in each clutch (Hynes 1955; McCahon and Pascoe 1988c). After approximately 5 moults (Gross *et al.* 2001) the amphipods achieve maturity at 6mm in length and can reach a size of 12mm in females and 16mm in males (Welton and Clarke 1980). Studies by Welton and Clarke (1980) to determine times of brood production and length of life were performed in laboratory incubators. Body length was also measured during growth from juvenile to adult. They discovered that there was a decrease in brood development time (between copulation and brood release) with increasing temperature recording mean times of 36.5 (10°C), 23.9 (15°C) and 19.2 (20°C) days, with no broods produced at 5°C. The time taken to reach maturity (6mm) was also found to be shorter at a higher temperature, with recorded means of 133 days (15°C) and 87.5 days (20°C) (Welton and Clarke 1980).

1.1.1 Habitat and life cycle

The optimal habitat of *G. pulex* (Figure 1.2) is usually a small stream with a moderate to strong current and sufficient cover, such as weeds, stones (Maitland 1966) and marginal vegetation (Welton 1979).

Figure 1.2: Typical *G. pulex* habitat

There are seasonal variations of brood production in *G. pulex*. It has been observed that breeding generally occurs from March to September and ceases during the winter months, triggered by a concomitant drop in temperature (Maitland 1966). Another reason for this variation is autumnal leaf fall, which is their primary food source. During autumn and winter, a quiescent period of extended intermoult is required for the accumulation of food and energy (Sutcliffe 1993).

In a U.K. study, Hynes (1955) observed that egg production started in December and then increased rapidly until peaking during April and May after which a slow declination was evident. This corresponded with the first neonates identified in early March until June. The neonates developed and reached maturity by July or August and produced a clutch of eggs before breeding declined in the autumn when there was a drop in temperature (Hynes 1955). During winter, a high mortality rate depleted the population (Hynes 1955; Welton 1979) which was augmented by the initiation of breeding in January to March. The juveniles that over-wintered without having reached maturity (born in late summer the previous year) developed and matured by April or May and produced their own clutch of eggs. Although the moult cycle length increased at lower temperatures, growth occurred at a reduced rate. Therefore, if the winter is particularly mild, *G. pulex* may be observed breeding in the winter months (Hynes 1955).

A study 24 years later by Welton (1979) confirmed the observations of Hynes (1955) by recording a maximum population level in September with a rapid decline from October to November, with a minimum in February. Low levels of large ovigerous females were present suggesting that females larger than 10.3mm died at the end of July after breeding. The maximum brood release occurred in July and juveniles were always present in greater numbers than mature adults (Welton 1979).

Hynes established that during the autumn months there were 3 times as many males as females. He believed this may be due to the faster development time of males and also a possibility that sex is dependent on temperature and therefore that males would be born earlier in the season than the females. He also observed that by winter there were equal numbers of males and females and that the majority of *G. pulex* observed had reached maturity (6mm). During spring, the number of females rose to give a

proportion of 1.25 females/males (Hynes 1955). Welton concurs with this observation by identifying an average proportion of between 0.7 and 1.4 mature males/females throughout the year and also identified an average proportion of between 1.4 and 3.3 immature males/immature females (Welton 1979).

1.1.2 Taxonomy

G. pulex belongs to the phylum Arthropoda, subphylum Crustacea, class Malacostraca and order Amphipoda as described in Figure 1.3. Arthropods constitute the largest animal phylum, and includes insects, arachnids and crustaceans. They possess a hardened exoskeleton (which is periodically shed (moulted) to allow for growth) consisting of the polysaccharide, chitin and jointed appendages, the characteristic from which the phylum was named in Greek.

Crustaceans are mostly aquatic species, though there are some terrestrial species such as woodlice e.g. *Porcellio scaber*. Malacostracan crustaceans are the largest class of crustacea and all contain 5 head segments, 8 thoracic segments and 6/7 (Eumalacostraca/Phyllocarida) abdominal segments. Amphipods are shrimp-like crustaceans that can be terrestrial or aquatic (marine or freshwater), exisiting benthically or pelagically. They have sessile eyes, 7 pairs of gnathapod/pereopods and lack a carapace (Figure 1.1). According to the National Center for Biotechnology Information (NCBI) only 1487 proteins and 14,442 nucleotide sequences (12,346 of which were generated from this project) have been identified in Amphipoda, compared to Diptera (flies) in which 192,783 proteins and 2,896,102 nucleotide sequences have been identified.

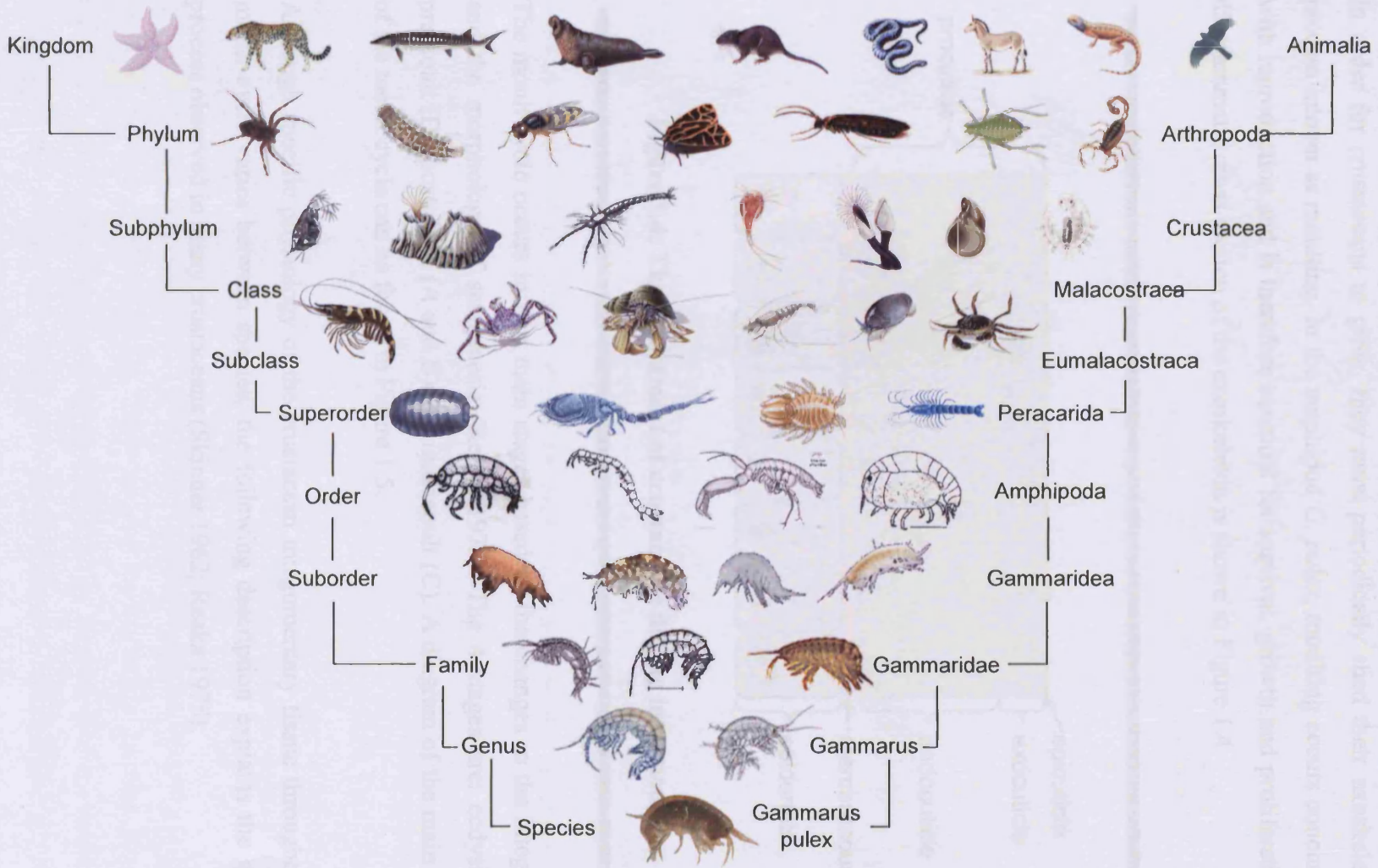


Figure 1.3: Taxonomy tree of *Gammarus pulex*

1.1.3 Moulting

In order for crustaceans to grow, they must periodically shed their exoskeleton, a process known as moulting. In the amphipod *G. pulex*, moulting occurs concurrently with reproduction and is therefore essential for survival, growth and proliferation. A diagrammatic cross section of the exoskeleton is shown in Figure 1.4.

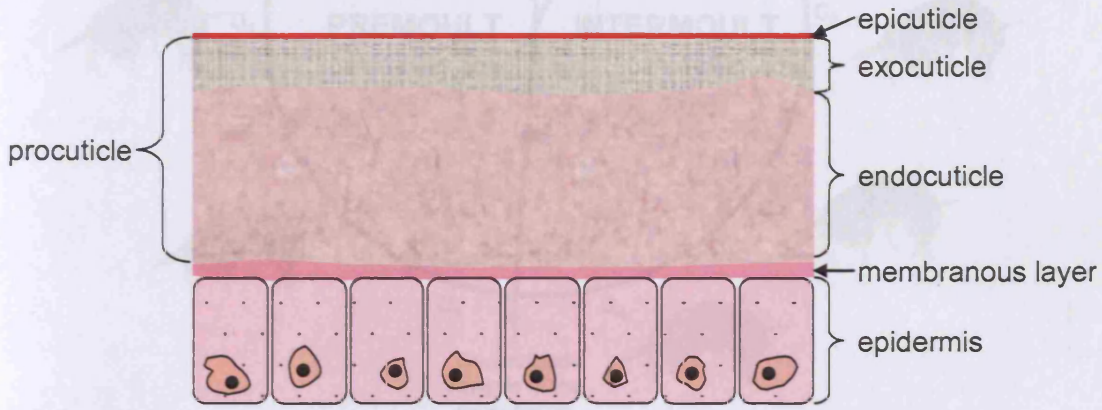


Figure 1.4: The integument of crustaceans during intermoult

The moult cycle occurs in four main stages based on the changes in the integument and the morphology of setagenesis (Reaka 1975). The 4 stages are: ecdysis (E), premoult (D), postmoult (A and B) and intermoult (C). A diagram of the main stages of the moult cycle can be found in Figure 1.5.

Although specific physiology of the crustacean integumentary tissue throughout the moult cycle varies between species, the following description explains the general process observed in many crustaceans (Skinner 1962; Reaka 1975).

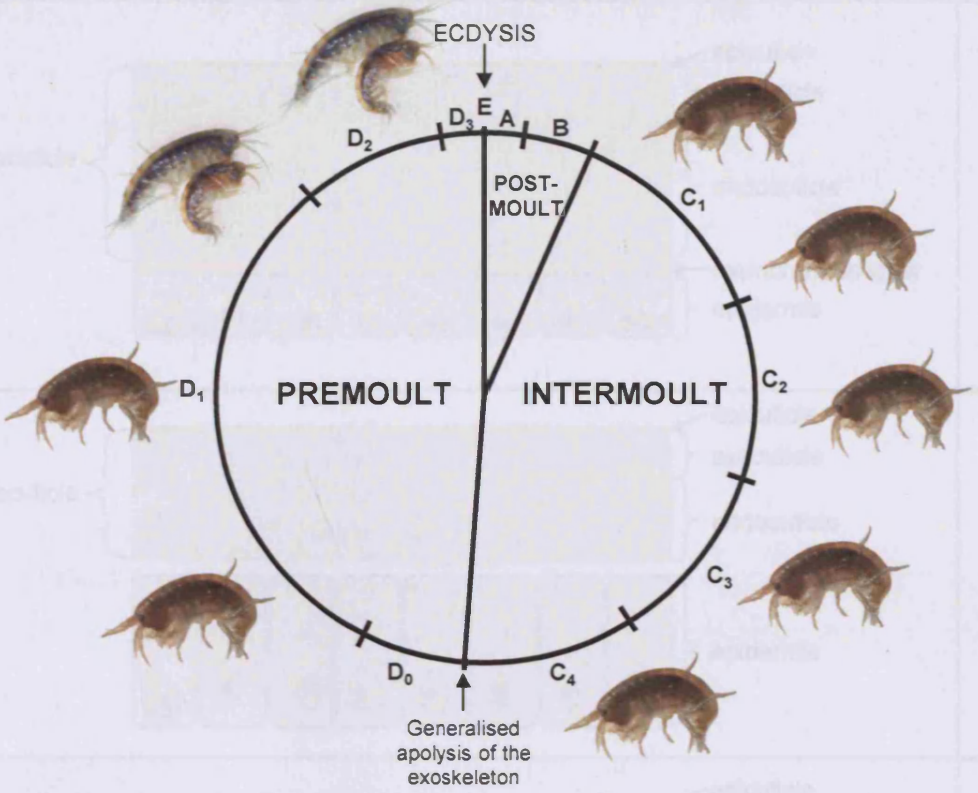


Figure 1.5: The moult cycle of *G. pulex*

The cycle is divided into post-moult (A and B), intermoult (C) and premoult (D) leading to ecdysis (E), the shedding of the exoskeleton. When males and females pair, the females are approximately nine days before ecdysis, a process which occurs to allow successful reproduction and somatic growth.

A breakdown in the membranous layer initiates premoult, causing the epidermis and exoskeleton to separate. Premoult involves the secretion of the epicuticle (outer layer) and exocuticle (middle layer) and the readsorption of the endocuticle. This is quickly followed by ecdysis (Reaka 1975). Ecdysis is the shedding of the old exoskeleton containing the epicuticle and exocuticle layers; this process is essential for the fertilisation of females and for somatic growth (Sutcliffe 1993; West 1997). At this point the animals are particularly vulnerable to predators due to the soft cuticle and impaired movement (West 1997). After ecdysis the epicuticle hardens (stage A) followed by the mineralisation of the exocuticle. The formation of the endocuticle (stage B) continues into intermoult with the membranous layer formed during late intermoult (stage C_4) (Skinner 1962). At this point in the moult cycle the thickness of the cuticle is at its maximum (Reaka 1975). The changes in the integumentary tissue are shown as diagrammatic cross sections in Figure 1.6.

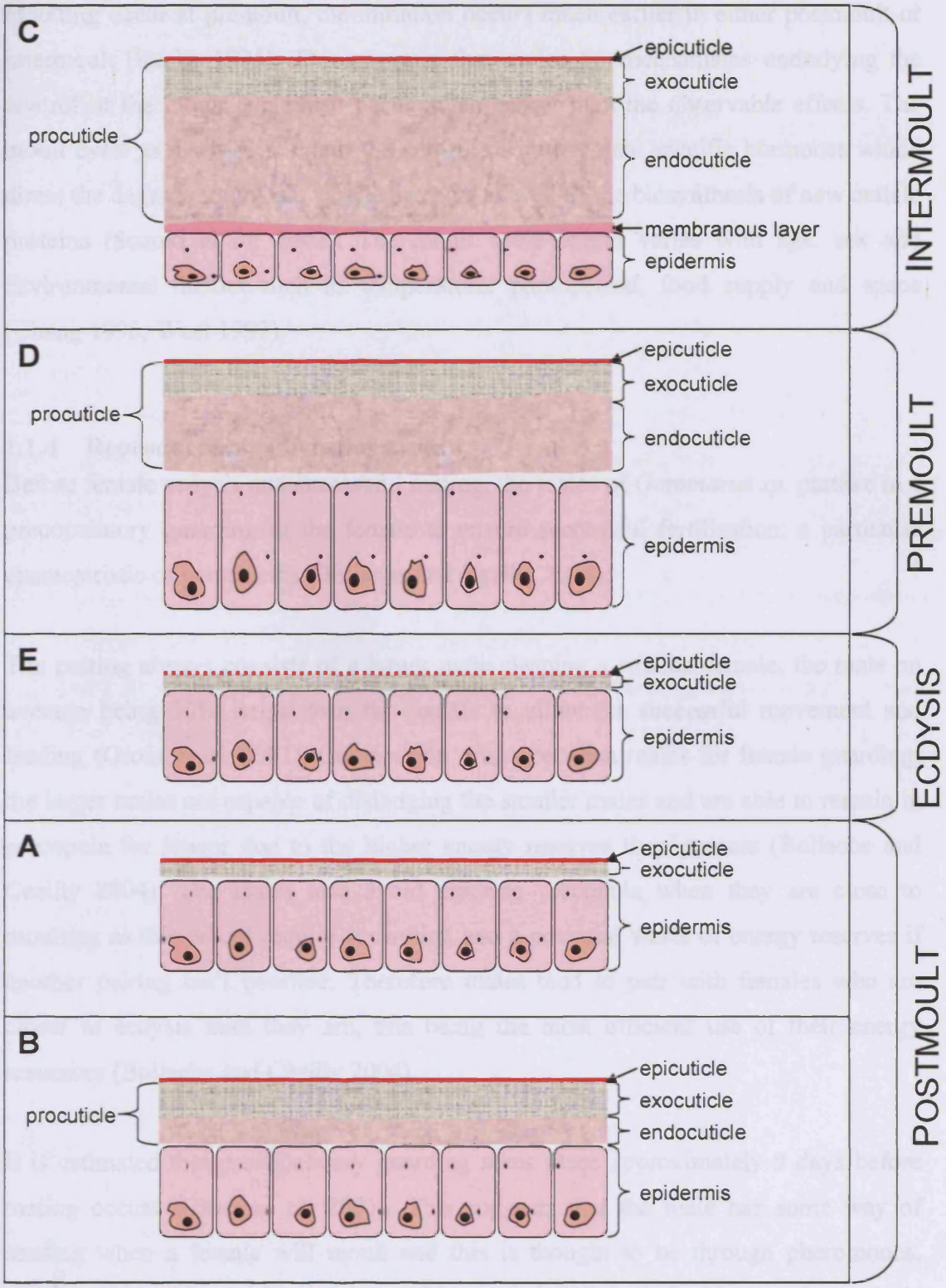


Figure 1.6: Changes in the integument of crustaceans throughout the moult cycle.

It has been observed in stomapod crustaceans that although the visible effects of moulting occur at premoult, the initiation occurs much earlier in either postmoult or intermoult (Reaka 1975). This suggests that molecular mechanisms underlying the control of the moult cycle also occur much earlier than the observable effects. The moult cycle as a whole is under the control of invertebrate specific hormones which direct the degradation of old cuticle proteins as well as the biosynthesis of new cuticle proteins (Suzuki *et al.* 2002). The moult cycle length varies with age, sex and environmental factors such as temperature, photoperiod, food supply and space (Chang 1995; West 1997).

1.1.4 Reproduction and vitellogenesis

Before female ecdysis and successful mating, the males of *Gammarus sp.* partake in a precopulatory guarding of the female to ensure successful fertilisation; a particular characteristic of gammarids (Bollache and Cezilly 2004).

The pairing always consists of a larger male clasping a smaller female, the male on average being 30% larger than the female to allow for successful movement and feeding (Gross *et al.* 2001). Competition arises between males for female guarding; the larger males are capable of dislodging the smaller males and are able to remain in precopula for longer due to the higher energy reserves they possess (Bollache and Cezilly 2004). The males also avoid entering precopula when they are close to moulting as this would require separation and a potential waste of energy reserves if another pairing isn't possible. Therefore males tend to pair with females who are closer to ecdysis than they are, this being the most efficient use of their energy resources (Bollache and Cezilly 2004).

It is estimated that precopulatory guarding takes place approximately 9 days before mating occurs (Gross *et al.* 2001). This suggests that the male has some way of sensing when a female will moult and this is thought to be through pheromones, possibly from urine or by acoustics (Watts *et al.* 2001). A few hours after ecdysis copulation takes place, the pair separates and the fertilised eggs are stored in the female's brood pouch until hatching (Cold and Forbes 2004). In the absence of a mate, females may postpone both ecdysis and egg-laying (Sutcliffe 1993).

The yolk of the egg contains proteins that provide nutrition to the developing embryo. During oocyte growth these proteins are synthesised and accumulated in a process termed vitellogenesis. The main yolk proteins in crustaceans are vitellogenin (Vtg) and vitellin (Vn). Vitellogenin is a precursor of vitellin and is thought to be synthesised in the hepatopancreas of arthropods during primary vitellogenesis (Chen *et al.* 1999). Vitellogenin is then transported to the ovaries via the haemolymph, the invertebrate circulatory system. During ovarian maturation and shortly before ovulation, vitellogenin is endocytosed into the developing oocytes, this process is termed secondary vitellogenesis (Charniaux-Cotton 1985; Tsang *et al.* 2003). The process of vitellogenesis is regulated by a complex system of hormones excreted from the X-organ/sinus gland complex, the Y-organ and the mandibular organ of crustaceans. These organs, their products and the hormonal control of vitellogenesis in crustaceans are explained in more detail in Section 1.2.

Understanding the reproductive and developmental biology of crustaceans is a prerequisite when studying endocrine disruption, as the mechanisms involved are dependent on comprehending the processes involved in vitellogenesis, moulting and reproduction (Chang 1995).

1.2 ENDOCRINOLOGY

The invertebrate endocrine system is responsible for the control of development, growth and reproduction and primarily consists of peptidic and lipidic hormones (Oberdorster and Cheek 2001). Comparisons can be drawn between insect (hexapodan) and crustacean endocrinology due to the presence of these hormones in both arthropod subphyla (Chang 1993).

Vertebrate steroid hormones, such as testosterone and oestrogen, have also been detected in some invertebrates although their endogenous origin is strongly disputed, at present having no known physiological role (LaFont 2000; Oberdorster and Cheek 2001; Segner *et al.* 2003). The evidence for vertebrate-type steroids in crustaceans is discussed in Section 1.2.2.3.

1.2.1 Glands and organs of the reproductive endocrinology system

There are 3 main organs that make up the reproductive endocrinology system in crustaceans: the mandibular organ and the X and Y organs. The location of these organs and their hormone products are described in Table 1.1.

Gland / Organ	Location	Hormones secreted	Effect
X-organ	Eyestalk / Head for sessile crustacea	Mandibular organ inhibiting hormone	**Negative** effect on ecdysteroid synthesis
		Vitellogenesis-inhibiting hormone	
		Moult-inhibiting hormone	
Y-organ	Anterior branchial chamber	Ecdysteroids	Growth, moulting and reproduction
Mandibular organ	Base of mandibular tendons	Methyl farnesoate	**Positive** effect on ecdysteroid synthesis

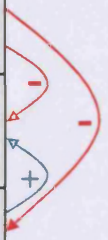


Table 1.1: Glands and organs of the reproductive endocrinology system
The table shows the different glands and organs involved in reproductive endocrinology in arthropods and the hormones they secrete. Each hormone has a direct or indirect physiological effect on growth, moulting and reproduction.

1.2.1.1 X-organ and sinus gland

The sinus gland is found in the eyestalk of decapod crustaceans or behind the eyes of crustaceans with sessile eyes, such as amphipods. It acts as a storage and release site for many hormones, generally by exocytosis from the axon terminals (Brodie and Halcrow 1977). In amphipods it contains neurosecretory axons, glial cells and the stromal sheath (Brodie and Halcrow 1977). The neurosecretory axon terminals of the sinus gland constitute the X-organ (Liu *et al.* 1997) which is responsible for the secretion of 5 major hormones including the peptidic moult-inhibiting hormone (MIH) that regulates the synthesis of ecdysone (Mattson and Spaziani 1985; Huberman 2000). The X-organ/sinus gland complex and its function can be seen in Figure 1.7.

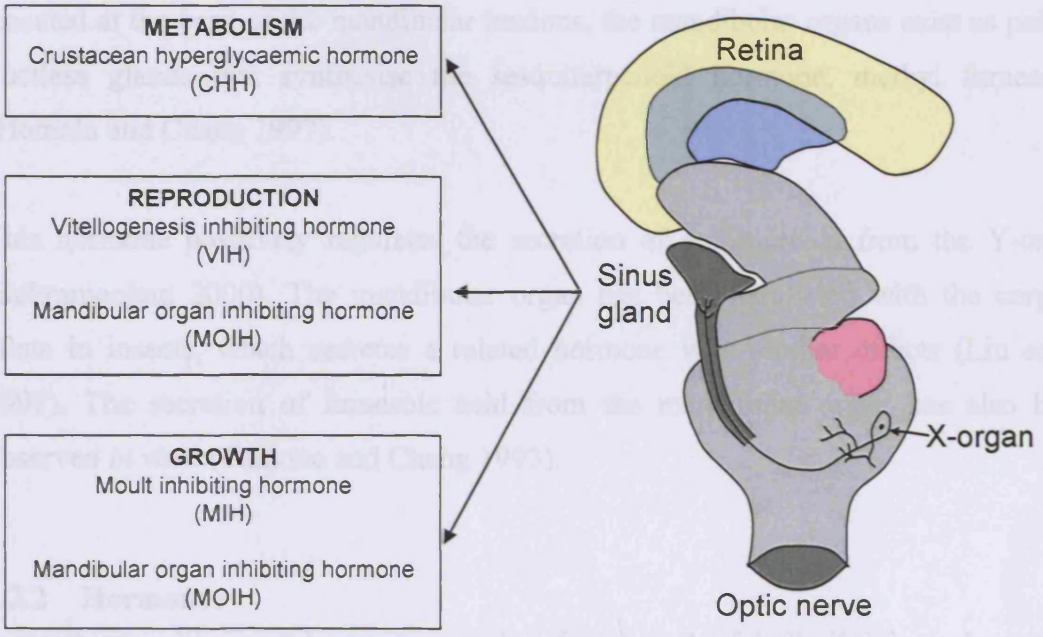


Figure 1.7: The eyestalk of decapod crustaceans

The crustacean eyestalk contains the X-organ/sinus gland complex which is responsible for 5 key hormones that regulate processes involved in metabolism, reproduction and growth (LaFont 2000). In sessile eyed crustaceans such as *G. pulex* these organs are located behind the eye.

1.2.1.2 Y-organ

The Y-organ is an endocrine gland located in the anterior branchial chamber (reviewed in Lachaise *et al.* 1993) that secretes the moult-controlling steroid ecdysone into the haemolymph. Crustaceans are thought to transport dietary cholesterol from the haemolymph into the Y-organ as required, where it is converted into ecdysone (Watson and Spaziani 1985; Fingerman 1987). It is present and active in both adults and juveniles and is homologous to the insect prothoracic glands (Homola and Chang 1997). The secretory activity of the Y-organ is positively regulated by products of the mandibular organ and negatively regulated by products of the X-organ/sinus gland complex (Subramoniam 2000). It has also been suggested that ecdysteroids may control MIH secretion from the X-organ/sinus gland complex by a feedback mechanism (Chung *et al.* 1998).

1.2.1.3 Mandibular organ

Located at the base of the mandibular tendons, the mandibular organs exist as paired, ductless glands that synthesise the sesquiterpenoid hormone, methyl farnesoate (Homola and Chang 1997).

This hormone positively regulates the secretion of ecdysteroids from the Y-organ (Subramoniam 2000). The mandibular organ has been paralleled with the corpora allata in insects, which secretes a related hormone with similar effects (Liu *et al.* 1997). The secretion of farnesoic acid from the mandibular organ has also been observed *in vitro* (Tamone and Chang 1993).

1.2.2 Hormones

Invertebrates, like vertebrates, are under the control of both lipidic and peptidic hormones. They may either be vertebrate-related or invertebrate-specific (LaFont 2000). The lipidic hormones are phylogenetically conserved steroids that act through cytoplasmic or nuclear receptors. Binding activates gene transcription in response to the specific steroid hormones.

1.2.2.1 Lipidic hormones

1.2.2.1.1 Ecdysteroids

The steroid hormones responsible for the control of growth and moulting are termed ecdysteroids (Lee *et al.* 2002). They act at the transcriptional level (Koelle *et al.* 1991) and can be found in both insects and crustaceans (Chang 1995; Oberdorster and Cheek 2001). The primary ecdysteroid, ecdysone, is derived from dietary cholesterol, (Watson and Spaziani 1985) catalysed by cytochrome P450 monoxygenases (Bollenbacher *et al.* 1977; Warren *et al.* 2002) and synthesised and secreted by the Y-organ in crustaceans (Chang 1995). It is hydroxylated to form the major ecdysteroid involved in moulting, 20-hydroxyecdysone (20E) (King and Siddall 1969). The functional species of ecdysone can be seen in Figure 1.8.

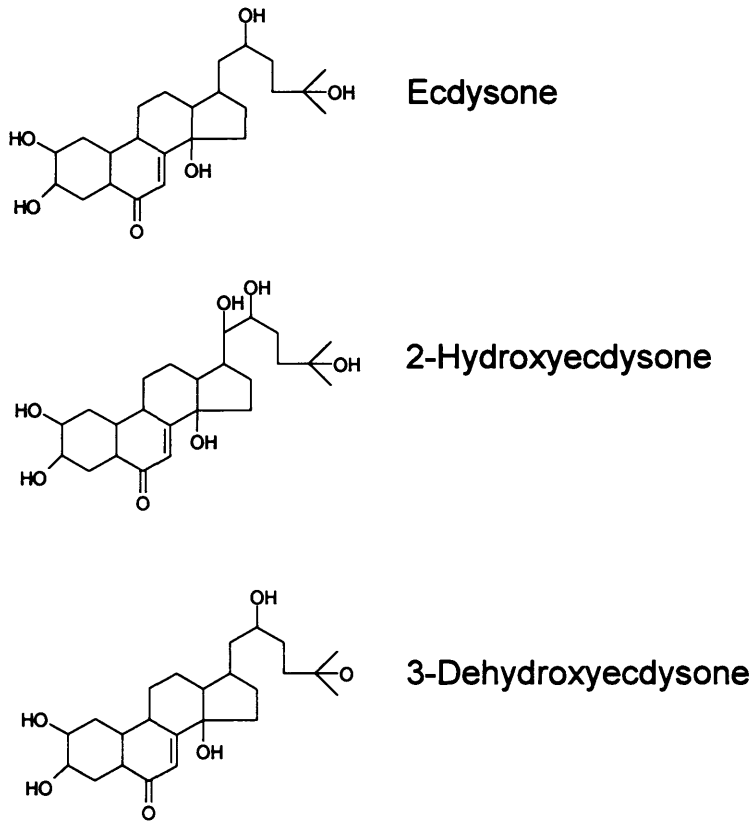


Figure 1.8: The functional species of ecdysone secreted by the Y-organ (Huberman 2000).

Ecdysone, 20-hydroxyecdysone and 3-dehydroecdysone are the 3 main products of the Y-organ. The major product is ecdysone which is involved in moulting and reproduction of insects and crustaceans.

The secretion of ecdysone from the Y-organ is negatively controlled by moult-inhibiting hormone (MIH) (Oberdorster and Cheek 2001) and positively controlled by methyl farnesoate (Block *et al.* 2003). Ecdysteroids have been associated with both the control of the moult cycle and vitellogenesis, the formation of the egg yolk.

During the moult cycle of crustaceans, there are elevated levels of 20E in premoult (El Haj *et al.* 1997) corresponding to ovarian maturation when moulting and reproduction are linked. Although no conclusive evidence exists, it has been suggested that ecdysteroids bind tightly to vitellogenin having a continued effect on the morphogenesis of the embryo until Y-organs are fully developed (Subramoniam 2000; Block *et al.* 2003). After ecdysis, 20E levels decline and synthesis is maintained at a basal level (Block *et al.* 2003). Ablation of the Y-organ results in a decreased synthesis of vitellogenin, evidently from the drop in ecdysteroid levels. This highlights the importance of ecdysone for vitellogenesis and ovarian maturation (Gross *et al.* 2001).

1.2.2.1.2 Methyl farnesoate and juvenile hormone

Methyl farnesoate (MF), is a sesquiterpenoid hormone secreted by the mandibular organ (Homola and Chang 1997; Liu *et al.* 1997) that has a positive effect on the synthesis and secretion of ecdysteroids from the Y-organ (Chang 1995; Subramoniam 2000). MF is a precursor of the insect juvenile hormone III (JH III) which is involved in development, reproduction (Liu *et al.* 1997; Bede *et al.* 2001) and moulting (Homola and Chang 1997). They are chemically related to the terpene family found in vertebrates that includes retinoic acid (RA) and the plant hormones abscissic acid and gibberellin (Jones and Sharp 1997).

In insects and crustaceans these sesquiterpenoid hormones are known to induce vitellogenin synthesis (Laufer *et al.* 1993) at the transcriptional level via nuclear receptors (Jones and Sharp 1997). They also promote ovarian maturation (Liu *et al.* 1997; Breitholtz and Bengtsson 2001) and are found at their highest concentration when the peak level of vitellogenin expression occurs (Laufer and Sagi 1991; Laufer *et al.* 1993). In comparison with the insect JH III, MF lacks epoxide function (Liu *et*

al. 1997). Whether the unepoxidated form is a prohormone that is epoxidated to an active form is unknown, although evidence suggests that crustacean mandibular organs are unable to epoxidise MF (Homola and Chang 1997). The different JH/MF homologues can be seen in Figure 1.9.

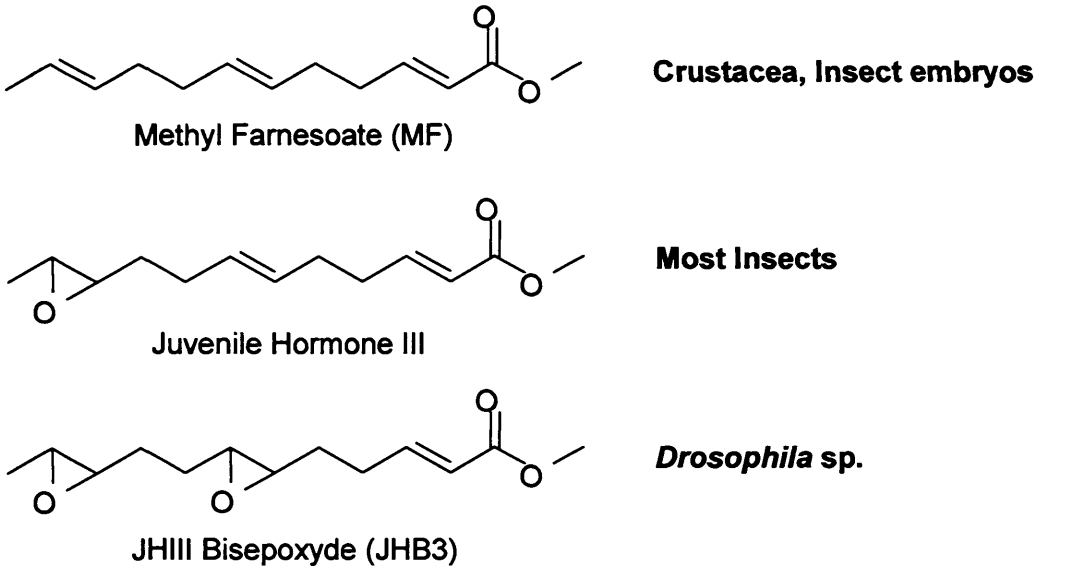


Figure 1.9: Juvenile Hormone (JH) and Methyl Farnesoate (MF) in insects and crustaceans (LaFont 2000).

These structurally similar hormones have comparable effects in their respective organisms in both reproduction and moulting through the promotion of ecdysteroid synthesis/secretion and induction of vitellogenesis.

Inhibitors of methyl farnesoate (MF) are thought to be present in the eyestalks of crustaceans and are known as mandibular organ inhibiting hormones (MOIHs) (Liu *et al.* 1997) and are described in more detail in Section 1.2.2.2. Juvenile hormones have been found to bind to ultraspiracle protein (USP) as ligands (Jones and Sharp 1997) and have a suppressive effect on the activation of the ecdysteroid receptor (EcR) by recruitment of histone deacetylase complexes (Maki *et al.* 2004). USP and EcR are described in more detail in Section 1.2.3.1.

1.2.2.2 Peptidic

Both invertebrate-specific and vertebrate related peptidic hormones have been identified in crustaceans.

1.2.2.2.1 Crustacean Hyperglycaemic Hormone family (CHHs)

The crustacean hyperglycaemic hormone (CHH) family are a group of neuropeptides secreted from the X-organ/sinus gland complex in the eyestalks of crustaceans. The ablation of the eyestalks, and therefore the removal of the CHH neuropeptides led to an increase in glucose levels in the haemolymph of the fiddler crab *Uca pugilator.* This observation was used to name the hormone family by describing the hyperglycaemic nature of the eyestalk neuropeptides (Abramowitz *et al.* 1944).

They are all 7-8kDa peptides that show strong sequence similarity, particularly in the conservation of cysteine residues (reviewed in LaFont 2000). They include crustacean hyperglycaemic hormone (CHH), moult-inhibiting hormone (MIH), vitellogenesis-inhibiting hormone (VIH) and mandibular organ-inhibiting hormone (MOIH). They are described in detail in the following sections. CHH family peptides are split into two groups depending on their primary structure, types I (CHHs) and II (MIH, MOIH and VIH) (Lacombe *et al.* 1999).

1.2.2.2.2 Crustacean Hyperglycaemic Hormone (CHH)

These are pleiotropic type I peptides which contain 72 residues, a pyroglutamate blocked N-termini and an amidated C-termini (Bulau *et al.* 2003; Ollivaux *et al.* 2006). They are known to be involved in carbohydrate metabolism (Abramowitz *et al.* 1944; Santos and Keller 1993), reproduction (de Kleijn *et al.* 1998) and moulting (Chung *et al.* 1999). Due to a peak in CHH expression at the end of premoult, it is thought to be involved in the completion of ecdysis by inducing a rapid uptake of water due to the role of CHHs in ion and water balance regulation (Chang 2001).

1.2.2.2.3 Moult Inhibiting Hormone (MIH)

In crustaceans, moulting and reproduction are controlled by ecdysteroids secreted from the Y-organ. A negative regulator of these steroids, the moult inhibiting hormone (MIH) neuropeptide is synthesised and released from the X-organ/sinus gland complex. It was discovered in 1905 by Charles Zeleny when the eyestalks of a crab (containing the X-organ/sinus gland complex) were ablated (Zeleny 1905). This ablation resulted in a decrease in moult cycle length, leading to the hypothesis that a moult inhibitor was secreted from the X-organ/sinus gland complex present in crustacean eyestalks.

MIH is thought to act on the Y-organ to inhibit ecdysteroid synthesis and secretion throughout the moult cycle, only declining in order for ecdysis to occur (Lee *et al.* 1998; Chen *et al.* 2007) at the same time as ecdysteroid levels peak in late premoult. This is contradicted by work studying the common shore crab, *Carcinus maenas*, in which a drop in MIH transcript levels was evident in late premoult (Chung and Webster 2003). It is possible these different observations are due to the presence of two different isoforms of MIH with different inhibitory potencies, suggesting different physiological roles as identified in penaeid shrimps (Gu *et al.* 2002; Chen *et al.* 2007).

1.2.2.2.4 Vitellogenesis and Gonad Inhibiting Hormones

Also secreted from the X-organ/sinus gland complex is vitellogenesis inhibiting hormone (VIH), which is sometimes referred to as gonad inhibiting hormone (GIH). This is known to inhibit vitellogenesis (Liu *et al.* 1997), identified through eyestalk ablation experiments. In prawns, secondary vitellogenesis (see Section 1.1.4) is inhibited by GIH causing genital rest in females and can also modulate the rate of secondary vitellogenesis. The mechanisms involved in this action are unknown (Charniaux-Cotton 1985). It has been isolated in the isopod, *Armadillidium vulgare*, and been shown to inhibit vitellogenesis in vitro and in vivo in females and is also thought to inhibit the androgenic gland in males (Greve *et al.* 1999).

1.2.2.2.5 Mandibular Organ Inhibiting Hormone (MOIH)

The mandibular organ inhibiting hormone (MOIH) is secreted by the X-organ/sinus gland complex and inhibits the action of the mandibular organ. It is the crustacean analogue of the insect allatostatin (Liu *et al.* 1997; Oberdorster and Cheek 2001). MOIH acts by inhibiting synthesis and secretion of methyl farnesoate (Wainwright *et al.* 1996) impeding vitellogenesis (Liu *et al.* 1997).

1.2.2.2.6 Androgenic gland hormone (AGH)

In malacostracan crustaceans it is impossible to determine the sex of neonates and the gender-specific characteristics only appear after several juvenile moults (Zou and Fingerman 1997a). Sexual differentiation in crustaceans is determined by the presence or absence of androgenic gland hormone (AGH), a peptide hormone secreted from the extragonadal androgenic glands only found in males (Zou and Fingerman 1997a). After androgen gland differentiation, the presence of AGH results in the development of male characteristics and the absence results in the development of female characteristics (Zou and Fingerman 1997b; Breitholtz and Bengtsson 2001; Rodgers-Gray *et al.* 2004).

1.2.2.2.7 Vitellogenesis Stimulating Ovarian Hormone (VSOH)

First identified by Junéra in 1977, the existence of a vitellogenesis stimulating ovarian hormone (VSOH) was discovered after androgenic gland ablation of male *Orchestia gammarella*. Despite the cessation of spermatogenesis in ablated males, vitellogenesis and oogenesis did not occur. However, when ovaries were transplanted into the males from females, vitellogenesis was initiated. This indicated the presence of factors in the ovaries essential for vitellogenesis. To confirm this, ovarian ablation was performed on females which halted vitellogenesis after a week; this was re-initiated by the implantation of new ovaries (Junera *et al.* 1977; Charniaux-Cotton 1985). More recent research has implicated a range of steroid hormones as being VSOH including progesterone and 17β-oestradiol (Summavielle *et al.* 2003).

1.2.2.3 Vertebrate-type steroids

The ability of some crustaceans to synthesise vertebrate-type steroids is undetermined, although progesterone and 17β-oestradiol have been detected in the black tiger shrimp *Penaeus monodon* (Fairs *et al.* 1989). The synthesis of oestrogen and testosterone has been observed in the gonads of molluscs (Spooner *et al.* 1991) and testosterone has been detected in the estuarine mysid *Neomysis integer* (Verslycke *et al.* 2002). Despite the fact that levels of progesterone and 17β-oestradiol fluctuate with the level of vitellogenin during ovarian maturation, (Quinitio *et al.* 1994) it has not been established whether they have a functional role as hormones in arthropods (Breitholtz and Bengtsson 2001; Verslycke *et al.* 2002). It has been suggested that vitellogenesis is controlled by vertebrate-type steroids, although there is no information regarding the effect of these hormones on the transcriptional regulation of vitellogenin (Subramoniam 2000).

1.2.3 Hormone Receptors and receptor-interacting proteins

Steroid hormones bind to apo-receptors in the cytoplasm which are activated and act as transcription factors in the nucleus for steroid-responsive genes (Fang *et al.* 1996). These receptors are also known to bind to coactivator and corepressor proteins which have a positive or negative effect on gene transcription (Haverinen *et al.* 2001).

Most steroid hormone receptors are nuclear receptors that are localised to the nucleus through nuclear localisation signals (NLSs) (Haverinen *et al.* 2001). Nuclear translocation is initiated after ligand binding in the cytoplasm dissociates the inactive receptor from its molecular chaperones transforming it into the active transcription factor (Fang *et al.* 1996).

Nuclear receptors contain 3 key domains, ligand-binding, DNA binding and the hypervariable N-terminal. The DNA binding domain (DBD) binds to hormone receptor elements (HREs) usually as a heterodimer with another receptor (Ribeiro *et al.* 1995). The domains of nuclear receptor proteins, their activation and gene induction are shown diagrammatically in Figure 1.10.

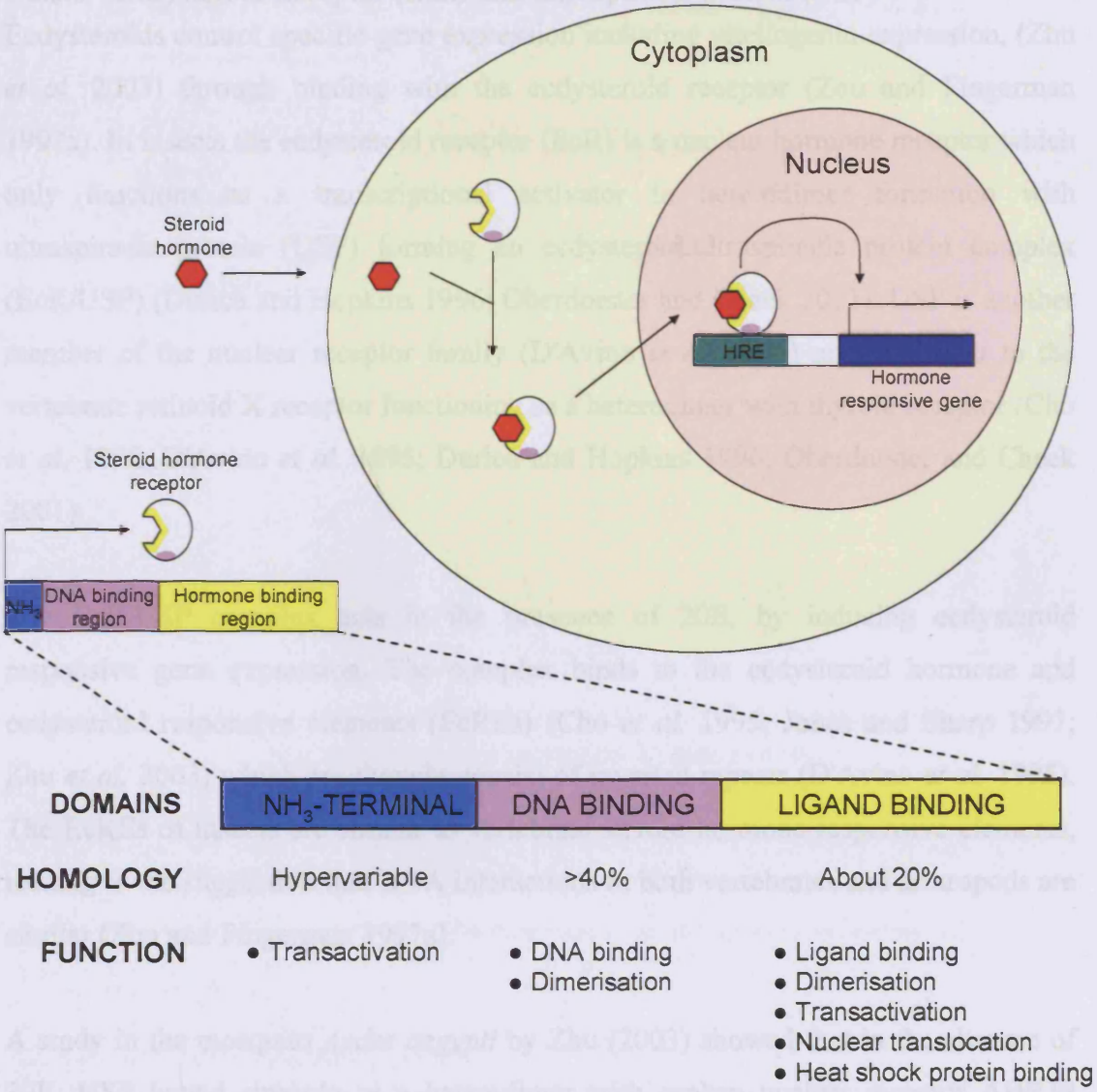


Figure 1.10: The domains of nuclear receptors (Ribeiro *et al.* 1995) and their induction of hormone responsive genes though activated hormone receptor binding to hormone responsive elements (HREs).

1.2.3.1 Ecdysteroid Receptor (EcR) and ultraspiracle protein (USP)

Ecdysteroids control specific gene expression including vitellogenin expression, (Zhu *et al.* 2003) through binding with the ecdysteroid receptor (Zou and Fingerman 1997a). In insects the ecdysteroid receptor (EcR) is a nuclear hormone receptor which only functions as a transcriptional activator in heterodimer formation with ultraspiracle protein (USP) forming an ecdysteroid:ultraspiracle protein complex (EcR/USP) (Durica and Hopkins 1996; Oberdorster and Cheek 2001). USP is another member of the nuclear receptor family (D'Avino *et al.* 1995) and is similar to the vertebrate retinoid X receptor functioning as a heterodimer with thyroid receptor (Cho *et al.* 1995; D'Avino *et al.* 1995; Durica and Hopkins 1996; Oberdorster and Cheek 2001).

The EcR/USP complex acts in the presence of 20E, by inducing ecdysteroid responsive gene expression. The complex binds to the ecdysteroid hormone and ecdysteroid responsive elements (EcREs) (Cho *et al.* 1995; Jones and Sharp 1997; Zhu *et al.* 2003) which are thought consist of inverted repeats (D'Avino *et al.* 1995). The EcREs of insects are similar to vertebrate steroid hormone responsive elements, leading to the suggestion that DNA interactions in both vertebrates and arthropods are similar (Zou and Fingerman 1997a).

A study in the mosquito *Aedes aegypti* by Zhu (2003) showed that in the absence of 20E, USP bound strongly as a heterodimer with orphan nuclear receptor AHR38 (homologous to *Drosophila melanogaster* DHR38 and vertebrate NGFI-B/Nurr1). This interaction prevented the transcriptional activation of ecdysteroid responsive genes by inhibiting the binding of USP to EcR. Therefore it was concluded that AHR38 functions by repressing the binding of EcR to the EcRE. Zhu also discovered SVP, another nuclear orphan receptor homologous to *D. melanogaster* Seven-up (Zhu *et al.* 2003). This process can be seen in Figure 1.11.

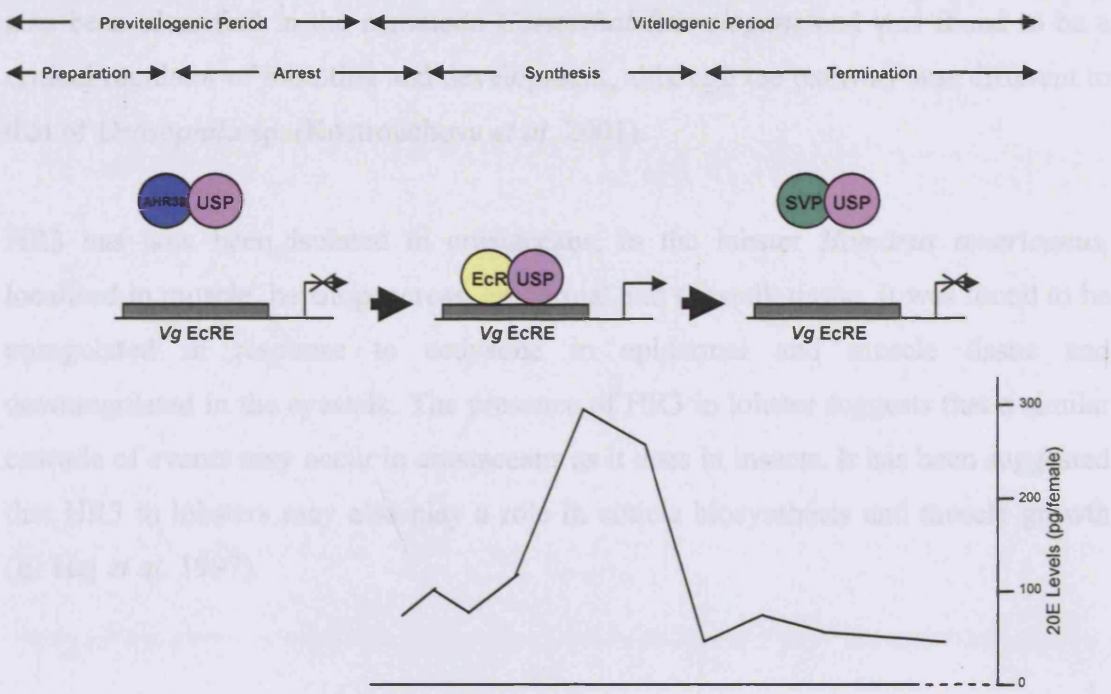


Figure 1.11: Action of ecdysteroid receptor and associated proteins in vitellogenin gene expression in mosquitoes.

After previtellogenesis, vitellogenesis is triggered by a blood meal. The presence of 20E triggers the dissociation of the USP/AHR38 complex and AHR38 is replaced by EcR which forms a complex with USP. SVP forms a dimer with USP when the levels of 20E drop in the termination stage of vitellogenesis preventing any further USP-based responses. (Zhu *et al.* 2003)

1.2.3.2 Hormone Receptor 3 (HR3)

In the fruit fly, *Drosophila* sp., the orphan nuclear receptor HR3 is an ecdysone inducible early-late gene which has been found to repress early genes activated by the late larval ecdysone pulse through a possible interaction with the EcR receptor. It has also been hypothesised that HR3 acts as an inducer of the pre-pupal factor βFTZF1, which regulates early gene response to a secondary pre-pupal ecdysone pulse. These combined responses initiate and control metamorphosis into the adult fly (White *et al.* 1997).

In the yellow fever mosquito *A. aegypti*, HR3 is expressed in the vitellogenic tissues and ovary of the females (Kapitskaya *et al.* 2000). In the domesticated silk moth, *Bombyx mori*, HR3 has been identified as a regulator of vitellogenesis in the ovary in response to ecdysone (Eystathioy *et al.* 2001). This research suggests that HR3 may

act as a key protein during ecdysone induced vitellogenesis. An HR3 homologue has also been identified in the nematode *Caenorhabditis elegans* and was found to be a critical regulator of moulting and development, although the pathway was different to that of *Drosophila* sp. (Kostrouchova *et al.* 2001).

HR3 has now been isolated in crustaceans, in the lobster *Homarus americanus,* localised in muscle, hepatopancreas, epidermal and eyestalk tissue. It was found to be upregulated in response to ecdysone in epidermal and muscle tissue and downregulated in the eyestalk. The presence of HR3 in lobster suggests that a similar cascade of events may occur in crustaceans as it does in insects. It has been suggested that HR3 in lobsters may also play a role in cuticle biosynthesis and muscle growth (El Haj *et al.* 1997).

1.2.3.3 Heat shock proteins

Heat shock proteins are molecular chaperones responsible for receptor protein conformation and folding (Fliss *et al.* 2000; Snyder and Mulder 2001). HSP90 was the first receptor associated protein (RAP) to be identified (Dougherty *et al.* 1984) and since then HSP70 has been implicated in the same role (Ribeiro *et al.* 1995). Heat shock proteins bind and stabilise the steroid hormone receptors in the absence of a ligand and act as a repressor of receptor activation (Fang *et al.* 1996; Snyder and Mulder 2001). Heat shock proteins hold the receptor in a suitable conformation for ligand binding, after which the mechanisms and circumstances of chaperone dissociation from the receptor is not fully understood.

Haverinen suggests that HSP90 may inhibit the transport of the steroid receptor into the nucleus by masking the NLS of progesterone receptor. Dissociation after ligand binding is required for nuclear transport (Haverinen *et al.* 2001). However, Georget suggests that HSP90 or associated proteins seem to play an important role in nuclear transfer of the androgen receptor even after agonist binding (Georget *et al.* 2002). This suggests that the chaperone may act in different ways dependent on the receptor type. The role of heat shock proteins is shown in Figure 1.12.

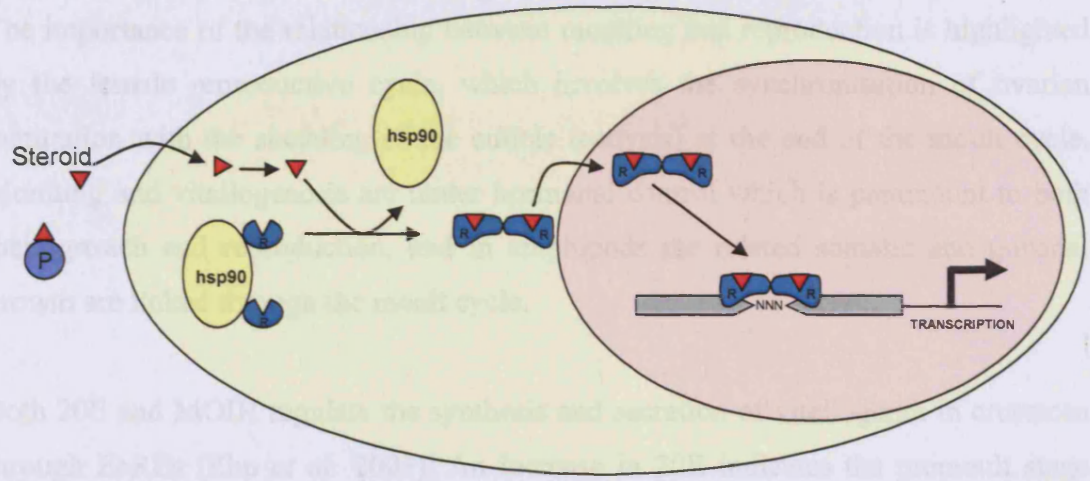


Figure 1.12: Action of HSP90 in chaperone receptor binding (Ribeiro *et al.* 1995)
P = Protein; R = Receptor

1.2.3.4 Androgen Receptor

In vertebrates, heat shock proteins are required to stabilise the receptor in an inactive conformation. The androgen receptor and the heat shock protein (HSP90) bind through the ligand binding domain (LBD) of the receptor protein. In the presence of androgen, the androgen receptor is transported into the nucleus still bound to HSP90. Following translocation into the nucleus, androgen-responsive genes are transcribed with the help of other recruited transcription factors (Georget *et al.* 2002).

1.2.3.5 Oestrogen Receptor

Although an oestrogen receptor orthologue has been isolated from the mollusc *Aplysia californica*, its function is unknown and its presence has not been identified in the majority of invertebrates to date (Thornton *et al.* 2003).

1.2.4 Hormonal control of moulting and vitellogenesis

The importance of the relationship between moulting and reproduction is highlighted by the female reproductive cycle, which involves the synchronisation of ovarian maturation with the shedding of the cuticle (ecdysis) at the end of the moult cycle. Moulting and vitellogenesis are under hormonal control which is paramount to both their growth and reproduction, and in amphipods the related somatic and gonadal growth are linked through the moult cycle.

Both 20E and MOIH regulate the synthesis and secretion of vitellogenin in crustacea through EcREs (Zhu *et al.* 2003). An increase in 20E indicates the premoult stage which combined with JH initiates cuticle protein synthesis in insects (Suzuki *et al.* 2002) and the formation of a new cuticle (Chan 1998). MF is known to act as a vitellogenin stimulant (Liu *et al.* 1997), acting in the same way as JH by stimulating synthesis in the fat body of the female of many insects (Homola and Chang 1997), as opposed to VIH and GIH which are thought to inhibit vitellogenin secretion (Oberdorster and Cheek 2001). EcR transcription is also induced at different levels through the vitellogenic cycle in mosquitoes, resulting in the accumulation of EcR during previtellogenesis (Cho *et al.* 1995).

The interaction of hormones and organs involved in reproduction and vitellogenesis are shown diagrammatically in Figure 1.13.

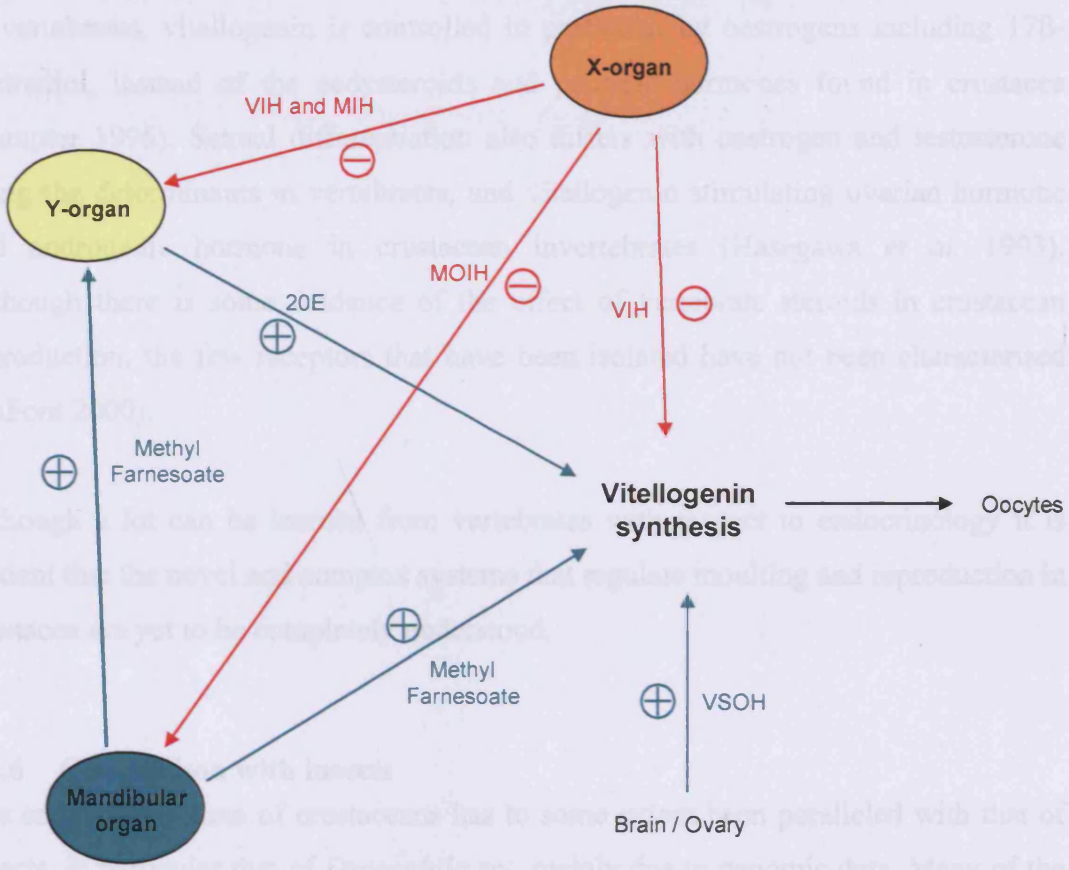


Figure 1.13: The reproductive endocrine system of crustaceans (Oberdorster and Cheek 2001).

The Y-organ produces 20E which induces vitellogenesis. The products of the X-organ have a negative effect on 20E from the Y-organ, MF from the mandibular organ and directly on vitellogenesis. MF, a product of the mandibular organ has a positive effect on 20E and directly on vitellogenesis.

1.2.5 Vertebrate / Invertebrate comparison

In vertebrates, vitellogenin is controlled in particular by oestrogens including 17β-oestradiol, instead of the ecdysteroids and peptidic hormones found in crustacea (Sumpter 1995). Sexual differentiation also differs with oestrogen and testosterone being the determinants in vertebrates, and vitellogenin stimulating ovarian hormone and androgenic hormone in crustacean invertebrates (Hasegawa *et al.* 1993). Although there is some evidence of the effect of vertebrate steroids in crustacean reproduction, the few receptors that have been isolated have not been characterised (LaFont 2000).

Although a lot can be learned from vertebrates with respect to endocrinology it is evident that the novel and complex systems that regulate moulting and reproduction in crustacea are yet to be completely understood.

1.2.6 Comparison with insects

The endocrine system of crustaceans has to some extent been paralleled with that of insects, in particular that of *Drosophila* sp., mainly due to genomic data. Many of the molecules involved are structurally similar, for example, the terpenoid MF is related to JH, the insect and crustacean EcR have very similar structures (Zou and Fingerman 1997a) and several shrimp cDNAs have been found to be homologous to insect cDNA. This has led to the hypothesis that at the molecular level both crustaceans and insects have a very similar regulatory mechanism in moult cycle control (Chan 1998; Huberman 2000).

Although crustaceans require the moult cycle in order for somatic growth to occur, insects also require a larval moult cycle that occurs until they are adults. These events are triggered by the same hormone, ecdysone, in both insects and crustaceans (Kreutzweiser *et al.* 1994; Durica and Hopkins 1996).

Although different inhibitor mechanisms are in place, the insect moulting process is regulated by a neuropeptide inhibitor that is unlike the regulation of crustacean moulting by the CHH family of neuropeptides (Chang 1993).

Comparisons have also been made between whole organs, with the corpora allata (CA) of insects producing JH, and the mandibular organ of crustaceans producing the analogous MF; the CA and mandibular organ are considered homologous organs. (Liu *et al.* 1997) The Y-organ has also been likened to the prothoracic hormone of insects (Homola and Chang 1997).

In both crustaceans and insects, ecdysteroids and terpenoids are responsible for the control of moulting, differentiation, reproduction and metamorphosis (Segner *et al.* 2003).

1.3 ENDOCRINE DISRUPTING CHEMICALS (EDCs)

Certain anthropogenic chemicals have been shown to act as hormone mimics by binding to steroid receptors in competition with endogenous ligands. These chemicals are termed "endocrine disrupting chemicals" (EDCs) and are defined as 'an exogenous substance that causes adverse effects in an intact organism, or its progeny, consequent to changes in endocrine function' (EC 1997). In exposed organisms, EDCs have the potential to affect hormone synthesis, secretion, transport, binding, action or elimination of the endogenous hormones and therefore disrupt homeostasis, development, reproduction and behaviour (Segner *et al.* 2003). They can act either as an agonist or antagonist on the native hormone receptor (Snyder and Mulder 2001).

Most of the EDCs causing concern at present are synthetic organic chemicals (Taylor *et al.* 1999) including organochlorine pesticides, insecticides, polychlorinated biphenyls (PCBs), polycyclic aromatic hydrocarbons (PAHs), polychlorinated dibenzo-p-dioxins (PCDDs), surfactants and plasticisers. These pollutants originate from soap, detergent, agriculture and the plastics and petrochemical industries (Sumpter 1995). In aquatic organisms exposure can be through the diet or water column (Sumpter 1998).

Of particular concern are the "xenoestrogens", the chemicals that specifically mimic oestrogen, binding to its receptor and have been linked to disruption of the male reproductive system (Watts *et al.* 2001). Aquatic organisms are exposed to xenoestrogens through sewage treatment works (STW) effluent which is discharged into rivers. STW effluent contains both synthetic and natural oestrogens and can be strongly oestrogenic to fish (Sumpter 1995). Therefore effluent consists of a mixture of chemicals and is likely to contain more than one EDC; the affect of effluent on aquatic organisms may be very different to those exposed to single chemicals. It has been suggested that a mixture is considerably more potent than individual chemicals, enhancing the effect of disruption (Sumpter 1995).

Individually, xenoestrogens present in effluent are weakly oestrogenic (Sumpter 1995) only causing disruption at high concentrations that are not normally present in river water. Ethinyloestradiol, an excreted contraceptive, has been found in river

water and in effluent. It was found to cause vitellogenin synthesis in male fish (Purdom *et al.* 1994), a protein usually detectable only in females (Copeland *et al.* 1986).

The quantity of effluent that contributes to the volume of a river varies considerably. It can often consist of 50% of river volume reaching 90% during periods of low rainfall (Sumpter 1998) and is dependent on the river size. Effluent concentration and quality influence the oestrogenicity of the river (Sumpter 1995).

The hydrophobicity of many of these chemicals increases the risk of bioaccumulation, bioconcentration (Sumpter 1995) and biomagnification (Vallack *et al.* 1998) throughout the aquatic ecosystem. This is due to the chemicals ability to adsorb to the sediment, often a key food source for benthic invertebrates such as *G. pulex* (Gross *et al.* 2001; Gross-Sorokin *et al.* 2003). The effect of a given EDC on the exposed organism is also dependent on the route of metabolism that the particular chemical takes and the resultant metabolite that is produced (Sumpter 1995).

The structural similarity between ecdysone and xenoestrogens suggest that the ecdysteroid receptor (EcR) may also be a target for EDC disruption, via blockage, and consequently affecting the moult cycle (Zou and Fingerman 1997a; Zou and Fingerman 1997b; Gross *et al.* 2001; Segner *et al.* 2003). The result of this would be a slowing of the moulting process and a reduction in growth (Gross *et al.* 2001). If oestrogens and xenoestrogens can bind the ecdysteroid receptors (EcR), the process of vitellogenesis may also be affected (Gross *et al.* 2001).

A review by Colborn highlights the widespread effect of endocrine disruption on wildlife, linking EDCs with physiological effects (Colborn *et al.* 1993). Details are shown in Table 1.2.

Effects	Species					
	Birds	Fish	Shellfish	Mammals	Turtles	Gastropods
Abnormal thyroid function	✓	✓				
Decreased fertility	✓	✓	✓	✓		
Decreased hatching success	✓	✓			✓	
Demasculisation and feminisation	✓	✓		✓		
Defeminisation and masculinisation	✓	✓				✓
Alteration of immune function	✓			✓		

Table 1.2: Effects of endocrine disruption on wildlife (Colborn *et al.* 1993).

The limited knowledge of invertebrate endocrinology restricts the understanding of the mechanisms involved in endocrine disruption.

1.3.1 Intersex / imposex

Intersexuality is defined as the presence of male and female sex characteristics in a single individual in a gonochoristic species (Morgan 1920). Exposure of eggs or juveniles to ethinyloestradiol is thought to produce a state of hermaphroditism (intersex) in fish (Sumpter 1995), an effect that was noticed 25 years ago when intersex fish were found at two STWs (Sumpter 1998), an occurrence that is still continuing today (Gross-Sorokin *et al.* 2006). In 1998, 20% of observed wild male flounder in the Mersey estuary, U.K. contained oocytes in their testes, thought to be due to exposure of fish larvae to oestrogenic chemicals during gonadogenesis. However there was no accumulation of vitellogenin in the oocytes, confirming the

concept that oestrogenic chemicals induce vitellogenin synthesis in the liver of fish but do not stimulate absorption into oocytes (Allen *et al.* 1999). The presence of intersex roach in British rivers has been correlated with predicted levels of oestrogenic chemicals, although the level of intersex did not correlate with vitellogenin levels in the observed fish. This unexpected result may be due to seasonal variation, or more likely due to the fact the study was on a wild population and the difference in exposure times (Jobling *et al.* 2006).

Imposex *Nucella lapillus* dogwhelks (true females with male organs) have been observed across the U.K. after exposure to tributyltin (TBT) causing malformation of the females' reproductive system and sterility (Gibbs *et al.* 1991; Oehlmann *et al.* 1996) thought to be partially due to DNA damage (Hagger *et al.* 2006). Studies on the freshwater ramshorn snail, *Marisa cornuarietis*, have shown the conversion of androstenedione to dihydroandrostenedione (DHA) by 5α-reductase in females to be inhibited after exposure to TBT and triphenyltin (TPT). DHA is the primary metabolite in females whereas in males, androstenedione is converted to 5α-dihydrotestosterone (DHT) and to a lesser extent testosterone. DHT is also produced in females, although at low levels, however after exposure to TBT and TPT, the ratio of DHA:DHT decreases inducing the development of imposex (Janer *et al.* 2006a). TBT exposure also causes the percentage of free testosterone in females to increase (Janer *et al.* 2006b). After TBT exposure, the gastropod *Bolinus brandaris* showed reduced levels of oestradiol in both males and females but no inhibition of aromatase, an enzyme that converts testosterone to oestradiol, was observed (Morcillo and Porte 1999). It has been ascertained that although aromatase inhibition can cause imposex development, it is not the primary mechanism and it has been suggested that a neuropeptide may also be responsible (Oberdorster and McClellan-Green 2002). The exact biochemical mechanisms involved in imposex induction in molluscs by TBT is still hotly debated (Oehlmann *et al.* 2007).

Intersex occurrence in crustaceans is known to be linked to the presence of the androgenic gland (AG) and secretion of the androgenic gland hormone (AGH). After AG ablation of juvenile crayfish, the resulting adults behave similarly to genotypic females, with female morphology, undertaking both primary and secondary vitellogenesis (Barki *et al.* 2006). A link between intersexuality and AGH has also

been observed in the amphipod *Echinogammarus marinus* with a putative AGH peptide only present in normal males (Ford *et al.* 2005). Intersexuality has been reported to be widespread in amphipods, including Gammaridae (Ford and Fernandes 2005). The consequences of intersexuality in amphipods was made evident in a study on *E. marinus* when it was observed that the morphological changes reduced the success of precopulatory pairing (Ford *et al.* 2004) and the females had smaller brood sizes and lower embryo survival (Ford *et al.* 2003). The occurrence of both male and female intersex specimens is increased at polluted sites although the causes are unknown. It has been hypothesised they may be directly due to pollution, through immunosuppression and consequent parasite infection or though a complex relationship of the above (Ford *et al.* 2004; Ford *et al.* 2006). The presence of intersex *Gammarus fossarum* has been discovered in Germany although the source of this phenomenon was not identified (Jungmann *et al.* 2004). It has been suggested that intersexuality in gammarids may be due to both environmental and parasitic factors after a study in Scotland of a *Gammarus duebeni* population. Like *E. marinus*, the probability of precopulatory pairing and the levels of fecundity were reduced in intersex gammarids (Kelly *et al.* 2004).

Ablation of the androgenic gland in intersex crayfish *Cherax quadricarinatus* (feminised true males) allowed maturation of ovaries and induced secondary vitellogenesis, supporting evidence that the androgenic gland contains a vitellogenesis inhibitory factor (Khalaila *et al.* 1999).

1.4 CURRENT APPROACHES

The effect of endocrine disruption on the aquatic environment requires studying a range of species representing different taxa within the community. Sentinel species that have been recommended include annelids, molluscs, crustaceans, insects, echinoderms, fish and aquatic mammals (Taylor *et al.* 1999).

Detection of endocrine disruption in fish by measuring their vitellogenin levels has been widely studied, and is now used as a biomarker of exposure to oestrogenic chemicals (Pickford *et al.* 2003; Jobling *et al.* 2004; Jobling *et al.* 2006). It was found that there was an increase in the plasma vitellogenin concentrations detected by antibody assay (Tyler *et al.* 1999), when fish were exposed to STW effluent (Jobling *et al.* 2004).

A range of crustaceans have been utilised to assess the effects of endocrine disruption. It is essential to have a good understanding of the effects of EDCs in a wide range of species from this phylum due to the differences in their biology. For example, the cladoceran, *Daphnia magna,* is a moulting parthogenic species and a filter feeder. In contrast, the barnacle, *Balanus amphitrite,* is a non-moulting hermaphroditic, intertidal species and a suspension feeder. In the same phylum is the malacostracan, *G. pulex*, a moulting, gonochoristic benthic detrivore.

A vitellogenin-like protein, cypris major protein (CMP), a larval storage protein of *B. amphitrite* barnacles has been suggested as a biomarker of endocrine disruption in crustaceans after an elevation in CMP levels following 4-nonylphenol exposure (Billinghurst *et al.* 2000). Other indicators such as the cytochrome P450 enzymes (CYP) and heat shock proteins (HSP) involved in the metabolism of xenobiotics have also been discovered following exposure of *H. americanus* larvae to the insecticide, heptachlor (Snyder and Mulder 2001). The highly publicised case of intersex molluscs in response to the anti-fouling paint, TBT (Gibbs *et al.* 1991), is suggested to act through inhibition of cytochrome P450 aromatase (reviewed in Matthiessen and Gibbs 1998; Alzieu 2000), another possible biomarker of disruption (Lee *et al.* 2002). The pesticide endosulfan was found to inhibit moulting in *D. magna.* It has been suggested

that this is due to interference with the ecdysteroid receptor (Zou and Fingerman 1997a).

It has been proposed that the way forward in detecting the change of endocrine disruption on aquatic invertebrates is to undertake a full life cycle study, which for *G. pulex* would be about 100 days, to reduce the chance of misinterpreting general toxicity for endocrine disruption. This would also allow the chemical to be present at the critical window of EDC sensitivity, which, due to the absence of knowledge of invertebrate endocrinology, is currently unknown (Segner *et al.* 2003).

1.4.1 *Gammarus pulex* as a biomarker of endocrine disruption

Species that are deemed useful indicators of environmental stressors are termed "sentinel species". Criteria for selecting these species are extensive and *G. pulex* meets the majority of these criteria and has previously been recommended as a test organism (McCahon and Pascoe 1988b; Poulton and Pascoe 1990; Pascoe *et al.* 1994). Sentinel species must be ubiquitous organisms that are ecologically important. They must live in a habitat considered to receive significant exposure to pollutants and reproduce sexually with distinct males and females. On the practical side, they must be readily cultured and acquired, preferably from a local source to ensure continuity (Taylor *et al.* 1999). As a sentinel species *G. pulex* has widespread use as a tool for biomonitoring the toxicity of pollutants (Watts *et al.* 2001).

There is currently a lack of knowledge of the molecular mechanisms underlying endocrine control in *G. pulex*. This must be overcome in order to assess the genetic effects of freshwater pollutants.

1.4.2 Ecotoxicity studies in *Gammarus pulex*

G. pulex has already been employed in a wide variety of ecotoxicity studies and currently the most commonly used indicators of endocrine disruption in *G. pulex* are changes in behaviour, population sex ratios, reproduction and effects on precopulatory pairing (Cold and Forbes 2004). However these approaches depend upon limited endpoints and do not provide information on the mechanisms involved. To date, no studies at the molecular level have been undertaken in this species despite its

widespread use in other sentinel organisms including fish such as rainbow trout, roach, minnows and bass (Harries *et al.* 1996; Denslow *et al.* 2001; Larkin *et al.* 2002b; Zerulla *et al.* 2002; Larkin *et al.* 2003; Pickford *et al.* 2003; Jobling *et al.* 2006), and less commonly other organisms including barnacles, *Daphnia* sp., lobsters and mussels (Billinghurst *et al.* 2000; Snyder and Mulder 2001; Bultelle *et al.* 2002; Soetaert *et al.* 2007).

Lindane, an organochlorine insecticide, was found to decrease the growth and feeding rate (Blockwell *et al.* 1996; Blockwell *et al.* 1998) and prematurely disrupt the precopulatory pairing (Malbouisson *et al.* 1994), therefore affecting reproduction of *G. pulex*. The environmental oestrogen, 17α-ethinylestradiol has been found to increase the rate of sexual maturation and increase the population level skewing towards a sex proportion of 2 (female/male) (Watts *et al.* 2002). The pyrethroid insecticide, esfenvalerate was found to impact on pairing, reproduction and survival at environmentally relevant concentrations (Cold and Forbes 2004).

Exposure of *G. pulex* to sewage treatment effluent below an STW outfall was found to induce morphological changes in females showing an abnormal structure of oocytes during vitellogenesis (Gross *et al.* 2001). After exposure to 4-Nonylphenol it was discovered that uptake of pollutants in *G. pulex* is not only by feeding on detritus but also directly through the water column (Gross-Sorokin *et al.* 2003). *Gammarus* sp. from polluted sites in Canada exhibited significant alteration in the levels of chitin, arthropodin, sclerotin, metallothioneins and intermediary glucose metabolism consequently affecting the exoskeleton and gametogenesis (Gagne *et al.* 2005).

The tolerance of *G. pulex* to cadmium has been assessed indicating that females are less tolerant to exposure and sensitivity varies with moult cycle with early post-moult animals begin most vulnerable (McCahon and Pascoe 1988a; McCahon and Pascoe 1988d).

1.4.3 Ecotoxicogenomics

Ecotoxicogenomics is a relatively new field created through the technological advances in molecular genetics. It combines ecotoxicology and genomics to produce a powerful tool to identify the molecular mechanisms and pathways affected after environmentally sensitive pollutant exposure. The key approach is utilising DNA microarrays thereby allowing simultaneous analysis of thousands of gene transcripts from an individual organism, organ or population. The comparison of exposed and non-exposed organisms gives the potential to create a substantial number of biomarkers by identifying changes in gene expression following exposure.

Recently, microarrays have frequently been used in this manner, advancing our understanding of the effects and consequences of pollutant exposure at the molecular level. It has been suggested that this technique is a valid approach to identifying the effects of EDC exposure (Nuwaysir *et al.* 1999; Francois *et al.* 2003; Iguchi *et al.* 2006; Iguchi *et al.* 2007). The feasibility of this approach has also been tested in minnows (Larkin *et al.* 2002a) and *D. magna* (Poynton *et al.* 2007).

In mammals, microarray technology was used to identify changes in gene expression in the uterus of immature rats following exposure to diethylstilbestrol (DES), octylphenol (OP), nonylphenol (NP), bisphenol-A (BPA) or genistein (Hong *et al.* 2006).

Comparisons between gene expression profiles of bass exposed to oestradiol (E2), 4-nonylphenol (4-NP) and *p,p'*-DDE have also been successfully carried out (Larkin *et al.* 2002b). Expression profiles have been compared following exposure of minnows to 17β-oestradiol (E2), 17α-ethinyloestradiol (EE2), diethylstilbestrol (DES), methoxychlor (MXC) and endosulfan (Larkin *et al.* 2003). Differential expression of genes after exposure of zebrafish to oestrogen were assessed in both males and females using microarrays giving a greater understanding of the effects of EDCs on gametogenesis in both sexes (Santos *et al.* 2007).

The effect of TBT on the ascidian, *Ciona intestinalis,* has been studied revealing more than 200 gene expression changes (Azumi *et al.* 2004). After exposure to

fluoranthene, clofibrate, atrazine, b-naphthoflavone or DES, 356 potential biomarkers were identified in *C. elegans* (Reichert and Menzel 2005). A range of potential biomarkers to propiconazole exposure in daphnids have also been identified (Soetaert *et al.* 2006).

A vast quantity of information is generated by microarray experimentation and a wide range of applications in the field of ecotoxicology can be tested and compared including different species, toxicants, exposure times and concentrations. These significant merits of microarray technology and the evidence and success of previous research make it an ideal approach to studying the molecular mechanisms and ecotoxicological effects of EDC exposure on *G. pulex*.

1.5 PROJECT AIMS

The aims of this project were to gain a better understanding of the endocrine system of *G. pulex* at the molecular level and establish the genetic effects of endocrine disrupting chemicals and intersex state through the study of *E. marinus*. The identification of potential molecular biomarkers was achieved through 2 main approaches, individual candidate genes and large scale genomics.

- EST sequences (~13,400) were generated from 2 *G. pulex* cDNA libraries, a directional full length cDNA representing a full spectrum of sex, age and moult stage and an adult male specific subtractive library. Sequences derived from both libraries were analysed for biological sequence similarities compared, annotated using gene ontology attributes, clustered and a database created (Chapter 3).

- Mixed, male and female *G. pulex* cDNA libraries, were screened for candidate genes involved in molecular mechanisms regulating the endocrinological system and moulting processes that hypothetically may be disrupted by EDC exposure (Chapter 4).

- Discovery of the full gene sequence of potential sex specific biomarkers was attempted using multiple molecular biological approaches including Rapid Amplification of cDNA Ends (RACE) and degenerate Polymerase Chain Reaction (PCR) (Chapter 4).

- Expression levels of 2 potential physiological biomarkers were ascertained in 'normal' and exposed *G. pulex* using real-time quantitative PCR. The sequence data for both cuticle protein and vitellogenin were partially known and were used to analyse changes in gene expression of animals exposed to 17β-oestradiol and testosterone (Chapter 5).

- Sequence data of vitellogenin and cuticle protein were extended using the vast genetic information generated by the *G. pulex* cDNA libraries (Chapter 5).

- A *G. pulex* 'microarray system' was developed and optimised to create a robust and precise tool for gene expression profiling (Chapter 6).

- Microarray technology was used to determine the differential gene expression of ~13,400 gene objects to give a large quantity of data on the molecular responses of *G. pulex* to give a genetic 'snapshot' of physiological status. The normal physiological genetic status of *G. pulex* was ascertained through the analysis of specimens at different moult cycle stages, at distinct life stages and of both sexes. The basal expression levels of the biological statuses indicated below, was assessed (Chapters 7 and 8).
  - Moult stage
    - Intermoult (C_2)
    - Late Intermoult / Early Premoult (C_4 / D_0)
    - Premoult (D_1)
  - Sex
    - Male (>7mm)
    - Female (>7mm)
  - Developmental stage
    - Neonate (~2mm)
    - Small Juvenile (~4mm)
    - Large Juvenile (~5.5mm)

- Microarray techniques were optimised for *E. marinus* in addition to *G. pulex* in order to produce the best ecotoxicogenomics 'tool' for gammarids (Chapter 9).

- Intersex and normal *E. marinus* amphipods of both sexes were analysed by cross-species microarray hybridisation onto *G. pulex* microarray to create distinct transcriptomic profiles for each 'sex'. Differentially expressed genes were identified between the four states, suggesting putative biomarkers of sex and intersex (Chapter 9). The four test groups are shown below.
  - Normal Female (NF)
  - Normal Male (NM)
  - Intersex Female (IF)
  - Intersex Male (IM)

CHAPTER 2

MATERIALS AND METHODS

MATERIALS AND REAGENTS

The sources of materials, reagents and buffers used in this study are summarised in Appendix A (Table 1 and Table 2). Reagents were of molecular biology grade, unless otherwise stated.

2.1. AUTOCLAVING CONDITIONS

Before any technique was performed using DNA, all equipment was sterilised by autoclaving at 120°C at 15psi for 20 minutes. For equipment required for use with RNA this process was repeated.

2.2. DNA MARKERS

The DNA markers used were lambda DNA digested with *Hind*III (λDNA/*Hin*dIII) and ϕX174 digested with *Hae*III (ϕX174/*Hae*III) available from Promega, Southampton, UK and the 100bp and 1kb ladders available from New England BioLabs (UK) Ltd. Herts., UK. The fragment sizes of these markers (bp) are given in Appendix A (Table 3).

2.3. WATER

Several distinct grades of water were used. Where the grade of water is not specified double distilled water (ddH_2O) was used. For all applications involving DNA manipulations autoclaved HPLC grade water was utilised.

2.4. VECTORS

The vectors used throughout this report are shown in Appendix A (Table 4).

2.5. MEDIA

LB (Luria-Bertani Broth) media and LB agar were prepared with ddH_2O from capsules according to manufacturer's instructions (Bio101, Vista, CA, USA) and autoclaved prior to use (Section 2.1). Following sterilisation, the medium was left to cool to 55°C and where required, ampicillin or kanamycin (both from Sigma-Aldrich, Gillingham, Dorset, UK) was added at a concentration of 100μg/ml or 40μg/ml respectively.

2.6. ANTIBIOTICS

Stock solutions of ampicillin (100mg/ml) and kanamycin (40mg/ml) were prepared in sterile ddH_2O and sterilised by passing through a 0.22μM Nucleopore™ filter (Whatman International Ltd., Maidstone Kent, UK) prior to being aliquoted and stored at -20°C.

2.7. BACTERIAL STRAINS

The genotypes of the *Escherichia coli* (*E. coli*) strains used during this study are detailed in Appendix A (Table 5).

2.8. REAGENTS AND BUFFERS

All routine laboratory solutions were prepared using ddH_2O. Sterilisation was achieved by autoclaving where required. Heat sensitive components were passed through 0.22μm Nucleopore™ (Whatman International Ltd., Maidstone Kent, UK) filters and added separately following autoclaving. Where required, β-mercaptoethanol and Dithiothreitol (DTT) were always added fresh to any solution. Ethylenediamine Tetraacetic Acid (EDTA), Tris-Acetate-EDTA buffer (TAE) and Tris EDTA buffer (TE) were made according to the protocol of Sambrook *et al.*, (Sambrook 1989). A list of routinely employed solutions and reagents is shown in Appendix A (Table 1 and Table 2). All reagents were molecular biology grade unless otherwise stated.

METHODS

2.9. ANIMAL HUSBANDRY

2.9.1. *Gammarus pulex* sample collection and culture

Gammarus pulex were collected from a shallow stream near Llanfrynach, Powys, Wales by kick sampling. A photograph of the sampling site can be seen in Figure 2.1. This method involved kicking up substrate and stones upstream of a flat bottomed net to capture benthic invertebrates.

Figure 2.1: *G. pulex* sampling site

The banks alongside the fast running stream were lined with alder trees, the leaves from which provide food and cover. A location as close to the source as possible was used to reduce the risk of contamination by agricultural runoff. Water was tested for turbidity, permanganate index, ammonia, nitrite, nitrate and phosphate in order to assess its suitability as a control site.

The catch was transported to the laboratory in site-sampled river water and sorted, separating *G. pulex* from the other invertebrates.

They were washed in distilled water and kept in a glass tank with 15cm depth of Kent Marine R/O Right (Aquatics Online Ltd., Bridgend, UK) treated double distilled water (1.5g in 10L ddH_2O) to replicate freshwater. Using treated ddH_2O eliminated the risk of changes in water quality throughout future experiments, as the quality of river water varies dependent on rainfall and effluent quality. Tap water also changes in quality and composition, and pesticides have previously been detected in tap water (Drinking Water Inspectorate 2002). These variations may affect baseline and control results and cause experiments to be statistically invalid.

The water was constantly aerated and internally filtered through filter floss (J&K Aquatics, Taunton, UK) and the amphipods maintained at 17°C under a 16hr light : 8hr dark regime. *G. pulex* were fed on common alder (*Alnus glutinosa*) leaves that had been conditioned for one week in tank water in order for partial decomposition to occur (Figure 2.2). These were provided as a substrate for bacterial and fungal growth as this has been determined to be an essential part of their diet (Willoughby and Sutcliffe 1976). In addition to detrital matter, live waterweed (*Elodea densa*) was added to increase shelter and supplement their diet.

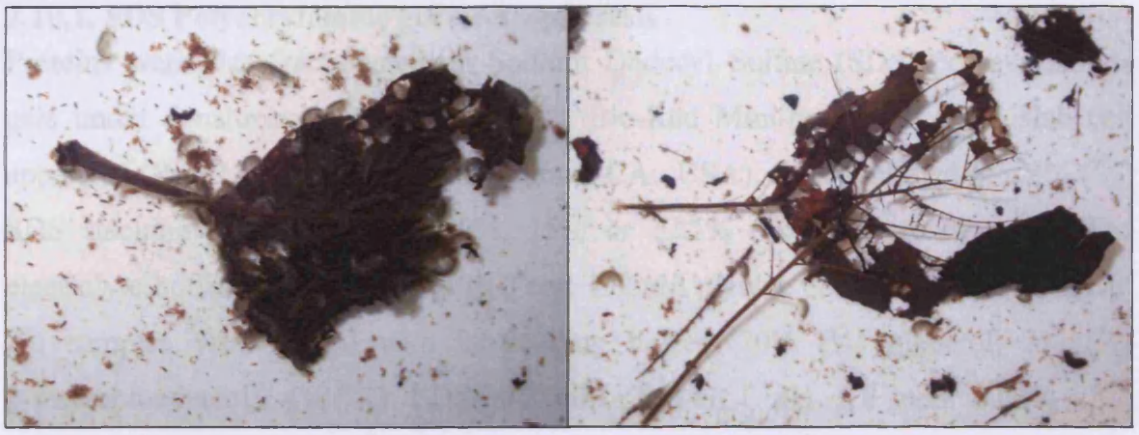
2.9.2. Exposure of *G. pulex* to 17β-oestradiol and testosterone

Exposures of *G. pulex* to 17β-oestradiol and testosterone were undertaken by WRc-NSF Ltd, Medmenham, Bucks., UK. The following information was taken from "Endocrine disruption in aquatic and terrestrial invertebrates", a DEFRA final report (Johnson *et al.* 2005). The animals were exposed to 0.01, 0.1, 1 and 10mg/L of each chemical for a 7 or 14 day period alongside controls. The environmental conditions were 25 ± 2°C, with a photoperiod of 16 hours light and 8 hours dark. For each exposure concentration a stock solution of the test substance was prepared in absolute ethanol and added to 30L groundwater in 50L all-glass aquaria. The stock solutions were prepared such that the same volume of ethanol was added to each treatment (0.1ml/L). The contents of each vessel were gently aerated using an air diffuser stone throughout the test period. Perspex sheets, with holes cut so that each sheet held 10

plastic medicine beakers, were then suspended over the solutions. The beakers were stacked in sets, with meshes positioned between them. One pre-copula pair of *G. pulex* was added to each stack of beakers; hence each test vessel contained ten replicate pre-copula pairs of test organisms at the start of each test. To provide the adults with food, alder leaf discs were added to each stack of beakers at test commencement. Additional discs (usually three) were added to each set of beakers as required (usually once) during the test.

After test initiation, measurements of water quality (temperature, pH, concentration of dissolved oxygen, total hardness and bicarbonate alkalinity) were made on the control and test solutions. These solutions were renewed each Monday, Wednesday and Friday. During the test the mobility of the test animals was assessed and recorded at 24-hour intervals. On day 7, five pre-copula pairs from each treatment were placed in RNA*later*® and stored (5°C) for analysis. On day 14, following observations of mobility on the test animals, measurements of water quality were made. Then the remaining pre-copula pairs from each treatment were placed in RNA*later*® and stored for analysis.

At test commencement, samples for chemical analysis were collected from each vessel and prepared for analysis. Then to assess stability of the test substance once extracted onto the C18 columns, duplicate samples were taken from each treatment at the start of the test. One set was submitted for analysis immediately, while the second set was foil wrapped and stored at 5°C in the dark for 48 hours prior to analysis. Analyses were carried out on samples where significant effects on biomarker responses and toxicity indices were recorded.



G. pulex devouring conditioned alder leaves.

Figure 2.2: Alder leaves (*A. glutinosa*)

The fallen leaves were collected during the autumn season and stored dry. They were rehydrated in mature tank water and conditioned for a week before being fed to *G. pulex*.

2.10. PROTEIN METHODS

2.10.1. SDS Polyacrylamide gel electrophoresis

Proteins were size fractionated on Sodium Dodecyl Sulfate (SDS) polyacrylamide gels under denaturing conditions using a Bio-Rad Mini-protean II Dual slab cell apparatus (Bio-Rad Laboratories, Richmond, CA., USA). Gels consisted of a 5% (w/v) SDS stacking gel overlying a 20%, 15% or 12.5% (w/v) SDS running gel. The electrolyte buffer was 25mM Tris-HCl and 192mM glycine containing 1% (w/v) SDS. All samples were mixed with 2x loading buffer (20% (v/v) glycerol, 2% (v/v) β-mercaptoethanol, 4% (w/v) SDS, 125mM Tris-HCl pH 6.8 and 0.02% (w/v) Bromophenol Blue) and heated for 10 minutes at 100°C immediately before electrophoresis. Pre-stained protein markers (9-194kDa, Invitrogen Ltd., Paisley, UK) were loaded along side the protein samples. The samples were fractionated at 200 volts for one hour or until the smallest marker reached the bottom of the gel.

To visualise proteins, the SDS-PAGE gels were soaked for one hour in Coomassie Brilliant Blue staining solution (0.1% (w/v) Coomassie Blue, 50% (v/v) methanol and 10% (v/v) acetic acid) and destained in 25% (v/v) methanol and 10% (v/v) acetic acid. Both staining and destaining were performed at room temperature under constant shaking (Bibby Stuart Platform Rocker, 40rpm).

2.10.2. Isolation and purification of *Taq* DNA polymerase

A glycerol stock of *E. coli* expressing *Taq* DNA polymerase was spread onto a LB agar plate containing ampicillin and incubated at 37°C for 16 hours. A colony was selected and grown for 16 hours in LB media containing ampicillin (10ml) shaking (225rpm) at 37°C.

The cation exchange resin (Bio-Rex 70, 100-200 mesh, Bio-Rad) was mixed with Buffer C (20mM HEPES (4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid) pH7.9, 1mM EDTA, 0.5mM PMSF (phenylmethanesulphonylfluoride), 0.5% (v/v) Tween-20, 0.5% (v/v) Igepal CA-630) in proportions to give a column volume of 17cm^3 resin. The resin was poured into a column (25mm diameter) and the pH equilibrated to the pH of Buffer C.

The *E. coli* culture was pelleted by centrifugation (1500g, 10 minutes) and resuspended in LB media containing ampicillin (10ml). Two flasks containing LB media with ampicillin (500ml) were each inoculated with the overnight culture (1ml) and grown shaking (225rpm) at 37°C until the optical density reached 0.2 representing the logarithmic phase of growth.

In order to induce expression, isopropyl-beta-D-thiogalactopyranoside (IPTG, 0.25ml of 1M) was added to each flask to give a concentration of 0.5mM. The cultures were grown in the same way for a further 19 hours. The cultures were centrifuged (4600g, 20 minutes) in a chilled rotor (Beckman Coulter JLA-10.500) using a Beckman Coulter Avanti® JE centrifuge at 4°C. The pellets were resuspended in 50ml Buffer A (50mM Tris HCl pH 7.9, 50mM dextrose, 1mM EDTA) and recentrifuged (4600g, 20 minutes, 4°C). The supernatant was discarded and the pellets each resuspended in Buffer A containing 0.4% ($^{w}/_{v}$) lysozyme (10ml).

The preparation was incubated at room temperature for 30 minutes after which 10ml Buffer B (10mM Tris-HCl pH 7.9, 50mM KCl, 1mM EDTA, 1mM PMSF, 0.5% ($^{v}/_{v}$) Tween-20, 0.5% ($^{v}/_{v}$) Igepal CA-630) was added to each flask and heated at 75°C for 20 minutes. The flasks were placed on ice for 30 minutes and PEI (Polyethyleneimine) added to a concentration of 0.15% ($^{v}/_{v}$) and the flasks returned to ice for a further 10 minutes. After the cells were pelleted (9000g, 20 minutes, 4°C) they were resuspended in Buffer C (10ml) containing 0.025M KCl and re-pelleted (12,000g, 20 minutes, 4°C) before being resuspended in Buffer C (10ml) containing 0.15M KCl. After centrifugation (12,000g, 20 minutes, 4°C) the pellets were resuspended in Buffer C (20ml).

The suspension was loaded onto the column and washed with Buffer C containing 0.05M KCl (100ml). The *Taq* DNA polymerase was eluted from the column with Buffer C containing 0.2M KCl and the eluant collected in 5ml fractions. The fractions were analysed by SDS-PAGE (Section 2.10) and the positive fractions pooled for purification. The pooled fractions were concentrated to ½ volume using a Centriprep Concentrator by centrifugation (3000g, 40 minutes). A 15cm section of dialysis tubing was boiled (10 minutes) and the concentrated fractions dialysed overnight in Buffer C containing 50% ($^{v}/_{v}$) glycerol at 4°C.

Serial dilutions of the purified *Taq* polymerase (diluted in Buffer C with 50% (v/v) glycerol) were tested in a PCR to determine titre. The optimised dilution will correspond to approximately 50U/ml. The neat *Taq* polymerase was then stored at -80°C with diluted aliquots at -20°C for immediate use.

2.11. MOLECULAR BIOLOGY TECHNIQUES

2.11.1. Genomic DNA extraction

Genomic DNA (gDNA) was extracted from *G. pulex* using the Wizard® Genomic DNA Purification Kit (Promega, Southampton, UK) according to manufacturer's instructions for the isolation of gDNA from animal tissues. *G. pulex* were ground under nitrogen and weighed. Into prechilled tubes containing ~20mg of frozen tissue, Nuclei Lysis Solution (600µl) was added and the samples incubated at 65°C for 30 minutes. RNase solution (3µl, 4mg/ml) was added to remove and RNA present in the samples and incubating for 30 minutes at 37°C and cooled to room temperature for 5 minutes.

Protein precipitation solution (200µl) was added to the samples, vortexed and chilled on ice for 5 minutes to precipitate all the protein. The samples were centrifuged at 16,000g for 4 minutes to pellet the proteins and other cell debris. The supernatant containing the DNA was transferred to a fresh tube containing isopropanol (600µl) to precipitate the DNA, which was pelleted by centrifugation at 16,000g for 1 minute. The DNA pellet was washed with 70% (%v) ethanol, air dried for 10 minutes, resuspended in HPLC water and stored at 4°C.

2.11.2. Plasmid DNA extraction

Plasmid DNA from *E. coli* cultures was extracted using a Wizard® Plus SV Miniprep (Promega, Southampton, UK) according to manufacturer's instructions. This kit utilises the alkaline lysis method for plasmid isolation. *E. coli* cultures were grown in 10ml LB with 100µg/ml ampicillin for 16 hours at 37°C in a shaking incubator (250rpm) and 5ml centrifuged at 9400g for 5 minutes.

The pellet was resuspended in cell resuspension solution, which includes the Mg^{2+} containing chelating agent EDTA. This inhibits DNase action and makes the bacterial membrane permeable allowing the cell membranes to be lysed with cell lysis solution. It also contains RNase A which degrades any RNA present. The cell lysis solution contains the detergent SDS which denatures bacterial proteins and lyses the cell membrane releasing the plasmids and NaOH which denatures chromosomal and plasmid DNA. The addition of alkaline protease before the incubation stage ensures

the breakdown of the bacterial proteases. The reaction is neutralised with neutralization solution that contains guanidine hydrochloride and potassium acetate which a precipitate with the chromosomal DNA and proteins leaving the plasmid DNA in solution. The acetic acid present neutralises the NaOH from the cell lysis buffer.

After centrifugation to pellet the chromosomal DNA, proteins and cell debris, the supernatant containing plasmid DNA was transferred to a Wizard® *Plus* SV miniprep Spin Column. In the presence of high salt, in this case potassium acetate, the plasmid DNA binds to the spin column resin. The column wash solution was used to remove contaminating salts and proteins and contains a small quantity of salt to remove mononucleotides leaving the plasmid DNA bound to the resin. Plasmid DNA was released from the resin by the addition of pre-warmed (65°C) nuclease-free water and collected by centrifugation of the spin column. The DNA was stored at 4°C for immediate use and -20°C for long term storage.

2.11.3. Total RNA extraction

Total RNA was extracted from *G. pulex* using the Tri® Reagent method (Sigma-Aldrich, Gillingham, Dorset, UK). All equipment was double autoclaved to insure inactivation of any RNase present. *G. pulex* were washed three times in ddH_2O, leaving for 5 minutes in each wash to remove any debris from the water and for any digested food to be excreted. They were dried on Whatman Grade 1 filter paper (Whatman plc, Brentford, UK) and dropped into a mortar containing liquid nitrogen. After grinding, the fine powder was transferred to a weighed pre-chilled 1.5ml microfuge tube and immediately transferred to a Dewar flask of liquid nitrogen. The tube was reweighed and a quantity of Tri® Reagent was added according to the manufacturer's instructions.

Acid-washed sterilized beads (Sigma-Aldrich, Gillingham, Dorset, UK, particle size 425-600 μm, 0.1g) were added and the sample vortexed for 30 seconds to homogenize the cells. The samples were centrifuged at 12,000g at 4°C for 10 minutes to pellet the cell debris and the supernatant transferred to a fresh tube and incubated at room temperature for 5 minutes. Chloroform was added in accordance with the manufacturer's instructions and shaken followed by a further room temperature

incubation of 15 minutes. The layers were separated by centrifugation at 12,000g at 4°C for 15 minutes and the aqueous phase transferred to a fresh tube.

RNA was precipitated by the addition of isopropanol and an incubation of 5 minutes at room temperature. The precipitated RNA was pelleted by centrifugation at 12,000g at 4°C for 10 minutes and the pellet washed with 75% (%v) ethanol, air dried and resuspended in sterilized filtered HPLC water preheated to 65°C. The quality and quantity was analysed by spectrophotometry (Section 2.11.6) and agarose gel electrophoresis (Section 2.11.11). RNA was stored at -80°C until required.

2.11.4. Messenger RNA (mRNA) extraction

Messenger RNA (mRNA) was extracted from total RNA using an mRNA Purification Kit (GE Healthcare UK Ltd., Bucks., UK). This form of RNA is usually only 1-5% of total RNA, so a high quantity of total RNA (~1mg) is required at the start of the protocol. Firstly the spin columns were prepared by resuspending the gel matrix, draining and washing with high salt buffer twice. Before loading onto the column, the total RNA was concentrated by ethanol precipitation and resuspended in 1ml elution buffer. The RNA was heated at 65°C for 5 minutes to remove any secondary structure and place on ice. Sample buffer was added to the RNA solution and mixed gently, then loaded onto the column with great care not to let any run down the sides of the gel plug. The column was left for 2 minutes at room temperature for the RNA to soak into the gel before centrifugation at 350g for 2 minutes.

High salt buffer was applied to the column and removed by centrifugation at 350g for 2 minutes. The column was washed three times with low salt buffer and the mRNA extracted by the addition of 0.25ml pre warmed elution buffer (65°C) which was collected by centrifugation at 150g for 2 minutes. This elution stage was repeated four times to give a total of 1ml.

The entire process was repeated with a fresh column to ensure that pure mRNA was isolated away from any total RNA. The purified mRNA was split into three aliquots and one concentrated by ethanol precipitation, resuspended in 16µl sterilized HPLC water and quantified by spectrophotometry. The remaining mRNA samples were stored in ethanol at -80°C.

2.11.5. Purification of Nucleic acids

2.11.5.1. Phenol:Chloroform Extraction of DNA and RNA

DNA and RNA was purified by phenol:chloroform extraction to remove traces of any proteins that may be present. An equal volume of phenol:chloroform (1:1; $^{v}/_{v}$) was added, with phenol used being at pH7.8 for DNA purification and pH4.8 for RNA purification. After vortexing the samples were centrifuged at 14,000g for 2 minutes and the upper aqueous layer containing the nucleic acids were transferred to a fresh tube. Traces of phenol were then removed by adding an equal volume of chloroform:*iso*amyl alcohol (24:1; $^{v}/_{v}$), the reaction vortexed and centrifuged at 14,000g for 2 minutes. The upper aqueous layer was removed to a fresh tube and the DNA or RNA concentrated by ethanol precipitation (Section 2.11.5.2)

2.11.5.2. Ethanol Precipitation of DNA

Nucleic acids were concentrated and remaining traces of phenol removed by ethanol precipitation after which, 10% ($^{v}/_{v}$) of sodium acetate (NaOAc, 3M) was added followed by a 2.5x volume of chilled 100% ethanol. When RNA was concentrated, a glycogen carrier (10mg/ml) was added at 1% ($^{w}/_{v}$) to facilitate precipitation. After gentle inversion the reaction was incubated overnight at -20°C. Precipitated nucleic acids were pelleted by centrifugation at 4°C for 15 minutes at 12,000g. The pellet was air dried and resuspended in sterilised HPLC water.

2.11.6. Reverse Transcription Polymerase Chain Reaction (RT-PCR)

Complementary DNA (cDNA) was synthesised from messenger RNA (mRNA) using reverse transcriptase. Total RNA (2µg) was heated at 65°C for 3 minutes to remove any secondary structure was combined with dNTPs (10pmols), oligo(dT) primer (50pmols), random hexamers (50pmols) and made up to 10µl with HPLC water. After incubating for 10 minutes at 70°C, the reaction was placed on ice for 10 minutes. DTT (dithiothreitol, 200mmols), RNasin (20U) and MMLV reverse transcriptase (400U) were added and the reaction buffered in 1x MMLV buffer in a total volume of 20µl. After gentle mixing the samples were incubated at room temperature for 10 minutes, followed by 50 minutes at 37°C in order for the first strand to be synthesised

and finally 10 minutes at 80°C to inactivate the enzymes. The cDNA samples were then stored at -20°C until required as a template for PCR.

2.11.7. Estimation of the quantity & purity of nucleic acids by spectrophotometry

The concentration of DNA and RNA can be determined by absorption at 260nm. The absorption of 1OD is equivalent to approximately 50µg/ml dsDNA, 33µg/ml ssDNA, 40µg/ml RNA and approximately 30µg/ml for oligonucleotides. The purity of nucleic acids can be assessed by measuring a ratio of $^{A260}/_{A280}$ and $^{A260}/_{A230}$. Protein contaminants absorb at 280nm and so pure DNA should have a ratio of 1.8 and RNA a ratio of approximately 2.0. Phenol and ethanol contamination is absorbed at 230nm and therefore the $^{A260}/_{A230}$ ratio of pure nucleic acids should be about 2.2. Absorbance was measured on a Labcrew Ultraspec 2100 Pro spectrophotometer or a Pharmacia GeneQuant spectrophotometer.

2.11.8. Estimation of the quantity of cDNA by PicoGreen® Assay

PicoGreen® (Invitrogen Ltd., Paisley, UK) is a selective dye that only binds to double stranded DNA and DNA:RNA hybrids. After reverse transcription the cDNA is still bound to the RNA, so this method is able to selectively quantify the cDNA quantity without interference from other nucleic acids such as primers and template RNA (single stranded). A solution of λDNA/*Hin*dIII marker (2µg/ml) (Promega Ltd., Southampton, UK.) was prepared by a 250-fold dilution of the 500µg/ml stock solution. Samples (1µl) were diluted with 1X TE (49µl).

Standards were prepared at 2x concentration to give a final well concentration of between 3.91 and 500ng/ml. The standards (50µl) and samples (50µl) were transferred to a black OptiPlate-96 (PerkinElmer LAS (UK) Ltd., Bucks., UK). For accuracy this was performed in duplicate or triplicate. PicoGreen® was diluted 1:400 with 1x TE in an amber tube. Diluted PicoGreen® (50µl) was added to each well to make a total volume of 100µl per well.

The plate was shaken in the DPC MicroMix® 5 plate shaker (Diagnostic Products Corporation, Los Angeles, California, USA) for 1 minute (Form 50, amplitude 9) and

then pulsed in a centrifuge to bring the contents to the bottom of the wells. The plate was read using a Flourocounter (FluoroCount™, PerkinElmer LAS (UK) Ltd., Bucks., UK) at the following wavelengths: excitation, 485nm and emission, 530nm. The standard curve was used to determine the original concentration of the cDNA.

2.11.9. Amplification of DNA by the Polymerase Chain Reaction (PCR)

2.11.9.1. Primer Design

Primers were designed using Primer Express® (Version 2.0, Applied Biosystems, Foster City, USA) software. The primers used throughout this report are shown in Appendix A (Table 6).

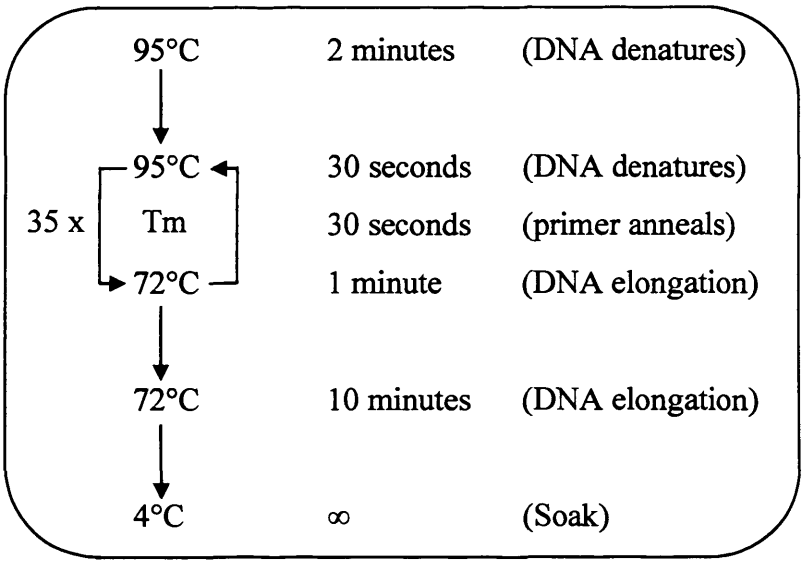
2.11.9.2. Degenerate PCR primer design

For each candidate gene, regions of high protein homology and conserved domains in closely related organisms were identified. The corresponding DNA sequences for each region were used to design the primers keeping the degeneracy as low as possible (<16) whilst still retaining the specificity. Degenerate bases used in primer design and degenerate primers used are shown in the appendix (Table 7 and Table 8 respectively).

2.11.9.3. PCR procedure

Polymerase chain reactions (PCRs) were performed in order to amplify genes of interest from cDNA, genomic DNA (gDNA) or plasmid DNA. For each reaction ~100ng DNA template (or a single colony) was amplified using 500nM M13F and M13R primers (Appendix A; Table 6) using 50U/ml *Taq* DNA polymerase buffered with 1x *Taq* PCR amplification buffer (Promega, Southampton, UK) and 0.2mM of each dNTP. Each reaction was supplemented with an optimised quantity of $MgCl_2$ of between 2 and 2.5mM.

The reaction was carried out in a Techne® Flexigene (Techne, Cambridge, UK), an MWG Primus 25 or 96 thermal cycler (MWG Biotech (UK) Ltd, Milton Keynes, UK) using the programme indicated below unless otherwise stated. PCR amplicons were analysed by agarose gel electrophoresis as described in Section 2.11.11.



2.11.10. Gene cloning

2.11.10.1. Extraction and purification of DNA from an agarose gel

After DNA electrophoresis (Section 2.11.11) the gel was visualised on a UV transilluminator and the band of interest excised from the gel with a sterile scalpel. The gel was placed on top of a 1ml filter tip cut just below the filter. The filter and gel were placed in a 1.5ml microtube and the liquid extracted by centrifugation (20,600g, 7 minutes). The liquid from the gel was collected at the base of the tube and purified by phenol:chloroform extraction (Section 2.11.5.1).

Pellet Paint® NF Co-Precipitant (Novagen, Merck Chemicals Limited, Nottingham, UK) was added (1µl) to the aqueous layer and was mixed with 10% (%v) of sodium acetate (NaOAc, 3M) followed by a 2.5x volume of 100% ethanol. The samples were incubated at room temperature for 5 minutes and the DNA pelleted by centrifugation (20,600g, 10 minutes). The pellet was washed firstly with 70% (%v) ethanol and then with 100% ethanol before being air dried and resuspended in an appropriate quantity of ddH_2O.

2.11.10.2. Ligation of PCR products into pGEM-T vector

PCR products were ligated into pGEM-T vector (Promega, Southampton, UK) according to the manufacturer's instructions. The insert DNA quantity used was determined from a 3:1 insert:vector ratio calculated using the following equation:

$$\frac{ng\ of\ vector \times size\ of\ insert\ (kb)}{vector\ size\ (kb)} \times \frac{3}{1} = ng\ of\ insert$$

2.11.10.3. Transformation of Plasmid DNA

In order to replicate genes of interest, plasmid DNA was transformed into *E. coli* DH5α™ cells (Invitrogen Ltd., Paisley, UK) as per manufacturer's instructions (Invitrogen Ltd., Paisley, UK, 2001 version). The transformation was incubated in LB medium for an hour at 37°C in an incubator shaking at 250rpm. After gentle mixing ensuring cells were in suspension, 200µl was spread evenly onto LB agar plate containing 100µg/ml ampicillin and incubated for 16 hours at 37°C.

2.11.10.4. PCR Screening of Transformation Colonies

Successful recombinants were identified by screening for the insert in randomly selected colonies by PCR. After a 16 hour incubation in 10ml LB containing 100µl/ml ampicillin (maintained at 37°C with orbital shaking at 250rpm), 5µl was diluted in 500µl of LB and mixed. The suspension was boiled for 5 minutes, cooled and 1µl used as a template in a 20µl PCR as described in Section 2.11.9.3. For pGEM-T and pBluescript® SK II+ clones, M13 primers (Appendix A; Table 6) were used.

2.11.11. Agarose Gel Electrophoresis

Separation of DNA and RNA was performed by electrophoresis. Before separation of RNA was performed, the gel tank, mould and combs were soaked in 1% ($^{w}/_{v}$) sodium dodecyl sulphate (SDS) overnight and the TAE and agarose double autoclaved. The agarose gel was prepared by dissolving agarose (1-4% ($^{w}/_{v}$) as specified in the text) in 1x TAE buffer in a microwave before cooling to a point where no more steam was produced. Ethidium bromide was then added at a concentration of 0.4µg/ml, the solution mixed and poured into the gel mould. After leaving the gel to set for at least 30 minutes, the DNA or RNA samples were mixed with blue loading dye and the samples loaded onto the gel alongside either *Hin*dIII digested lambda DNA (λDNA/*Hin*dIII) or *Hae*III digested ϕX174 (ϕX174/*Hae*III) DNA molecular weight marker. The samples were then resolved by electrophoresis at 100V for approximately 30 minutes in a Pharmacia GNA-100 tank in 1x TAE buffer. Nucleic acid bands were then visualised on a UV gel documentation system.

2.11.11.1. Agarose minigel electrophoresis of mRNA and cDNA

A 1% ($^{w}/_{v}$) agarose solution (in 1x TAE) was prepared and heated until the agarose was dissolved. The agarose was cooled to 55°C and poured into a minigel mould and covered with a microscope slide until the agarose had set (~10 minutes). The gel was transferred to a clean microscope slide and loaded with a 1:1 ($^{v}/_{v}$) mixture of loading dye and nucleic acid to a total volume of 2µl. The nucleic acids were separated by electrophoresis in a Pharmacia GNA-100 tank in 1x TAE buffer at 120V for 20 minutes.

A 1:2000 dilution ($^{v}/_{v}$) of SYBR® Green I or II gel stain (Molecular Probes Europe BV, Leiden, The Netherlands) in 1x TAE buffer was prepared and covered in foil. The gel was washed in 1x TAE and stained by incubating the gel in the SYBR® Green solution wrapped in foil at room temperature, rocking at 50 cycles a minute for 15 minutes. The bands were visualised using a Genomic Solutions® GeneTAC LS IV laser scanner at 560nm using the Green filter, set to Cy3 (Genomic Solutions Ltd., Huntingdon, UK).

2.11.12. DNA Sequencing

DNA was sequenced using ABI PRISM® BigDye v3.1 Terminator Sequencing technology on the ABI PRISM® 3100 DNA Sequencer run by the Cardiff University Molecular Biology Support Unit. After quantification by spectrophotometry (Section 2.11.7), DNA volume required was calculated by the following equation:

$$\text{Plasmid DNA volume (µl)} = \frac{290}{\textit{Concentration } (\mu g / ml)}$$

$$\text{PCR product volume (µl)} = \frac{60}{\textit{Concentration } (\mu g / ml)}$$

The volume of DNA was made up to 7.5µl. A 1:4 ratio of BigDye:water was made and buffered in 1x BigDye buffer, of this mix; 6µl was added to the 7.5µl of DNA with 0.2mM of each M13 primer (Appendix A; Table 6). Following this the insert was amplified by PCR using the following conditions:

	96°C	3 minutes	(DNA denatures)
35 x	96°C	15 seconds	(DNA denatures)
	50°C	10 seconds	(primer anneals)
	72°C	3 minutes	(DNA elongation)
	4°C	∞	(Soak)

The PCR reactions are purified by isopropanol precipitation. Isopropanol (90µl of 70% (v/v)) was added to each reaction, mixed and incubated in the dark for 15 minutes at room temperature. The tubes were centrifuged at 12,000g for 30 minutes and the DNA pellet washed with 180µl 70% (v/v) isopropanol, centrifuged at 12,000g for 15 minutes and the pellet air dried in the dark at room temperature for 10 minutes. The pellet was then resuspended in HiDye formamide (15µl) and sequenced on a the ABI PRISM® 3100 DNA Sequencer run by the Cardiff University Molecular Biology Support Unit.

2.11.13. TaqMan® Quantitative Polymerase Chain Reaction (QPCR)

Quantitative PCR is used to determine the level of expression of a specific gene. It works by using a sequence specific probe that contains a fluorescent dye (FAM; carboxyfluorescein) tagged at the 5' end and a quencher (TAMRA; carboxytetramethylrhodamine) at the 3' end. The probe is designed to bind to cDNA representing the gene of interest. Fluorescence only occurs when FAM is released from the probe and consequently detached from the TAMRA quencher. This release occurs through the 5' exonuclease activity of the *Taq* polymerase that extends the second strand from a pair of primers based either side of the probe. A scheme of how this process works is shown in Figure 2.3.

Primers were designed (Dr. Judith Richards) using the following parameters in Primer Express:

- Primers and probes were selected in a region with G/C contents of 20-80%.
- Primers were designed to amplify short segments of DNA within the target sequence. The amplicon size was between 50 and 150 bp.
- If possible at least one primer and/or probe should cross an exon junction amplifying cDNA but not genomic DNA. This isn't possible with *G. pulex* due to the lack of sufficient genomic data.
- The melting temperature (Tm) of the probe was between 68 and 70°C.
- Runs of an identical nucleotides were avoided in both primers and the probe, especially G, where runs of 4 or more should be avoided.
- Gs on the 5' end of the probe were avoided.
- The strand that gives the probe more Cs than Gs was selected.
- Primers were designed to be as close to the probe as possible.
- The melting temperature (Tm) of the primer was between 58 and 60°C.
- The total number of Gs and Cs in the last five nucleotides at the 3' end of the primers did not exceed two.

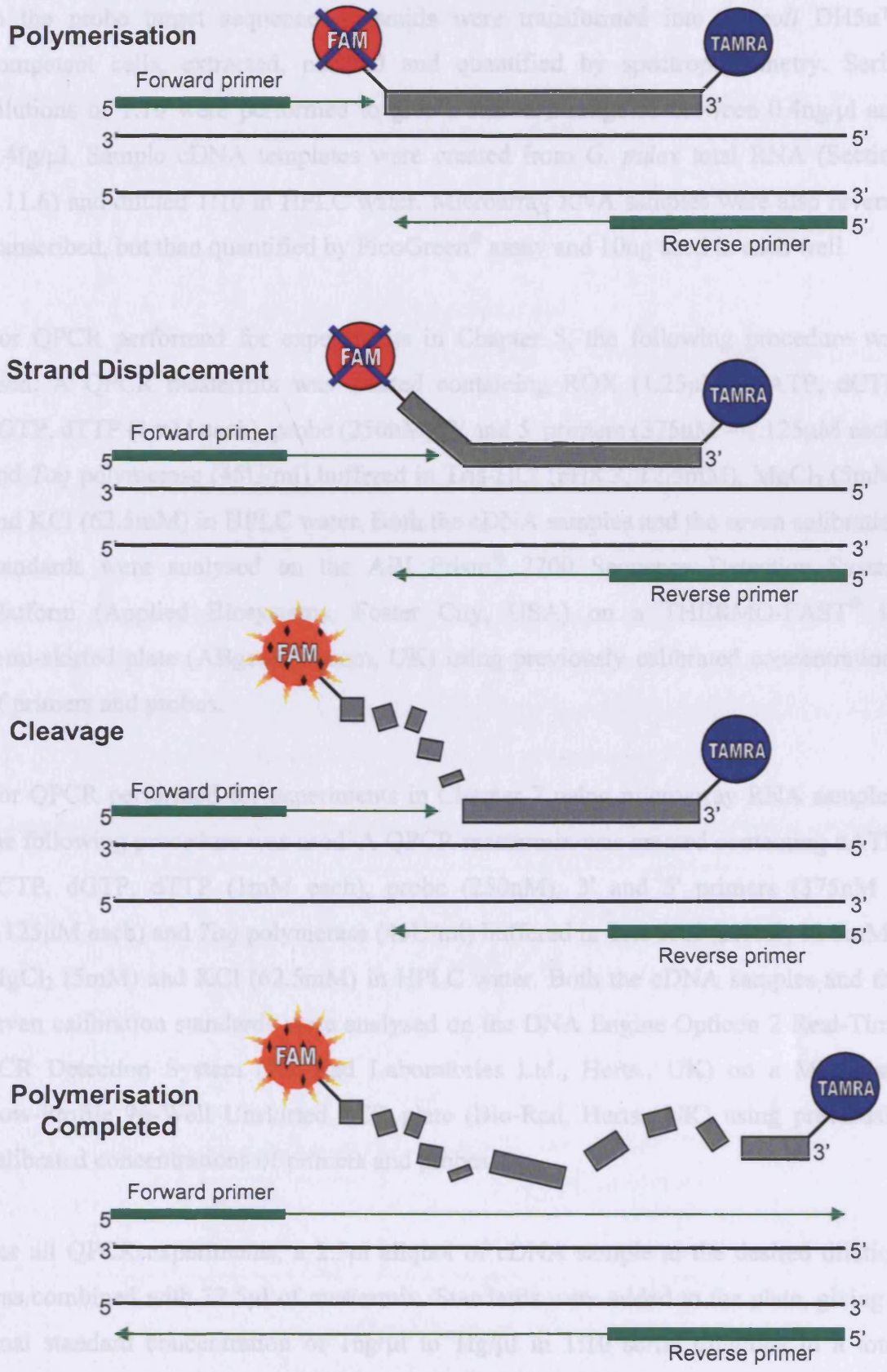


Figure 2.3: The principles of QPCR.

Standard templates for calibration consisted of a plasmid with an insert corresponding to the probe target sequence. Plasmids were transformed into *E. coli* DH5α™ competent cells, extracted, purified and quantified by spectrophotometry. Serial dilutions of 1:10 were performed to give a standard range of between 0.4ng/µl and 0.4fg/µl. Sample cDNA templates were created from *G. pulex* total RNA (Section 2.11.6) and diluted 1:10 in HPLC water. Microarray RNA samples were also reverse transcribed, but then quantified by PicoGreen® assay and 10ng used in each well.

For QPCR performed for experiments in Chapter 5, the following procedure was used. A QPCR mastermix was created containing ROX (1.25µM), dATP, dCTP, dGTP, dTTP (1mM each), probe (250nM), 3' and 5' primers (375nM – 1.125µM each) and *Taq* polymerase (45U/ml) buffered in Tris-HCl (pH8.3, 12.5mM), $MgCl_2$ (5mM) and KCl (62.5mM) in HPLC water. Both the cDNA samples and the seven calibration standards were analysed on the ABI Prism® 7700 Sequence Detection System Platform (Applied Biosystems, Foster City, USA) on a THERMO-FAST® 96 semi-skirted plate (ABgene, Epsom, UK) using previously calibrated concentrations of primers and probes.

For QPCR performed for experiments in Chapter 7 using microarray RNA samples, the following procedure was used. A QPCR mastermix was created containing dATP, dCTP, dGTP, dTTP (1mM each), probe (250nM), 3' and 5' primers (375nM – 1.125µM each) and *Taq* polymerase (45U/ml) buffered in Tris-HCl (pH8.3, 12.5mM), $MgCl_2$ (5mM) and KCl (62.5mM) in HPLC water. Both the cDNA samples and the seven calibration standards were analysed on the DNA Engine Opticon 2 Real-Time PCR Detection System (Bio-Rad Laboratories Ltd., Herts., UK) on a Multiplate Low-Profile 96-Well Unskirted PCR plate (Bio-Rad, Herts., UK) using previously calibrated concentrations of primers and probes.

For all QPCR experiments, a 2.5µl aliquot of cDNA sample at the desired dilution was combined with 22.5µl of mastermix. Standards were added to the plate, giving a final standard concentration of 1ng/µl to 1fg/µl in 1:10 serial dilutions in a total volume of 25µl in all wells. The plate was sealed with ultra clear cap strips (ABgene, Epsom, UK) to ensure fluorescence detection, and contents pulsed in a centrifuge. Standards were analysed in triplicate and samples in duplicate.

The plates were analysed on the ABI Prism® 7700 Sequence Detection System Platform which uses the thermal cycling programme shown below.

	50°C	2 minutes	(Polymerase activation)
	95°C	10 minutes	(DNA denatures)
40 x	95°C	15 seconds	(DNA denatures)
	60°C	1 minute	(Primer anneals, DNA elongation and fluorescence)

The level of gene expression is determined from the level of target cDNA at the start of PCR, and therefore the quantity of fluorogenic FAM released and detected. The reporter fluorescence signal is normalised against the passive reference dye, ROX (carboxy-X-rhodamine) to eliminate well variation from plastic imperfections, pipetting error and well location. Specific gene expression is represented by the cycle threshold, which is determined in the exponential phase of PCR, calculated as the middle of the linear part of the amplification on a logarithmic graph of normalised reporter signal against cycle number. This graph can be seen in Figure 2.4.

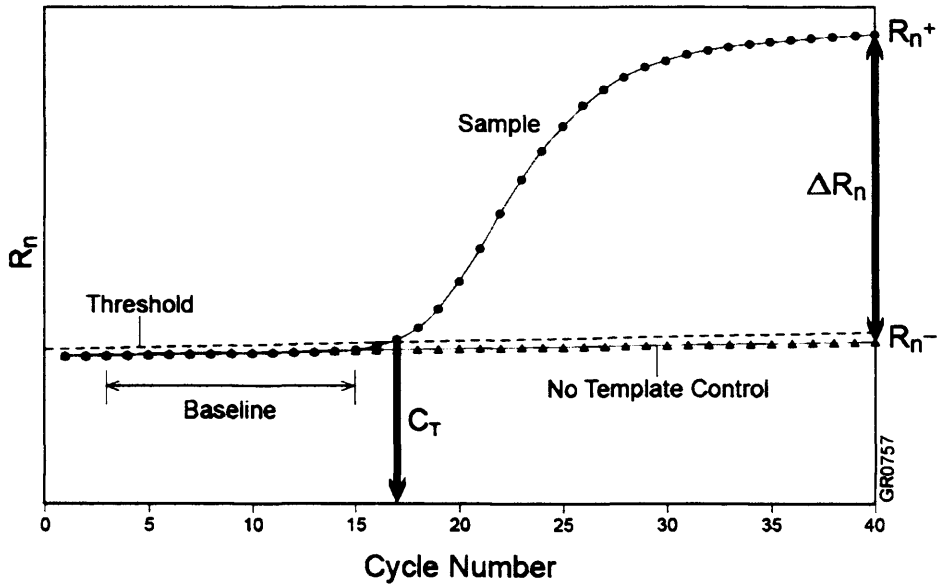


Figure 2.4: Real time detection

The cycle threshold (C_T) is the cycle number needed for the fluorescence to pass the threshold. Quantities were calculated from regression analysis of the calibration standards.

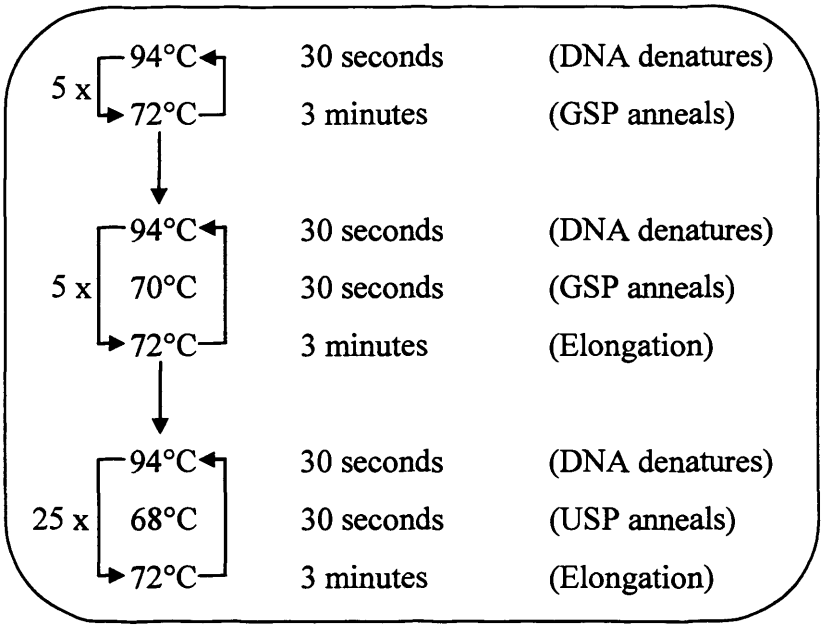
2.11.14. Rapid Amplification of cDNA Ends (RACE)

2.11.14.1. Primer design

Primers were designed for Rapid Amplification of cDNA Ends (RACE) PCR (Gene specific primers; GSPs) using Primer Express® (Version 2.0, Applied Biosystems, Foster City, USA) software. Successful primer sequences were between 23 and 28 nucleotides long, with between 50 and 70% GC (guanine and cytosine) content with a melting temperature of greater than 70°C with an overlap of known sequence of around 100bp. RACE primers used can be seen in Appendix A (Table 9).

2.11.14.2. RACE PCR

Rapid Amplification of cDNA Ends (RACE) allows both the 3' and 5' ends of cDNA synthesised from an mRNA population to be amplified and sequenced, and to potentially clone the full length cDNA. The *G. pulex* RACE library was generated by Dr. Judith Richards using a mixed adult and juvenile population as described by the manufacturer (Marathon™ cDNA Amplification Kit, Clontech-Takara Bio Europe, Saint-Germain-en-Laye, France). Amplification was performed using the SMART™ RACE cDNA Amplification Kit (BD Biosciences Clontech, Palo Alto, USA) according to the manufacturer's instructions. The PCR reaction contained 0.2mM of each dNTP, 2.5µl RACE-Ready cDNA, 5µl 10x Universal Primer Mix (UPM), 0.2mM GSP and 1µl BD Advantage 2 Polymerase mix buffered in 1x BD Advantage 2 PCR Buffer. The reaction was overlaid with a drop of mineral oil before amplification in a PerkinElmer DNA Thermal Cycler 480 (PerkinElmer LAS (UK) Ltd., Bucks., UK) programmed with the thermal cycling conditions shown below.



After amplification the PCR reaction was transferred to a fresh tube to separate it form the mineral oil. The reaction was then analysed by agarose electrophoresis (Section 2.11.11) and positive DNA bands removed and the DNA extracted and purified (2.11.10.1).

2.11.15. Sequence analysis

Analysis of DNA sequences was mostly undertaken using Discovery Studio™ Gene software (Accelrys, Cambridge, UK). Alignments were created using MultAlin, a multiple sequence alignment created by Florence Corpet which can be found at http://prodes.toulouse.inra.fr/multalin/multalin.html (Corpet 1988). Nucleotide and protein homology was identified using the Basic Local Alignment Search Tool (BLASTN or BLASTX respectively (Altschul 1997), on National Center for Biotechnology Information (NCBI) website at http://www.ncbi.nlm.nih.gov/BLAST/. Consensus sequences were created using BioEdit (Ibis Therapeutics, Carlsbad, USA) and DNAsis (Hitachi Software Engineering Europe AG, Berlin, Germany).

2.12. cDNA LIBRARY METHODS

2.12.1. Construction

A cDNA library was constructed using the pBluescript® II XR cDNA Library Construction Kit (Stratagene Europe, Amsterdam, The Netherlands).

2.12.1.1. First-Strand cDNA synthesis

First-strand methyl nucleotide mixture (0.2mM DATP, dGTP, dTTP and 0.1mM 5-methyl dCTP) linker-primer (56µg/ml) and RNase Block Ribonuclease Inhibitor (800U/ml) were combined in 1x first-strand buffer in an RNase-free microcentrifuge tube. The contents were mixed and the mRNA purified in Section 2.11.4. (5µg) was resuspended in 37.5µl HPLC water, added to the first strand synthesis reaction and remixed gently. The reaction was incubated at room temperature for 10 minutes. Stratascript RT (1500U/ml) was then added; the reaction mixed gently and pulsed to the bottom of the tube. As a control, 5µl was transferred to a separate tube containing first-strand methyl nucleotide mixture (5nmols dATP, dGTP, dTTP and 2.5nmols 5-methyl dCTP) in an RNase free tube. Both reactions were incubated at 42°C for 1 hour and the control reaction stored at -20°C.

2.12.1.2. Second-Strand cDNA synthesis

The first-strand synthesis reaction was placed on ice and combined with second-strand dNTP mixture (0.4mM dATP, dGTP, dTTP and 1mM 5-methyl dCTP), RNase H (15U/ml) and DNA polymerase I (495U/ml) buffered in 1x second-strand buffer.

2.12.1.3. Blunting the cDNA Termini

The reaction was gently mixed, spun down and incubated at 16°C for 2.5 hours. After being placed on ice the reaction was added to blunting dNTP mix (0.25mM each dNTP) and cloned *Pfu* DNA polymerase (22U/ml).

The reaction was gently mixed, spun down and incubated at 72°C for 30 minutes. To purify the cDNA, 200µl phenol-chloroform (1:1 (v/v); phenol was equilibrated to pH7.9) was added to the reaction and vortexed before being centrifuged at full speed

for 2 minutes at room temperature. The upper aqueous layer was transferred to a fresh tube and 200µl chloroform added, vortexed and spun at full speed for 2 minutes at room temperature. The upper aqueous layer was again transferred to a fresh tube and added to sodium acetate (0.1M) and ethanol (65% $^{v}/_{v}$). The mixture was vortexed and precipitated overnight at -80°C.

The precipitated cDNA was centrifuged at 21,500g for 60 minutes at 4°C and the ethanol removed and discarded. The pellet was washed by adding 500µl of 70% ($^{v}/_{v}$) ethanol and then spun at maximum speed for 2 minutes at room temperature. Again the ethanol was removed and discarded and the pellet dried by vacuum centrifugation. The DNA pellet was resuspended by adding 3.6µg of *Eco*RI adapters and then incubated at 4°C for 30 minutes. The cDNA suspension was transferred to a PCR tube and as a control a 1µl aliquot removed to run on an agarose gel.

2.12.1.4. Ligating the EcoRI Adapters

The blunted cDNA and *Eco*RI adapters were added to rATP (0.8mM) and T4 DNA ligase (333U/ml) buffered in 1x ligase buffer. The ligation reaction was pulsed down and incubated overnight at 8°C. An aliquot (1µl) was removed and used as a control.

2.12.1.5. Minigel electrophoresis of First and Second strand synthesis

The first-strand and second-strand synthesis controls were analysed by minigel electrophoresis using 1% ($^{w}/_{v}$) agarose, as described in Section 2.11.11.1 and stained using SYBR Green. The gel was visualised using a GeneTAC LSIV laser scanner (Genomics Solutions, Huntingdon, Cambs., UK) as described in Section 2.11.11.1.

2.12.1.6. Phosphorylating the EcoRI Ends

The ligase was heat inactivated at 70°C for 30 minutes, pulsed down and cooled at room temperature for 5 minutes. The ligation mix was added to rATP (0.9mM) and T4 polynucleotide kinase (455U/ml) buffered in 1x ligase buffer to phosphorylate the adapter ends. The reaction was incubated for 30 minutes at 37°C after which the kinase was heat inactivated for 30 minutes at 70°C.

2.12.1.7. Digesting with XhoI

The tube contents were pulsed down and equilibrated to room temperature for five minutes, added to *Xho*I (2265U/ml) buffered in *Xho*I buffer supplement and incubated at 37°C for 1.5 hours. To precipitate the cDNA, 5µl 10x STE buffer (0.1 M NaCl, 1 mM EDTA, 50 mM Tris-HCl at pH 6.9) and 125µl of 100% ethanol were added to the tube and incubated overnight at -20°C.

2.12.1.8. cDNA separation by Sepharose gel filtration

The cDNA was separated by Sepharose CL-4B gel filtration using SizeSep™ 400 Spun Columns (GE Healthcare UK Ltd., Bucks., UK) to extract the molecules of >400bp for cloning. The precipitated DNA was centrifuged at 21,500g for 60 minutes at 4°C and the pellet resuspended in 100µl of 1x STE buffer. The Sepharose CL-4B gel was resuspended and washed three times with 1x STE buffer. The column was centrifuged at 400g for 2 minutes and the cDNA sample added slowly to the centre of the flat surface at the top of the compacted bed being careful not to allow the sample to flow down the sides or through any cracks.

The larger size cDNA molecules were collected by centrifugation at 400g for 2 minutes and purified by phenol:chloroform extraction as described in Section 2.17. The cDNA was precipitate with ethanol as described in Section 2.18 and incubated overnight at -20°C.

The large precipitated cDNA molecules were centrifuged at 21,500g for 60 minutes at 4°C and the supernatant centrifuged for a further 30 minutes at the same speed and temperature to ensure all the cDNA was retained. The pellet was washed wit 200µl of 80% (v/v) ethanol without disturbing the pellet, the sample was spun at 15,000g at room temperature for 2 minutes and the pellet air dried by vacuum evaporation for 5 minutes. The library cDNA was resuspended in 4µl sterile HPLC water and 0.5µl removed for a spot assay.

2.12.1.9. cDNA Spot Assay

The DNA marker ϕX174/*Hae*III was diluted to produce a standard curve of 25, 12.5, 6.25, 3.13 and 1.56ng/μl and 1μl of these were pipetted onto a slide and combined with 1μl ethidium bromide (2μg/ml). The cDNA removed for assaying (0.5μl) was diluted 1:4 in HPLC water and 1μl pipetted onto the slide and combined with 1μl ethidium bromide (2μg/ml). The slide was then visualised under UV light and the cDNA quantified from the standard curve.

2.12.1.10. Ligation of library cDNA into pBluescript SK II vector

A ligation was prepared containing 5ng library cDNA, rATP (1mM; pH7.5), pBluescript SK II vector (10ng) and 1U DNA ligase buffered in 1x ligase buffer in a 2.5μl reaction. The ligation was incubated at 4°C for 2 days.

2.12.1.11. Transformation of library plasmids

The ligated library was transformed into XL10-Gold® ultracompetent cells according to the manufacturer's instructions. After incubation at 37°C for 1 hour, 1μl, 10μl and 100μl were spread onto ampicillin plates and incubated for 16 hours at 37°C to determine primary transformation efficiency. The number of transformants was determined using the following equation:

$$\frac{\text{No. colonies on 10μl plate}}{1000} = \text{total cfu (colony forming units)}$$

2.12.2. Amplification

In order to produce a large and stable quantity, the cDNA library was amplified by using the semi-solid method. Using this method the library is amplified in suspension producing three-dimensional uniform growth of colonies and reduces the risk of under-representation of less abundant clones by overgrowth of fast growing clones.

A low melting point agarose (1.35g, SeaPrep) was combined with 2x LB (450ml, 20g NaCl, 20g tryptone, 10g yeast extract, adjusted to pH 7.0 with 5M NaOH) and autoclaved (Section 2.1). The bottle was cooled to 37°C and ampicillin added whilst

stirred. Up to $5x10^5$cfu was added and stirred for a further 3 minutes. The bottle was capped and incubated on ice for 1 hour. The cap was loosened and the bottle incubated at 30°C for 43 hours. The contents were divided into two and spun (10,000g, 20 minutes) at room temperature and each pellet was resuspended in 25ml 2x LB-glycerol (12.5% (v/v)). The library was aliquoted into 1ml volumes and stored at -80°C. With 100µl of the library, serial dilutions were performed in order to estimate the titre.

2.12.3. Screening and analysis

A scraping of glycerol stock was dropped into 15ml LB and between 10 and 100µl spread onto either MacConkey agar no. 3 (Oxoid Ltd., Basingstoke, Hants., UK) or S-Gal™ agar (Sigma Aldrich, Gillingham, Dorset, UK). The plates were incubated for 16 hours at 37°C. Colonies were picked and dropped into a 200µl LB amp containing 8% (v/v) glycerol in 96-well plate wells. The plates were sealed and the clones were grown statically for 16 hours at 37°C.

2.12.3.1. PCR amplification

The plates were vortexed for 3 minutes and 5µl used in a PCR reaction to screen for and quantify insert size. The library PCR mastermix was 10x Triton® free PCR Buffer, 2.5mM $MgCl_2$, 200nM M13F and M13R primers (Appendix A; Table 6), 200µM of dATP, dCTP, dTTP, dGTP and 50U/ml. The thermal cycling programme used is as described in Section 2.11.9.3 with an annealing temperature of 56°C.

2.12.3.2. Library storage

After PCR (Section 2.12.3.1), the remaining library cultures were plates were split over two plates with 145µl in plate A and 50µl in plate B. To both plates, 50µl of 60% (v/v) glycerol was added to each well, the plates covered and stored at -80°C.

2.12.3.3. E-Gel agarose electrophoresis

The PCR reaction was analysed by agarose gel electrophoresis on an E-Gel® (Invitrogen Ltd., Paisley, UK). To each well 10µl dH_2O was added followed by 10µl of PCR product combined with library loading dye (1:1). The samples were loaded in parallel with E-Gel® Low Range molecular marker (Invitrogen Ltd., Paisley, UK). The gel was visualised under UV and successful clones identified.

2.12.3.4. Cherry picking and purification

The gel was edited using E-Editor™ (Invitrogen Ltd., Paisley, UK) and the image analysed using Microsoft Excel. PCR amplicon bands were tagged with the following label, high (4), low (3), double (2) and no band (1). The high bands were exported to a comma delimited file to be used by the MultiPROBE® II HT EX liquid handling system (PerkinElmer LAS (UK) Ltd., Bucks., UK) for cherry picking using the associated WinPrep® software. The composite plate products were purified by using the Montage® Multiscreen $PCR_{\mu 96}$ cleanup plates (Millipore UK Ltd., Watford, Hertfordshire, U.K) by vacuum filtration (10 minutes) and resuspending the DNA in 30µl sterile water which was aggravated on a DPC MicroMix® 5 plate shaker (Diagnostic Products Corporation, Los Angeles, California, USA) for 8 minutes (Form 50, amplitude 9).

2.12.4. BigDye 3 Sequencing

Sequencing mastermix was 1.6µM M13R primer (Appendix A; Table 6) 5x sequencing buffer, BigDye 3 and sterile water in a ratio of 2:5:22.5. Clean PCR product was diluted $^1/_{10}$ with sterile water and 2µl combined with 3µl mastermix before PCR amplification using the programme described in 2.11.12. The sequencing reaction was purified using the Montage® SEQ_{96} Sequencing Reaction Cleanup Kit (Millipore UK Ltd., Watford, Herts., UK). The reaction was combined with injection solution (20µl) and loaded onto the plate followed by vacuum filtration, washing the membrane (injection solution, 25µl) and resuspension of the sequencing product (injection solution, 15µl). HiDye formamide (15µl) was added to the samples which were analysed by the ABI PRISM® 3100 DNA Sequencer run by the Cardiff University Molecular Biology Support Unit (MBSU).

2.12.4.1. Sequence analysis

The sequences were all named according to the NERC environmental genomics scheme to allow for bioinformatics analysis. The raw trace chromatograms from the sequencing reaction were processed using trace2dbEST (Parkinson *et al.* 2004) which contains a base calling component (phred) and a sequence trimming component (cross_match). This software produces good quality EST sequences formatted for submitting to NCBI dbEST (http://www.ncbi.nlm.nih.gov/dbEST).

The EST sequences were clustered using CLOBB (Parkinson *et al.* 2002) and a consensus putative gene sequence contig derived by phrap. The EST sequences were then processed by the software package PartiGene (Parkinson *et al.* 2004) found at http://www.nematodes.org/PartiGene.

The biological process, cellular component and molecular function of the gene sequences were described where possible, by defining their Gene Ontology (GO) classification using GOblet (http://goblet.molgen.mpg.de) (Groth *et al.* 2004), AmiGO (http://www.godatabase.org) (Ashburner *et al.* 2000) or blast2GO (http://www.blast2go.de/) (Conesa *et al.* 2005).

The first 500 clones were sequenced by the Cardiff University MBSU and the total 13,440 clones were sequenced by the University of Edinburgh School of Biological Sciences Sequencing Service (SBSSS).

2.13. cDNA ARRAY METHODS

2.13.1. Transfer of 96 well composite plates to 384 well plates

The 96-well composite plates containing the purified PCR products were transferred (5µl) to 384 well plates containing DMSO (5µl) by the MultiPROBE® II HT EX liquid handling system (PerkinElmer LAS (UK) Ltd., Bucks., UK) using the associated WinPrep® software. This transfer was required for subsequent array printing on the SpotArray72 Microarray Printing System (PerkinElmer LAS (UK) Ltd., Bucks., UK) or the Flexys Robotic Workstation (Genomic Solutions Ltd., Cambs., UK).

2.13.2. Redundancy screen

2.13.2.1. Direct labelling

PCR products representing the highest abundance clusters were labelled using the CyDye™ fluorescently labelled nucleotide Cy5-dUTP (GE Healthcare UK Ltd., Bucks., UK). A PCR mix was created containing 0.3mM dCTP, dGTP and dATP, 0.18mM dTTP, 0.5µM Libn_for and Libn_rev primer, 2.3mM $MgCl_2$, 50U/ml *Taq*, 1µl Cy5-dUTP and ~20nmols of each PCR product buffered in $^1/_{10}$ of 10x Triton® free buffer. Up to five PCR products were labelled in each reaction tube. The DNA was amplified and labelled using the PCR conditions shown below:

	95°C	5 minutes	(DNA denatures)
33 x	95°C	1 minute	(DNA denatures)
	Tm	1 minute	(primer anneals)
	72°C	1 minute	(DNA elongation)
	72°C	10 minutes	(DNA elongation)
	4°C	∞	(Soak)

The labelled PCR products were purified to remove nucleotides and unincorporated dye using a GFX column (GE Healthcare UK Ltd., Bucks., UK) as described in the manufacturer's instructions. Successful labelling was determined by agarose minigel electrophoresis as described in Section 2.11.11.1 except that the samples were loaded with glycerol (50%, $^{v}/_{v}$) in a 1:1 ratio and no staining of the gel after electrophoresis was necessary. The gel was visualised using a Genomic Solutions® GeneTAC LS IV laser scanner at 675nm, set to Cy5 (Genomic Solutions Ltd., Huntingdon, UK).

2.13.2.2. Low density microarray printing (redundancy screen)

Products of seven composite 384-well plates and the unlabelled redundancy PCR products were printed onto UltraGAPS™ coated slides (Corning B.B., The Netherlands) using the Flexys Robotic Workstation (Genomic Solutions Ltd., Cambs., UK) by the Cardiff University Molecular Biology Support Unit (MBSU).

2.13.2.3. Blocking redundancy slides

Blocking buffer (1% ($^{w}/_{v}$) BSA, 0.1% ($^{w}/_{v}$) SDS, 5x SSC Buffer) was warmed to 55°C in a Coplin Jar; the slides were added to the solution and incubated for 45 minutes. The slides were then washed twice in distilled water and once in isopropanol before being rapidly dried with compressed air.

2.13.2.4. Hybridisation of labelled probe to redundancy microarray

For each slide 5μM of each primer (Libn_for and Libn_rev) and labelled probes (20%, $^{v}/_{v}$ representing 20 labelled probes) were combined in distilled water and denatured for 3 minutes at 95°C. They were then added to 2x hybridisation buffer (50% ($^{v}/_{v}$) formamide, 0.2% ($^{w}/_{v}$) SDS, 10x SSC Buffer) in a 1:1 ratio.

The hybridisation mixture was added to the microarray slide which was overlaid with a blank microscope slide. The slides were placed on a rack in an airtight box containing water to ensure a high level of humidity, although the slides did not come into contact with the water at any time. They were incubated at 55°C and hybridised for 18 hours.

2.13.2.5. Washing redundancy slides after hybridisation

The unbound label was removed from the slides by washing in Wash Solution A (1x SSC, 0.2% (w/v) SDS) at 55°C for 10 minutes, then in Wash Solution B (0.1x SSC, 0.2% (w/v) SDS) at 55°C for 10 minutes, and again in fresh Wash Solution B at 55°C for a further 10 minutes. The slides were rinsed in 0.1x SSC at room temperature and dried quickly using a compressed air line. The slide was stored in the dark at room temperature until ready to scan.

2.13.2.6. Redundancy microarray scanning and analysis

The slide was scanned using a ScanArray™ Express HT scanner (PerkinElmer LAS (UK) Ltd., Bucks., UK) at a resolution of 10μm. The image was analysed using the software package ImaGene™ Version 5 (BioDiscovery Inc., USA).

2.13.3. Microarray printing

Before every print run the printing room was cleaned. The water and humidity tank was soaked in Vircon overnight, the pins removed and sonicated in TeleChem Micro Cleaning Solution (1x, TeleChem International, Inc., CA, U.S.A.) using the manufacturer's directions. The pins were left at room temperature for at least 2 hours after being air dried to ensure they were completely dry to prevent pins sticking to the print head. The floors of the printing room were cleaned with Vircon and the surfaces with isopropanol wipes. The bed of the printer was vacuum cleaned and a fresh sticky mat placed outside the clean room door. Shoe covers and lab coats were worn at all times.

Each split pin was checked using a light microscope to check the capillary gap was clean and free from blockages. Any blockages were removed using a small piece of foil. The pin wash base was cleaned with a dental brush, dipped in isopropanol, pushed into each pin well until the brush was clean. Vircon was washed through the system to clean all the pipes and left overnight. After an overnight soak the Vircon solution was removed from the water and humidity tank and rinsed thoroughly with distilled autoclaved water until all the foam was removed. Distilled autoclaved water was then cycled through the machine using the wash/dry pins procedure whilst the

print head was empty, until there was no Vircon present in the pipes. Pins were returned to the print head carefully making sure all of them were aligned in the same direction. The humidity tank was filled to the recommended volume with distilled autoclaved water. The pins were then washed with distilled autoclaved water three times. The printer was now ready to be used.

Slides were removed from the vacuum pack only on the day of printing and handled with nitrile gloves. The bottom left of the slides were etched with GP (for *G. pulex*), and the print run number followed by the slide number, for example on the first print run the slides were labelled GP1/1 to GP1/50. The Nexterion Slide E microarrays are coated on both slides, but the face of the slide pointing towards the centre was the side chosen for printing. Any glass particles produced during etching were removed with compressed air. Slides were placed on the bed of a SpotArray72 Microrray Printing System (PerkinElmer LAS (UK) Ltd., Bucks., UK) as shown in Figure 2.5.

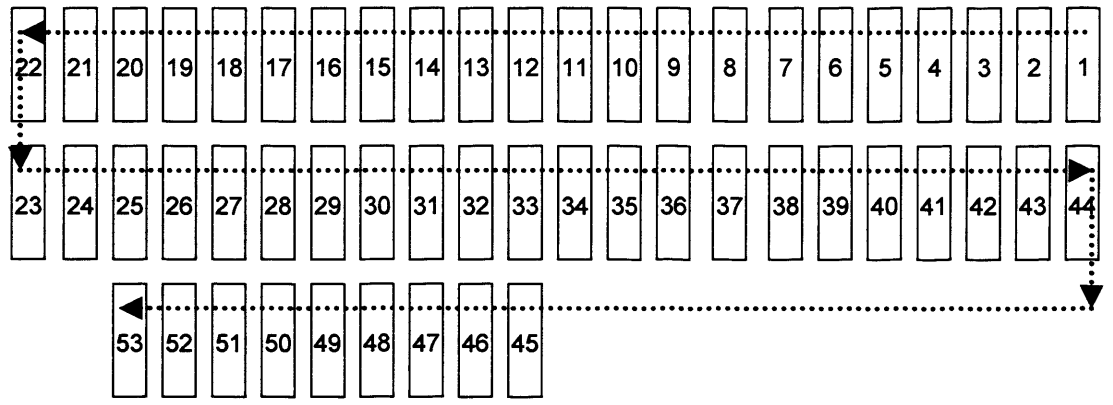


Figure 2.5: Print bed layout showing slide order and position

Dotted lines show the movement of the print head in all the print runs. For the first 2 print runs, the top row of slides was reversed starting with slide 22. For print run 1 the bottom row started tight to the left and ran from slide 45 – 50 from left to right. In print run 2, the bottom row slide layout was the same, but ran from 45- 50. In the third and final print run, the slide layout was exactly as above and the first 3 slides were old Nexterion E test slides and were technically used as extra blotting slides as the maximum number of blots available on the slides in the standard blotting positions were at a maximum.

The pins were always acclimatised to the relevant buffers by printing onto slides using the chosen buffer and repeating until the spots produced were of good morphology and size. Pins were pre-blotted 10 times to remove residual spotting solution and each wash cycle consisted of a wash for 3 seconds followed by a 3 second dry. This was repeated 3 times. Blotting slides were turned over or replaced as required and washed with distilled autoclaved water at 65°C between uses.

Purified PCR products in 384-well plates were diluted by ½ in Schott Nexterion Spot Modified Buffer (SCHOTT Jenaer Glas GmbH, Germany) under a laminar flow hood (InterMed, Microflow Pathfinder Ltd, Fleet, Hampshire, UK) to reduce the risk of contamination. The plates were stored at 4°C until they were required for printing. The microarrays were printed on Schott Nexterion Slide E (SCHOTT Jenaer Glas GmbH, Germany) using a SpotArray72 Microrray Printing System (PerkinElmer LAS (UK) Ltd., Bucks., UK) under a sterile environment. Each PCR product was printed once and the Landmark/Vtg/Lucidea Microarray ScoreCard (GE Healthcare UK Ltd., Bucks., UK) plate (Figure 2.6) was printed 5 times producing slides containing 13,728 spots in a 4 x 12 metagrid, with each subgrid in a 16 x 18 format containing 286 spots.

After printing, the slides were humidified to 75% on the printer bed for 30 minutes followed by baking at 120°C for 30 minutes to immobilise the DNA onto the slide. After the slides had returned to room temperature, they were reflective scanned using the ScanArray™ Express HT scanner (PerkinElmer LAS (UK) Ltd., Bucks., UK) set on 'RedReflect' (excitation/emission: 633nm) to determine the correct spotting onto the slides. Once scanned the slides were stored in the dark inside a desiccator until required for blocking and hybridisation.

2.13.3.1. Carryover

Before the start of each print run, a carryover test was performed to ensure cleanliness of pins and to ensure none of them were blocked. For this test the landmark/scorecard plate was used (Figure 2.6).

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
A	LM	LM	LM	LM	LM	LM	LM	LM	LM	LM	LM	LM	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	N1	N2
B	LM	LM	LM	LM	LM	LM	LM	LM	LM	LM	LM	LM	R1	R2	R3	R4	R5	R6	R7	R8	Vtg1	Vtg2	U1	U2
C	LM	LM	LM	LM	LM	LM	LM	LM	LM	LM	LM	LM	N1	N2	C10	C9	C8	C7	C6	C5	C4	C3	C2	C1
D	LM	LM	LM	LM	LM	LM	LM	LM	LM	LM	LM	LM	U1	U2	U3	U3	R8	R7	R6	R5	R4	R3	R2	R1
E	Buf	Buf	Buf	Buf	Buf	Buf	Buf	Buf	Buf	Buf	Buf	Buf												
F	Buf	Buf	Buf	Buf	Buf	Buf	Buf	Buf	Buf	Buf	Buf	Buf												
G	Buf	Buf	Buf	Buf	Buf	Buf	Buf	Buf	Buf	Buf	Buf	Buf												
H	Buf	Buf	Buf	Buf	Buf	Buf	Buf	Buf	Buf	Buf	Buf	Buf												
I																								
J																								
K																								
L																								
M																								
N																								
O																								
P																								

Figure 2.6: Landmark/Scorecard plate layout

LM = landmark; Buf = Nexterion Modified Spotting Solution (1x); C = calibration control; N = negative control; U = utility control; R = ratio control; Vtg = vitellogenin. More details about the Lucidea Microarray ScoreCard are available from GE Healthcare UK Ltd., Bucks., UK.

All 48 pins were used to print the landmark 10 times, followed by a wash and dry procedure, then the pins picked up 1x spotting buffer and a further 10 spots were printed. These spots were printed with the maximum spot spacing allowed by the printer to prevent dragging of the spots into each other as had been previously seen by the duplication of spots printed with NXT spotting solution. If any carryover or blockages were evident then the relevant pin was cleaned, the wells affected were replaced with fresh solutions and the carryover test repeated until all the pins were printing cleanly.

2.13.4. Sample collection for microarrays

RNA*later*®-ICE was chilled at -80°C for 30 minutes. *G. pulex* were sacrificed in liquid nitrogen and then submerged in ~10x volume chilled RNA*later*®-ICE per individual sample. The frozen sample was transferred from -20°C to thaw overnight. This procedure allows the instant sacrifice of the amphipods and the RNA*later*®-ICE softens the cuticle to allow effective homogenisation.

2.13.4.1. Naming System

A naming system was developed for each sample so that all relevant information was included in the sample name (Figure 2.7).

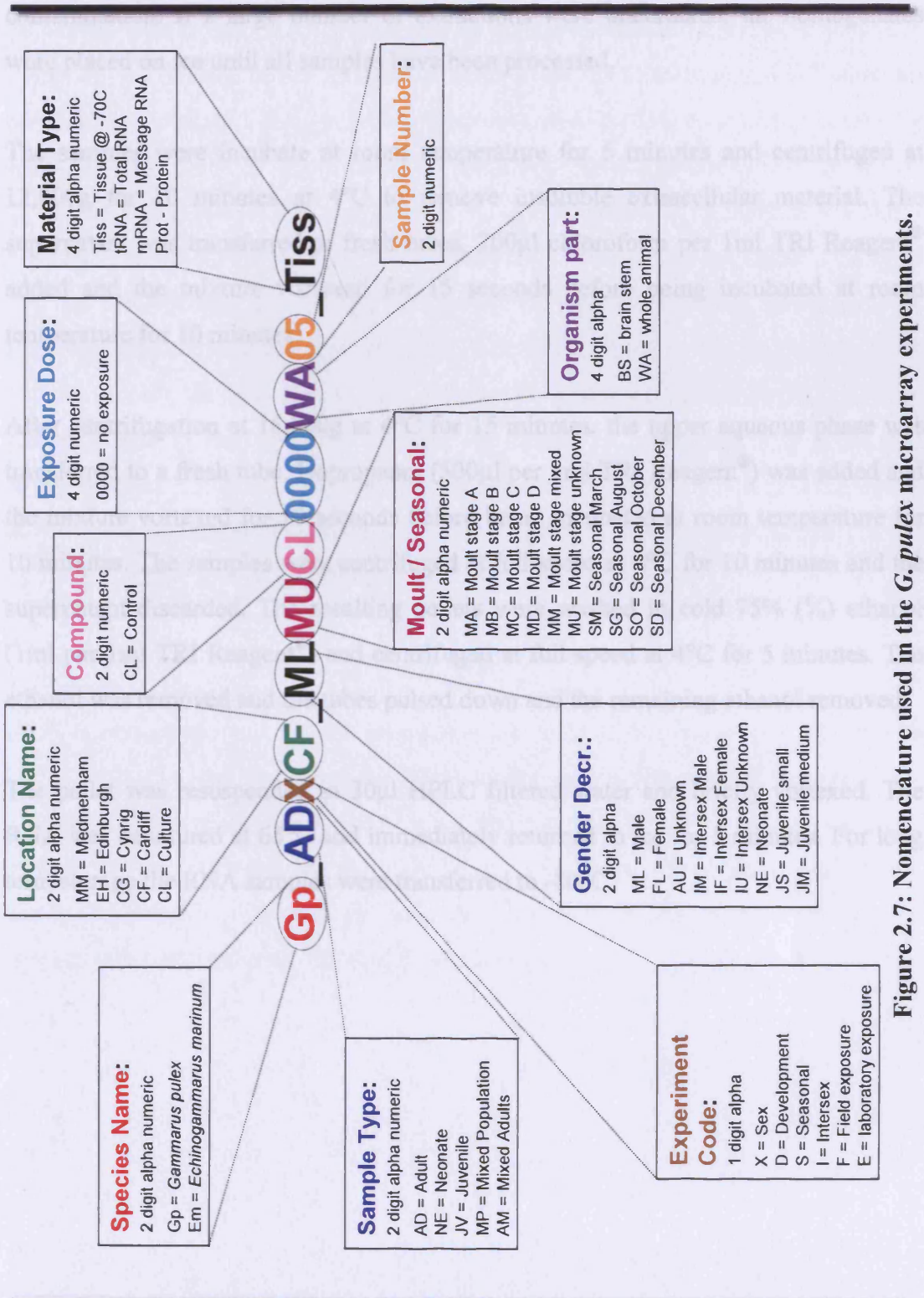


Figure 2.7: Nomenclature used in the *G. pulex* microarray experiments.

2.13.5. RNA extraction for microarrays

Animals were transferred to ~10x volume TRI Reagent® and homogenised with a T8, 8mm stator dispersing element (Ultra-Turrax®, IKA® Werke GmbH & Co. KG, Staufen, Germany); using less Tri Reagent increases the likelihood of genomic DNA contamination. If a large number of extractions were undertaken, the homogenates were placed on ice until all samples have been processed.

The samples were incubate at room temperature for 5 minutes and centrifuged at 12,000g for 10 minutes at 4°C to remove insoluble extracellular material. The supernatant was transferred to fresh tubes, 200µl chloroform per 1ml TRI Reagent® added and the mixture vortexed for 15 seconds before being incubated at room temperature for 10 minutes.

After centrifugation at 16,000g at 4°C for 15 minutes, the upper aqueous phase was transferred to a fresh tube. Isopropanol (500µl per 1ml TRI Reagent®) was added and the mixture vortexed for 10 seconds before being incubated at room temperature for 10 minutes. The samples were centrifuged at full speed at 4°C for 10 minutes and the supernatant discarded. The resulting pellets were washed in cold 75% (v/v) ethanol (1ml per 1ml TRI Reagent®) and centrifuged at full speed at 4°C for 5 minutes. The ethanol was removed and the tubes pulsed down and the remaining ethanol removed.

The pellet was resuspended in 30µl HPLC filtered water and briefly vortexed. The RNA was denatured at 65°C and immediately returned to ice for 5 minutes. For long term storage the RNA samples were transferred to -80°C.

2.13.6. Purification of RNA for microarrays

2.13.6.1. Optional LiCl precipitation

LiCl precipitation solution (3x, 7.5M LiCl, 50mM EDTA) was added in a 1:3 ratio (LiCl:RNA) to the RNA sample so the resultant RNA concentration in solution is >0.4μg/μl with a final LiCl concentration of 2.5M. The contents were vortexed, pulsed to the bottom of the tube and incubated at -20°C overnight. LiCl does not efficiently precipitate DNA, proteins or carbohydrates and is the method of choice for removing inhibitors of downstream applications including cDNA synthesis and CyDye labelling. The RNA was pelleted by centrifugation at 16,000g at 4°C for 1 hour and the pellet washed with 75% ($^{v}/_{v}$) ethanol. After centrifugation at 16,000g at 4°C for 15 minutes the ethanol was removed, the tubes pulsed down and the remaining ethanol removed. For this stage there was no need to dry the pelleted RNA out completely, as any minor residual ethanol would not greatly affect RNeasy column clean-up and very dry samples are harder to resuspend. The pellet was resuspended in HPLC filtered water (88.5μl) and briefly vortexed before denaturation at 65°C and an immediate return to ice for 5 minutes. RNeasy column purification (QIAGEN, Crawley, UK) is not required.

2.13.6.2. Genomic contamination

If LiCl precipitation was not performed, the total volume was made up to 88.5μl with HPLC filtered water. To assess the RNA quality and to determine if there was any genomic DNA contamination, 1μl of RNA was run on a 1.5% ($^{w}/_{v}$) agarose gel using equipment that had been soaked with 1% ($^{w}/_{v}$) SDS and rinsed with sterile water Sterile TAE buffer was used for both the tank and gel. An example of samples contaminated with genomic DNA in shown in Figure 2.8.

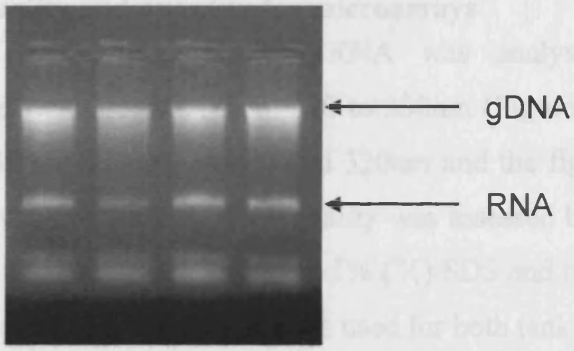


Figure 2.8: Genomic DNA contamination of RNA samples

If there was genomic contamination, it was removed by using QIAGEN RNase free DNase kit. If this procedure was used, all sample s involved in the microarray experiment were treated in the same way. DNase I was resuspended in the RNase-free water supplied (550µl) by transferring the RNase-free water to a sterile microfuge tube and using a sterile syringe to pick up the correct volume. The metal tab on the DNase I vial was pulled back and the rubber seal pierced. The vial was carefully inverted to dissolve the powder. The following components were combined by inversion: RNA solution (87.5µl contaminated with genomic DNA), Buffer RDD (10µl, supplied by the manufacturer) and DNase I stock solution (2.5µl, supplied by the manufacturer. The mixture was incubated at room temperature for 10 minutes and cleaned by RNeasy column purification (QIAGEN, Crawley, UK).

2.13.6.3. RNeasy column purification of RNA for microarrays

If no genomic contamination was evident, RNase-free water (12.5µl, QIAGEN supplied) was added to each sample to make up to 100µl and cleaned by RNeasy column purification (QIAGEN, Crawley, UK). The RNA requires further purification to remove proteins. Therefore RNA was thawed (65°C for 30 seconds until thawed and put straight back on ice) and purified using a QIAGEN RNeasy kit according to the manufacturer's instructions. The RNA was eluted from the column by adding RNase free water (30µl) onto the membrane surface and centrifuging for 1 minute at 8000g. To increase the concentration, the RNA was eluted a second time using the eluate from the first elution. This time the column was centrifuged for 1 minute at 16,000g.

2.13.7. Assessment of RNA quality and quantity for microarrays

To quantify the RNA, a 1:50 dilution of the RNA was analysed on a spectrophotometer. The samples were scanned from 230 to 330nm (Figure 2.9). The absorbance was recorded at 230nm, 260nm, 280nm and 320nm and the figures. The RNA concentration and ratios were calculated. RNA quality was assessed by running 1µl on a 1.5% ($^{w}/_{v}$) agarose gel using tanks are soaked in 1% ($^{w}/_{v}$) SDS and rinsed with sterile water. Sterile (double autoclaved) TAE buffer was used for both tank and gel.

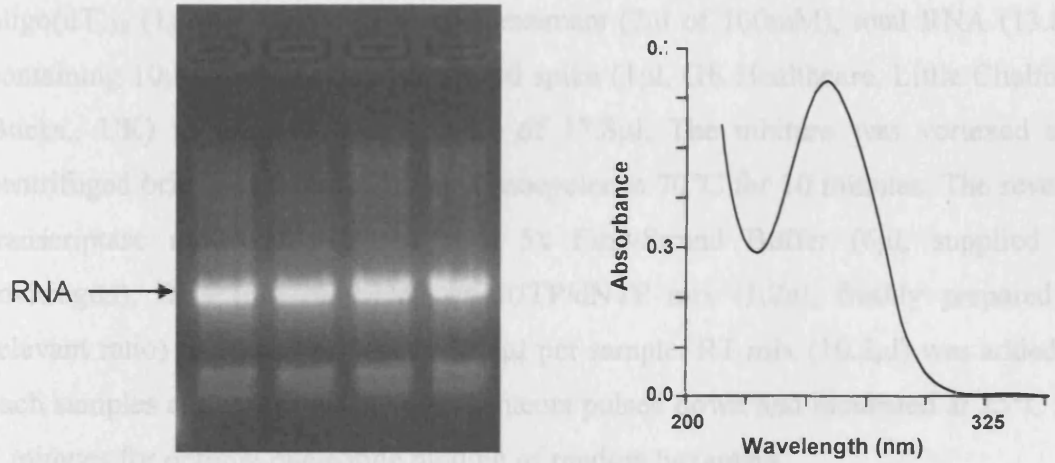


Figure 2.9: RNA quality and quantity assessment

After ascertaining that both $^{260}/_{230}$ and $^{260}/_{280}$ ratios are close to or above 1.8 and the quantity of RNA required for 10µg was calculated. This was less than 13.8µl which is >0.5µg/ml. If not, the RNA was concentrated by vacuum centrifugation at 65°C.

2.13.8. cDNA labelling

Before starting, the aa-dUTP:dTTP ratio was determined for the individual species used as this can vary depending on GC content. These 7 ratios were tested (**U:T:ACG**) in the following way:

+ve control	1	2	3	4	5	6	7	-ve control
0:5:5	1:4:5	3:7:10	2:3:5	5:5:10	3:2:5	7:3:5	4:1:5	5:0:5

In PCR 200µl thin walled tube the following components were mixed, anchored oligo(dT)$_{18}$ (1µl of 100mM), random hexamers (2µl of 100mM), total RNA (13.8µl containing 10µg) and Lucidea ScoreCard spike (1µl, GE Healthcare, Little Chalfont, Bucks., UK) to make a total volume of 17.8µl. The mixture was vortexed and centrifuged briefly and heated in a thermocycler to 70°C for 10 minutes. The reverse transcriptase mix was prepared with 5x First-Strand Buffer (6µl, supplied by Invitrogen), DTT (3µl of 0.1M), aa-dUTP/dNTP mix (1.2µl, freshly prepared at relevant ratio) to a total volume of 10.2µl per sample. RT mix (10.2µl) was added to each samples and vortexed; the tube contents pulsed down and incubated at 25°C for 2 minutes for optimal nucleotide binding of random hexamers.

Superscript II RT (2µl) was added to each reaction tube to give a total volume of 30µl. The contents were gently mixed by flicking the tube and pulsing down the contents. After an incubation in the thermocycler at 42°C for 3 hours NaOH (10µl of 1M) and EDTA (10µl of 0.5M, pH 8.0) were added and briefly vortexed and centrifuged to give a total volume of 50µl. The EDTA compromises the activity of the RT and the following incubation completely inactivates the enzyme. The samples were heated to 65°C and incubated for 15 minutes. To neutralise the solution, HEPES buffer (25µl of 1M, pH 7.0) was added, and the contents briefly vortexed and centrifuged giving a total volume of 75µl.

2.13.8.1. cDNA purification for microarrays - QIAGEN column cleanup method

This method is a modification of the QIAquick PCR purification kit (QIAGEN, Crawley, UK) The phosphate and wash buffers are substituted for the buffers supplied by the manufacturer because they contain free amines which compete with the CyDye coupling reaction. Buffer PB (375µl, 5x, supplied by the manufacturer) was added to the cDNA and transferred to a QIAquick column. The column was placed in the supplied collection tube and centrifuged at ~13,000g for 1 minute. The collection tube was emptied and phosphate wash buffer (750µl) added to the column and centrifuged at 13,000g for 1 minute. The collection tube was emptied and the wash and centrifugation step repeated. The collection tube was emptied again and the column centrifuged for 1 minute at 16,000g. The column was transferred to a new 1.5ml microfuge tube and RNase-free water (30µl) carefully added to the centre of the column membrane. The cDNA was incubated for 1 minute at room temperature and eluted by centrifugation at ~13,000g for 1 minute. A second elution was performed into the same tube by repeating the elution with fresh RNase-free water to give a final elution volume of 60µl.

The samples were dried down in the SpeedVac (Eppendorf UK Ltd., Cambridge, UK) to a volume of 3µl. If all the liquid had evaporated, the pellets were resuspended in 3µl of RNase-free water, this was difficult and resuspension was assisted by heat shocking at 65°C for 1 minute and then returning the samples to ice for 5 minutes. Complete resuspension was ensured before continuing as it is essential to get good labelling efficiency.

2.13.8.2. Coupling

Samples were adjusted to 3µl with RNase-free water and briefly vortexed and centrifuged. Sodium bicarbonate (1.5µl, 0.3M, pH9.0, $NaHCO_3$) was added to give a total volume of 4.5µl and the sample vortexed followed by a brief centrifugation.

The CyDye pack was warmed to room temperature after which dimethyl sulfoxide (DMSO) was added (38µl). This was done immediately before adding it to the cDNA. The CyDye was fully resuspended by vortexing combined with brief centrifugation to make sure no dye remained in the lid of the tube. CyDye (4.5µl) was added to the

cDNA resuspension to give a total volume of 9µl. The mixture was briefly vortexed and centrifuged to gather all reactants to the bottom of the tube, this was followed by an overnight incubation in the dark.

2.13.8.3. Label purification

To each reaction, NaOAc (35µl of 100mM, pH5.2) was added followed by Buffer PB (250µl, 5x, supplied by the manufacturer). A QIAquick PCR spin column was placed in a collection tube (supplied by the manufacturer) and the sample added to the column and centrifuged at 13,000g for 1 minute. The collection tube was emptied and the column washed by the addition of Buffer PE (750µl, supplied by the manufacturer) which was subsequently centrifuged at 13,000g for 1 minute. The collection tube was emptied and the column centrifuged for an additional 1 minute at 16,000g. The column was transferred to a fresh microfuge tube, and then Buffer EB (30µl) was added to the top of the column followed by and incubation at room temperature for 1 minute. After centrifugation at 13,000g for 1 minute, the eluant was retained. A second elution was performed to give a total volume of 60µl.

2.13.8.4. Assessment of label quantity and quality

After purification, labelled cDNA was assessed for quantity, purity and incorporation efficiency, using stringent specifications to assess its suitability for hybridisation. Labelled cDNA quantification was performed on a Jenway 6505 spectrophotometer using the spectral scan and recording the absorbance at 230nm, 260nm, 280nm, 320nm, 550nm and 650nm. Quality was assessed using minigel electrophoresis; the resulting gel was scanned on a GeneTAC™ LS IV scanner (560nm).

A "minigel" was prepared by pouring 1.5% ($^{w}/_{v}$) agarose into the mould and then once set turned out onto a microscope slide. The gel was placed in an electrophoresis tank and covered with 1x TAE (all tanks were rinsed thoroughly first on order to remove all traces of ethidium bromide). Labelled cDNA (1µl) was combined with glycerol (1µl, 50% ($^{v}/_{v}$)) and loaded into the well of the gel which was run for 30 minutes at 100V. The excess liquid was drained from the gel and scanned in a LSIV GeneTac carousel scanner (Genomic Solutions, Huntingdon, Cambs., UK) to assess the size distribution of the labelled product. At this stage the label was assessed as good if it

has a large size distribution through the gel and was not degraded and didn't contain unincorporated product which appears as a smear at the base of the gel (Figure 2.10). Additionally the amount of protein contamination could also be assessed as this appears stuck in the well. If the entire label was protein and not a cDNA distribution then the sample should be abandoned and the RNA further purified before attempting to label again.

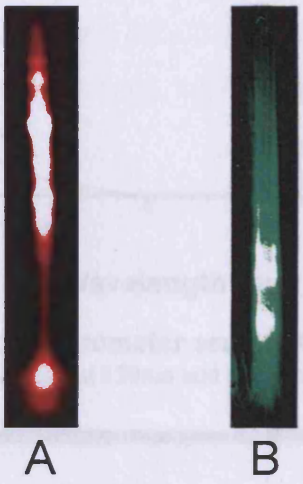


Figure 2.10: Minigel electrophoresis of labelled cDNA
A: The cDNA is labelled with Cy5 and contains unincorporated dye as shown by the blob at the base of the gel. This needs to be re-run through a QIAGEN purification column. B: The cDNA is labelled with Cy3 and is good enough to hybridise showing large size distribution.

Spectrophotometer readings were taken of the entire 60µl sample (no dilution) to provide an indication of the amount of cDNA present (260nm) and the amount of Cy3 (550nm) present. The samples were scanned between 220nm and 650nm to determine the cDNA concentration and the dye incorporation using the Jenway 6505 spectrophotometer (Barloworld Scientific Ltd., T/As Jenway, Essex, UK). A graph showing the spectra of Cy3-labelled cDNA can be seen in Figure 2.11. The frequency of incorporation was calculated which should ideally have been between 20 and 50. Further calculation were performed to determine the quantity of sample required to add to the hybridisation for >50pmols of label.

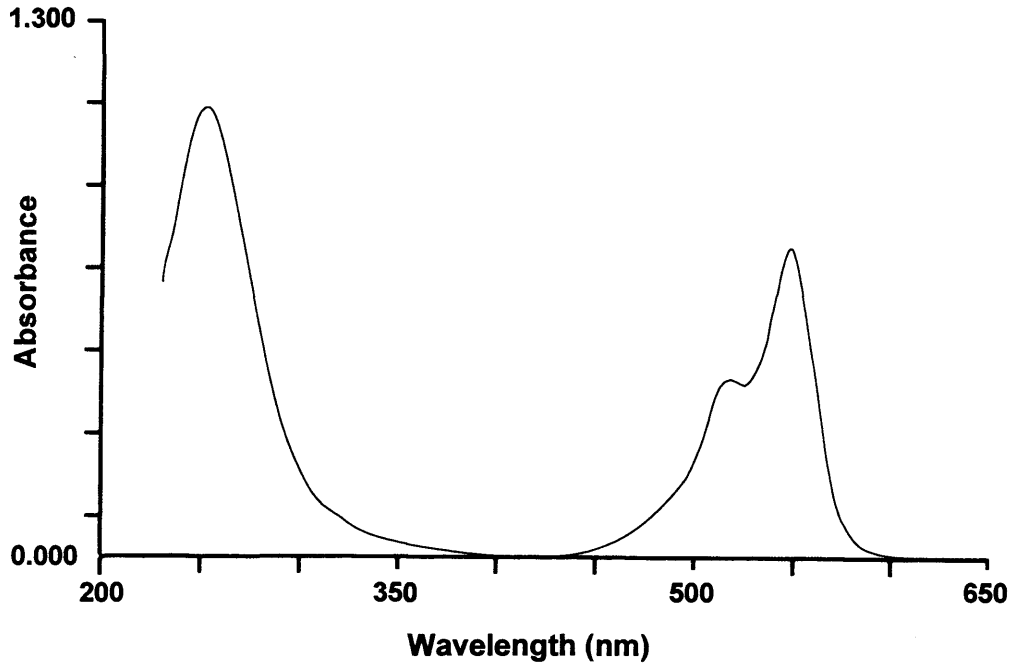


Figure 2.11: Spectrophotometer scan of Cy3 labelled cDNA
The scan shows a Cy3 peak at 550nm and a nucleic acid peak at 260nm

2.13.9. Microarray hybridisation

2.13.9.1. Preparing probes for hybridisation of the microarray

All samples were labelled with Cy3 and combined with Cy5 labelled oligonucleotide mixture (pGEM-T & pBluescript mix). The appropriate amount of label was added to a fresh microfuge tube and dried down in the SpeedVac at 65°C in the dark. Samples were dried down to a volume of 29.5µl or less and made up to 29.5µl with HPLC grade water if required. A blocker, polyA (2µl of 100µM) and reference oligonucleotide mixture (3.5µl = 3.5pmols) were added to the probe to give a total volume of 35µl. The contents were pulsed to the bottom of the tube.

2.13.9.2. Preparing substrates for hybridisation of the microarray

Pre-hybridisation washing is desirable as any unbound DNA is removed, which may compete with the bound DNA during hybridisation to the probe molecules. Initial washing of slides was as described by the manufacturer, using a combination of 3 wash buffers and water to thoroughly wash away any salts and prepare the slide surface for hybridisation. The 3 wash buffers were Triton® X-100, HCl and KCl containing solutions. Blocking was performed using 2 buffers; a BSA based buffer and the proprietary blocking buffer, Nexterion Block E (SCHOTT Jenaer Glas GmbH, Jena, Germany). This 2-step block seemed to provide successful slide blocking and give the least background after hybridisation.

2.13.9.2.1. Schott Nexterion Slide E

Whilst the samples were drying, the slides were washed to remove unbound probe molecules and buffer substances to avoid interference with subsequent hybridisation experiments. The slides were rinsed at room temperature, once in Triton®X-100 (5 minutes, 0.1% (%v)), twice in HCl (2 x 2 minutes, 1mM), once in KCl (10 minutes, 10mM) and finally once in deionised water (1 minute). This was followed immediately by blocking; blocking buffer was prepared freshly each time. The slides were incubated slides in Coplin jars (3 slides maximum per jar) or 200 ml glass troughs (6 slides maximum), firstly in Nexterion Block E (1x, 15 minutes, 50°C), rinsed in deionised water (1 minute) and then in a BSA-based buffer (5x SSC, 0.1% (%w) SDS, 1% (%w) BSA) for 30 minutes at 42°C. The slides were finally washed by dipping in four changes of filtered sterile water (4 x 1 minute). If white streaks or marks were seen on the slide then the water washes were repeated. The slides were quickly dried using compressed air and kept in the dark until required for hybridisation (less than 1 hour).

2.13.9.2.2. UltraGAPS™ slides

Slides were incubated in a freshly made BSA-based blocking buffer (5x SSC, 0.1% (%w) SDS, 1% (%w) BSA) in Coplin jars (3 slides maximum per jar) or a 200ml glass trough (6 slides maximum) for 45 minutes at 42°C. They were then rinsed by dipping in four changes of filtered sterile water (4 x 1 minute) and dipped in

isopropanol. The slides were dried quickly using compressed air and kept in the dark until required for hybridisation (less than 1 hour).

2.13.9.3. Hybridisation of the microarray

The hybridisation buffer (50% (v/v) formamide, 10x SSC, 0.2% (w/v) SDS) was prepared freshly each time and warmed at 37°C until the SDS was in solution. The labelled probes (29.5µl) were heat denatured in a heat block at 95°C for 3 minutes and centrifuged for 1 minute to remove condensation from the lid. Hybridisation buffer (35µl) was added to the labelled probe and mixed by pipetting up and down briefly. A LifterSlip was placed on soft tissue on the bench and the full 70µl of the sample pipetted in a line down the longest side. The blocked microarray was placed facing down (side with etching or label downwards) onto the slide so that the hybridisation mixture was slowly drawn across the slide. The slide sandwich was turned over so that the microarray slide was on the bottom and the LifterSlip (Erie Scientific 24x60I-2-4733, VWR, Poole, UK) on the top. They were placed in a humidity chamber at 42°C for 24 hours (or 37°C for 48 hours for cross-species hybridisations). For *G. pulex* microarray slides, 24 hour hybridisations were undertaken, for cross-species *Echinogammarus marinus*, 48 hour hybridisations were used due to the differences in gene sequences between the species; longer time gives the probe more time to hybridise to the target spots. For *G. pulex* microarray slides, a temperature of 42°C was used, for cross-species *E. marinus*; the temperature was decreased to 37°C. This change in protocol for *E. marinus* hybridisations reduced the specificity of the probe:target binding to allow for differences in gene sequences. There was always some water present in the bottom of the humidity chamber and the platform for the slides was completely dry before the slides were placed inside.

2.13.9.3.1. Washing the probe from the microarray

The slides were separated in Wash Buffer 1 (1x SSC, 0.2% (w/v) SDS) at room temperature before being incubated in a fresh aliquot of Wash Buffer 1 in Coplin jars (3 slides maximum per jar) or 200ml glass troughs (6 slides maximum) at 55°C for 10 minutes. The slides were agitated 5 times and transferred to Wash Buffer 2 (0.1x SSC, 0.1% (w/v) SDS) at 55°C for 10 minutes. After agitation, the slides were transferred to

fresh at 55°C for 10 minutes; this was repeated once more to give a total of 3 washes in Wash Buffer 2. After further agitation the slides were transferred to Wash Buffer 3 (0.1x SSC) 1 minute at room temperature and this step was then repeated. The slides were dried quickly using compressed air and stored in the dark until ready for scanning.

2.13.10. Scanning

Slides were scanned using a ScanArray™ Express HT scanner (PerkinElmer LAS (UK) Ltd., Bucks., UK) using a fixed PMT and laser power at 66% for the Cy5 oligonucleotide channel and 70% for the Cy3 sample channel. The resolution was set to 5μm and saved as a tiff file for subsequent image analysis. A maximum of three slides were scanned in each batch to prevent stray light from the laser bleaching the slides held in the rack.

2.13.11. Image analysis

Images were analysed using SPOT (Version 2.0, CSIRO Mathematical and Information Sciences, Australia (Yang *et al.* 2001; Yang *et al.* 2002)) using the default settings.

2.13.12. Statistics

SPOT data files were merged with a *G. pulex* GAL (GenePix Array List) file specific to the *G. pulex* arrays. The GAL file provides information on the name and location of each printed gene. By merging this file with SPOT files, quantitative information of each spot was linked to the gene name and location on the slide. The merged data files were imported into GeneSpring, excluding the control and blank data (Version GX 7.3; Agilent Technologies UK Ltd., Cheshire, UK) and per spot, per gene and per chip normalisation performed. Data was normalised per spot by dividing the test channel (Cy3) by the control channel (Cy5); data was normalised per gene to median with a raw data cutoff of 100; between chip variation was normalised to the 50th percentile with a raw data cutoff of 100. Slide data was compared using box plots and those with poor distribution removed; this was performed as a screening process for bad slides and the samples set re-normalised as described above. Raw data (log ratio) was

filtered between 100 and the maximum in the number of conditions that equalled the minimum sample size for each group within the experiment. Using the screened raw data, genes showing 2-fold changes (up and down) by analysis of normalised data were isolated and probabilities assessed by statistics suited to the experiment. Principal component analysis (PCA) was performed on the statistically valid data. This allows visualisation of a data set within a coordinate system, axis are based on variance, selecting the characteristics that contribute the most variance. Dendrograms (hierarchical clustering) of statistically relevant genes with a 2 fold change in expression were created. Potential biomarkers were identified from this dataset.

CHAPTER 3

DEVELOPMENT OF A *GAMMARUS PULEX* cDNA LIBRARY

3.1 INTRODUCTION

A cDNA library of *Gammarus pulex* genes could potentially allow the identification of novel genes in which expression is altered by endocrine disruption. As well as disrupted gene expression, the basal levels of gene expression during varying moult stage, developmental stages and gender specific gene expression could be characterised by use of microarray technology.

Suppressive subtractive hybridisation (SSH) is a technique used to enrich a specific gene population. Combining this technique with microarray technology allows the identification of differentially expressed genes and the profiling of their expression pattern in different physiological, biological or environmental states. This combination has already been put into practice (Yang *et al.* 1999), successfully being used to produce gene pools representing specific life-stages of *Daphnia magna* which were subsequently utilised within a microarray in order to further describe life-stage specific perturbations in gene expression after exposure to the pesticide, propiconazole (Soetaert *et al.* 2005). The process involved in producing SSH libraries means the resulting PCR (Polymerase Chain Reaction) products are small in size (200 - 400bp) and therefore rarely represent entire transcripts. Furthermore, these SSH fragments are randomly distributed throughout the cDNA representing both coding and non-coding elements with equal frequency. Use of SSH reduces the number of housekeeping genes, leaving mainly the genes of interest, however, as this is an enrichment technique it does not completely eliminate common transcripts from the SSH library.

In order to confirm that animals collected for cDNA library construction were *G. pulex*, other than by visual observation, the RNA was reverse transcribed and PCR performed on the resultant cDNA to amplify the cytochrome oxidase I gene (COI). The primers used were COIa-H and COI-Gf (Appendix A, Table 6) known to effectively amplify COI in gammarids (Meyran *et al.* 1997). The amplified PCR product was cloned into pGEM-T vector (Promega Ltd., Southampton, UK) and

sequenced using M13 primers (Appendix A, Table 6). A homology search of GenBank using BLASTN (Section 2.11.15), revealed a 100% identity with the COI sequence of *G. pulex* haplotype Gp1.

3.2 *G. PULEX* MIXED LIBRARY

A library made from animals of mixed sex, age and moult stage was created to represent a wide range of genes expressed in *G. pulex*.

LIBRARY	MALE			FEMALE		OTHER
	PAIRED		UNPAIRED	PAIRED	UNPAIRED	
	C	D	C	D	C	
mxA	18	4	6	22	12	Juveniles/Neonates
mxB	90 adults + 5 paired adults (moult staged unknown)					~50 Juveniles/Neonates

Table 3.1: Number of animals used in mixed libraries

3.2.1 Library B (mxB)

A *G. pulex* cDNA library was constructed as described in Chapter 2, Section 2.12. Both the first and second strand syntheses were successful and a picture of the minigel can be seen in Figure 3.1. A spot assay (Chapter 2, Section 2.12.1.9) was necessary before ligation into pBluescript® SK II vector to determine the presence of cDNA after Sepharose chromatography. The resulting spot assay was photographed under UV light (Figure 3.2).

The calibration standard that generated equivalent fluorescence under UV light to a ¼ dilution of cDNA as determined by eye was 3.13ng/µl and therefore original library cDNA could be estimated to have a concentration of ~10ng/µl; consequently 0.5µl contained the prerequisite quantity (5ng) for library construction. After ligation, transformation into XL10-Gold ultracompetent cells yielded a low level of transformants with only 750cfu (colony forming units). However, 96 colonies were picked and screened by PCR with M13 primers before being separated on an E-Gel® (Invitrogen Ltd., Paisley, UK; Figure 3.3). A band represents a positive vector insert and therefore a successfully cloned library product, the lack of a positive band indicates an unsuccessful PCR amplification and therefore a negative clone. The gel was analysed using E-Editor (Invitrogen Ltd., Paisley, UK) and the successful clones

selected. The positive clones were identified by strong clear single bands with a size greater than 300bp. Those clones with two or more bands, smeared bands or with a size of less than 300bp were disregarded. The clones were rated on a scale of 1 to 4 as described in Chapter 2, Section 2.12.3.4.

Figure 3.1: cDNA synthesis minigel

The gel was stained with SYBR Green and shows the successful primary and secondary cDNA synthesis as smears. Lane 1 is primary synthesis product; Lane 2 is secondary synthesis product.

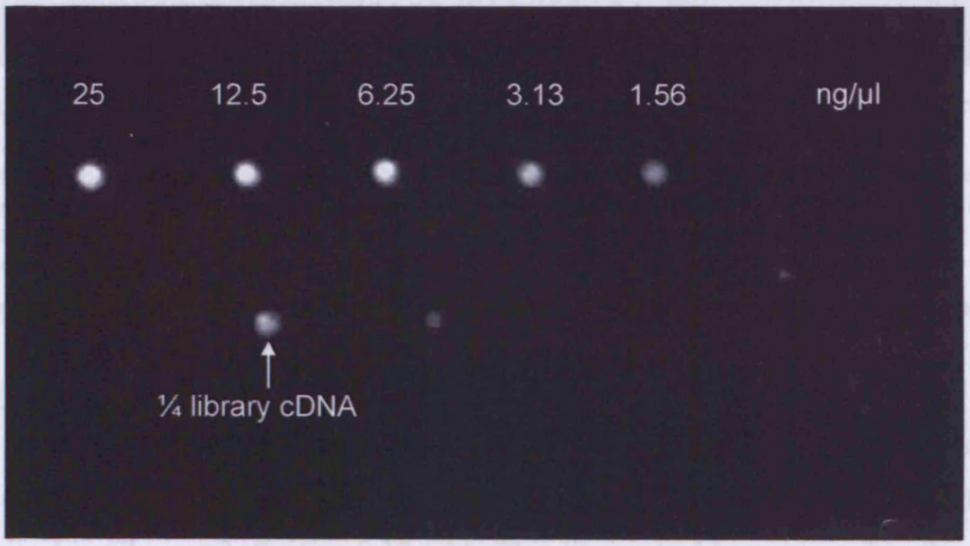


Figure 3.2: Spot assay of *G. pulex* library cDNA

The image depicts the fluorescence generated under UV light of DNA samples mixed with an equal quantity of ethidium bromide (2µg/ml). The upper row of spots represents a set of standards at the concentrations indicated, whilst the lower row corresponds to a ¼ dilution of size fractionated cDNA. The spot identified by the arrow is indicative of an approximate concentration of 3ng/µl when compared to DNA standards.

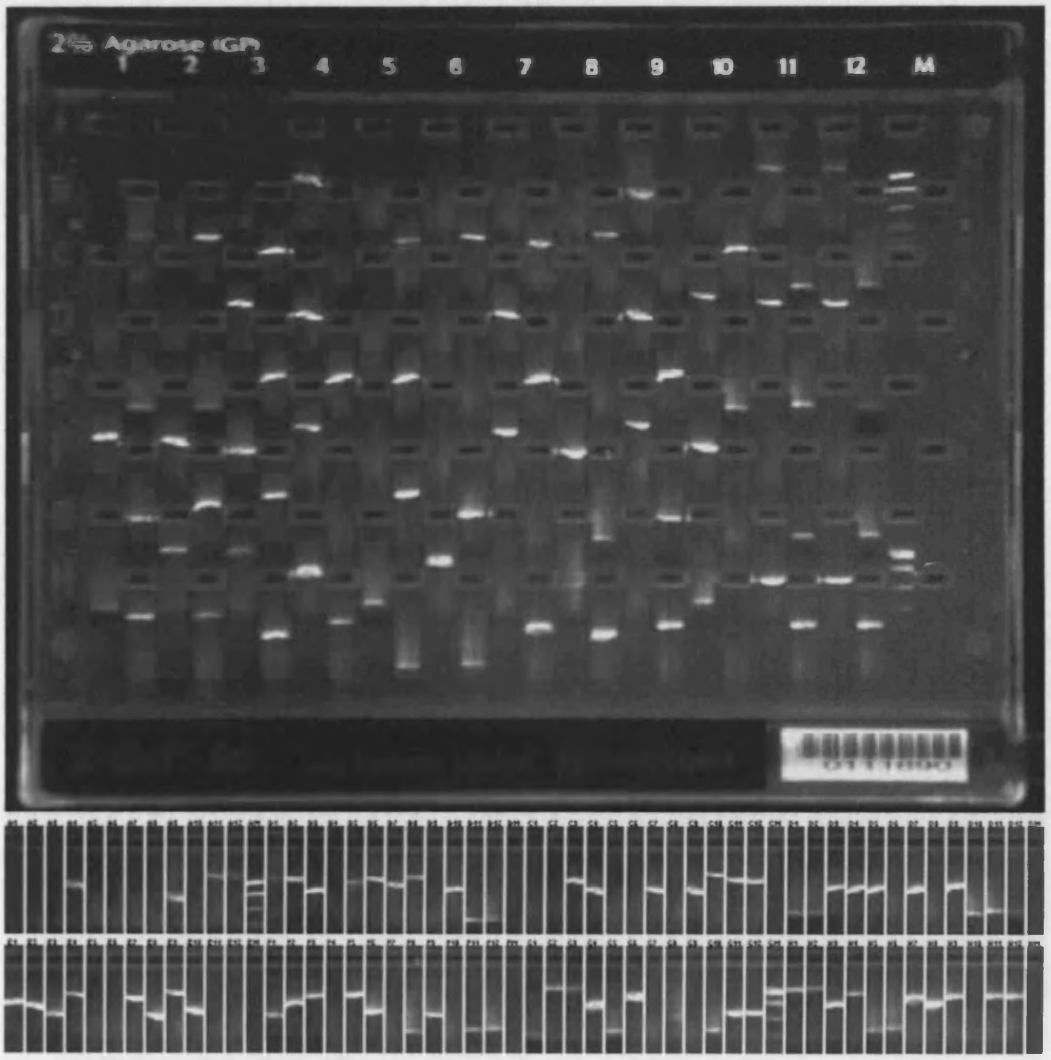


Figure 3.3: The 96-well mxAB cDNA library PCR amplification on an Invitrogen E-Gel®.

The pre-cast E-Gel® system was used for high-throughput screening of PCR products. Each gel cassette allowed the simultaneous analysis of 96 samples and 8 size markers (although only 2 are used in this circumstance). Associated software (E-Editor) allowed a 52 x 2 layout image to be generated for easier size comparison and quality control.

3.2.2 Library A (mxA)

Due to the low titre of the library mxB, a previously generated cDNA library (mxA; Dr Judith Richards) was ligated and transformed in XL-10 ultracompetent cells as described in Chapter 2, Section 2.12.1.11. This library had been created in the same way as mxB library but using animals from WRc-NSF, Medmenham, Bucks., UK, rather than a local source. The primary transformation yielded 3 x 10^5 transformants and after amplification this number was expanded to 3.4 x 10^8, a 1000-fold increase. A photograph of E-Gel® analysis of this library can be seen in Figure 3.4.

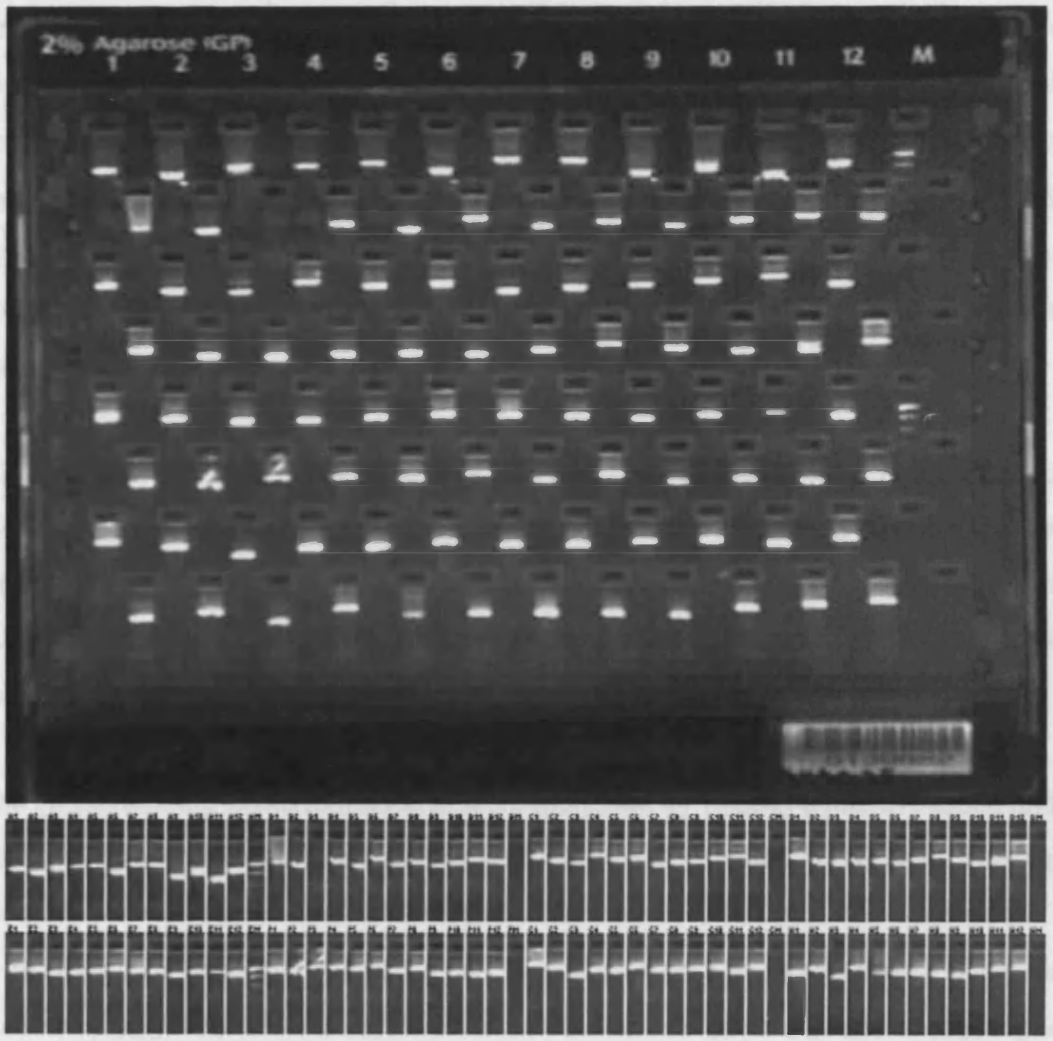


Figure 3.4: E-Gel® analysis of the mxA library (Plate mxAA25)
Plate mxAA25 demonstrates the increase in successful clones compared to the mxAB library E-Gel® shown in Figure 3.3.

From the mxA primary transformation, 1056 clones were picked and screened by PCR (mxAB) with a success of 91%, yielding 957 positive clones. From the amplified transformation (mxAA), 50% of the picked clones developed into a cloudy culture with no PCR product. This suggested that the amplified library had been contaminated at some point during the 43 hour incubation at 30°C. The culture viewed after 24 hours was clear with tiny distinct colonies in suspension, however after the full 43 hours the culture was cloudy and this was originally put down to overgrowth. In an attempt to screen out contaminating bacteria, a positive *Escherichia coli* clone and a negative clone were plated onto S-Gal™ agar.

This screening agar discriminates between bacteria containing a functioning *lacZ* gene and the ability to express β-galactosidase and those genetically modified (GM) bacteria with a *lacZ* gene interrupted with a gene insert. The chemical S-Gal™ contained within the agar is metabolised by β-galactosidase yielding a black colony and those bacteria that cannot express β-galactosidase yield white colonies. Therefore the GM bacteria containing inserts should give white colonies, and those contaminating bacteria would produce black colonies if they contained a functioning *lacZ* gene.

Although this method of screening was successful, when the glycerol was plated onto the agar, the contaminating bacteria grew at a faster rate than the *E. coli* and smothered the desired clones. Due to the differing rates of growth, a selection that inhibited growth of the contaminating bacteria had to be developed for the successful selection of a large quantity of clones.

The contaminating bacterium species were identified by plating a positive clone and a negative cloudy clone onto chromogenic urinary tract infection (UTI) agar. This agar contains chromogenic substrates that are metabolised in the presence of β-galactosidase (Red-Gal substrate) and β-glucosidase (X-Gluc substrate). They also contain phenylalanine and tryptophan which act as an indicator of tryptophan deaminase (TDA) activity. This test yielded blue/green colonies indicative of an enterococci bacterium. Although wild type *E. coli* should produce pink colonies on this agar, white colonies were seen due the inhibited β-galactosidase function of successfully transformed GM *E. coli*. The plates can be seen in Figure 3.5.

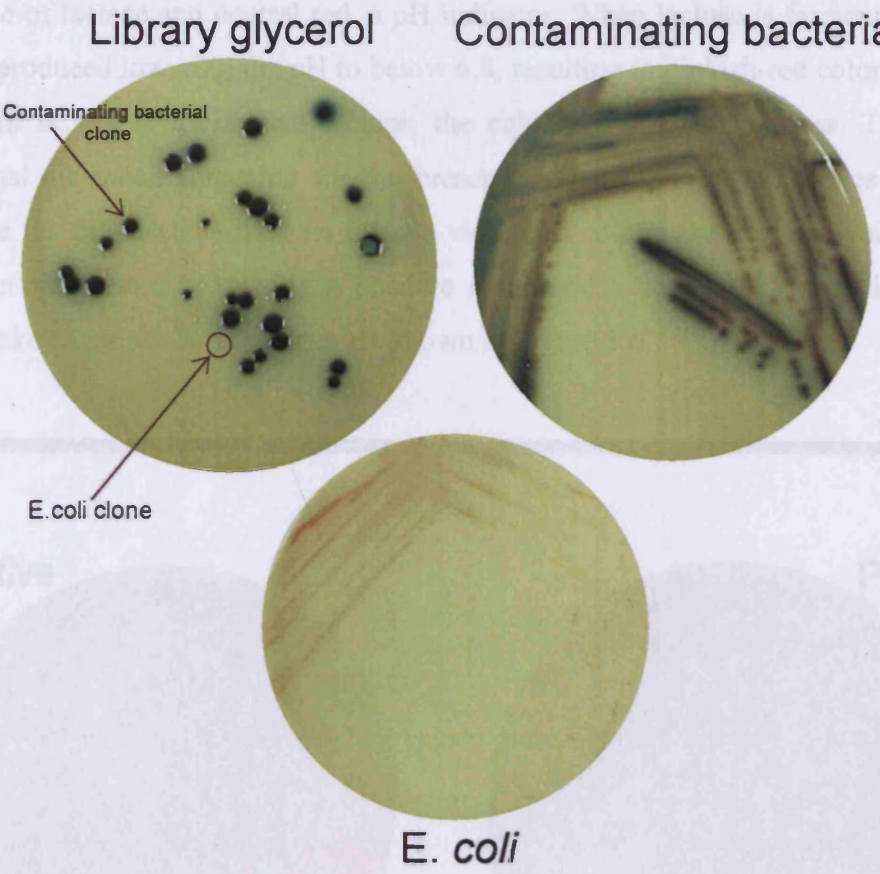


Figure 3.5: Chromogenic UTI screening of mxAA library

Library glycerol stock was grown on chromogenic UTI agar in order to determine the species of contaminating bacteria present. The *E. coli* clones were small, distinct and pink in appearance; the contaminating bacteria were large and blue-green in colour. A colony of *E. coli* and contaminating bacteria from an LB plate used for library screening were cultured in LB broth and streaked onto chromogenic UTI agar for comparison. The blue-green colour of the contaminating bacteria was indicative of Enterococci.

A distinguishable difference between *Escherichia* sp. and *Enterococcus* sp. strains are their gram staining and tolerance to certain fractions of bile salts. A selective media that can be used to distinguish between gram-positive and gram-negative bacterium is MacConkey. *Enterococcus* sp. differ from most gram positive bacterium in that they can grow in the presence of crystal violet and certain bile salts. A modification of the traditional MacConkey agar, MacConkey agar no. 3 (Oxoid Ltd., Basingstoke, Hants., UK) not only has the ability to inhibit most gram-positive bacterium due to the presence of crystal violet but it also contains a special fraction of bile salts that specifically inhibit *Enterococcus* sp. MacConkey agar is also differential with the

ability to distinguish between lactose and nonlactose fermenters. This is due to the presence of lactose and neutral red, a pH indicator. When lactose is fermented, lactic acid is produced lowering the pH to below 6.8, resulting in pinkish-red colonies. If an organism is unable to ferment lactose, the colonies remain colourless. This is an additional aid when screening for the presence of an insert in the clones as those negative for the insert will be an intense violet red colour and can be avoided. The selection was tested by plating a positive clone and a negative cloudy clone onto MacConkey agar no. 3. The plates are shown in Figure 3.6.

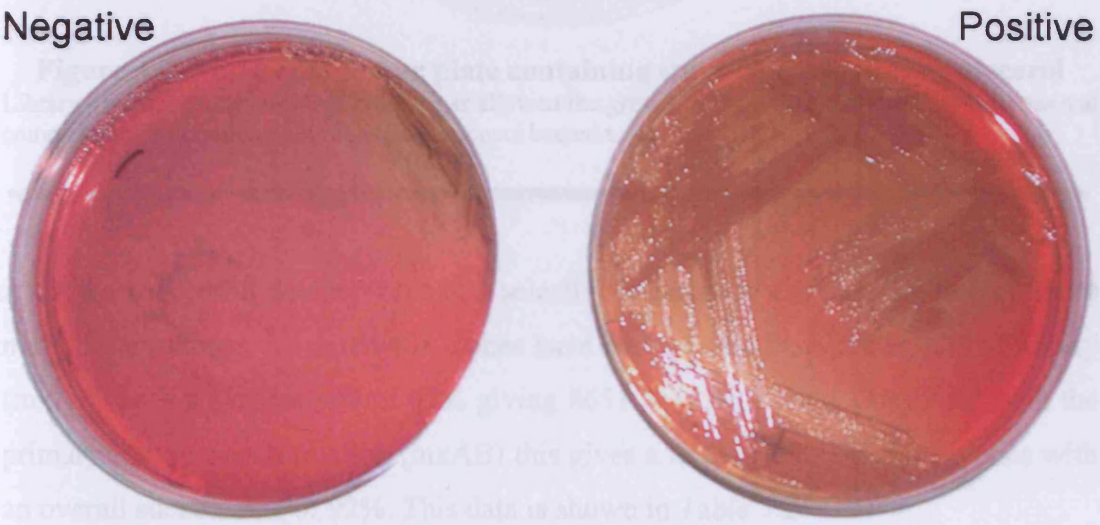


Figure 3.6: MacConkey agar selection of mxAA library
A negative contaminated colony and a positive *E. coli* colony were cultured in LB broth and streaked onto MacConkey agar to determine if there was positive selection for *E. coli* against *Enterococcus* sp. Growth of contaminating enterococci bacteria was inhibited on MacConkey agar and therefore deemed suitable for library screening.

After successful inhibition of the contaminating bacteria, through the use of MacConkey agar no. 3, the glycerol stock was spread onto plates as described in Chapter 2, Section 2.12.3. The resulting plate can be seen in Figure 3.7.

Figure 3.7: MacConkey agar plate containing spread mxAA library glycerol
Library glycerol grown on MacConkey agar allowed the growth of small and distinct *E. coli* clones and complete inhibition of contaminating enterococci bacteria.

After the successful development of a selective method, it was possible to pick more mxA library clones. To date 9408 clones have been picked from the amplified library (mxAA) with a success rate of 92% giving 8651 positive clones. Combined with the primary library transformation (mxAB) this gives a total of 9608 positive clones with an overall success rate of 92%. This data is shown in Table 3.2.

TRANSFORMATION	NUMBER PICKED	POSITIVE CLONES	PERCENTAGE SUCCESS
mxAB	1056	957	91%
mxAA	9408	8651	92%
TOTAL	**10,464**	**9608**	**92%**

Table 3.2: Current number of positive clones isolated from the mixed cDNA library.

3.3 *G. PULEX* MALE SUBTRACTIVE LIBRARY

3.3.1 maSA

The male subtractive library (courtesy of Dr. Judith Richards) was generated using the PCR-Select™ cDNA Subtraction Kit (Clontech-Takara Bio Europe, Saint-Germain-en-Laye, France) using paired males separated from the females; 3 animals from each gender were used to make the library. A subtractive library generates products that are significantly higher abundance in one population than another, producing an enriched library. In this case, male abundant genes were enriched by a subtraction with females. In contrast to the mixed library the male subtractive library (maS) was ligated into pGEM-T vector as described in Chapter 2, Section 2.11.10.2. S-Gal™ agar was used to screen the transformants which yields white clones when an insert is present and black when the vector is empty. This black/white screening was only used for the maSA library which was processed after the maSB and maSC libraries.

Primary transformation into XL-10 ultracompetent cells yielded a titre of $0.87x10^5$cfu. This was amplified as previously described to give a final titre of $3x10^9$cfu, an amplification factor of 34,000 and was named maSA. In an initial test of this amplified library 288 clones (3 x 96-well plates) were picked from S-Gal™ plates, with 251 being positive for an insert, an 87% success rate. The library was then amplified as described in Chapter 2, Section 2.12.2. However, subsequent retrieval of selected clones stored as glycerol stocks at -70°C generated PCR products which exhibited significant smearing when analysed on E-Gels. This was determined to be issue related to a degraded glycerol stock compromised due to too many freeze/thaw cycles. After switching to a new glycerol stock, the PCR amplifications were successful. From this point on, a scraping of glycerol was used rather than repeatedly thawing and diluting a volume of the glycerol stock, to prevent degradation. To date a total 3840 clones have been picked from this transformation yielding 3341 positives which is a success rate of 87%.

3.3.2 maSB

A second transformation of the maS library (maSB) yielded a primary transformation titre of $1.3x10^4$cfu. This was too low to undertake semi-solid clonal amplification; the minimum titre required for this process is $2.0 x10^4$cfu to avoid over-representation of individual clones. From this primary transformation 576 clones were picked from LB agar and 379 were positive, a 66% success rate.

3.3.3 maSC

A third and final transformation of the maS library (maSC) was performed giving a primary transformation titre of $2.8x10^3$cfu. A total of 480 clones were picked from this primary transformation from LB agar plates with a success rate of 63% giving 301 positive clones. Again the titre was too low for amplification so the maSA amplification was used for all future library screening. To date 4896 colonies have been picked from all the subtractive libraries of which 4021 clones were positive giving an overall success rate of 82%. This data can be seen in Table 3.3.

TRANSFORMATION	NUMBER PICKED	POSITIVE CLONES	PERCENTAGE SUCCESS
maSA	3840	3341	87%
maSB	576	379	66%
maSC	480	301	63%
TOTAL	**4896**	**4021**	**82%**

Table 3.3: Number of positive clones isolated from the male subtractive cDNA library.

3.4 SEQUENCE ANALYSIS

After PCR product purification using Montage® Multiscreen $PCR_{\mu 96}$ cleanup plates and transferral into a cherry-picked composite plate, 5 plates were sequenced as described in Chapter 2, Section 2.12.4. Three plates represented the mixed library (Gp_mxAA_001-003) and 2 represented the male subtractive library (Gp_maSA_01-02). These sequences were analysed using the bioinformatics software trace2dbest (Parkinson *et al.* 2004) which base checks the sequence file and rejects any poor sequences.

The 408 successful sequences were submitted to the NCBI expressed sequence tag (EST) database (GenBank accession numbers: mxA: DN966600 - DN966824; maS: DN976214 - DN976396) and clustered using the CLOBB component of the programme PartiGene (Parkinson *et al.* 2004). The 428 resultant clusters were analysed for homologous genes using BLASTN and BLASTX queries, the results of which were integrated into the *G. pulex* wwwPartiGene database and can be accessed at http://ecoworm.bios.cf.ac.uk through "*Gammarus pulex* (1)". These sequences were in addition to the 157 previously sequenced clones (sequenced by Judith Richards from library mxA) that were submitted in the same way totalling 564. The sequences were checked for redundancy by comparing the number of singletons to those in clusters. A table of the initial plate sequencing cluster results can be seen in Table 3.4.

ABUNDANCE	NUMBER OF CLUSTERS	PERCENTAGE
43	1	0.23%
11	1	0.23%
8	1	0.23%
5	1	0.23%
4	3	0.70%
3	11	2.57%
2	25	5.84%
1	385	89.95%
TOTAL	428	100.00%

Table 3.4: Initial plate sequencing cluster results

After confirmation that the library contained ~90% singletons, the PCR products (100 composite mxA and 40 maS plates) were prepared and sent to the University of Edinburgh, School of Biological Sciences Sequencing Service (SBSSS) totalling 13,440 sequences. The raw files were processed through the same quality pipeline as before, rejecting 953. They were submitted to the NCBI expressed sequence tag (EST) database, with newer sequences overwriting older only if they had a greater length or better quality. This process left 325 of the original sequences on NCBI, replaced 239 and added a further 12,080, totalling 12,644 *G. pulex* ESTs (157 original mxA, 168 Cardiff sequenced, 12,319 Edinburgh sequenced). These were clustered and uploaded into the Arthropoda PartiGene Database at http://www.nematodes.org/NeglectedGenomes/ARTHROPODA/Crustacea.html. This resulted in 3917 clusters. A table of the final sequencing cluster results can be seen in Table 3.5.

ABUNDANCE	NUMBER OF CLUSTERS	PERCENTAGE
1031	1	0.03%
839	1	0.03%
146	1	0.03%
107	1	0.03%
104	1	0.03%
103	1	0.03%
94	1	0.03%
82	1	0.03%
77	2	0.05%
73	2	0.05%
71	1	0.03%
64	1	0.03%
54	1	0.03%
53	1	0.03%
50	2	0.05%
48	1	0.03%
47	1	0.03%
45	1	0.03%
44	1	0.03%
43	2	0.05%
42	2	0.05%
41	1	0.03%
40	1	0.03%
39	1	0.03%
37	2	0.05%
36	1	0.03%
35	2	0.05%
34	5	0.13%
33	3	0.08%
31	3	0.08%
30	1	0.03%

ABUNDANCE	NUMBER OF CLUSTERS	PERCENTAGE
29	3	0.08%
28	1	0.03%
27	2	0.05%
26	6	0.15%
25	5	0.13%
24	5	0.13%
23	3	0.08%
22	5	0.13%
21	6	0.15%
20	5	0.13%
19	4	0.10%
18	6	0.15%
17	9	0.23%
16	7	0.18%
15	5	0.13%
14	8	0.20%
13	7	0.18%
12	12	0.31%
11	6	0.15%
10	10	0.26%
9	26	0.66%
8	27	0.69%
7	23	0.59%
6	34	0.87%
5	74	1.89%
4	118	3.01%
3	217	5.54%
2	595	15.19%
1	2642	67.45%
TOTAL	3917	100.00%

Table 3.5: Final plate sequencing cluster results

3.5 GENE ONTOLOGY ANNOTATION

Gene ontology annotations are used to annotate genes and gene products with a list of attributes. Gene products are annotated into three principles; cellular component, molecular function and biological process. The mixed and male subtractive library sequences were analysed separately for each of the three groups and then compared. Cellular component indicates in which component of a cell or protein group the gene product is likely to be found. Molecular function describes activity at the molecular level and biological process describes a series of events by one or more groups of molecular functions. Biological process is similar to molecular function but the process describes a series of steps rather than individual function (Ashburner *et al.* 2000).

Due to the lack of amphipod sequencing projects these *G. pulex* libraries yielded many sequences with an unknown function. For this reason the annotations were assigned after an InterPro scan of each individual clone. This was performed using Blast2GO (Conesa *et al.* 2005), which performed a BLASTX of each sequence, followed by InterPro scan and finally assigned full GO annotations to each clone. These results were used to determine the distribution of active genes at the cellular level, the proportion of genes involved in specific molecular functions and the biological processes involved. GOSlim annotations were not used for the individual libraries due to lack of data on this particular organism, the generic terms were not deemed descriptive enough. In the following sections, definitions determined by the Gene Ontology Consortium (http://www.geneontology.org) are used to describe the annotations; all material quoted from the Gene Ontology Consortium is provided between quotation marks and in *italic*. These quotations are not paraphrased or abbreviated since they represent agreed definitions whose absolute meaning may be misrepresented if misquoted.

As a general guide to the distribution of molecular function, cellular component and biological process represented on the microarray, GOSlim annotations (GOA - generic) were assigned and are shown in Section 3.5.4. GOSlim annotations are a reduced version of GO annotations and give a wider overview making them useful to describe a summary of the *G. pulex* microarray. In addition to individual annotations,

the contigs representing each cluster (Section 3.5.5) were also analysed using Blast2GO in the same way and assigned GOSlim annotations (see Section 3.5.5). This describes the genes representing particular molecular functions, cellular components and biological processes that have been identified through this project, assigning GOSlim annotation to individual genes rather than clones. Not all sequences were able to be annotated with a gene ontology description despite using both a BLASTX search and InterPro scan in order to identify homologous sequences. Gene objects were annotated at multiple levels, increasing with specificity as the level number increases. In this analysis, only Levels 2 and 3 were interpreted; for GOSlim annotation, by default, the highest descriptive level is used. The proportions of sequences annotated in Level 2 and 3 are displayed in Figure 3.8.

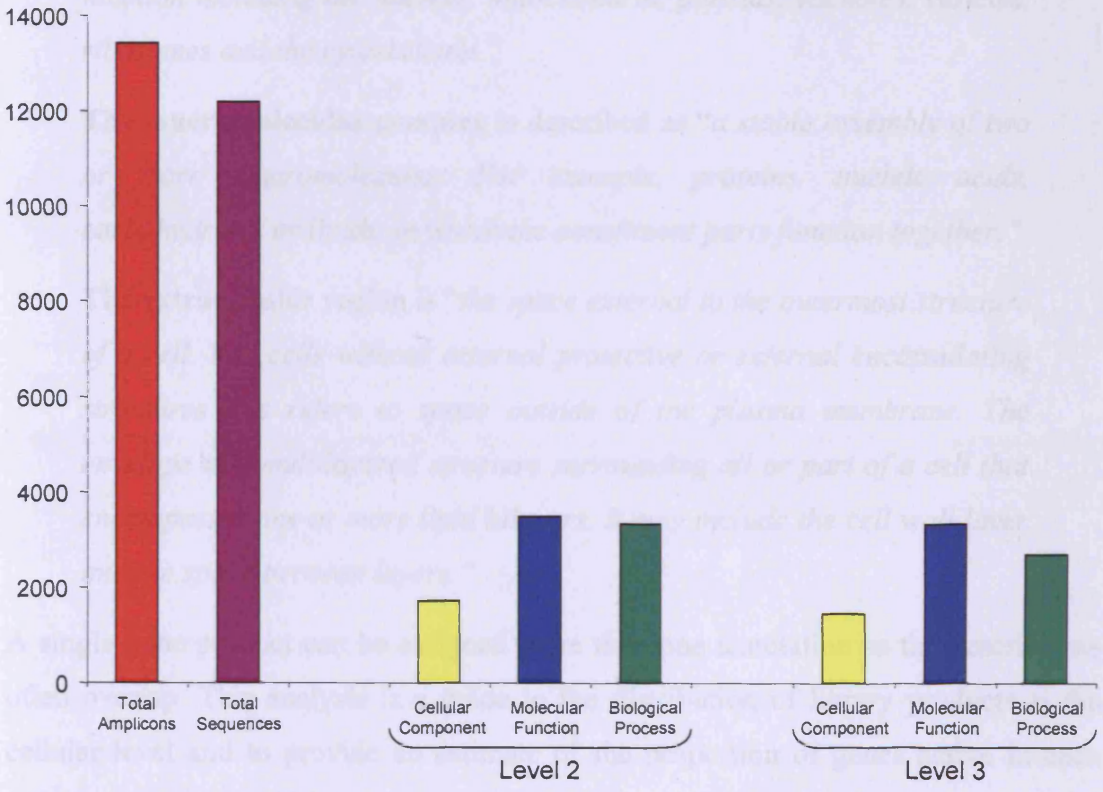


Figure 3.8: Number of sequences annotated with a gene ontology description

3.5.1 Cellular Component

3.5.1.1 Mixed (mxA) library

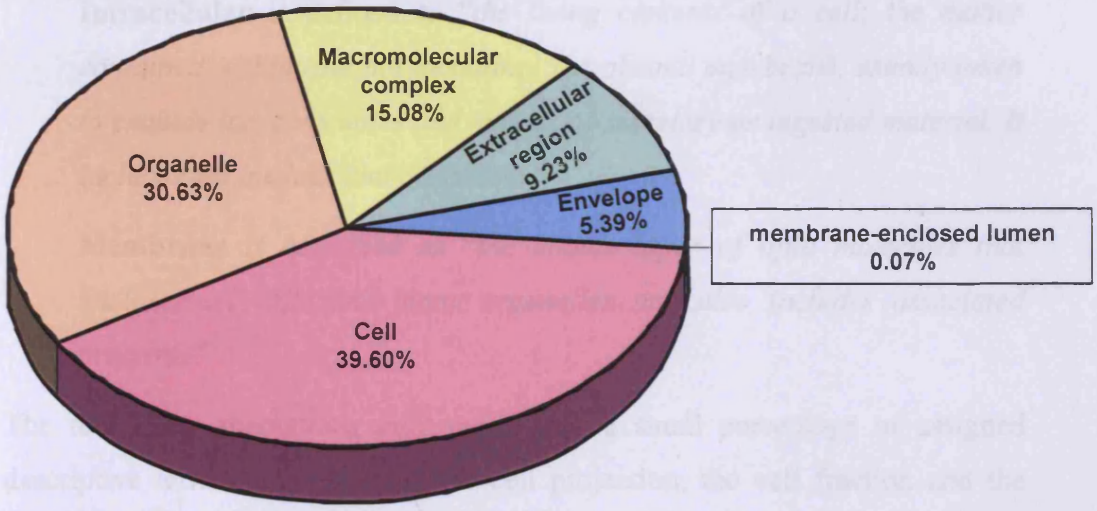
The majority of mxA library gene products, annotated to the term cellular component at Level 2, are associated with the cell, organelles, part of a macromolecular complex, in the extracellular region or within the envelope. A small number were within the membrane-enclosed lumen. The five main components, cell, organelle, macromolecular complex, extracellular region and envelope, are described below.

Those products annotated with "cell" include the plasma membrane and any external encapsulating structures such as the cell wall and cell envelope. The following definitions are provided to illume as to the precise nature of the major categories[1];

> **The organelle** is "*an organised structure of distinctive morphology and function including the nucleus, mitochondria, plastids, vacuoles, vesicles, ribosomes and the cytoskeleton.*"
>
> **The macromolecular complex** is described as "*a stable assembly of two or more macromolecules. For example, proteins, nucleic acids, carbohydrates or lipids, in which the constituent parts function together.*"
>
> **The extracellular region** is "*the space external to the outermost structure of a cell. For cells without external protective or external encapsulating structures this refers to space outside of the plasma membrane. The envelope is a multilayered structure surrounding all or part of a cell that encompasses one or more lipid bilayers. It may include the cell wall layer and the space between layers.*"

A single gene product can be assigned more than one annotation as the descriptions often overlap. This analysis is a guide to the distribution of library products at the cellular level and to provide an estimate of the proportion of genes active in each cellular component. Using a pie chart to display the Level 2 annotations shows the relative frequency of annotations and is a useful method to assess the likelihood of a given annotation being awarded to a gene product in a particular gene set. The absolute frequencies are the true percentage of gene products assigned a particular annotation, taking into account that each product can be annotated with more than one descriptor. This data can be seen in Figure 3.9.

[1] Definitions taken from Gene Ontology Consortium (http://www.geneontology.org)



Term name	Percent gene products with cellular component annotation
Cell	81.31%
Organelle	62.89%
Macromolecular complex	30.97%
Extracellular region	18.96%
Envelope	11.07%
Membrane-enclosed lumen	0.13%

Figure 3.9: Mixed library Level 2 annotation under cellular component.
Pie chart representing the relative frequency of Level 2 annotations under cellular component for the mixed library. The table below shows the absolute frequencies for the same gene set. Only gene fragments with "known" ontology were included in the chart.

GO annotation assigned at Level 3 for cellular component are shown in Figure 3.10 and Table 3.6. Sequences assigned to "cell" were divided between five sublevels, intracellular, membrane, cell projection, cell fraction and external encapsulating structure. The most common assignment was "intracellular", also known as protoplasm, followed by "membrane" with remaining annotations only constituted a small percentage of assigned descriptive terms. This includes the cell projection, the cell fraction and the external encapsulating structure.

The definitions of the most common compartments are given below[1];

> **Intracellular** is defined as *"the living contents of a cell; the matter contained within (but not including) the plasma membrane, usually taken to exclude large vacuoles and masses of secretory or ingested material. It includes the nucleus and cytoplasm."*
>
> **Membrane** is described as *"the double layer of lipid molecules that encloses all cells and many organelles and also includes associated proteins."*

The remaining annotations only constituted a small percentage of assigned descriptive terms. This includes the cell projection, the cell fraction and the external encapsulating structure.

Within the organelle, there were six sublevels divided between intracellular organelle, non-membrane bound organelle, membrane-bound organelle, organelle membrane, and to a lesser extent organelle lumen and vesicle. Descriptions of the top three attributes are as follows[1];

> The **intracellular organelle** is described in the same way as organelle (see above).
>
> **Non-membrane bound organelle** is *"an organised structure of distinctive morphology and function and not bounded by a lipid bilayer membrane. It includes ribosomes, the cytoskeleton and chromosomes."*
>
> Contrastingly **membrane-bound organelles** are *"organised structures of distinctive morphology and function, bounded by a single or double lipid bilayer membrane. They include the nucleus, mitochondria, plastids, vacuoles, and vesicles but exclude the plasma membrane. The organelle membrane is the lipid bilayer surrounding an organelle."*

[1] Definitions taken from Gene Ontology Consortium (http://www.geneontology.org)

Within the macromolecular complex, there were two sublevels; protein complex and ribonucleoprotein complex, these are defined as follows[1];

> The **protein complex** is described as "*any protein group composed of two or more subunits, which may or may not be identical. Protein complexes may have other associated non-protein prosthetic groups, such as nucleic acids, metal ions or carbohydrate groups.*"
>
> The **ribonucleoprotein complex** is described as "*a macromolecular complex containing both protein and RNA molecules.*"

The extracellular region was divided between four sublevels; extracellular space, collagen, proteinaceous extracellular matrix and apoplast. The relevant descriptions for these categories are provided below[1];

> The **extracellular space** is defined as "*that part of a multicellular organism outside the cells proper, usually taken to be outside the plasma membranes, and occupied by fluid.*"
>
> **Collagen** is defined as *any of the various assemblies in which collagen chains form a left-handed triple helix, they may assemble into higher order structures.*"
>
> **Proteinaceous extracellular matrix** is defined as "*a layer consisting mainly of proteins (especially collagen) and glycosaminoglycans (mostly as proteoglycans) that forms a sheet underlying or overlying cells such as endothelial and epithelial cells. The proteins are secreted by cells in the vicinity. Water and solutes pass freely through it.*"
>
> **Apoplasts** are "*the protoplasts of cells in a plant are connected through plasmodesmata, and plants may be described as having two major compartments: the living symplast and the non-living apoplast. The apoplast is external to the plasma membrane and includes cell walls, intercellular spaces and the lumen of dead structures such as xylem vessels. Water and solutes pass freely through it.*"

[1] Definitions taken from Gene Ontology Consortium (http://www.geneontology.org)

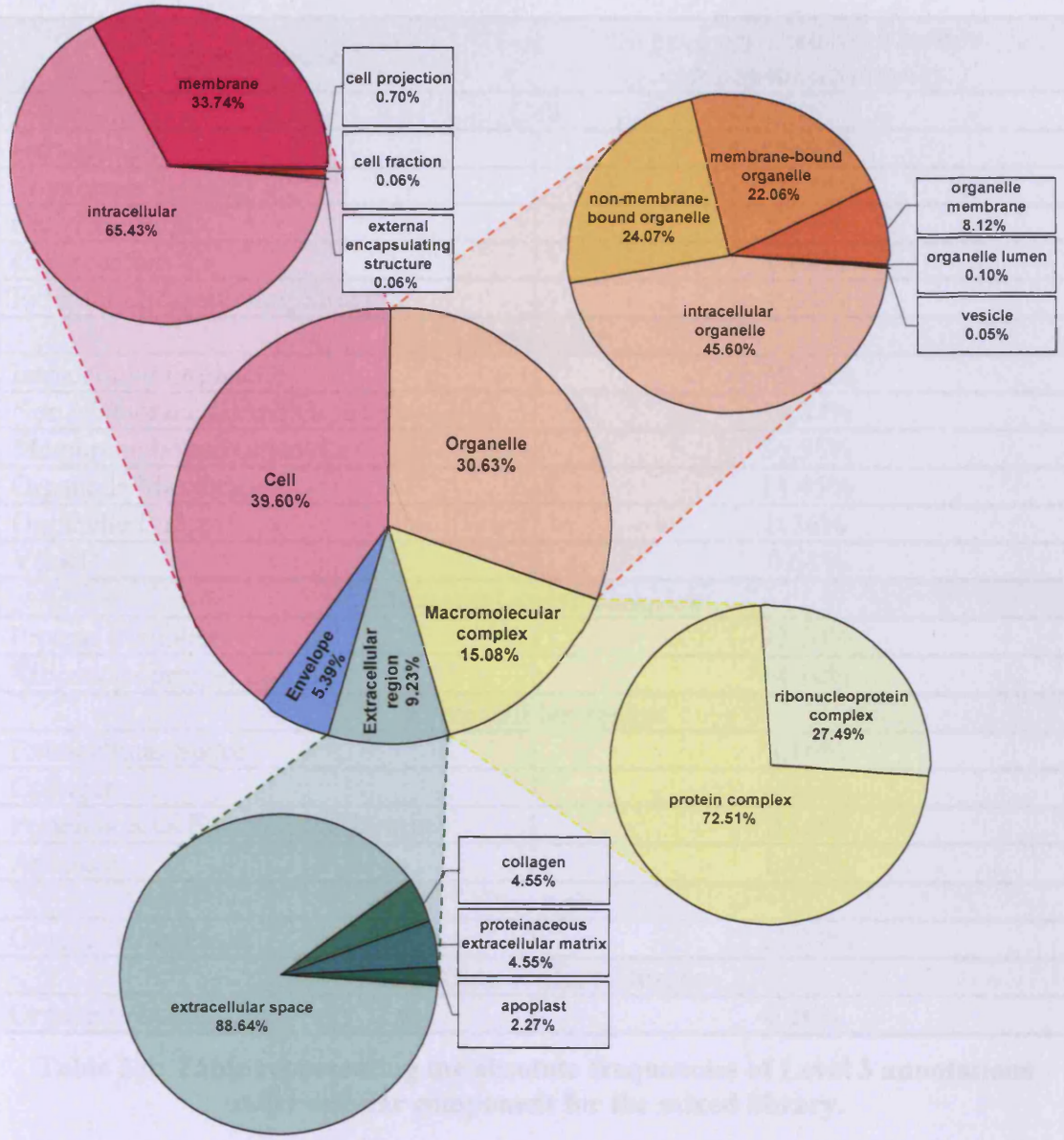


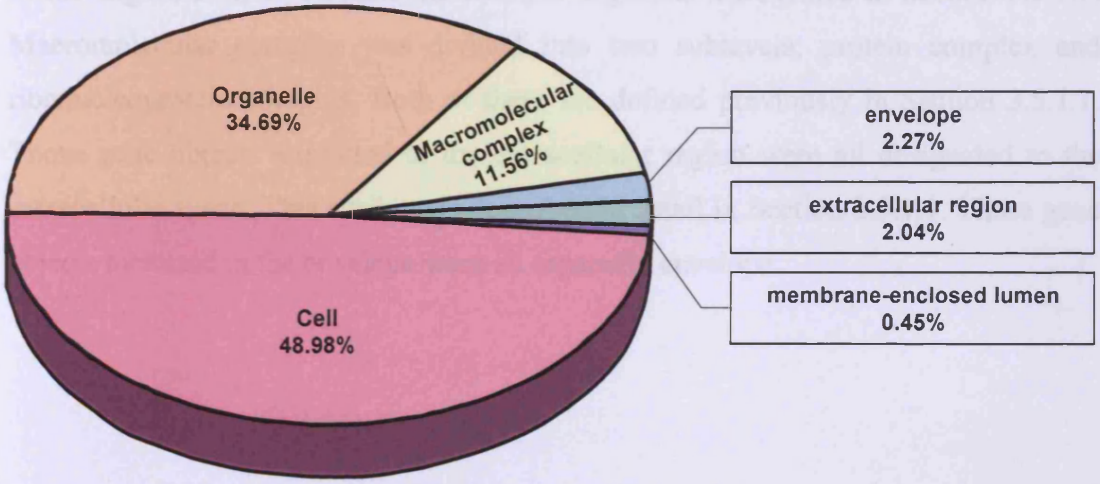
Figure 3.10: Level 3 mixed library cellular component gene ontology annotation
Pie chart representing the level 3 cellular components of the mxA cDNA library. Only gene fragments with "known" ontology were included in the chart.

Term name	% gene products with cellular component annotation
Cell	
Intracellular	82.98%
Membrane	42.79%
Cell Projection	0.89%
Cell Fraction	0.08%
External Encapsulating Structure	0.08%
Organelle	
Intracellular Organelle	75.53%
Non Membrane-bound Organelle	39.87%
Membrane-bound organelle	36.55%
Organelle Membrane	13.45%
Organelle Lumen	0.16%
Vesicle	0.08%
Macromolecular complex	
Protein Complex	37.20%
Ribonucleoprotein Complex	14.10%
Extracellular region	
Extracellular Space	3.16%
Collagen	0.16%
Proteinaceous Extracellular Matrix	0.16%
Apoplast	0.08%
Envelope	
Organelle Envelope	13.29%
Membrane-enclosed lumen	
Organelle Lumen	0.16%

Table 3.6: Table representing the absolute frequencies of Level 3 annotations under cellular component for the mixed library.

3.5.1.2 Male subtractive (maS) library

The majority of gene products from the maS library annotated within cellular component are associated with the cell, with the remaining inside organelles as part of a macromolecular complex, within the envelope, in the extracellular region or, to a lesser extent, within the membrane-enclosed lumen. This data can be seen in Figure 3.11. A definition of these attributes has been described in Section 3.5.1.1.



Term name	Percent gene products with cellular component annotation
Cell	95.15%
Organelle	67.40%
Macromolecular complex	22.47%
Envelope	4.41%
Extracellular region	3.96%
Membrane-enclosed lumen	0.88%

Figure 3.11: Male subtractive library Level 2 annotation under cellular component.

Pie chart representing the relative frequency of Level 2 annotations under cellular component for the male subtractive library. The table below shows the absolute frequencies for the same gene set.

Four of these attributes; cell, organelle, macromolecular complex and extracellular region, were further divided as shown in Figure 3.12 and Table 3.7. Sequences assigned to the cell were divided into three sublevels; intracellular, membrane and external encapsulating structure. A definition of the two major categories, intracellular and membrane are previously described in Section 3.5.1.1. Gene objects localised within organelles were divided into five sublevels; intracellular organelle, membrane-bound organelle, non-membrane bound organelle, organelle membrane and organelle lumen. The three major annotation descriptions, intracellular organelle, membrane-bound organelle and non-membrane bound organelle are defined in Section 3.5.1.1. Macromolecular complex was divided into two sublevels; protein complex and ribonucleoprotein complex. Both of these are defined previously in Section 3.5.1.1. Those gene objects attributed to the extracellular region were all designated to the extracellular space. This attribute is described in detail in Section 3.5.1.1. Those gene objects localised in the envelope were all organelle envelope.

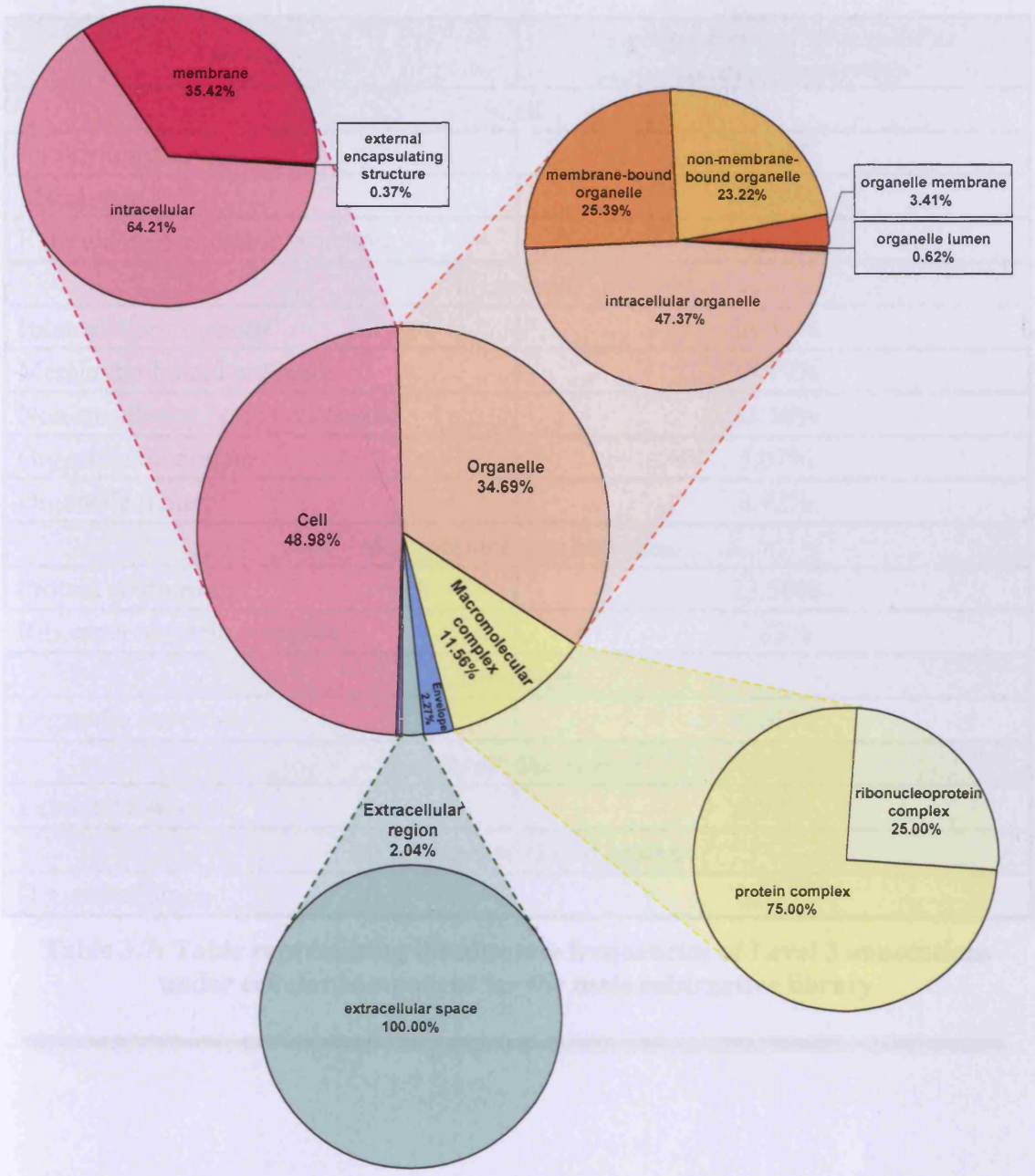


Figure 3.12: Level 3 male subtractive library cell component gene ontology annotation

Only gene fragments with "known" ontology were included in the chart.

Term name	% gene products with cellular component annotation
Cell	
Intracellular	80.18%
Membrane	44.24%
External encapsulating structure	0.46%
Organelle	
Intracellular organelle	70.51%
Membrane-bound organelle	37.79%
Non-membrane bound organelle	34.56%
Organelle membrane	5.07%
Organelle lumen	0.92%
Macromolecular complex	
Protein complex	23.50%
Ribonucleoprotein complex	7.83%
Envelope	
Organelle envelope	4.61%
Extracellular region	
Extracellular space	0.92%
Membrane-enclosed lumen	
Organelle lumen	0.92%

Table 3.7: Table representing the absolute frequencies of Level 3 annotations under cellular component for the male subtractive library

3.5.1.3 Comparison of cellular component gene ontology annotation between mxA and maS libraries

A doughnut chart displaying comparative information can be seen in Figure 3.13 and a bar chart displaying a comparison of the absolute frequencies is shown in Figure 3.14.

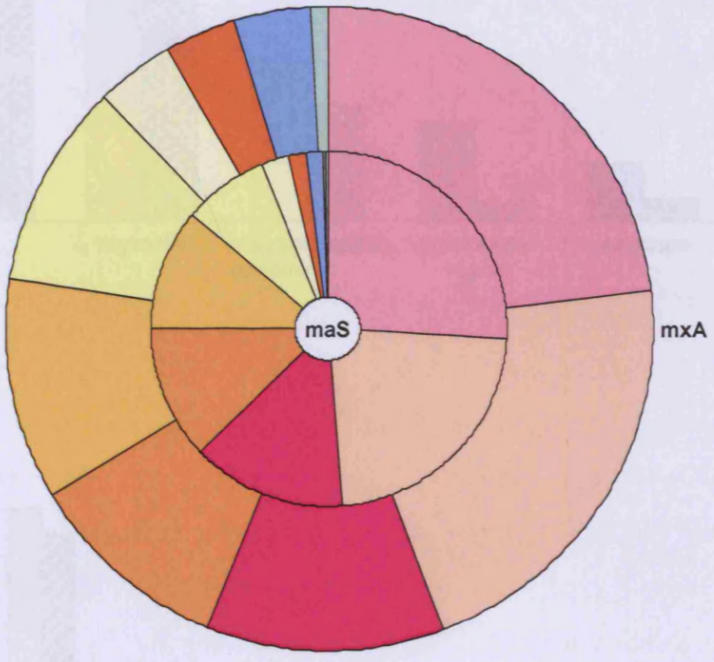


Figure 3.13: Top 99.5% cellular component gene ontology annotation using relative annotation frequencies

mxA = mixed library; maS = male subtractive library. Only gene fragments with "known" ontology were included in the chart.

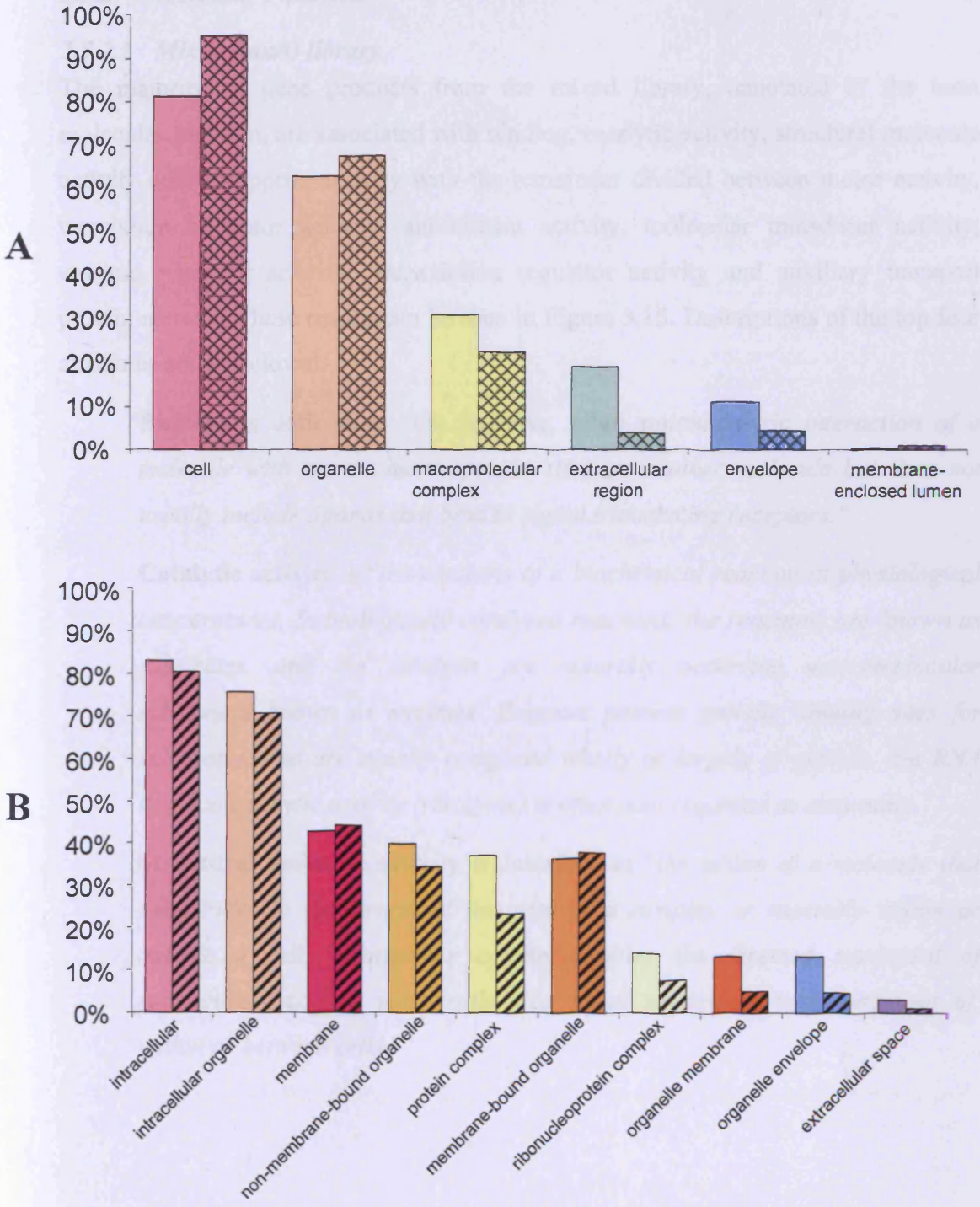


Figure 3.14: Comparison of the mixed and male subtractive libraries under cellular component for Level 2 and 3.

The bar charts represent the absolute frequencies of annotation. Open bars represent the mixed library; hatched bars represent the male subtractive library. A: Comparison of Level 2 annotations. B: Comparison of Level 3 annotations. Only gene fragments with "known" ontology were included in the chart.

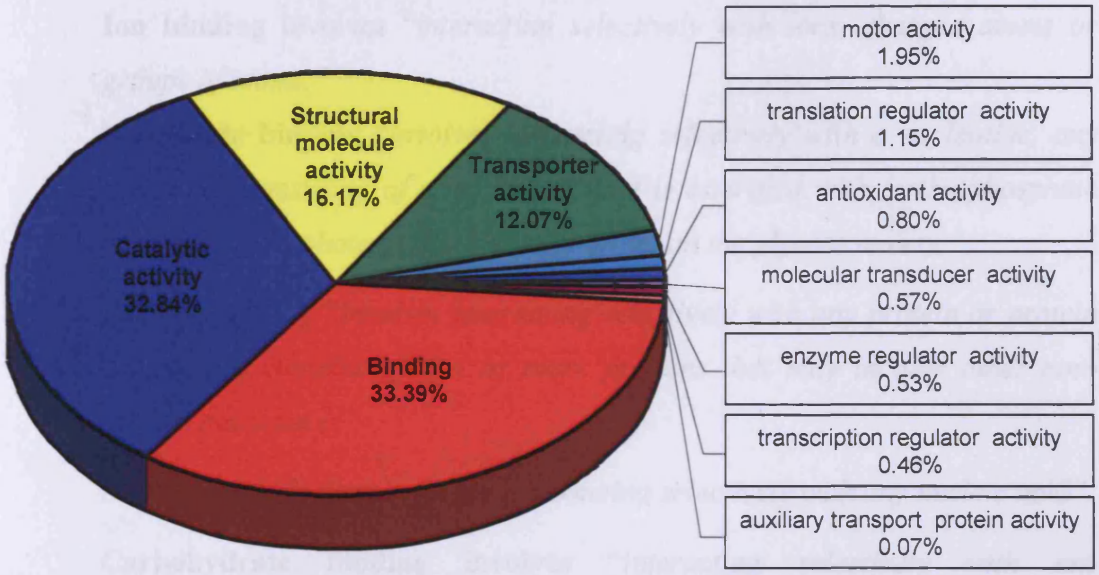
3.5.2 Molecular Function

3.5.2.1 Mixed (mxA) library

The majority of gene products from the mixed library, annotated to the term molecular function, are associated with binding, catalytic activity, structural molecule activity and transporter activity with the remainder divided between motor activity, translation regulator activity, antioxidant activity, molecular transducer activity, enzyme regulator activity, transcription regulator activity and auxiliary transport protein activity. These results can be seen in Figure 3.15. Descriptions of the top four attributes are as follows[1];

> **Binding** is defined as "*the selective, often stoichiometric interaction of a molecule with one or more specific sites on another molecule but does not usually include ligands that bind to signal transducing receptors.*"
>
> **Catalytic activity** is "*the catalysis of a biochemical reaction at physiological temperatures. In biologically catalysed reactions, the reactants are known as substrates, and the catalysts are naturally occurring macromolecular substances known as enzymes. Enzymes possess specific binding sites for substrates, and are usually composed wholly or largely of protein, but RNA that has catalytic activity (ribozyme) is often also regarded as enzymatic.*"
>
> **Structural molecule activity** is described as "*the action of a molecule that contributes to the structural integrity of a complex or assembly within or outside a cell. Transporter activity enables the directed movement of substances (such as macromolecules, small molecules, ions) into, out of, within or between cells.*"

[1] Definitions taken from Gene Ontology Consortium (http://www.geneontology.org)



Term name	Percent gene products with molecular function annotation
Binding	51.62%
Catalytic Activity	50.77%
Structural Molecule Activity	24.99%
Transporter Activity	18.67%
Motor Activity	3.01%
Translation Regulator Activity	1.78%
Antioxidant Activity	1.23%
Molecular Transducer Activity	0.89%
Enzyme Regulator Activity	0.82%
Transcription Regulator Activity	0.72%
Auxiliary Transport Protein Activity	0.10%

Figure 3.15: Mixed library Level 2 annotation under molecular function.
Pie chart representing the relative frequency of Level 2 annotations under molecular function for the mixed library. The table shows the absolute frequencies for the same gene set. Only gene fragments with "known" ontology were included in the chart.

Four of these attributes; binding, catalytic activity, structural molecule activity and transporter activity, were further divided as shown in Figure 3.16. Binding was divided into twelve subgroups; ion, nucleotide, protein, nucleic acid, carbohydrate, lipid, cofactor, metal cluster, vitamin and peptide binding.

The top four attributes are described in more detail below[1];

> **Ion binding** involves "*interacting selectively with ions, charged atoms or groups of atoms.*"
>
> **Nucleotide binding** "*involves interacting selectively with a nucleotide, any compound consisting of a nucleoside that is esterified with (ortho)phosphate or an oligophosphate at any hydroxyl group on the glycose moiety.*"
>
> **Protein binding** "*involves interacting selectively with any protein or protein complex (a complex of two or more proteins that may include other non-protein molecules).*"
>
> **Nucleic acid binding** involves "*interacting selectively with any nucleic acid*".
>
> **Carbohydrate binding** involves "*interacting selectively with any carbohydrate.*"

Catalytic activity was divided between nine sublevels; hydrolase, oxidoreductase, transferase, lyase, ligase, helicase, isomerase, deaminase and integrase activity. The top three attributes are described[1];

> **Hydrolase activity** is the "*catalysis of the hydrolysis of various bonds. Hydrolase is the systematic name for any enzyme of EC class 3.*"
>
> **Oxidoreductase activity** is the "*catalysis of an oxidation-reduction (redox) reaction, a reversible chemical reaction in which the oxidation state of an atom or atoms within a molecule is altered. One substrate acts as a hydrogen or electron donor and becomes oxidised, while the other acts as hydrogen or electron acceptor and becomes reduced.*"
>
> **Transferase activity** is the "*catalysis of the transfer of a group from one compound (generally regarded as the donor) to another compound (generally regarded as the acceptor). Transferase is the systematic name for any enzyme of EC class 2.*"

Structural molecule activity was divided between four sublevels; structural constituent of cuticle, structural constituent of ribosome and, to a lesser extent, extracellular matrix structural constituent and structural constituent of cell wall. The top two attributes are described[1];

[1] Definitions taken from Gene Ontology Consortium (http://www.geneontology.org)

The **structural constituent of cuticle** is defined as "*the action of a molecule that contributes to the structural integrity of a cuticle.*"

The **structural constituent of ribosome** is defined as "*the action of a molecule that contributes to the structural integrity of the ribosome and may be used to annotate ribosomal RNAs as well as ribosomal proteins.*"

Transporter activity was divided between six sublevels; ion transporter activity, oxygen transporter activity, carrier activity, ATPase activity coupled to movement of substances, carbohydrate transporter activity and channel or pore class transporter activity. The top two attributes are described[1];

Ion transporter activity "*enables the directed movement of charged atoms or small charged molecules into, out of, within or between cells.*"

Oxygen transporter activity "*enables the directed movement of oxygen into, out of, within or between cells.*"

[1] Definitions taken from Gene Ontology Consortium (http://www.geneontology.org)

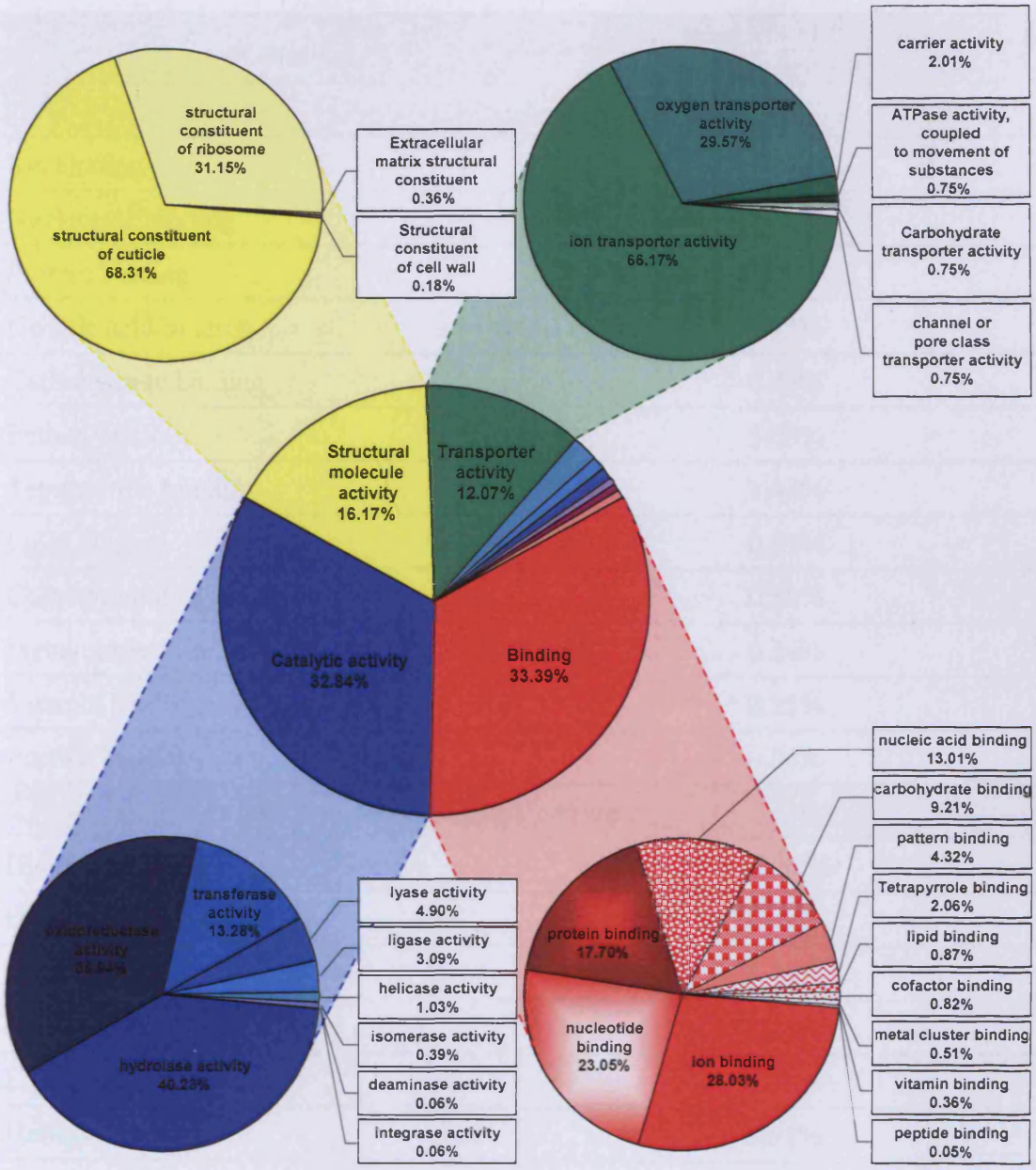


Figure 3.16: Mixed Library molecular function gene ontology annotation
Only gene fragments with "known" ontology were included in the chart.

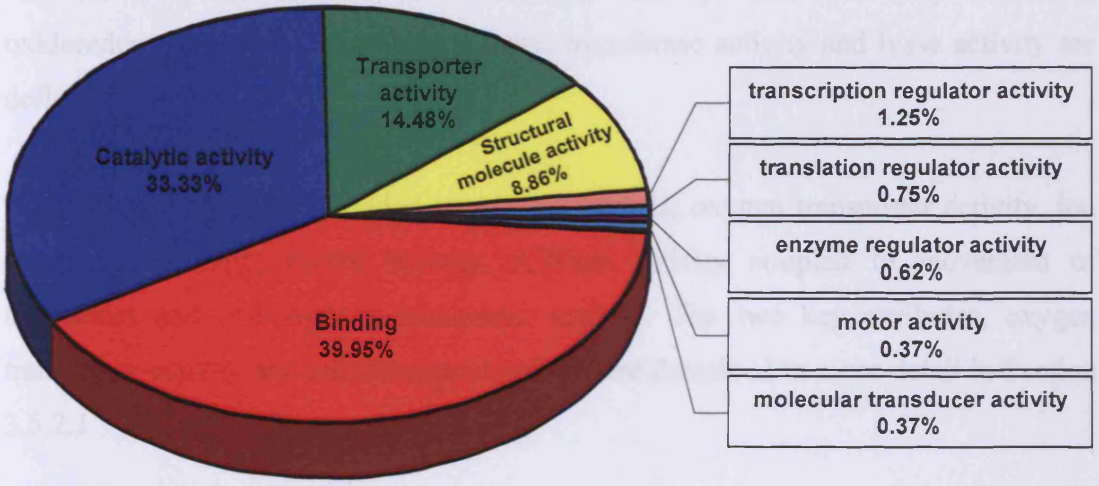
Term name	% gene products with molecular function annotation
Binding	
Ion binding	19.68%
Nucleotide binding	16.17%
Protein binding	12.42%
Nucleic acid binding	9.13%
Carbohydrate binding	6.46%
Pattern binding	3.03%
Tetrapyrrole binding	1.44%
Lipid binding	0.61%
Cofactor binding	0.58%
Metal cluster binding	0.36%
Vitamin binding	0.25%
Peptide binding	0.04%
Catalytic activity	
Hydrolase activity	22.52%
Oxidoreductase activity	20.69%
Transferase activity	7.44%
Lyase activity	2.74%
Ligase activity	1.73%
Helicase activity	0.58%
Isomerase activity	0.22%
Deaminase activity	0.04%
Integrase activity	0.04%
Structural molecule activity	
Structural constituent of cuticle	13.54%
Structural constituent of ribosome	6.17%
Extracellular matrix structural constituent	0.07%
Structural constituent of cell wall	0.04%

Transporter activity	
Ion transporter activity	9.53%
Oxygen transporter activity	4.26%
Carrier activity	0.29%
ATPase activity, coupled to movement of substances	0.11%
Carbohydrate transporter activity	0.11%
Channel or pore class transporter activity	0.11%
Motor activity	
Microtubule motor activity	0.36%
Translation regulator activity	
Translation factor activity, nucleic acid binding	1.88%
Antioxidant activity	
Peroxidase activity	1.30%
Molecular transducer activity	
Signal transducer activity	0.94%
Enzyme regulator activity	
Enzyme inhibitor activity	0.61%
GTPase regulator activity	0.22%
Ornithine decarboxylase regulator activity	0.04%
Transcription regulator activity	
Transcription cofactor activity	0.04%
Transcription factor activity	0.65%
Transcription activator activity	0.04%
Auxiliary protein activity	
Channel regulator activity	0.11%

Table 3.8: Table representing the absolute frequencies of Level 3 annotations under molecular function for mixed library

3.5.2.2 Male subtractive (maS) library

The majority of gene products from the male subtractive library, annotated to the term molecular function, are associated with binding, catalytic activity and transporter activity, with the remainder divided between structural molecule, transcription regulator, translation regulator, enzyme regulator, motor and molecular transducer activity. This can be seen in Figure 3.17. Definitions of the four key attributes have been previously described in Section 3.5.2.1.



Term name	Percent gene products with molecular function annotation
Binding	58.84%
Catalytic activity	47.42%
Transporter activity	20.60%
Structural molecule activity	12.61%
Transcription regulator activity	1.78%
Translation regulator activity	1.07%
Enzyme regulator activity	0.89%
Motor activity	0.53%
Molecular transducer activity	0.53%

Figure 3.17: Male subtractive library Level 2 annotation under molecular function.

Pie chart representing the relative frequency of Level 2 annotations under molecular function for the male subtractive library. The table shows the absolute frequencies for the same gene set. Only gene fragments with "known" ontology were included in the chart.

Four of the attributes, binding, catalytic activity, transporter activity and structural molecule activity were further divided as described in Figure 3.18 and Table 3.9. Gene objects involved in binding were divided into eleven sublevels; nucleotide-, protein-, ion-, nucleic acid-, carbohydrate-, cofactor-, pattern-, tetrapyrrole-, vitamin-, lipid- and steroid-binding. Definitions of the five key attributes, nucleotide binding, protein binding, ion binding, nucleic acid binding and carbohydrate binding are previously described in Section 3.5.2.1.

Catalytic activity was divided into six sublevels; oxidoreductase, hydrolase, transferase, lyase, helicase and isomerase activity. The four key attributes, oxidoreductase activity, hydrolase activity, transferase activity and lyase activity are defined in Section 3.5.2.1

Transporter activity was divided into five sublevels; oxygen transporter activity, ion transporter activity, carrier activity, ATPase activity coupled to movement of substances and carbohydrate transporter activity. The two key attributes, oxygen transporter activity and ion transporter activity are described in more detail in Section 3.5.2.1

Structural molecule activity was divided into two sublevels; structural constituent of ribosome and structural constituent of cuticle. Both of these attributes are described in more detail in Section 3.5.2.1.

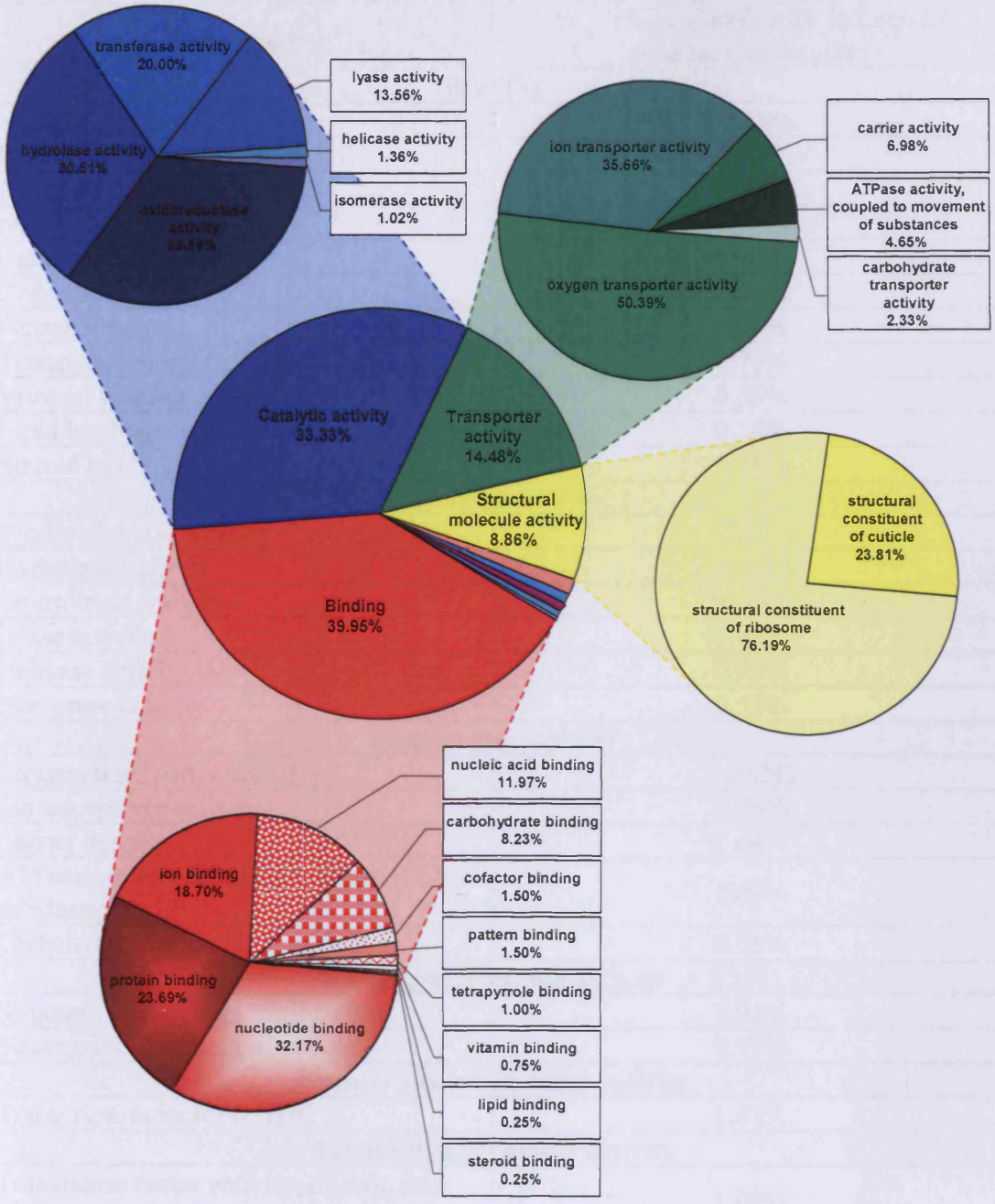


Figure 3.18: Level 3 male subtractive molecular function gene ontology annotation

Only gene fragments with "known" ontology were included in the chart.

Term name	% gene products with molecular function annotation
Binding	
Nucleotide binding	23.45%
Protein binding	17.27%
Ion binding	13.64%
Nucleic acid binding	8.73%
Carbohydrate binding	6.00%
Cofactor binding	1.09%
Pattern binding	1.09%
Tetrapyrrole binding	0.73%
Vitamin binding	0.55%
Lipid binding	0.18%
Steroid binding	0.18%
Catalytic activity	
Oxidoreductase activity	18.00%
Hydrolase activity	16.36%
Transferase activity	10.73%
Lyase activity	7.27%
Helicase activity	0.73%
Isomerase activity	0.55%
Transporter activity	
Oxygen transporter activity	11.82%
Ion transporter activity	8.36%
Carrier activity	1.64%
ATPase activity, coupled to movement of substances	1.09%
Carbohydrate transporter activity	0.55%
Structural molecule activity	
Structural constituent of ribosome	2.91%
Structural constituent of cuticle	0.91%
Transcription regulator activity	
Transcription factor activity	1.82%
Translation regulator activity	
Translation factor activity, nucleic acid binding	1.09%
Enzyme regulator activity	
Enzyme inhibitor activity	0.91%
Molecular transducer activity	
Signal transducer activity	0.55%

Table 3.9: Table representing the absolute frequencies of Level 3 annotations under molecular function for male subtractive library

3.5.2.3 *Comparison of molecular function gene ontology annotation between mxA and maS libraries*

A doughnut chart displaying comparative information can be seen in Figure 3.19 and a bar chart displaying a comparison of the absolute frequencies is shown in Figure 3.20.

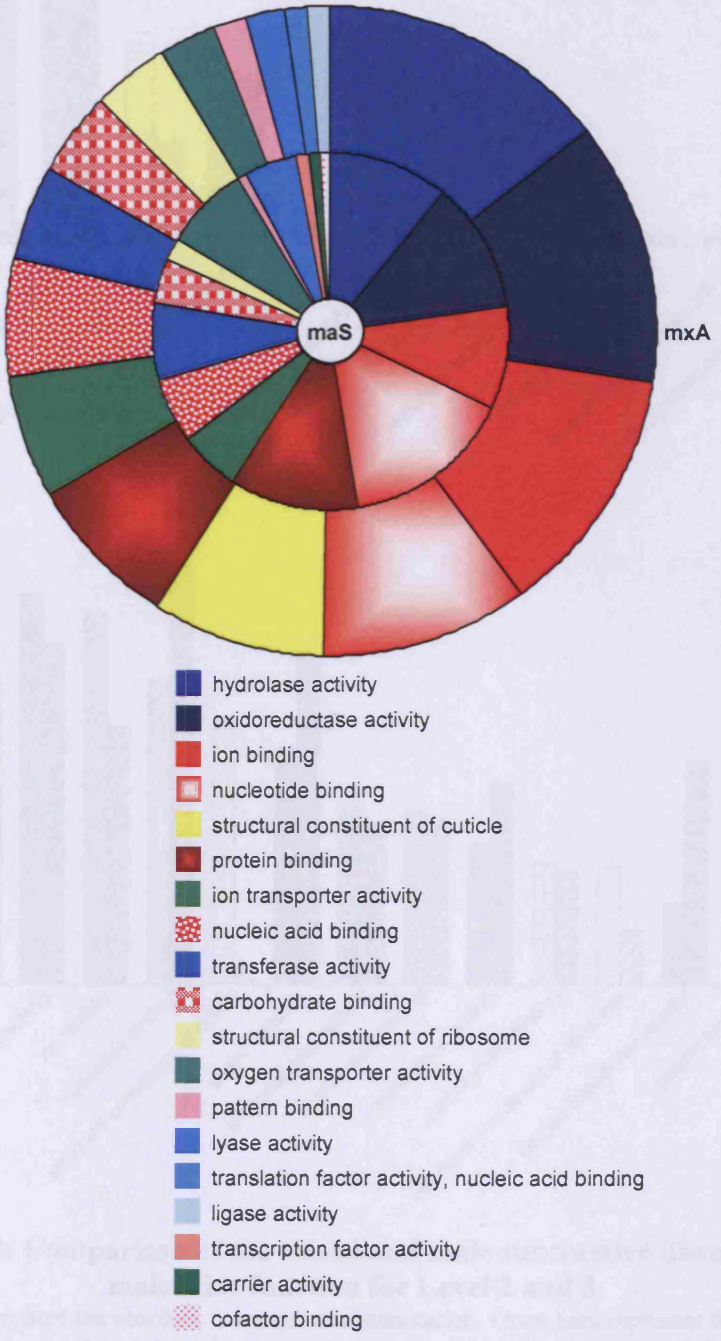


Figure 3.19: Top 95% molecular function gene ontology annotation
mxA = mixed library; maS = male subtractive library. Only gene fragments with "known" ontology were included in the chart. An absence of cuticle constituents are evident in the mixed library.

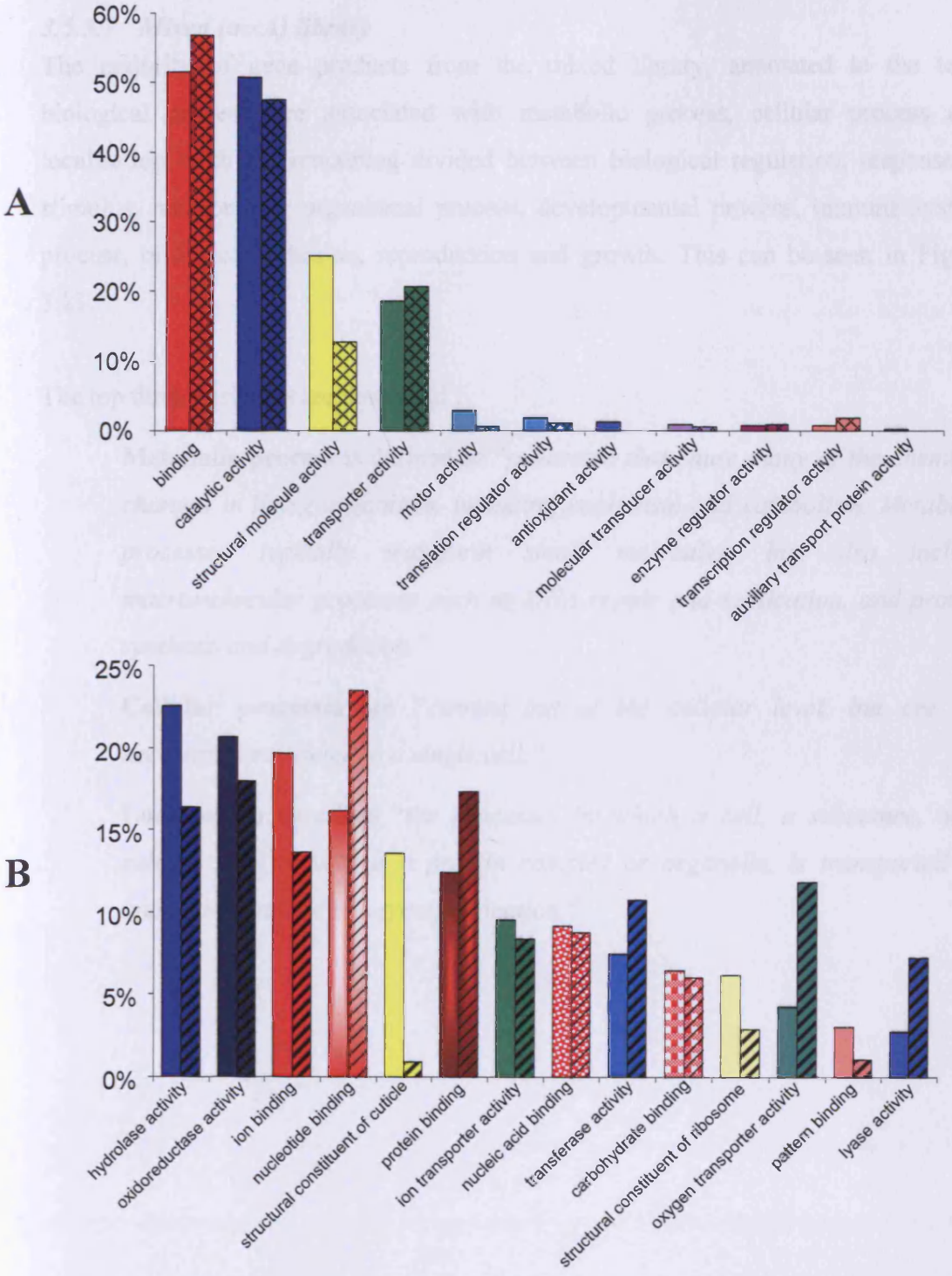


Figure 3.20: Comparison of the mixed and male subtractive libraries under molecular function for Level 2 and 3.

The bar charts represent the absolute frequencies of annotation. Open bars represent the mixed library; hatched bars represent the male subtractive library. A: Comparison of Level 2 annotations. B: Comparison of Level 3 annotations. Only gene fragments with "known" ontology were included in the chart.

3.5.3 Biological Process

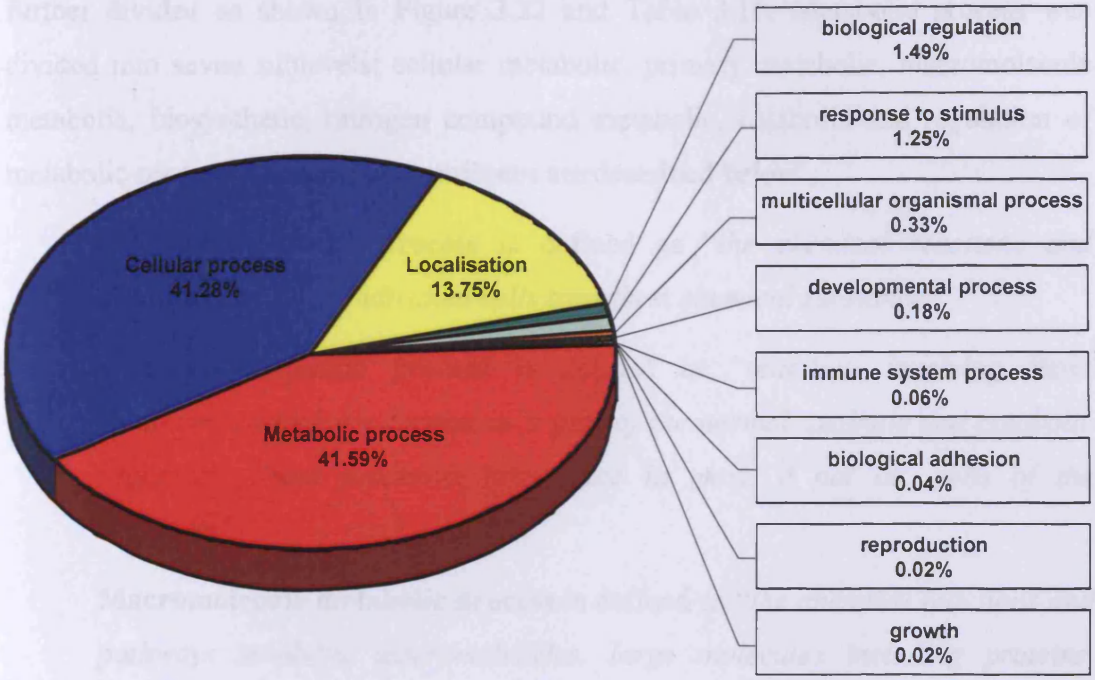
3.5.3.1 Mixed (mxA) library

The majority of gene products from the mixed library, annotated to the term biological process, are associated with metabolic process, cellular process and localisation, with the remaining divided between biological regulation, response to stimulus, multicellular organismal process, developmental process, immune system process, biological adhesion, reproduction and growth. This can be seen in Figure 3.21.

The top three attributes are described[1];

> **Metabolic process** is defined as *"processes that cause many of the chemical changes in living organisms, including anabolism and catabolism. Metabolic processes typically transform small molecules, but also include macromolecular processes such as DNA repair and replication, and protein synthesis and degradation."*
>
> **Cellular processes** are *"carried out at the cellular level, but are not necessarily restricted to a single cell."*
>
> **Localisation** describes *"the processes by which a cell, a substance, or a cellular entity, such as a protein complex or organelle, is transported to, and/or maintained in a specific location."*

[1] Definitions taken from Gene Ontology Consortium (http://www.geneontology.org)



Term name	Percent gene products with biological process annotation
Metabolic process	81.91%
Cellular process	81.31%
Localisation	27.07%
Biological regulation	2.94%
Response to stimulus	2.46%
Multicellular organismal process	0.64%
Developmental process	0.36%
Immune system process	0.12%
Biological adhesion	0.08%
Growth	0.04%
Reproduction	0.04%

Figure 3.21: Mixed library Level 2 annotation under biological process.
Pie chart representing the relative frequency of Level 2 annotations under biological process for the mixed library. The table shows the absolute frequencies for the same gene set. Only gene fragments with "known" ontology were included in the chart.

Three of these attributes; metabolic process, cellular process and localisation, were further divided as shown in Figure 3.22 and Table 3.10. Metabolic process was divided into seven sublevels; cellular metabolic, primary metabolic, macromolecule metabolic, biosynthetic, nitrogen compound metabolic, catabolic and regulation of metabolic process. The top three attributes are described below[1];

> **Cellular metabolic process** is defined as "*the chemical reactions and pathways by which individual cells transform chemical substances.*"
>
> **Primary metabolic process** is defined as "*reactions involving those compounds which are formed as a part of the normal anabolic and catabolic processes. These processes take place in most, if not all, cells of the organism.*"
>
> **Macromolecule metabolic process** is defined as "*the chemical reactions and pathways involving macromolecules, large molecules including proteins, nucleic acids and carbohydrates.*"

Cellular process was divided into fifteen sublevels; cellular metabolic process, cell organisation and biogenesis, cell homeostasis, cell communication, regulation of cellular process, cell motility, cell cycle, detection of stimulus, cell adhesion, cell cycle process, cell development, cellular developmental process, cell growth, cell proliferation and membrane docking. The major attribute, cellular metabolic process is also a daughter term of metabolic process and the definition explained above.

Localisation was divided into four sublevels; establishment of localisation, cellular localisation, localisation of cell and macromolecule localisation.

> The major annotation, **establishment of localisation** is defined as "*the directed movement of a cell, substance or cellular entity, such as a protein complex or organelle, to a specific location.*"

[1] Definitions taken from Gene Ontology Consortium (http://www.geneontology.org)

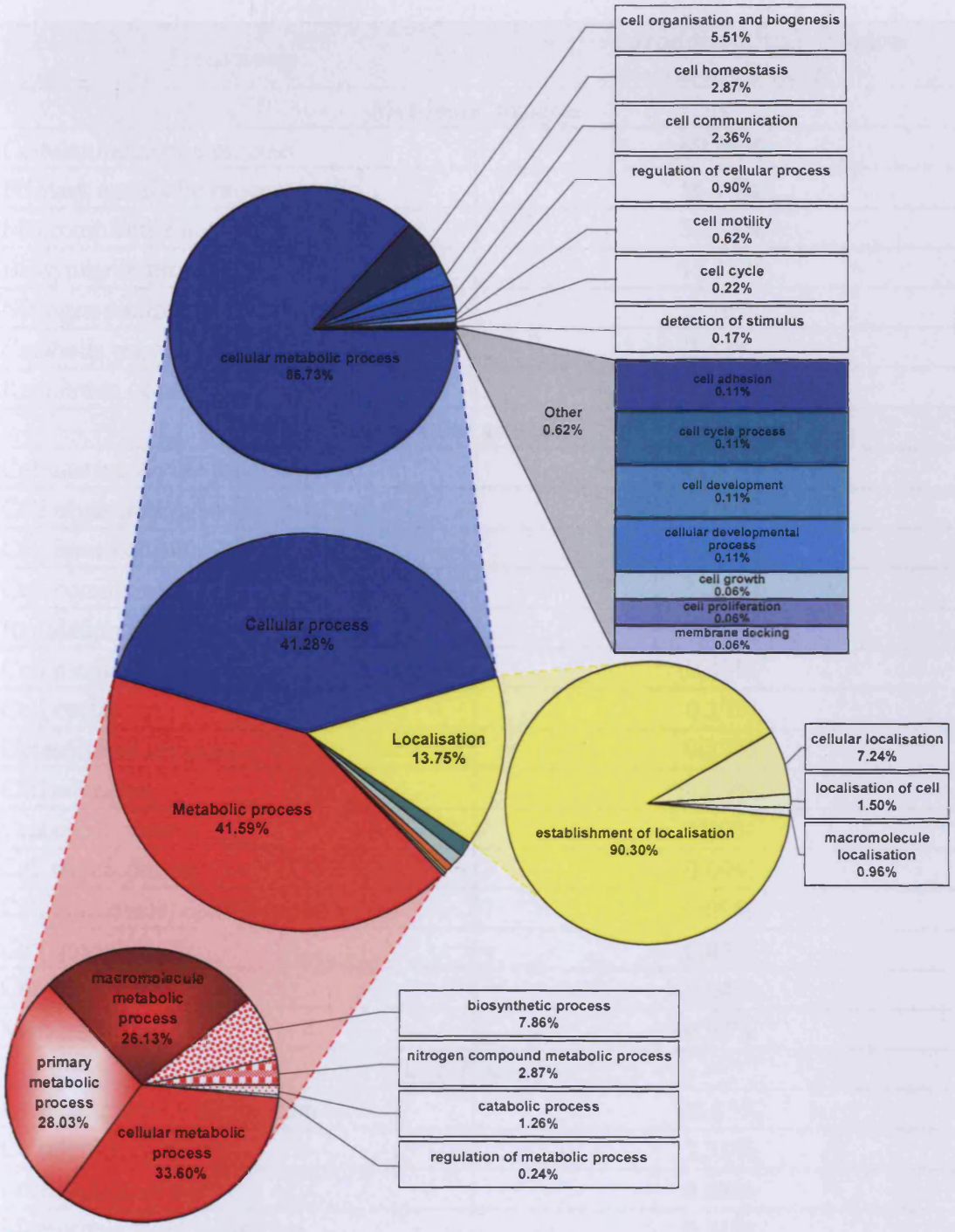


Figure 3.22: Level 3 mixed library biological process gene ontology annotation
Only gene fragments with "known" ontology were included in the chart.

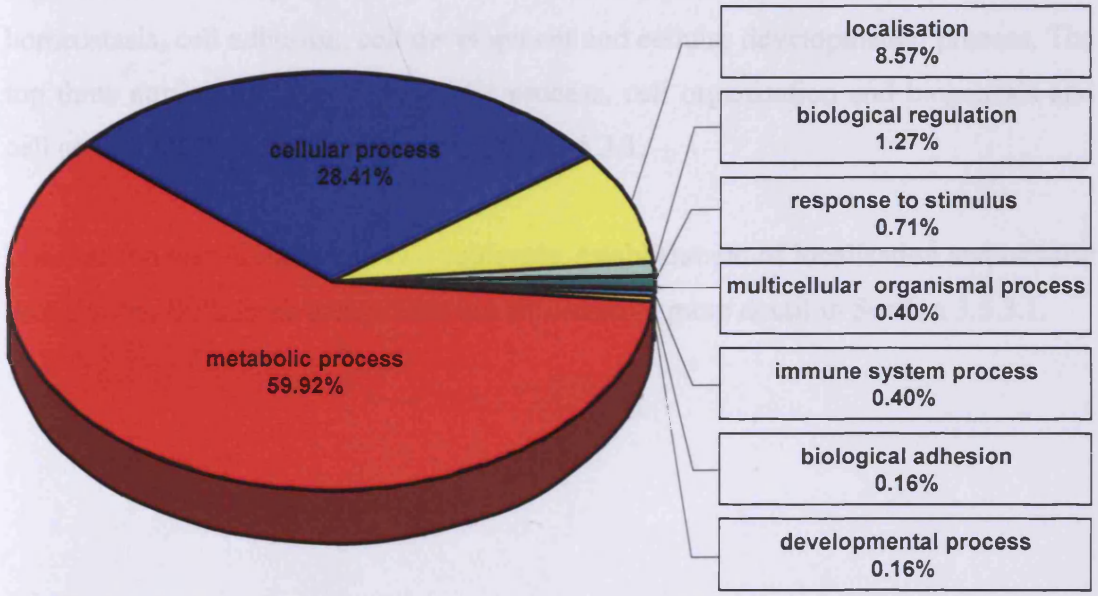
Term name	% gene products with biological process annotation
Metabolic process	
Cellular metabolic process	67.35%
Primary metabolic process	56.18%
Macromolecule metabolic process	52.38%
Biosynthetic process	15.76%
Nitrogen compound metabolic process	5.76%
Catabolic process	2.53%
Regulation of metabolic process	0.48%
Cellular process	
Cellular metabolic process	67.35%
Cell organisation and biogenesis	4.28%
Cell homeostasis	2.23%
Cell communication	1.83%
Regulation of cellular process	0.70%
Cell motility	0.48%
Cell cycle	0.17%
Detection of stimulus	0.13%
Cell adhesion	0.09%
Cell cycle process	0.09%
Cell development	0.09%
Cellular developmental process	0.09%
Cell growth	0.04%
Cell proliferation	0.04%
Membrane docking	0.04%
Localisation	
Establishment of localisation	28.85%
Cellular localisation	2.31%
Localisation of cell	0.48%
Macromolecule localisation	0.31%
Biological regulation	
Regulation of biological quality	2.27%
Regulation of biological process	0.83%
Regulation of molecular function	0.13%

Response to stimulus	
Response to chemical stimulus	1.96%
Response to stress	1.83%
Response to biotic stimulus	0.65%
Defence response	0.44%
Behaviour	0.17%
Response to external stimulus	0.17%
Detection of stimulus	0.13%
Immune response	0.13%
Response to abiotic stimulus	0.13%
Response to endogenous stimulus	0.04%
Multicellular organismal process	
Multicellular organismal development	0.31%
Digestion	0.13%
Neurological process	0.13%
Developmental process	
Multicellular organismal development	0.31%
Post-embryonic development	0.22%
Anatomical structure development	0.17%
Cellular developmental process	0.09%
Death	0.09%
Anatomical structure morphogenesis	0.04%
Immune system process	
Immune response	0.13%
Immune effector process	0.04%
Biological adhesion	
Cell adhesion	0.09%
Growth	
Cell growth	0.04%
Regulation of growth	0.04%
Reproduction	
Sexual reproduction	0.04%

Table 3.10: Table representing the absolute frequencies of Level 3 annotations under biological process for the mixed library.

3.5.3.2 Male subtractive (maS) library

The majority of gene products from the male subtractive library, annotated to the term biological process, are associated with metabolic and cellular processes and localisation with a small quantity responsible for biological regulation, response to stimulus, multicellular organismal process, immune system process, biological adhesion and developmental process. This can be seen in Figure 3.23. The three key attributes, metabolic process, cellular process and localisation are defined in Section 3.5.3.1.



Term name	Percent gene products with biological process annotation
Metabolic process	89.24%
Cellular process	42.32%
Localisation	12.77%
Biological regulation	1.89%
Response to stimulus	1.06%
Multicellular organismal process	0.59%
Immune system process	0.59%
Biological adhesion	0.24%
Developmental process	0.24%

Figure 3.23: Male subtractive library Level 2 annotation under biological process.

Pie chart representing the relative frequency of Level 2 annotations under biological process for the male subtractive library. The table shows the absolute frequencies for the same gene set. Only gene fragments with "known" ontology were included in the chart.

Three of the attributes; metabolic process, cellular process and localisation, were further divided as shown in Figure 3.24 and Table 3.11. Within the metabolic process category, the majority were involved in cellular, primary and macromolecule metabolic processes, with the remaining divided between catabolic, regulation of metabolic and nitrogen compound metabolic processes. The definitions of the three key attributes, cellular metabolic process, primary metabolic process and macromolecule metabolic process, are described in Section 3.5.3.1.

Cellular process was divided into eight sublevels; cellular metabolic process, cell organisation and biogenesis, regulation of cellular process, cell communication, cell homeostasis, cell adhesion, cell development and cellular developmental process. The top three attributes, cellular metabolic process, cell organisation and biogenesis and cell communication, are defined in Section 3.5.3.1.

Localisation was divided into two sublevels; establishment of localisation and cellular localisation. Both these annotations are described in more detail in Section 3.5.3.1.

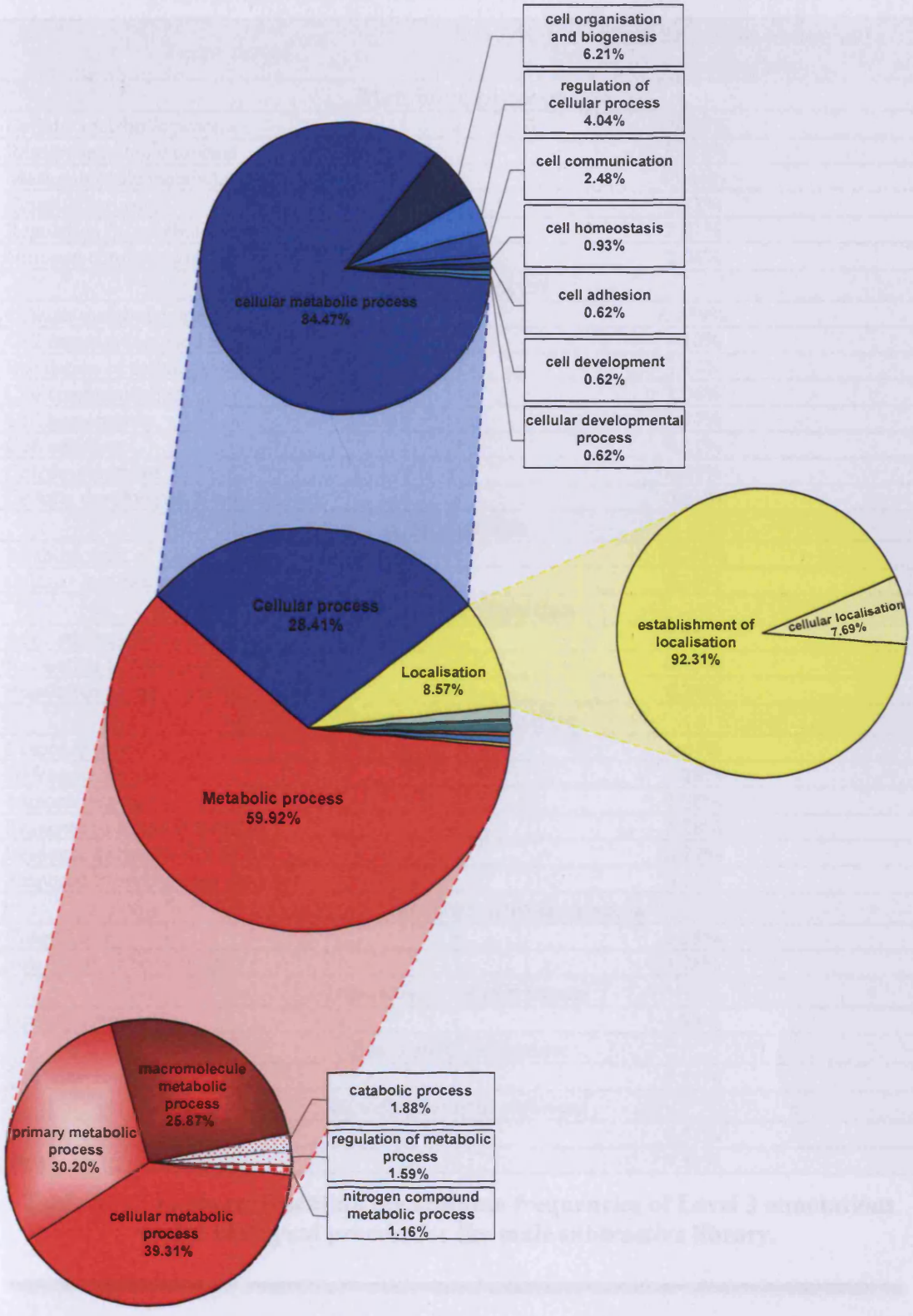


Figure 3.24: Level 3 male subtractive library biological process gene ontology annotation

Only gene fragments with "known" ontology were included in the chart.

Term name	% gene products with biological process annotation
Metabolic process	
Cellular metabolic process	69.39%
Primary metabolic process	53.32%
Macromolecule metabolic process	45.66%
Catabolic process	3.32%
Regulation of metabolic process	2.81%
Nitrogen compound metabolic process	2.04%
Cellular process	
Cellular metabolic process	69.39%
Cell organisation and biogenesis	5.10%
Regulation of cellular process	3.32%
Cell communication	2.04%
Cell homeostasis	0.77%
Cell adhesion	0.51%
Cell development	0.51%
Cellular developmental process	0.51%
Localisation	
Establishment of localisation	27.55%
Cellular localisation	2.30%
Biological regulation	
Regulation of biological process	3.32%
Regulation of biological quality	0.77%
Regulation of molecular function	0.26%
Response to stimulus	
Response to stress	2.04%
Defence response	1.28%
Immune response	1.28%
Response to external stimulus	1.28%
Response to chemical stimulus	0.77%
Response to endogenous stimulus	0.51%
Multicellular organismal process	
Coagulation	1.28%
Regulation of body fluids	1.28%
Immune system process	
Immune response	1.28%
Biological adhesion	
Cell adhesion	0.51%
Developmental process	
Cellular developmental process	0.51%
Death	0.51%

Table 3.11: Table representing the absolute frequencies of Level 3 annotations under biological process for the male subtractive library.

3.5.3.3 Comparison of biological process gene ontology annotation between mxA and maS libraries

A doughnut chart displaying comparative information can be seen in Figure 3.25 and a bar chart displaying a comparison of the absolute frequencies is shown in Figure 3.26.

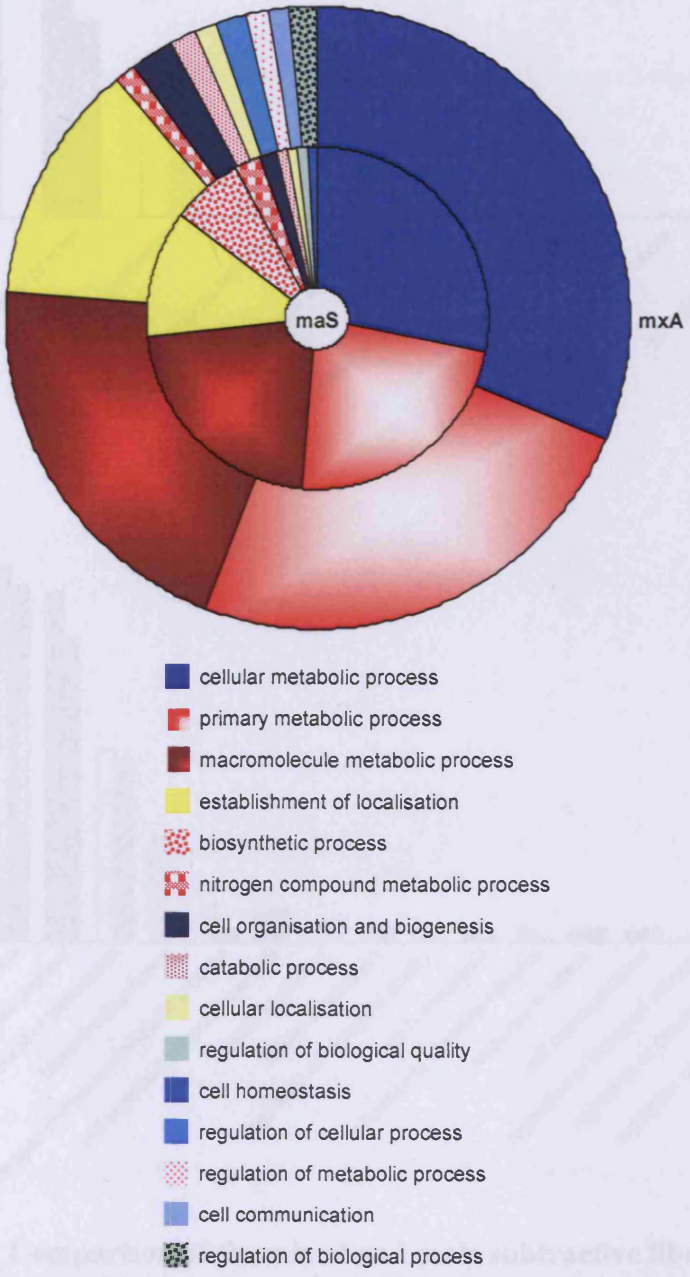


Figure 3.25: Top 95% biological process gene ontology annotation

mxA = mixed library; maS = male subtractive library. Only gene fragments with "known" ontology were included in the chart. The representation of biosynthetic processes is evident on the male subtractive library and a higher representation of regulatory-related gene objects are observed in the mixed library.

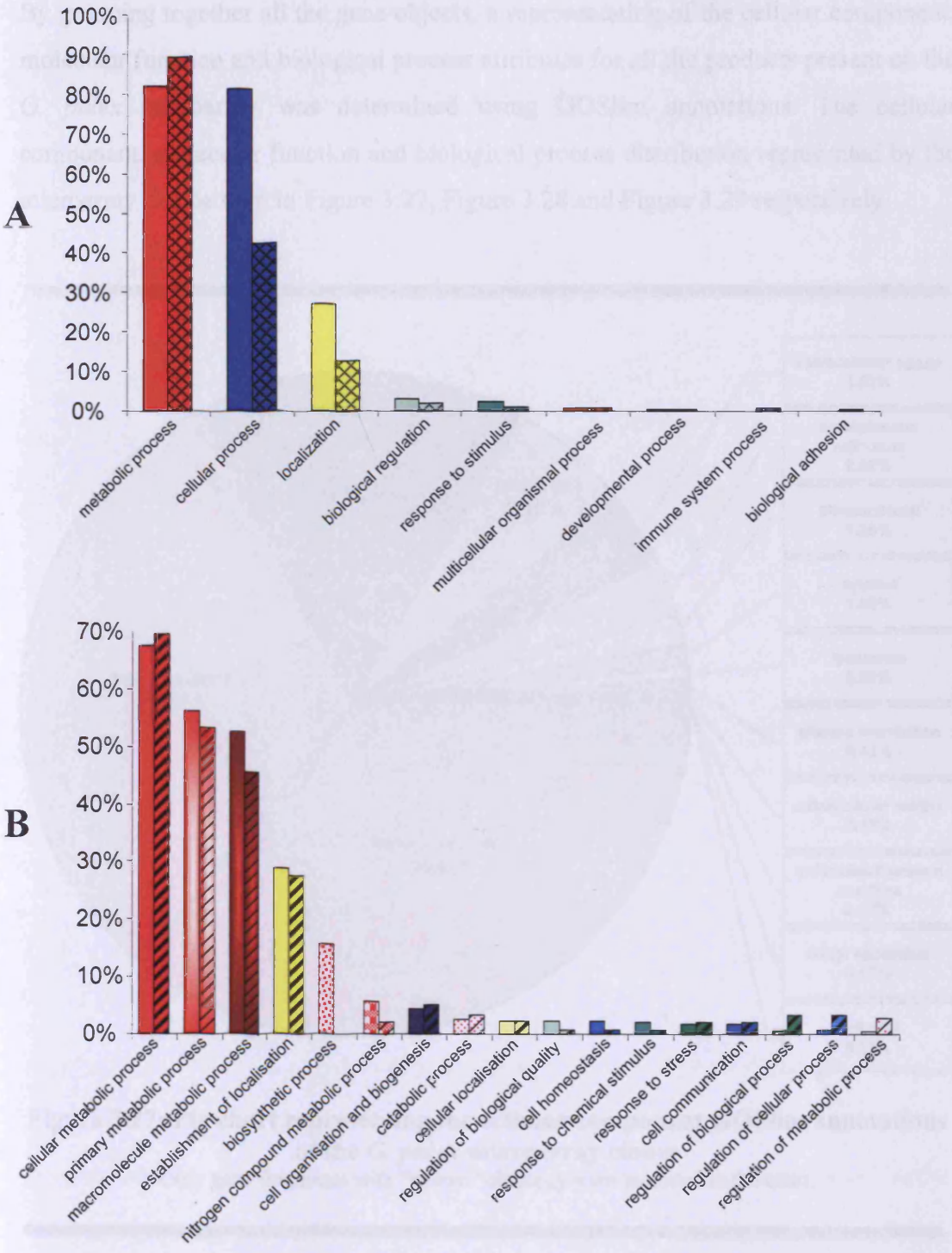


Figure 3.26: Comparison of the mixed and male subtractive libraries under biological process for Level 2 and 3.

The bar charts represent the absolute frequencies of annotation. Open bars represent the mixed library; hatched bars represent the male subtractive library. A: Comparison of Level 2 annotations. B: Comparison of Level 3 annotations. Only gene fragments with "known" ontology were included in the chart.

3.5.4 Microarray representation (GOSlim)

By grouping together all the gene objects, a representation of the cellular component, molecular function and biological process attributes for all the products present on the *G. pulex* microarray was determined using GOSlim annotations. The cellular component, molecular function and biological process distribution represented by the microarray can be seen in Figure 3.27, Figure 3.28 and Figure 3.29 respectively.

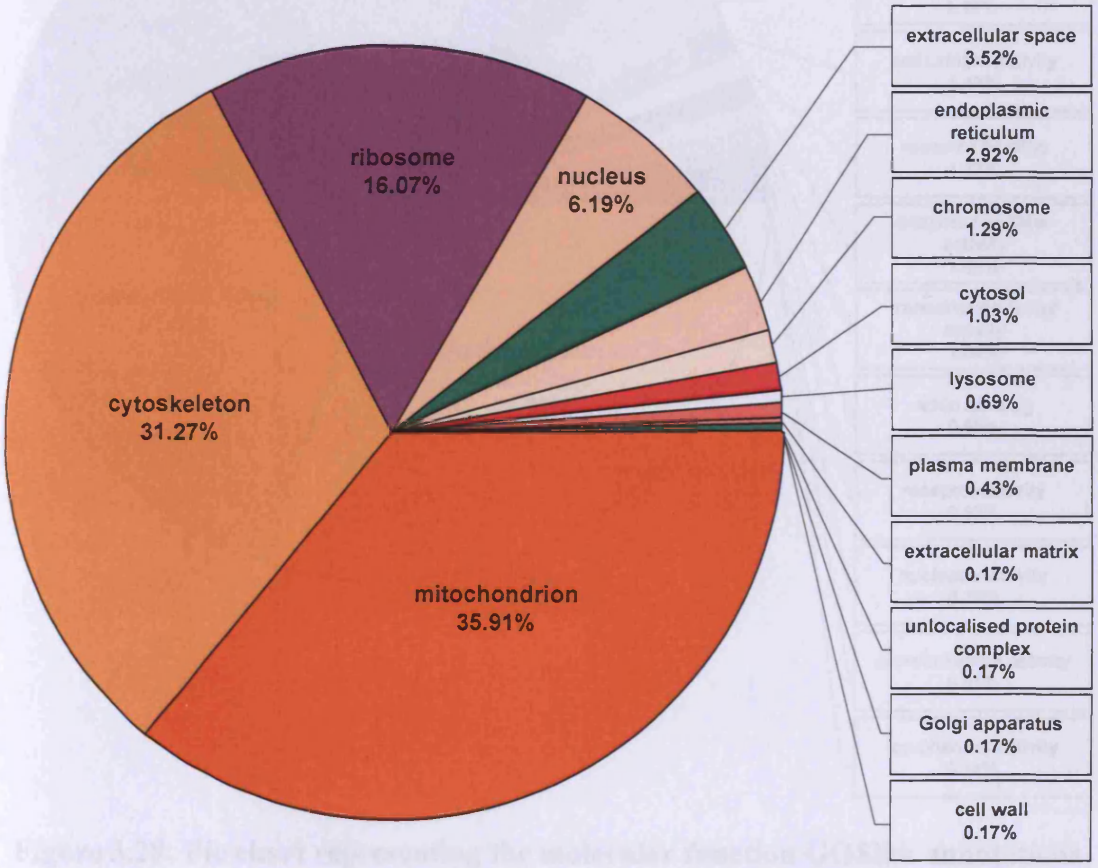


Figure 3.27: Pie chart representing the cellular component GOSlim annotations of the *G. pulex* microarray clones

Only gene fragments with "known" ontology were included in the chart.

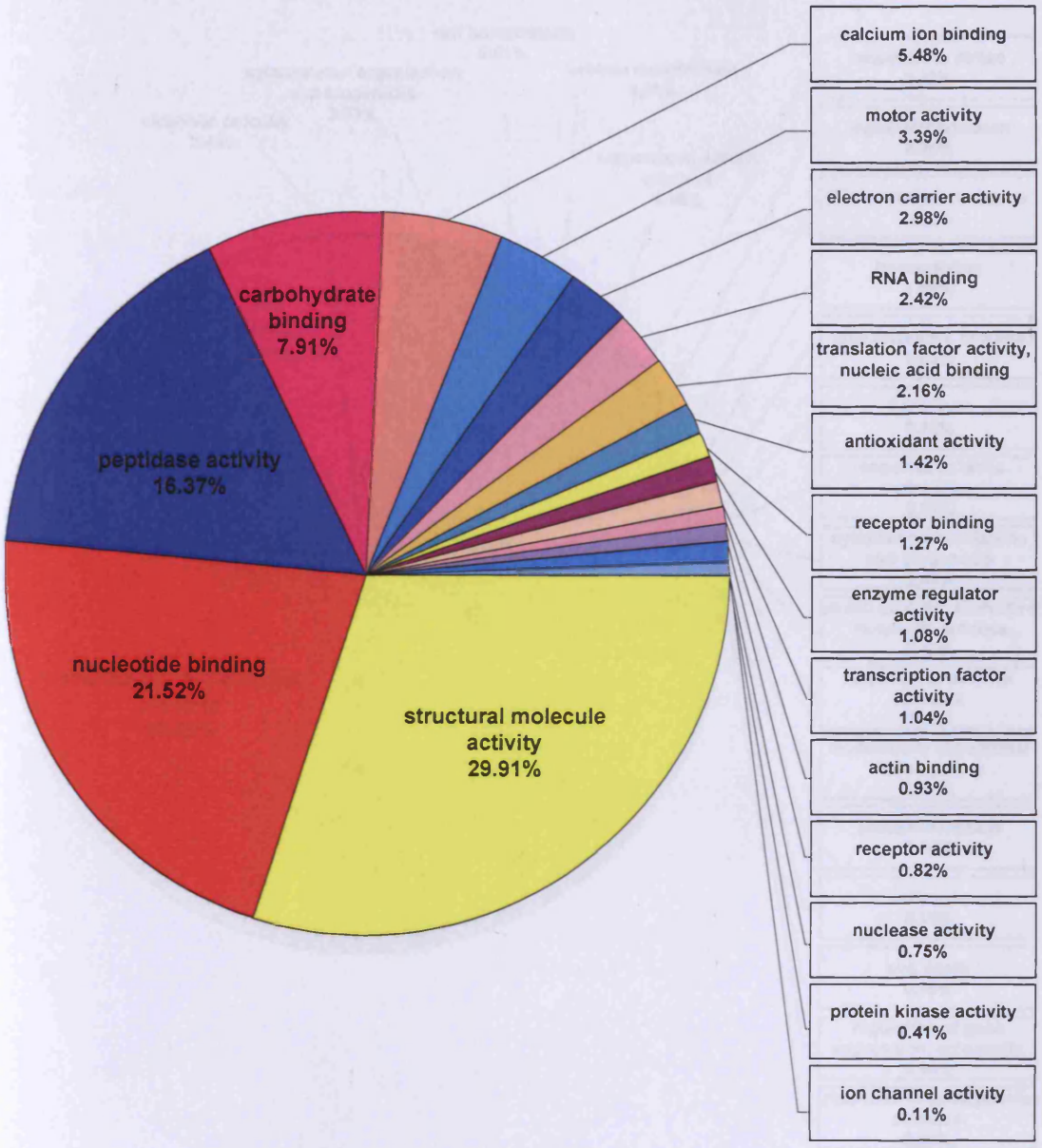


Figure 3.28: Pie chart representing the molecular function GOSlim annotations of the *G. pulex* microarray clones

Only gene fragments with "known" ontology were included in the chart.

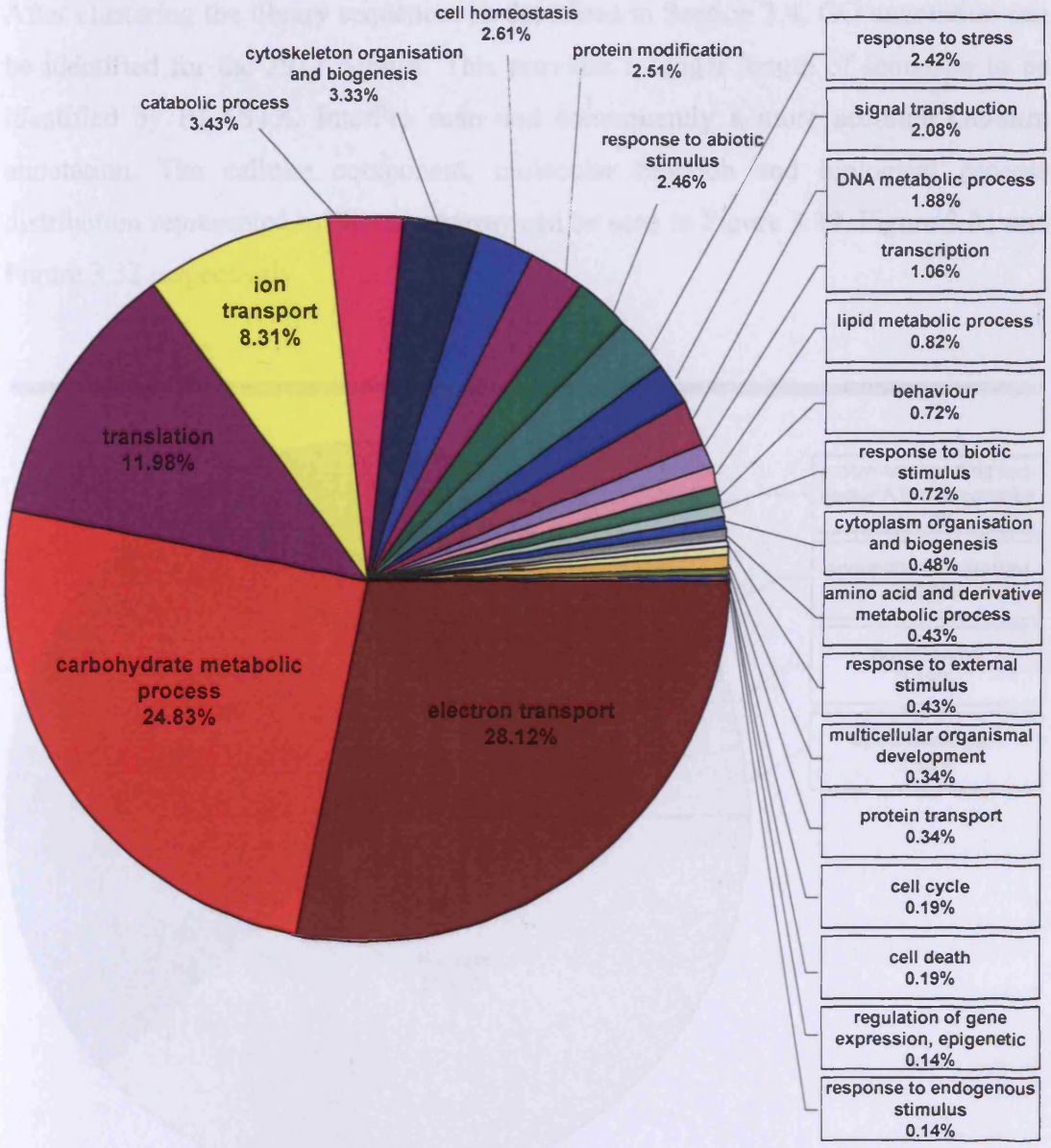


Figure 3.29: Pie chart representing the biological process GOSlim annotations of the *G. pulex* microarray clones

Only gene fragments with "known" ontology were included in the chart.

3.5.5 Contig representation

After clustering the library sequences as described in Section 3.4, GO annotation can be identified for the 3917 contigs. This provides a longer length of sequence to be identified by BLASTX, InterPro scan and consequently a more accurate GOSlim annotation. The cellular component, molecular function and biological process distribution represented by the microarray can be seen in Figure 3.30, Figure 3.31 and Figure 3.32 respectively.

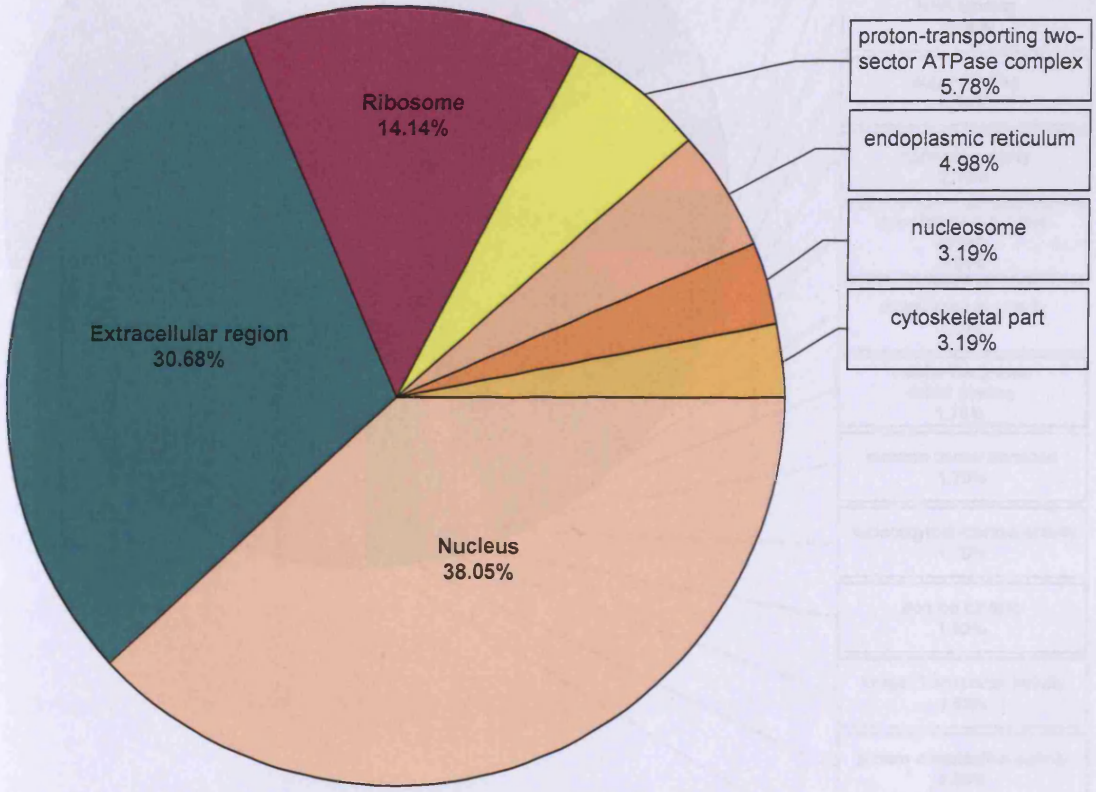


Figure 3.30: Pie chart representing the cellular component GOSlim annotations of the *G. pulex* microarray clusters

Only gene fragments with "known" ontology were included in the chart. Sequence number cut-off was set to 15.

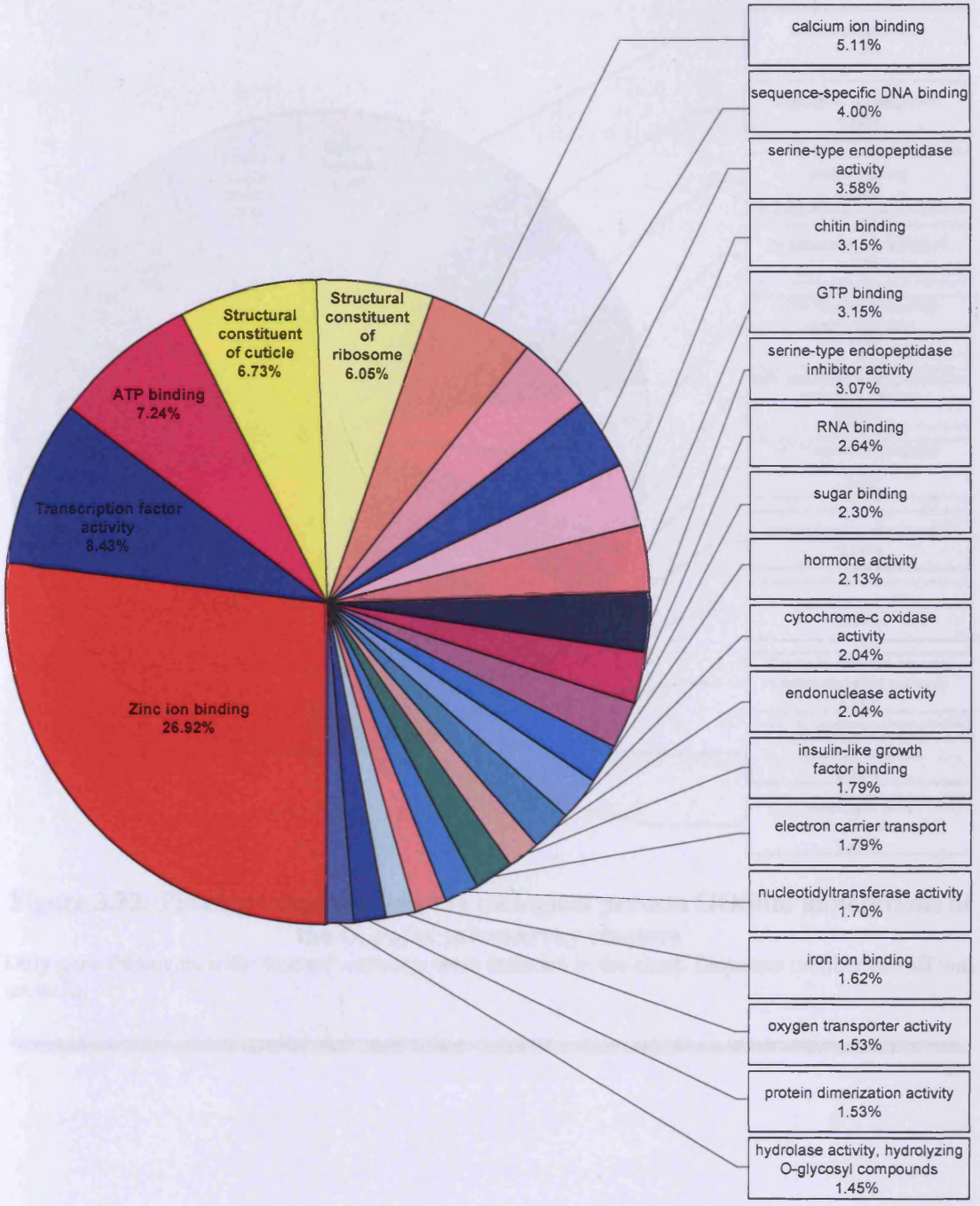


Figure 3.31: Pie chart representing the molecular function GOSlim annotations of the *G. pulex* microarray clusters

Only gene fragments with "known" ontology were included in the chart. Sequence number cut-off was set to 15.

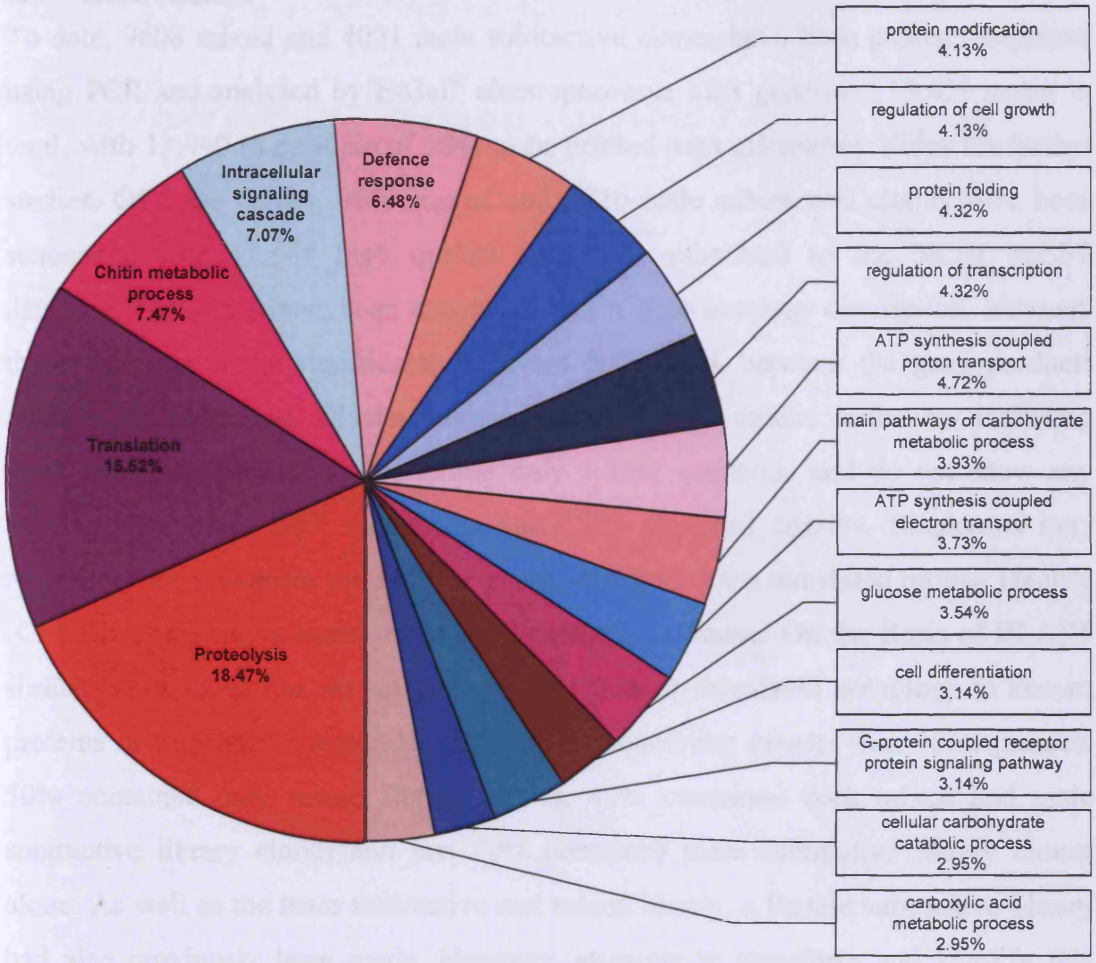


Figure 3.32: Pie chart representing the biological process GOSlim annotations of the *G. pulex* microarray clusters

Only gene fragments with "known" ontology were included in the chart. Sequence number cut-off was set to 15.

3.6 DISCUSSION

To date, 9608 mixed and 4021 male subtractive clones have been picked, amplified using PCR and analysed by E-Gel® electrophoresis. This generated 13,629 genes in total, with 13,440 (a multiple of 384) to be printed onto microarray slides for further studies. Of these clones, 9600 mixed and 3820 male subtractive clones have been sequenced with 12,664 high quality sequences submitted to the NCBI dbEST database. The genes have been annotated with a gene ontology description, although there were not many significantly relevant differences between the gene products annotation distribution. Cluster analysis revealed 3917 unique sequences including 2642 singletons ('clusters' containing only 1 EST sequence and do not show any overlap with other EST sequences) and 1275 clustered contigs. Singletons may represent EST sequences too small to accurately predict the translated protein identity (CLOBB clustering is based on BLAST similarity: **CL**uster **O**n the **B**asis of **BL**AST similarity) or those that are novel sequences with no translated homology to known proteins or translated nucleotides. Of clusters containing greater than 10 sequences, 50% contained only mixed library clones, 40% contained both mixed and male subtractive library clones and just 10% contained male subtractive library clones alone. As well as the male subtractive and mixed library, a female subtractive library had also previously been made. However, attempts to transform and amplify this library were unsuccessful.

3.6.1 cDNA library generation

A mixed cDNA library was created in order to represent all moult stages, developmental stages and both males and females. The male subtractive library was generated to try and identify male-specific genes that could be used as a biomarker of gender. Vitellogenin had already been successfully confirmed as a positive indicator for females; however, there was no similar male biomarker in common usage. The most likely candidate for a male-specific biomarker would be a gene expressed in response to the development or regulation of the androgenic gland and the secreted androgenic gland hormone. Some preliminary studies in male crayfish have revealed an insulin-like gene that is only expressed in the androgenic gland, therefore being a possible candidate for this role (Manor *et al.* 2007). Although sequencing both libraries did not identify any obvious potential male-specific biomarkers, by spotting

the library amplicons onto microarray slides, gene expression profiles for male and female *G. pulex* can be generated and compared, leading to novel male biomarker discovery.

There has been an exceedingly high success rate for both the mixed and male subtractive library with respect to number of clones picked compared to number of positive inserts. The success rate of the maS libraries is lower than that of the mixed library due to the nature of the different library methods and because the transformants were picked from non-selective agar plates. The subtractive library produces a range of PCR products with non-specific ends that are ligated in a random orientation into pGEM-T vector, whereas the mixed library produces double stranded cDNA with specific *Eco*RI and *Xho*I restriction sites at each end and are ligated directionally into pBluescript® SKII+. The relatively lower level of success in the male subtractive library is therefore due to the nature of the subtraction protocol, giving lower ligation efficiency and a shorter insert length.

The reason behind the low titre of mixed library mxB was later identified (Miss Jane Andre, personal communication). After similarly low titre for a *Lumbricus rubellus* cDNA library, Miss Andre conducted further studies; the cDNA size separation step using Sepharose chromatography was found to be producing low yields of cDNA. Repeated elutions were performed and the fractions analysed by electrophoresis. Despite manufacturer's instructions stating the desired cDNA is eluted on the first fraction, this was not the case when used with a specific bucket-rotor centrifuge. The desired fraction was the second; once this fraction was used, Miss Andre's cDNA library was successful. By both using the same equipment, cDNA library kit and competent cells (although different batches) in the same laboratory, it is highly likely that this is the cause of the low titre of library mxB.

3.6.2 Gene Ontology annotations

Of the 13,440 amplicons, 12,187 sequences were successfully generated, 49% of these showed homology to other protein sequences available in GenBank (BLASTX analysis; non-redundant (nr) database; expect-value < 0.1) and GO annotations assigned if possible. Only 14% of total sequences were assigned an annotation to the term cellular component. At 27%, slightly more sequences were annotated to the term

biological process followed by 29% annotated to molecular function. Overall, 35% of all individual sequences were assigned at least 1 GO annotation. Therefore, 65% of all sequences were undefined and neither the cellular location, function nor process of the translated protein is currently known. Of the 3917 unique sequences identified through cluster analysis, 30% showed homology to other protein sequences available in GenBank (BLASTX analysis; non-redundant (nr) database; expect-value < 0.1), 18% were assigned an annotation to the term cellular component, 32% sequences were annotated to molecular function, 22% sequences were annotated to the term biological process and overall 37% were annotated with at least 1 gene ontology. This reflects the limited genetic information available for amphipods and other invertebrates including 17% of all gene objects in the springtail, *Folsomia candida* (Timmermans *et al.* 2007), 24% in the eastern oyster, *Crassostrea virginica* (Quilang *et al.* 2007), 40% in the nematode, *Caenorhabditis elegans* (Wren 2006) and 44% in the decapod *Penaeus monodon* (Leu *et al.* 2007).

There was no distinctive difference between the cell component annotation of gene products in the mixed and male subtractive library with the majority of products being described as being located within the cell and organelle. The mixed library contained a slightly larger percentage of annotations to extracellular region and envelope.

As with cellular component, the molecular function seemed to be relatively evenly divided amongst the main attributes. Both libraries consist of gene products mainly responsible for binding and catalytic activity. One of the main differences is the lower number of structural attributes assigned to the male subtractive library also reflected in the higher level of gene objects annotated to structural constituent of cuticle.

The main differences between the mixed and male subtractive library with respect to biological process is the presence of annotations associated with biosynthetic process in the mixed subtractive library, which also contains a greater number of annotations responsible for cellular process. The male subtractive library contains a higher level of annotation for the regulation of cellular process and those thought to be involved in regulation of biological process. There is also an increase of annotations associated with metabolic process in the male subtractive library.

The majority of products represented on the microarray slides with cellular component GOSlim annotations are localised within the mitochondria and the cytoskeleton, with a large number contained within the ribosomes and nucleus. The molecular function attributes of the microarray slides were mainly distributed between structural molecule activity, nucleotide binding, peptidase activity and carbohydrate binding. The biological processes represented include electron transport, carbohydrate metabolic process, translation and ion transport.

Cluster analysis identified the largest number of unique sequences to be located in the nucleus, in the extracellular matrix or ribosomes. The majority of clusters annotated for molecular function were involved in zinc ion binding, transcription factor activity, ATP binding, structural constituent of cuticle and structural constituent of ribosome which together, were responsible for over 50% of the annotations allocated for molecular function. Biological processes represented by the unique sequences were mainly proteolysis, translation, chitin metabolic process, intracellular signalling cascade and defence response.

This work has resulted in the submission of 12,664 new *G. pulex* EST sequences and the generation of the first *G. pulex* online genetic database. Work described in Chapter 6 will build on these results to create an effective and novel tool for identifying molecular responses in *G. pulex*.

CHAPTER 4

IDENTIFICATION OF CANDIDATE GENES

4.1 INTRODUCTION

Candidate genes were pursued that would give a genetic 'snapshot' of the endocrine status of individual animal and therefore assess the impact that endocrine disrupting chemicals (EDCs) have on the freshwater shrimp *Gammarus pulex* (*G. pulex*). This was adopted to complement the high throughput genomic scale microarray approach in search for potential biomarkers for EDC analysis.

By reviewing current literature, four genes were selected for identification and further molecular characterisation. Those genes were hormone receptor 3 (HR3), oestrogen receptor (ER), ecdysteroid receptor (EcR) and ultraspiracle protein (USP). These genes all encode proteins essential to the reproductive endocrine system and represent key points in the cascade that leads to vitellogenesis and moulting.

4.1.1 Hormone Receptor 3 (HR3)

In both insects and crustaceans, HR3 is indirectly regulated by ecdysone and plays a critical role in control of moulting, vitellogenesis and, consequently, reproduction. It has previously been identified in flies, cockroach, moths, nematodes and lobster (El Haj *et al.* 1997; White *et al.* 1997; Kapitskaya *et al.* 2000; Eystathioy *et al.* 2001; Kostrouchova *et al.* 2001; Cruz *et al.* 2007). It has been suggested that HR3 may regulate expression of cuticle genes indirectly in the cockroach, *Blattella germanica* (Cruz *et al.* 2007) due to its presence in the epidermis, which has also been observed in *Helicoverpa armigera* and *Manduca sexta* (Langelan *et al.* 2000; Zhao *et al.* 2004). HR3 has a direct role in moulting in the nematode *Caenorhabditis elegans* identified by incomplete ecdysis following gene disruption (Kostrouchova *et al.* 1998; Kostrouchova *et al.* 2001). Both *B. germanica* nymphs and *C. elegans* were unable to shed their cuticle completely, but were capable of creating a new cuticle despite the inhibition of ecdysis (Kostrouchova *et al.* 1998; Kostrouchova *et al.* 2001; Cruz *et al.* 2007), suggestive of a role in cuticle detachment.

4.1.2 Ecdysteroid Receptor (EcR) and Ultraspiracle Protein (USP)

In insects, ecdysteroid hormones trigger moulting by binding to EcR which is subsequently activated by dimerisation with USP. A moulting cascade is initialised by this complex inducing further transcription factors that regulate the moult cycle (Zhao *et al.* 2004). EcR and USP have also been identified in the fiddler crab, *Uca pugilator* (Durica and Hopkins 1996) in regenerating limb buds, gills, eyestalks, hypodermis, hepatopancreas, muscle from non-regenerating walking legs and the large cheliped during late intermoult and throughout premoult (Chung *et al.* 1998b). Both EcR and USP are of utmost importance to moult regulation in crustaceans and therefore to successful reproduction. They could act as potential biomarkers of endocrine disruption and therefore isolation of the gene encoding these proteins is of great interest.

Identifying the expression levels of these key genes will allow us to further understand the molecular mechanisms involved in the endocrine pathways of *G. pulex*. Figure 4.1 provides a summary of the roles that these candidate genes play in moulting and vitellogenesis.

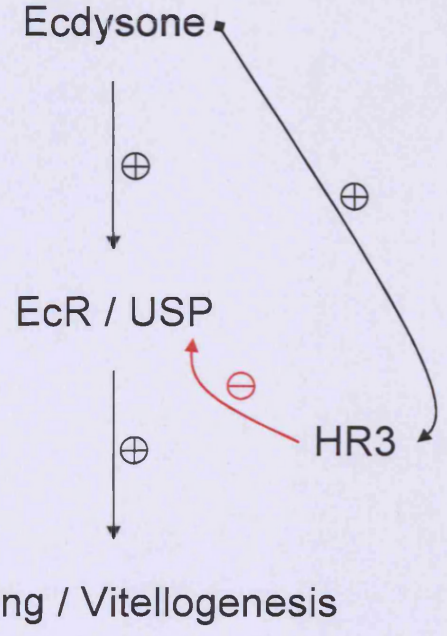


Figure 4.1: The putative roles of EcR, USP and HR3 in *G. pulex*

4.1.3 Oestrogen Receptor (ER)

The presence of oestrogen receptor in invertebrates is strongly debated, but recent evidence has suggested that some phyla may have evolved to retain this receptor, although not necessarily in an active form, and others have lost the ancestral gene. It is found in deuterostomes and protostomes but not in species such as *Drosophila melanogaster* or *C. elegans* and therefore it may have been lost in the Ecdysozoan lineage (Thornton *et al.* 2003). Its identification and characterisation in *G. pulex* is of great interest and importance with respect to the role it may play in normal and disrupted endocrine system.

Due to lack of genetic data available for this organism, a combination of degenerate and Rapid Amplification of cDNA Ends (RACE) amplification approaches were utilised in order to detect and obtain full length sequence of these genes.

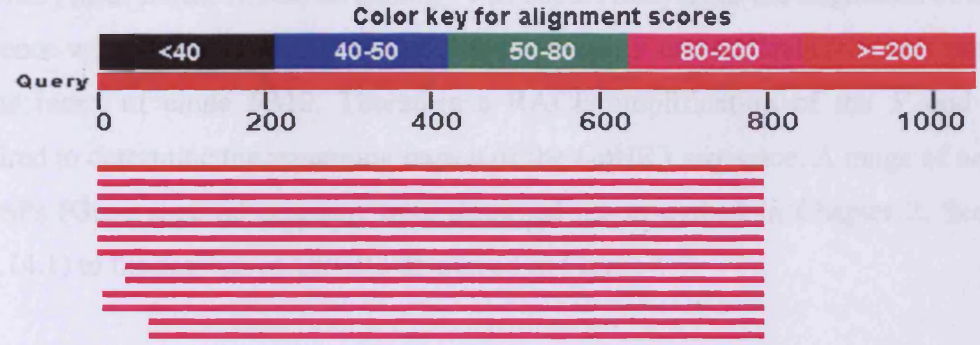
4.2 IDENTIFICATION OF HORMONE RECEPTOR 3

A 260bp gene fragment thought to encode a moult-regulating transcription factor was identified from the 5′ EST (expressed sequence tag) of a *G. pulex* cDNA library by a BLASTX search (Altschul 1997). Clone 5B12 from the mixed *G. pulex* library (Gp_mxA_05B12) was found to be homologous to the moult-regulating transcription factor HR3 in the cotton bollworm *H. armigera* (A/N: AAK14384) (Zhao *et al.* 2004). By resequencing from both the 3′ and 5′ end of the vector, a 1045bp consensus sequence was determined which represented the Gp_mxA_05B12 clone insert.

A BLASTX search against the consensus 5B12 sequence yielded the highest level of homology with a predicted protein from the honey bee *Apis mellifera* and nuclear receptor 3 from the yellow fever mosquito *Aedes aegypti*. The results from this BLAST search can be seen in Figure 4.2. These and other BLASTX results showing a good homology to the consensus 5B12 sequence are shown in Table 4.1.

ORGANISM	COMMON NAME	PROTEIN	ACCESSION NUMBER	E-VALUE
Apis mellifera	Honey bee	predicted protein	XP_392128	$8e^{-51}$
Aedes aegypti	Yellow fever mosquito	nuclear receptor 3	AAF36970	$1e^{-42}$
Drosophila melanogaster	Fruit fly	steroid receptor homolog	A46146	$8e^{-40}$
Manduca sexta	Tobacco hornworm	probable nuclear hormone receptor HR3 (MHR3)	Q08882	$6e^{-35}$
Bombyx mori	Domestic silkworm	nuclear receptor HR3 isoform A	AAK62804	$5e^{-34}$
Galleria mellonella	Greater wax moth	probable nuclear hormone receptor HR3 (GHR3)	P49868	$3e^{-33}$

Table 4.1: BLASTX hits from the HR3 sequence of *G. pulex*



gi|66500261|ref|XP_392128.2| PREDICTED: similar to ENSANGP00000017216 [Apis mellifera]
Length=500

```
Score =  203 bits (516),  Expect = 8e-51
Identities = 117/269 (43%), Positives = 165/269 (61%), Gaps = 9/269 (3%)
Frame = +3

Query  3    TRRQTASILYDDIISGFETDEISVVLAKLLGSAMQESVEGMTKQVDSQEEYRK-----KV  167
            T+ Q A  +  D  S      IS +L+K +  A   +    I+Q+  QE +RK     ++
Sbjct  232  TQTQVADTVAPDTPSAGVAAGISELLSKTIADAHARTCLLSTEQI--QESFRKPHDLSRL  289

Query  168  QYYENMHHEDIWMSAAHTLADMIKMIIEFAKMVPGFRKFDQTDQMSLLKRGGGFEIALVW  347
             YY+NM HE +W+  A  L  +I+ IIEFAKMVPGF K  Q DQ+ LLK  G FE+A++
Sbjct  290  IYYKNMAHEQLWLECAQKLTTVIQQIIEFAKMVPGFMKLSQDDQIVLLK-AGSFELAVLR  348

Query  348  MCRNFDLERNCVRYNECMLPIESFKTSEDIERQLVNSVFEFGRSVAELRLNEQEYGLLAA  527
            M R  DL++NCV Y + MLP ++F T++  E +LV+ VFE  RS+AEL+L E E  L +A
Sbjct  349  MSRYLDLQQNCVLYGDTMLPQDAFYTIDTAEMKLVSCVFELARSIAELKLTETELALYSA  408

Query  528  IVLLQADRPHVQNKTEINKLADACKKALCLMLSETHPHPVKGDVTAYETLLSKVRQLREL  707
             VLL  DRP ++   EI +L+ A  +AL   L   H  P+KGDVT  + +L+K+ QLRE+
Sbjct  409  AVLLSPDRPGLKGLAEITRLSQAVIRALRSELDRNHVSPIKGDVTVCDAILAKIPQLREI  468

Query  708  SALHLAALRRFRSTAGLMLKFPPLHDELF  794
            S LH+ AL +F+ +    L+FP LH ELF
Sbjct  469  SLLHMDALAKFKRSQP-HLEFPALHKELF  496
```

gi|7110506|gb|AAF36970.1| nuclear receptor 3 [Aedes aegypti]
Length=468

```
Score =  176 bits (445),  Expect = 1e-42
Identities = 110/270 (40%), Positives = 155/270 (57%), Gaps = 13/270 (4%)
Frame = +3

Query  3    TRRQTASILYDDIISGFETDEISVVLAKLLGSAMQ------ESVEGMTKQVDSQEEYRKK  164
            T    ++I+  D ISG    +I+ VL K L  A        E V  M ++  SQ+  R
Sbjct  196  TYEPRSTIIDSDFISGHTEGDINDVLIKTLAEAHANTNHKLEIVHDMFRK--SQDVTR--  251

Query  165  VQYYENMHHEDIWMSAAHTLADMIKMIIEFAKMVPGFRKFDQTDQMSLLKRGGGFEIALV  344
            + YY+NM  E++W+  A  L  MI+ IIEFAK++PGF +  Q DQ+ LLK G  FE+A+V
Sbjct  252  IMYYKNMSQEELWLDCAEKLTAMIQQIIEFAKLIPGFMRLSQDDQILLLKTGS-FELAIV  310

Query  345  WMCRNFDLERNCVRYNECMLPIESFKTSEDIERQLVNSVFEFGRSVAELRLNEQEYGLLA  524
             M R  DL  N V Y + MLP E F TS+  E +LV  +FE  +S+AEL+L E E  L
Sbjct  311  RMSRLMDLSTNSVLYGDIMLPQEVFYTSDSFEMKLVACIFETAKSIAELKLTETELALYQ  370

Query  525  AIVLLQADRPHVQNKTEINKLADACKKALCLMLSETHPHPVKGDVTAYETLLSKVRQLRE  704
            ++VLL  +R  V+  TEI +L +    A+   +   H  P+KGDVT  E LL+K+   RE
Sbjct  371  SLVLLWPERNGVRGNTEIQRLFEMSMSAIRQEIEANHA-PLKGDVTVLEILLNKIPIFRE  429

Query  705  LSALHLAALRRFRSTAGLMLKFPPLHDELF  794
            LS +H+ AL++F+      + FP L+ ELF
Sbjct  430  LSIMHMEALQKFKQDHPQYV-FPALYKELF  458
```

Figure 4.2: BLASTX results of consensus HR3 sequence.

The proteins with the closest homology to the 5B12 clone were aligned as a result of the BLASTX search. The alignment show that approximately the first 200 amino acids (~600bp) are missing from the 5′ end of the cDNA insert.

In order to determine the full length sequence of this putative hormone receptor (GpHR3) in *G. pulex,* RACE technology was used. Analysis of the alignment of 5B12 sequence with known HHR3s revealed that the entire 3′ end of the cDNA was present in the insert of clone 5B12. Therefore a RACE amplification of the 5′ end was required to determine the remaining region of the GpHR3 sequence. A range of nested 5′ GSPs (Gene specific primers) were designed (as described in Chapter 2, Section 2.11.14.1) to the regions of GpHR3 displayed in Figure 4.3.

5'GCACGAGGCGGCAGACGGCTTCCATCCTCTACGACGACATTATCTCGGGTTTCGAGACCGACGAGAT
CAGTGTGGTGCTGGCCAAGCTGCTGGGCAGCGCCATGCAGGAGTCCGTCGAGGGCATGACCAAGCAGG
TTGACAGCCAGGAGGAGTACAGGAAGAAAGTGCAATACTACGAGAACATGCATCACGAAGACATCTGGA
TGAGCGCCGCCCACACTTTGGCAGACATGATCAAGATGATCATCGAGTTCGCCAAGATGGTGCCCGGCT
TCAGGAAGTTTGACCAGACGGACCAGATGAGTTTGCTCAAACGCGGAGGTGGTTTCGAGATCGCGCTGG
TATGGATGTGCCGTAACTTTGACTTGGAGCGCAACTGCGTTCGCTACAACGAGTGCATGCTGCCCATTG
AGTCCTTCAAGACGAGCGAAGACATCGAGCGGCAGCTGGTGAACAGCGTATTCGAGTTTGGGCGGTCCG
TGGCCGAGCTGCGGCTCAACGAGCAGGAGTATGGCCTCCTGGCGGCTATAGTGCTTCTGCAAGCAGACC
GACCTCACGTGCAGAACAAGACTGAGATCAACAAGCTTGCCGACGCCTGTAAGAAGGCTCTCTGTTTGA
TGCTCAGCGAAACGCACCCGCATCCCGTGAAGGGAGATGTCACCGCCTACGAAACCCTCCTCTCTAAAG
TCCGCCAGCTTAGGGAGCTGAGCGCCCTCCACCTCGCCGCCCTGCGTCGCTTCAGGTCCACCGCGGGCC
TTATGCTCAAGTTCCCCCCCTTGCACGACGAGCTCTTCCCCCCTACTGACGATATCACTGACATGGTTG
AAGATTAAGCTCGCAGTGGCTACACCTGGTGCTACGGGGCCCCTTTGCATTGGGCTTCAGGTCGCCAGT
CCCAGAAACACTGAGTTCCCCTTCCGGTGGTGGTGTTTGCCATCCTTGGACCACTCCAACACTCTCTGC
TTACAGTGTGTCTGACTGGCTCGACGCGTGCAGCGCATTGTACTGCTACTAGAGCACATGTTCGCCACA
GCGATGATTAGTT 3'

Figure 4.3: DNA sequence of the 5B12 clone insert (GpHR3).
Primer sequences are highlighted, with an arrow indicating the direction of chain extension.

TGCTTGGTCATGCCCTCGACGGAC = GpHR3-5_1

CTCGACGGACTCCTGCATGGCGC = GpHR3-5_2

CCAGCAGCTTGGCCAGCACCACAC = GpHR3-5_3

Two primary reactions were performed using 5′ RACE-ready cDNA, prepared as described in Section 2.11.14.1 using mRNA purified from a population of mixed adults and juveniles. Two secondary nested reactions were performed using the primary amplification products. This procedure is shown in the diagram in Figure 4.4.

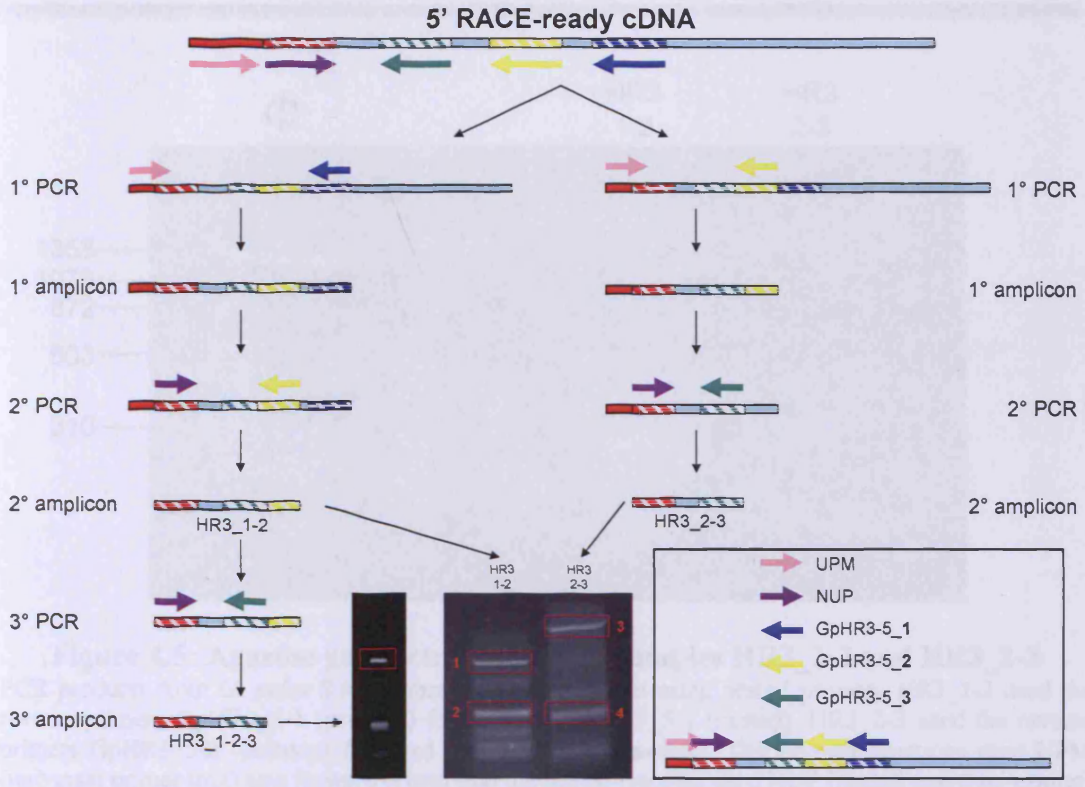


Figure 4.4: Schematic diagram of HR3 RACE reaction procedure.
This diagram showing the primary and nested RACE PCRs performed and the resultant amplicons. The resultant amplification products were visualised on 2% agarose gels, which are shown for illustrative purposes as an insert at the bottom of the figure. Expanded high resolution images are provided in Figure 4.5 and Figure 4.7. Primers GpHR3-5_1 to 3 were designed against the sequence of clone 5B12 (see Figure 4.3) whilst UPM and NUP primer sites (Appendix A, Table 6) are introduced at the 5′ of the cDNA during the RACE reaction (Section 2.11.14.1).

The resulting PCR products HR3_1-2 (generated from 1° amplification with UPM/ GpHR3-5_1 and 2° amplification with NUP/ GpHR3-5_2) and HR3_2-3 (generated from 1° amplification with UPM/ GpHR3-5_2 and 2° amplification with NUP/ GpHR3-5_3) were analysed by agarose gel electrophoresis. The gel was visualized under UV light and can be seen in Figure 4.5.

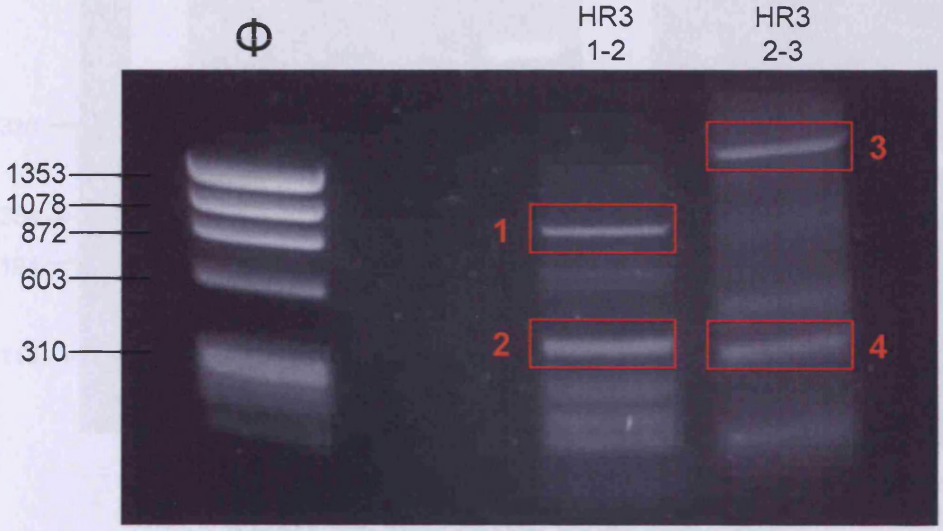


Figure 4.5: Agarose gel electrophoresis of samples HR3_1-2 and HR3_2-3
PCR products from *G. pulex* RACE library amplified twice using nested primers. HR3_1-2 used the reverse primers GpHR3_5-1 (primary) followed by GpHR3_5-1 (nested). HR3_2-3 used the reverse primers GpHR3_5-2 (primary) followed by GpHR3_5-3 (nested). The primary reactions used UPM (universal primer mix) as a forward primer and the nested reaction used NUP (nested universal primer) both purchased from Clontech (Clontech UK, Basingstoke, UK). Primer sequences are described in Chapter 2, Section 2.11.14.1.

After visualising the gel under UV, bands were selected that represented the longest 5′ RACE products (Figure 4.5, Bands 1 and 3) together with those that whose size was equivalent between the two RACE reactions (Figure 4.5, Bands 2 and 4) as these were the most intense and discrete. These DNA bands were excised and purified as described in Chapter 2, Section 2.11.10.1. The extracted DNA was cloned into pGEM-T vector and transformed into *Escherichia coli* DH5α™ cells. Clones were screened for recombinants by PCR using M13 primers (Appendix A, Table 6), visualised by agarose gel electrophoresis (Figure 4.6), and successful clones grown overnight prior to DNA extraction.

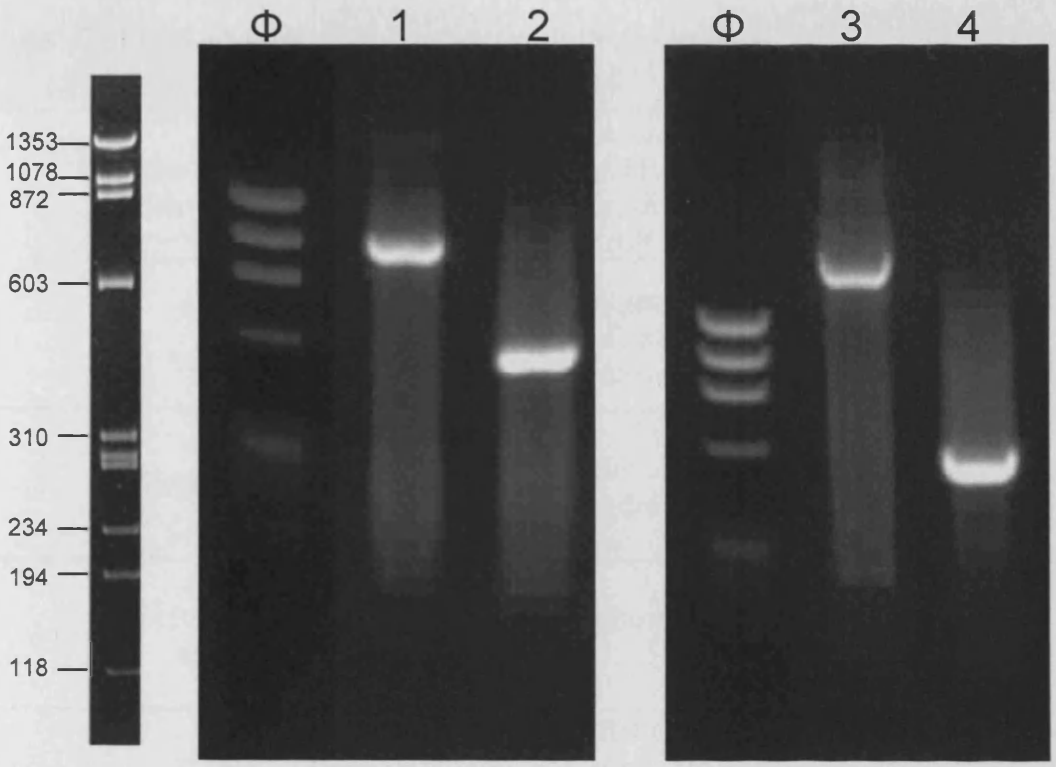


Figure 4.6: HR3 RACE clones 1, 2, 3 and 4

The bands represent DNA excised from gel in Figure 4.5 inserted into pGEM-T vector and screened with M13 primers. These four clones were sequenced using BigDye 3 reaction described in Chapter 2, Section 2.12.4.

Plasmid DNA was extracted and inserts sequenced from 5′ end. Resulting sequences were compared to BLASTX and BLASTN databases and those with the highest homology from BLASTX search are shown in Table 4.2. This analysis showed that none of the putative HR3 clones 1 - 4 showed significant homology to target HR3 sequence. To address the apparent non-specificity of these amplifications, a third round of amplification was introduced using NUP and GpHR3-5_3 primers and the product of the 2° HR3_1-2 reaction as a template (see Figure 4.4). Agarose gel electrophoresis was performed and resulting gel viewed under UV light to give the image shown in Figure 4.7.

CLONE	ORGANISM	COMMON NAME	PROTEIN	ACCESSION NUMBER	E-VALUE
1	*Kineococcus radiotolerans*	Bacteria	ATP-binding region, ATPase-like:Histidine kinase, HAMP region	EAM75113	0.025
	Homo sapiens	Human	scavenger receptor class F, member 1 isoform 1 precursor	NP_003684	9
2	*Salmo salar*	Atlantic salmon	glutamate dehydrogenase	CAD89353	0.53
3	*Leishmania major*	Protozoan parasite	proteophosphoglycan 5	AAZ14281	0.11
4	*Rattus norvegicus*	Norway rat	PREDICTED: similar to QUAKING isoform 6	XP_574316	0.086

Table 4.2: BLASTX database hits from DNA sequence of potential GpHR3 clones.

Figure 4.7: PCR products of RACE HR3_1-2-3

Products of *G. pulex* RACE library amplified three times with nested primers GpHR3_5-1 followed by GpHR3_5-2 and finally GpHR3_5-3. Universal primer mix (UPM) was forward primer in primary reaction and subsequent reaction used Nested Universal Primer (NUP), both purchased from Clontech (Clontech UK, Basingstoke, UK). A band of ~900bp is highlighted by a red box. Primer sequences are described in Chapter 2, Section 2.11.14.1.

The HR3_1-2-3 band was excised, purified and cloned into pGEM-T vector before transformation into *E. coli* DH5α™ cells. Vector inserts from two clones were sequenced from the 5′-end and homologous sequences identified by a BLASTX search. The results are shown in Table 4.3.

CLONE	ORGANISM	COMMON NAME	PROTEIN	ACCESSION NUMBER	E-VALUE
22RACEA29	*Aedes aegypti*	yellow fever mosquito	nuclear receptor 3	AAF36970	7e-04
35RACEA35	*Rattus norvegicus*	Norway rat	PREDICTED: similar to QUAKING isoform 6	XP_574316	0.15

Table 4.3: BLASTX results from RACE HR3_1-2-3 clone 5′ insert sequence.

The insert of clone 22RACEA29 was found to be homologous to the 5′ end of the nuclear receptor 3 gene of *A. aegypti* as determined by BLASTX analysis. It was therefore concluded it represented 5′ end of the *G. pulex* HR3 gene. The full insert was sequenced using both M13 forward and reverse and resulting sequence merged with original 5B12 sequence to give a composite sequence of 1874bp. This sequence along with the protein resulting from translation can be seen in Figure 4.8.

A BLASTX homology search was used to identify the protein sequence resulting from a translation of the composite DNA sequence. The gene was found to be homologous to nuclear receptor 3 of *A. aegypti* (GenBank accession number: AAF36970) with a significance of $5e^{-42}$. However, after alignment analysis of the predicted coding region of HR3 it was evident from *A. aegypti* and other species that the highly conserved DNA binding domain usually found within the first 200 amino acids of the protein was not present (Figure 4.9).

GpHR3 composite sequence

ACGCGGGAGAGCCAGATGAGGGCGGCCTCGGGACAGGAGACGTCGCCAGACTCCTCCGTGTATGAGTCCTCCACGCCCACCT
CTCCTGACGCCTACTCAGCGTACCGCTACGGCAGCAGTGACCTGAGCTACAACAACAACTACAGCTTCAGTCCCACCCCCGC
CGCCAACACCGTAGCCAATTTCGATATCCCCAGCGCCGACTCCACAACCTACGACCCACACAACAGAGGCGGCATTGACGCC
CACAGCCAAAATTCCCATGATATGCCCGACTCTGGAGCGCTCTCTCCCTCTCTCATAAACCCGTGCAGAAGCGACAGACCAT
GGAAAGATTCCGAAGCAGCATCTTTGTCGTGGGATTATCGAAAGTCCTTGTACGGCCACAGAGGACCATCTTTGGATGGTCA
GGACGACCTCAGCGGGGGCCTCGACTCCAAAGTCGGCATGAGCAGCGGTATGATGGAGCGGAGTAACATAAGCCATGACATG
GACATCATCATCAAGCAGGAGGTCGAGGGACTCGAGTTTGACACGTTACAGTCCAACAGAATATCAGCGGACAGCAGAGGAG
GCCTTAACAGATTATCCTCCGATCATATGAACTCCGTCGACAGTACAACTTACCTCCAAGGCCCACAGGCGGCACCAGTGGC
CTCTCCGTCCCCACAGTCGCTGTGTTCCCCACAGATGTCCCTGCCCCTGCATCAACACCAGCCGCAGGATCTGCTACCCCAG
GACAGCAGCCAGATGATGCTTCCTGGCCAGTCCCGCATGCAGATGGACAGGAGAGACTTCCAGAGCGAGCCCGCGCCCCTCA
CAGATCTGGAGCASGMSSGGCAGACGGCYKCCATCCTCTACGACGACATTATCTCGGGTTTCGAGACCGACGAGATCAGTGT
GGTGCTGGCCAAGCTGCTGGGCAGCGCCATGCAGGAGTCCGTCGAGGGCATGACCAAGCAGGTTGACAGCCAGGAGGAGTAC
AGGAAGAAAGTGCAATACTACGAGAACATGCATCACGAAGACATCTGGATGAGCGCCGCCCACACTTTGGCAGACATGATCA
AGATGATCATCGAGTTCGCCAAGATGGTGCCCGGCTTCAGGAAGTTTGACCAGACGGACCAGATGAGTTTGCTCAAACGCGG
AGGTGGTTTCGAGATCGCGCTGGTATGGATGTGCCGTAACTTTGACTTGGAGCGCAACTGCGTTCGCTACAACGAGTGCATG
CTGCCCATTGAGTCCTTCAAGACGAGCGAAGACATCGAGCGGCAGCTGGTGAACAGCGTATTCGAGTTTGGGCGGTCCGTGG
CCGAGCTGCGGCTCAACGAGCAGGAGTATGGCCTCCTGGCGGCTATAGTGCTTCTGCAAGCAGACCGACCTCACGTGCAGAA
CAAGACTGAGATCAACAAGCTTGCCGACGCCTGTAAGAAGGCTCTCTGTTTGATGCTCAGCGAAACGCACCCGCATCCCGTG
AAGGGAGATGTCACCGCCTACGAAACCCTCCTCTCTAAAGTCCGCCAGCTTAGGGAGCTGAGCGCCCTCCACCTCGCCGCCC
TGCGTCGCTTCAGGTCCACCGCGGGCCTTATGCTCAAGTTCCCCCCCTTGCACGACGAGCTCTTCCCCCCTACTGACGATAT
CACTGACATGGTTGAAGATTAAGCTCGCAGTGGCTACACCTGGTGCTACGGGGCCCCTTTGCATTGGGCTTCAGGTCGCCAG
TCCCAGAAACACTGAGTTCCCCTTCCGGTGGTGGTGTTTGCCATCCTTGGACCACTCCAACACTCTCTGCTTACAGTGTGTC
TGACTGGCTCGACGCGTGCAGCGCATTGTACTGCTACTAGAGCACATGTTCGCCACAGCGATGATTAGTT

GpHR3 protein sequence

TRESQMRAASGQETSPDSSVYESSTPTSPDAYSAYRYGSSDLSYNNNYSFSPTPAANTVANFDI
PSADSTTYDPHNRGGIDAHSQNSHDMPDSGALSPSLINPCRSDRPWKDSEAASLSWDYRKSLYG
HRGPSLDGQDDLSGGLDSKVGMSSGMMERSNISHDMDIIIKQEVEGLEFDTLQSNRISADSRGG
LNRLSSDHMNSVDSTTYLQGPQAAPVASPSPQSLCSPQMSLPLHQHQPQDLLPQDSSQMMLPGQ
SRMQMDRRDFQSEPAPLTDLEXXXQTXXILYDDIISGFETDEISVVLAKLLGSAMQESVEGMTK
QVDSQEEYRKKVQYYENMHHEDIWMSAAHTLADMIKMIIEFAKMVPGFRKFDQTDQMSLLKRGG
GFEIALVWMCRNFDLERNCVRYNECMLPIESFKTSEDIERQLVNSVFEFGRSVAELRLNEQEYG
LLAAIVLLQADRPHVQNKTEINKLADACKKALCLMLSETHPHPVKGDVTAYETLLSKVRQLREL
SALHLAALRRFRSTAGLMLKFPPLHDELFPPTDDITDMVED

Figure 4.8: The 1874bp section of the HR3 sequence known to date.

The yellow sequence highlights region of overlap between 5B12 forward and reverse sequences. The green sequence highlights region of overlap between 5B12 sequence and HR3 RACE fragment. The red bases are those that differed between 5B12 and RACE sequences.

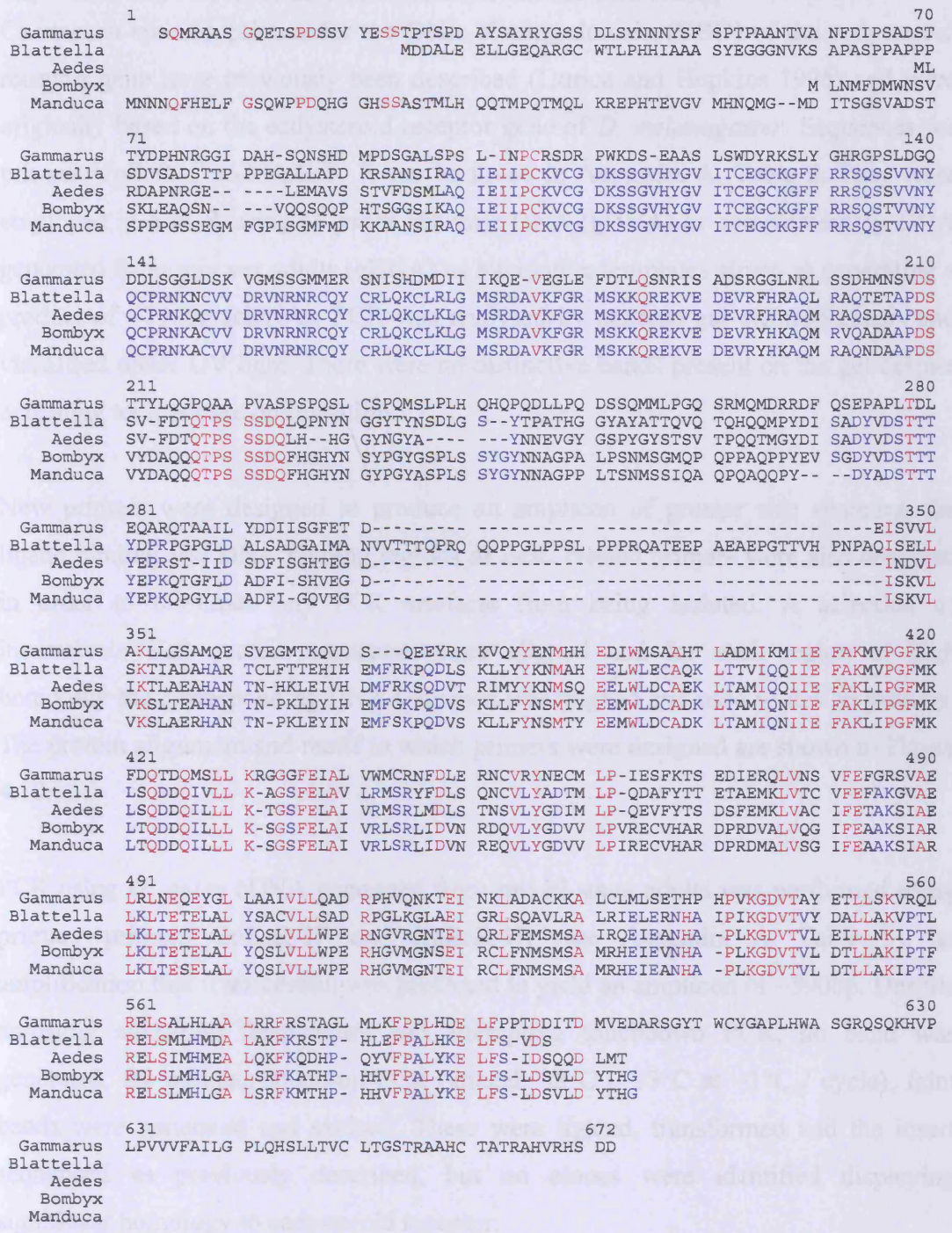


Figure 4.9: Amino acid alignment of the identified 1874bp (620aa) HR3 fragment
Gammarus = *G. pulex*, Blattella = *Blattella germanica*; Aedes = *A. aegypti*; Bombyx = *B. mori*; Manduca = *M. sexta*. Red amino acids are conserved in all 5 species; blue amino acids are conserved in the 4 non-gammarid species.

4.3 IDENTIFICATION OF ECDYSTEROID RECEPTOR GENE

Crustacean specific primers for the DNA binding domain (DBD) of the ecdysteroid receptor gene have previously been described (Durica and Hopkins 1996) and were originally based on the ecdysteroid receptor gene of *D. melanogaster*. Sequences for primers UpEcR_F and UpEcR_R can be found in Appendix A, Table 8. They were employed in a PCR using *G. pulex* genomic DNA (gDNA) or complementary DNA generated from mix sex adults (cDNA) as alternative templates aimed at generating a product of ~190bp (cDNA). PCR was analysed by agarose gel electrophoresis and visualised under UV light. There were no distinctive bands present on the gel despite annealing temperature optimisation.

New primers were designed to produce an amplicon of greater size spanning the ligand binding and DNA binding regions of EcR. Nested primers were also designed in order to eliminate any PCR artefacts from being isolated. A selection of invertebrate EcR protein sequences were aligned and for each region of high homology the corresponding DNA sequence was aligned and used to design primers. The protein alignment and motif to which primers were designed are shown in Figure 4.10.

PCR using *G. pulex* cDNA generated from mixed stage adults was performed using primary primers GpEcR-1F and GpEcR-1R (see Appendix A, Table 8), an amplification that if successful was predicted to yield an amplicon of ~590bp. Despite adjusting annealing temperature and attempting touchdown PCR, no band was generated. By utilising touch-up PCR (anneal: 38°C – 53°C at +1°C / cycle), faint bands were generated and excised. These were ligated, transformed and the insert sequenced as previously described, but no clones were identified displaying significant homology to ecdysteroid receptor.

The products from each of the primary PCRs were further amplified by secondary primers GpEcR-2F and GpEcR-2R (see Appendix, Table 8), however, yet again no distinct bands were evident on the resulting agarose gels. New primers were designed to alternative EcR protein motifs shown in Figure 4.11.

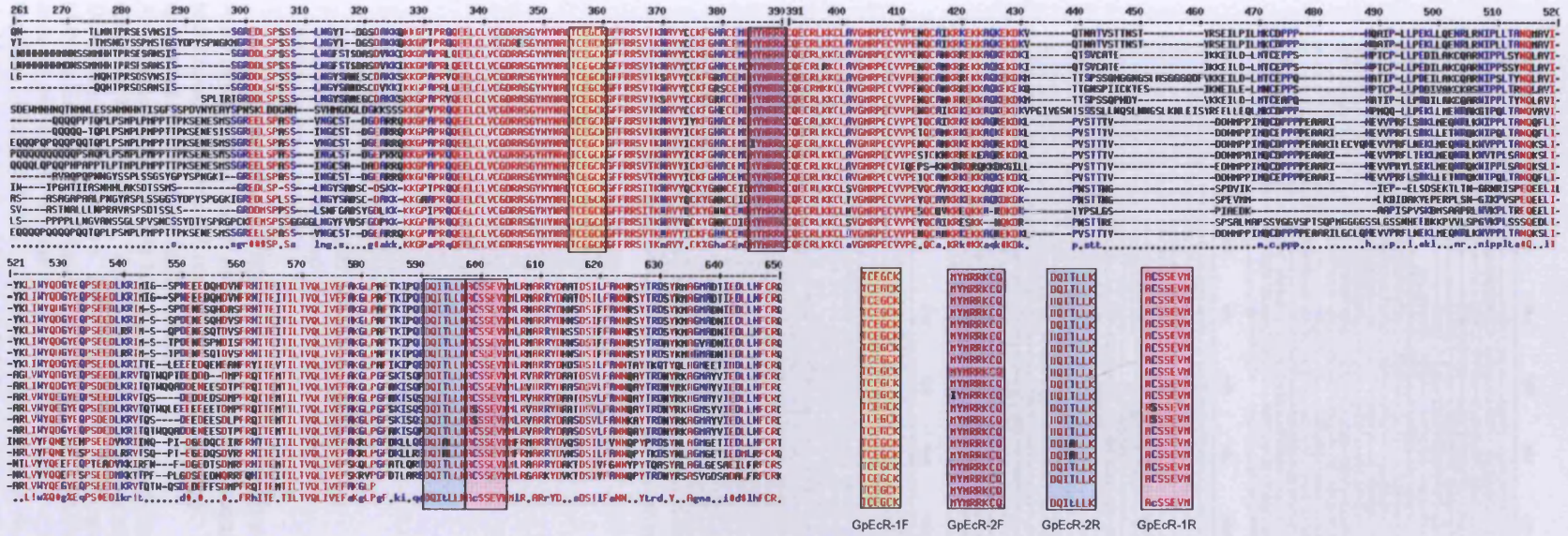


Figure 4.10: Protein alignment showing the locations that the new EcR primers were designed to.

The sequences used for alignment are from the species as follows from top to bottom:
Aedes albopictus (A/N: AAF19032); *Aedes aegypti* (A/N:AAQ23183); *Lucilia cuprina* (A/N:O18531); *Calliphora* vicina (A/N:AAG46050); *Drospophila melanogaster* (A/N:NP724460); *Ceratitis capitata* (A/N:CAA11907); *Drosophila pseudoobscura* (A/N:EAL26530); *Chironomos tentans* (A/N:P49882) *Plodia interpunctella* (A/N:AAR84611); *Choristoneura fumiferana* (A/N:AAC36491); *Heliothis virescens* (A/N:O18473); *Manduca sexta* (A/N:P49883) *Bombyx mori* (A/N:BAA07890); *Choristoneura fumiferana* (A/N:AAC61596); *Tenebrio molitor* (A/N:CAA72296); *Locusta migratoria* (A/N:AAD19828); *Celuca pugilator* (A/N:AAC33432); *Amblyomma americanum* (A/N:AAB94567); *Spodoptera frugiperda* (A/N: AAM54494).

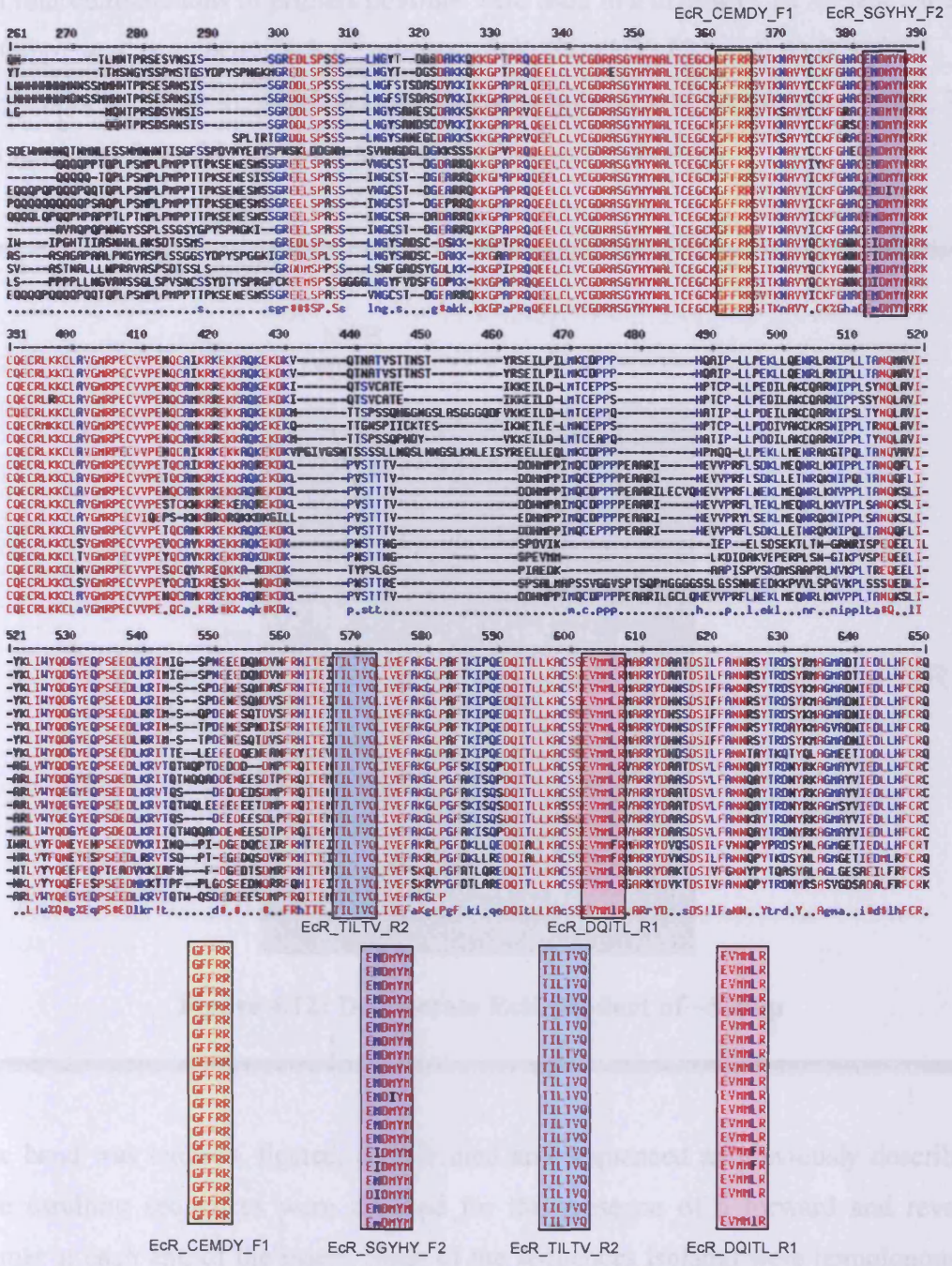


Figure 4.11: Alternative EcR primers design

Alternative primers were designed to the cDNA sequence corresponding to the highlighted amino acid sequences. They are all motifs of high homology. The sequences used for alignment are from the species as follows from top to bottom:

Aedes albopictus (A/N: AAF19032); *Aedes aegypti* (A/N:AAQ23183); *Lucilia cuprina* (A/N:O18531); *Calliphora vicina* (A/N:AAG46050); *Drosophila melanogaster* (A/N:NP724460); *Ceratitis capitata* (A/N:CAA11907); *Drosophila pseudoobscura* (A/N:EAL26530); *Chironomos tentans* (A/N:P49882) *Plodia interpunctella* (A/N:AAR84611); *Choristoneura fumiferana* (A/N:AAC36491); *Heliothis virescens* (A/N:O18473); *Manduca sexta* (A/N:P49883) *Bombyx mori* (A/N:BAA07890); *Choristoneura fumiferana* (A/N:AAC61596); *Tenebrio molitor* (A/N:CAA72296); *Locusta migratoria* (A/N:AAD19828); *Celuca pugilator* (A/N:AAC33432); *Amblyomma americanum* (A/N:AAB94567); *Spodoptera frugiperda* (A/N: AAM54494).

Once more primers were designed to support both primary and nested amplifications. All four combinations of primers possible were used in a cDNA PCR. A clear band of ~500bp was produced using primers EcR_SGYHY_F2 and EcR_DQITL_R1 (Appendix A, Table 8), which was the correct amplicon size for this primer pair. The gel produced from this amplification can be seen in Figure 4.12.

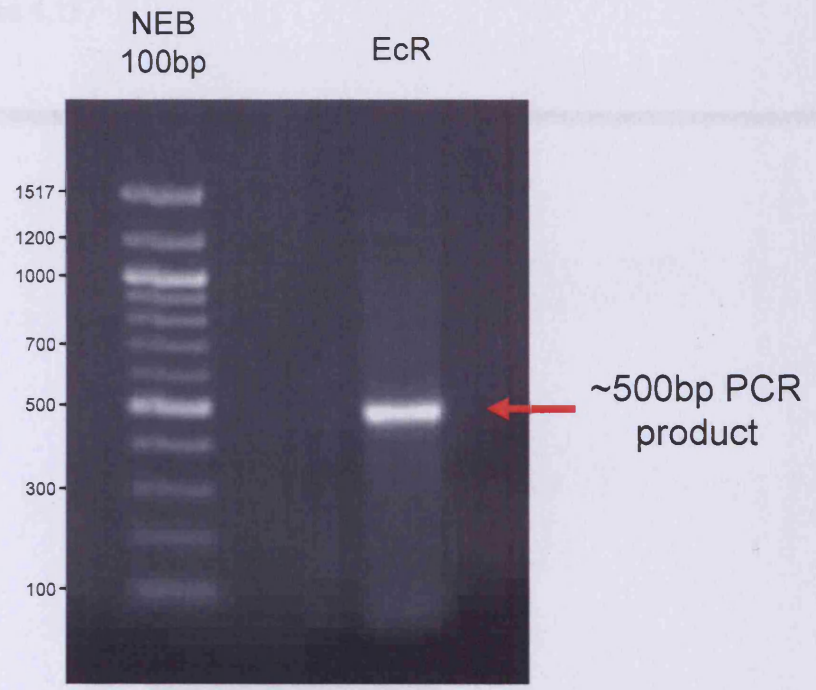


Figure 4.12: Degenerate EcR product of ~500bp

The band was excised, ligated, transformed and sequenced as previously described. The resulting sequences were checked for the presence of a forward and reverse primer at each end of the insert. None of the sequences isolated were homologous to EcR despite the fact that 8 clones were sequenced, 7 of which represented the same sequence which showed significant homology to an allatotropin neuropeptide precursor from the fall armyworm, *Spodoptera frugiperda*.

4.4 IDENTIFICATION OF THE ULTRASPIRACLE PROTEIN HOMOLOGUE

As well as EcR, crustacean specific primers for the DNA binding domain (DBD) of the ultraspiracle protein (USP) were described by Durica and Hopkins (Durica and Hopkins 1996) and were also based on *D. melanogaster*. The sequences for primers UpUSP_F and UpUSP_R can be found in the Appendix A, Table 8. They were used in PCR with *G. pulex* genomic DNA (gDNA) or complementary DNA (cDNA). PCR was analysed by agarose gel electrophoresis and visualised under UV light. The gel can be seen in Figure 4.13.

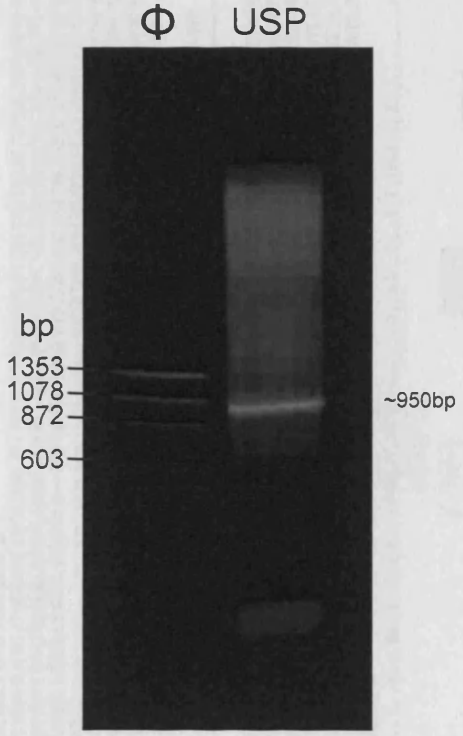


Figure 4.13: Agarose gel photograph under UV.
Indicates a possible USP gene band at ~950bp using cDNA as a template.

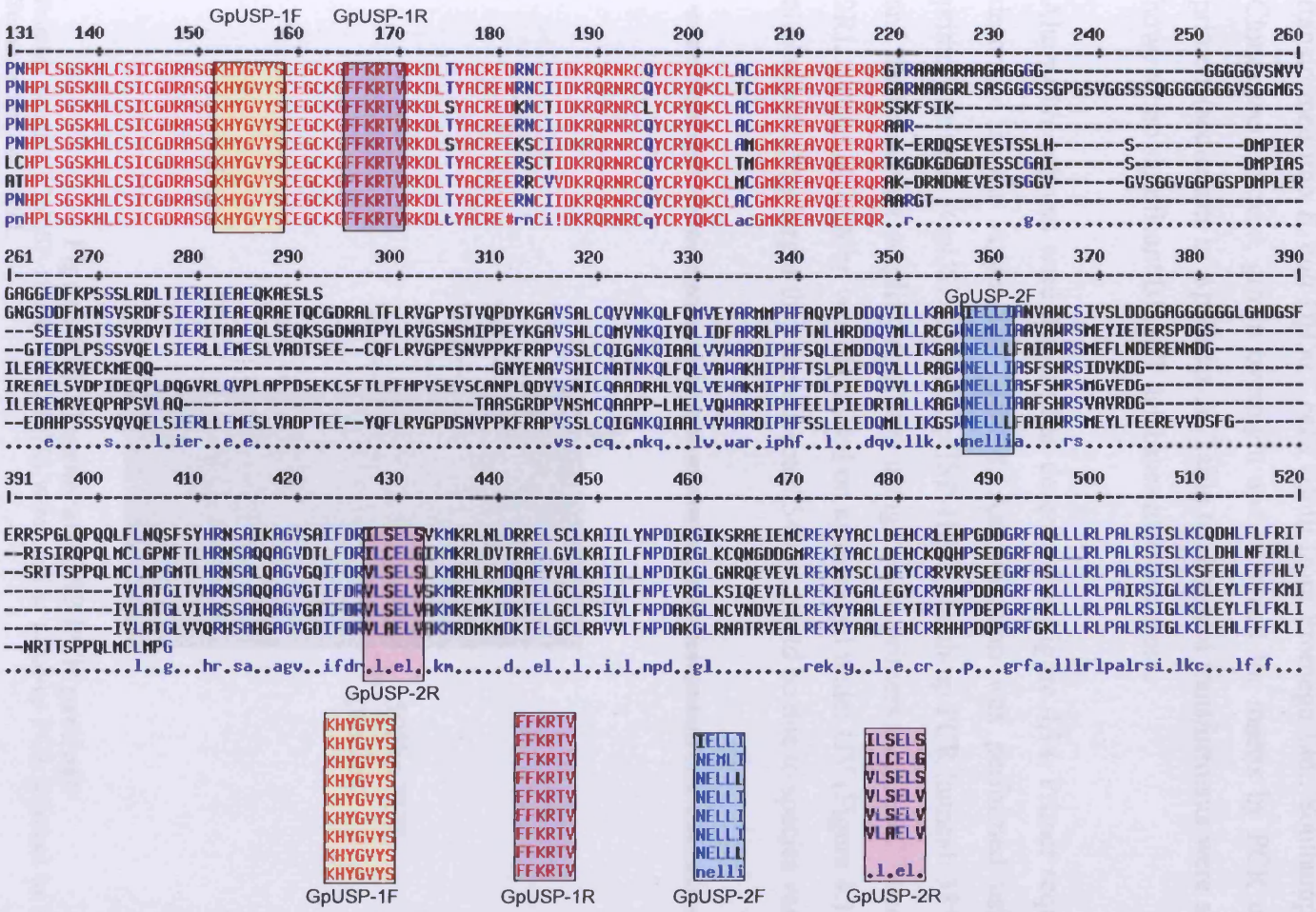


Figure 4.14: Ultraspiracle protein alignment for degenerate primer design.

Protein sequences used in alignment are as follows from top to bottom:
Lucilia sericata (A/N: BAD12053); *Drosophila* sp. (A/N: P20153); *Apis mellifera* (A/N: AAP33487); *Chilo suppressalis* (A/N: BAC53670); *Spodoptera frugiperda* (A/N: AAM54495); *Aedes albopictus* (A/N: AAF19033); *Celuca pugilator* (A/N: AAC32789); *Amblyomma americanum* (A/N: AAC15589).

The band was cut and DNA extracted, purified, ligated into pGEM-T vector and transformed into *E. coli* DH5α™ cells (Both from Promega Ltd., Southampton, UK). Clones were picked, grown overnight and screened for inserts by PCR using M13 primers (sequences in Appendix A, Table 6). Positive transformants were sequenced; however no significant BLAST databases matches arose.

Alternative primers were designed as described in Figure 4.14. Primer sequences are described in the Appendix, Table 8. Amplification was performed using firstly primary primers (GpUSP-1F and GpUSP-1R) in touch-up PCR (anneal: 38°C - 53°C) and then a further amplification step using nested primers (GpUSP-2F and GpUSP-2R). A band of ~590bp was visualised on agarose gel under UV (Figure 4.15), which, although slightly larger than the expected 540bp, could be due to species variation.

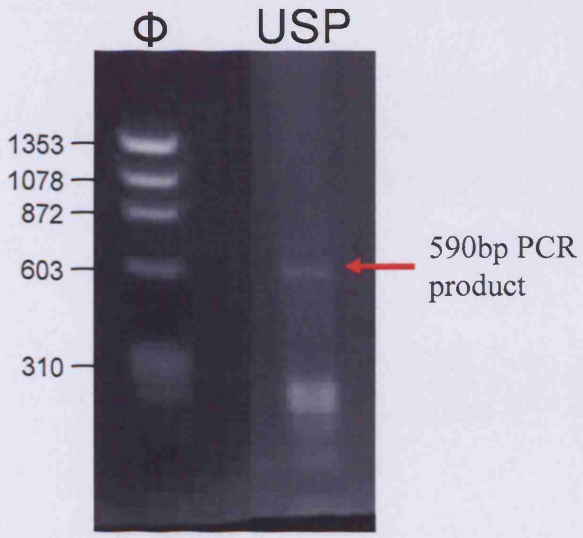


Figure 4.15: Degenerate USP PCR product.
Primary primers (GpUSP-1F and GpUSP-1R) were used in touch-up PCR followed by nested PCR. See main text for details.

The band was excised, ligated into pGEM-T vector (Promega Ltd., Southampton, UK), transformed and the insert sequenced as previously described. The insert sequences were not found to be homologous to USP.

4.5 IDENTIFICATION OF THE OESTROGEN RECEPTOR

Primers to the homologous regions of oestrogen receptor proteins were designed using an alignment of vertebrate oestrogen receptor sequences and can be seen in Figure 4.16.

No bands of the correct size were produced either with primary or nested PCR and therefore new primers were designed against an alignment of invertebrate oestrogen and oestrogen-related receptors as described in Figure 4.17. These invertebrate ER primers also did not yield any correct size bands.

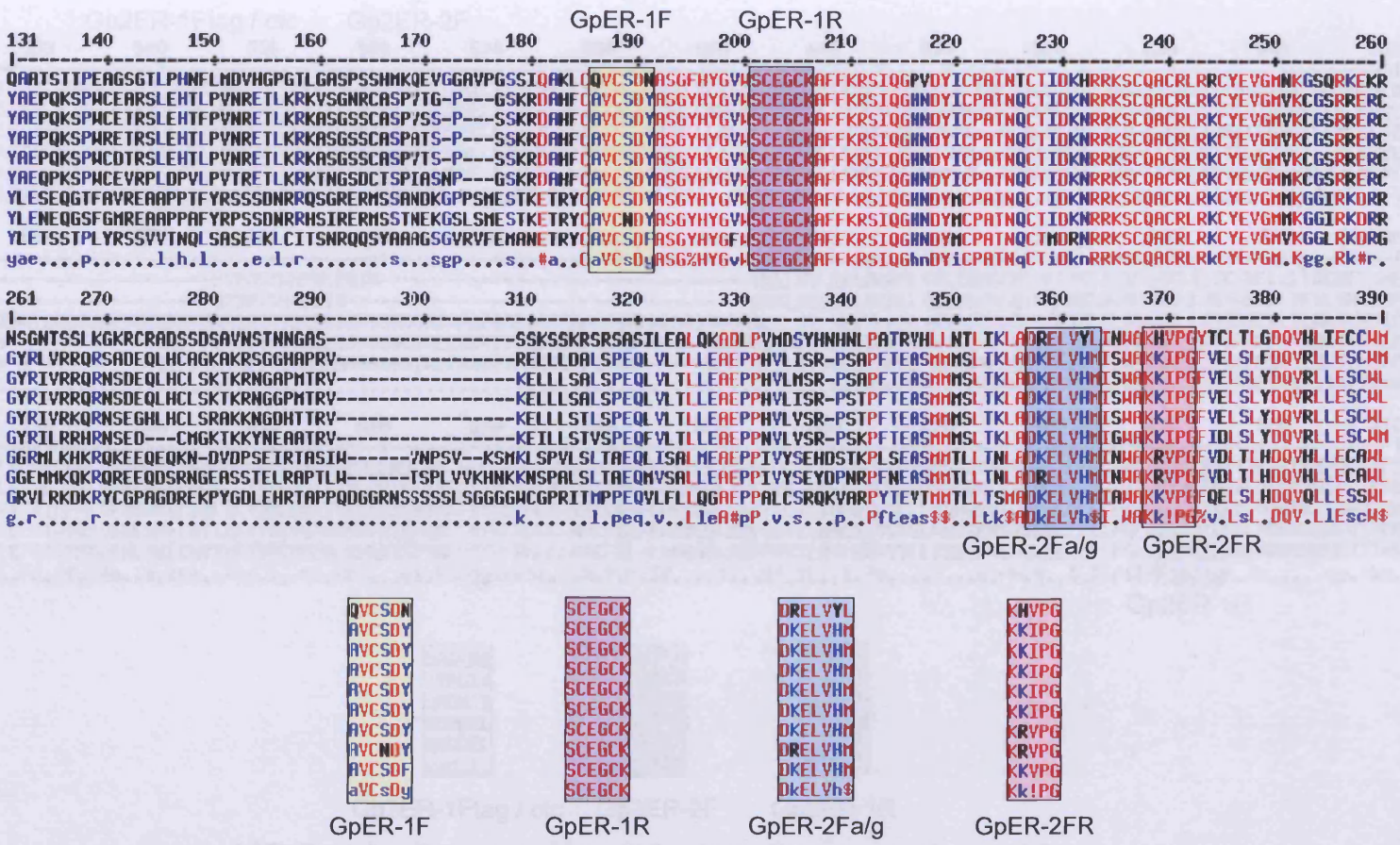


Figure 4.16: Protein alignment of vertebrate oestrogen receptor protein sequences.

Primers were designed to the homologous sequences to produce a pair of primary and a pair of nested primers. Two primers were designed for the nested reverse to reduce degeneracy. The proteins aligned are as follows from top to bottom:

Aplysia californica (A/N: AAQ95045); *Homo sapiens* (A/N: BAA24953); *Ovis aries* (A/N: AAD55772); *Bos taurus* (A/N: AAD24432); *Sus scrofa* (A/N: AAK15151); *Xenopus laevis* (A/N: AAQ84782); *Gallus gallus* (A/N: NP_990514); *Gallus gallus* (A/N: NP_990125); *Salmo salar* (A/N: CAA61999).

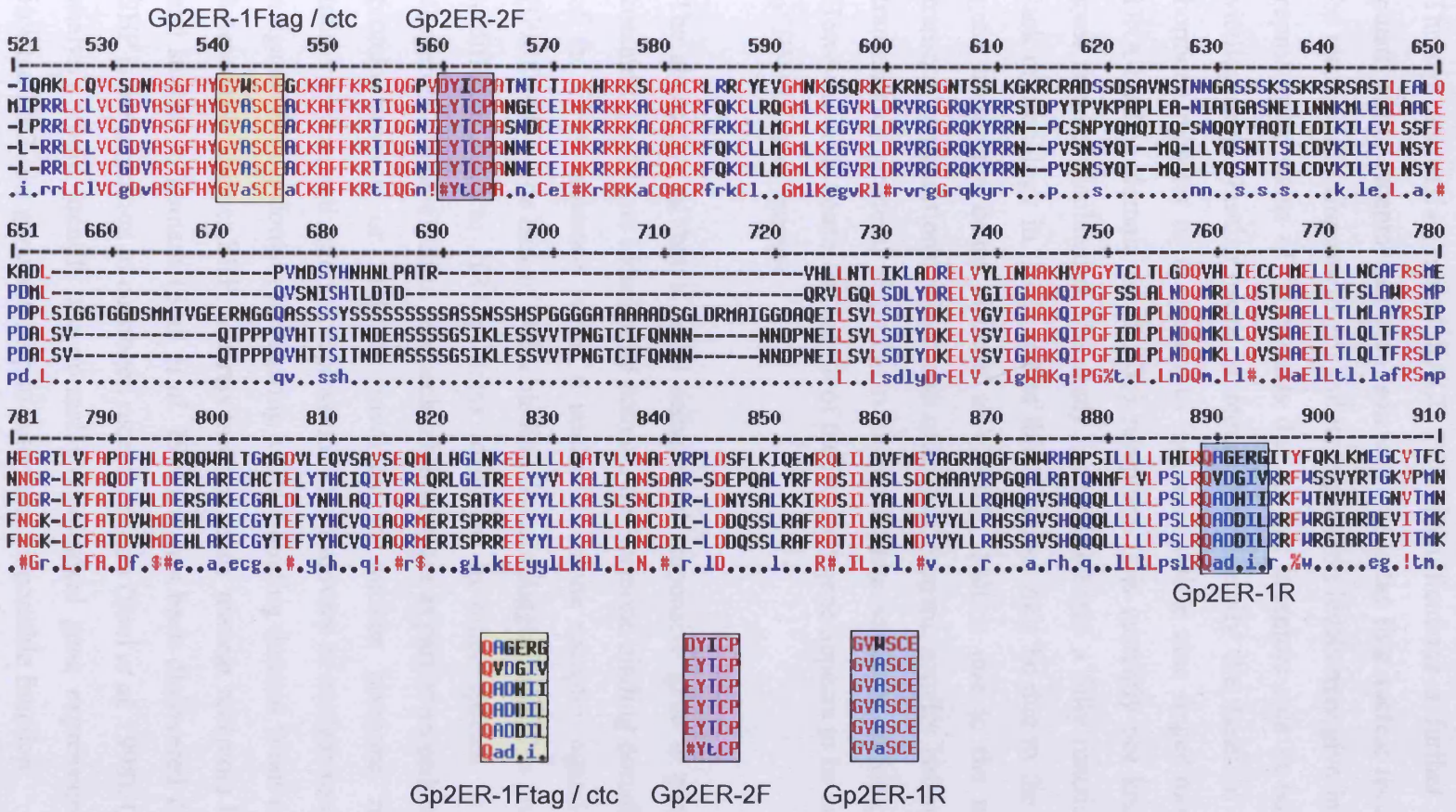


Figure 4.17: Protein alignment of invertebrate oestrogen receptor protein sequences.

Primers were designed to the homologous sequences to produce a pair of primary and a pair of nested primers. Two primers were designed for the nested reverse to reduce degeneracy. The aligned protein sequences are as follows from top to bottom:

Aplysia californica (A/N: AAQ95045); *Apis mellifera* (A/N: XP_392385); *Anopheles gambiae* (A/N: XP_321343); *Drosophila melanogaster* (A/N: AAF50473); *Drosophila melanogaster* (A/N: AAN12015).

4.6 DISCUSSION

Through successive nested RACE PCR amplifications a further fragment of the putative HR3 receptor has been isolated. This is the first nuclear receptor sequence to be identified in *Gammarus* sp. and its expression levels may give information on the reproductive status of individuals due to this receptors role in both moulting and vitellogenesis. Further work is required to identify the function of this putative hormone receptor as it appears to lack a complete zinc finger motif, found in the DNA-binding domain of all HR3 receptors. It is currently not known whether the gene fragment isolated in this study is translated into a fully functional protein. The lack of homology in the 5′ end of this sequence may be due to the problem of long gene transcripts being difficult to identify. This is due to the nature of reverse transcription reactions, which can often halt enzyme activity before completing the transcription process, resulting in an incomplete sequence lacking the full 5′ end. However, in this case, the length of the isolated gene appears to be comparable to that of HR3 in other species.

The alternative is that a novel ecdysteroid responsive gene is present in *G. pulex* containing only one identifiable domain, the hormone binding domain. InterPro scans of the DNA sequence reveal a nuclear hormone receptor ligand-binding domain (IPR008946). The lack of DNA binding domain suggests that this protein functions in a different way to HR3 proteins identified in other species. It could act as a chaperone, be involved in feedback inhibition, act as part of an ecdysteroid responsive protein cascade or may be involved in nuclear hormone receptor-mediated transcriptional activation or repression. The discovery of nuclear receptors containing a ligand binding domain but lacking a DNA binding domain is not unique. The small heterodimer partner SHP, a corepressor of many nuclear receptors has been isolated and studied in humans (Seol *et al.* 1996). It has been discovered that one action of SHP is the inhibition of oestrogen receptor action (Seol *et al.* 1998). Quantitative PCR analysis would identify the normal physiological gene expression patterns of the isolated HR3-like gene and help to determine its possible function.

So far, attempts to isolate ecdysteroid receptor, ultraspiracle protein and oestrogen receptor through degenerate PCR have failed, further optimisation and use of

alternative cDNA populations as DNA templates may allow the amplification of the correct sequence. A balance between specificity of primers to the gene sequence and flexibility due to codon differences needs to be made in order to successful isolate these genes.

The presence of both EcR and USP are very likely in *G. pulex* due to their presence in closely related arthropods. These include: the fruit fly, *D. melanogaster* (Diptera); the midge, *Chironomus tentans* (Diptera); the sheep blowfly, *Lucilia cuprina* (Diptera); the yellow fever mosquito, *A. aegypti* (Diptera); the tobacco hornworm, *M. sexta* (Lepidoptera); the spruce budworm, *Choristoneura fumiferana* (Lepidoptera); the rice striped stem borer, *Chilo suppressalis* (Lepidoptera); the tobacco budworm, *H. armigera* (Lepidoptera); the green peach aphid, *Myzus persicae* (Homoptera); the sweet potato whitefly, *Bemisia tabaci* (Hemiptera); the Colorado potato beetle, *Leptinotarsa decemlineata* (Coleoptera); the water flea, *Daphnia magna* (Cladocera) and the Atlantic sand fiddler crab, *U. pugilator* (Decapoda) (Koelle *et al.* 1991; Imhof *et al.* 1993; Henrich *et al.* 1994; Truman *et al.* 1994; Durica and Hopkins 1996; Kapitskaya *et al.* 1996; Jones and Sharp 1997; Chung *et al.* 1998a; Perera *et al.* 1998; Elke *et al.* 1999; Vogtli *et al.* 1999; Henrich *et al.* 2000; Durica *et al.* 2002; Minakuchi *et al.* 2003; Wu *et al.* 2004; Ogura *et al.* 2005; Graham *et al.* 2007; Kato *et al.* 2007; Parthasarathy and Palli 2007). However the presence of an oestrogen receptor gene is not certain and therefore the identification of EcR and USP sequences were given a higher priority.

The presence of an active form of oestrogen receptor in *G. pulex* cannot be determined. The identification of an oestrogen receptor or oestrogen-related receptor in this species would provide information on the endocrine axis and possibly whether a primordial inactive gene is still present within the genome. However, the evidence available does not justify the candidate gene approach, and a broader genomics approach should be utilised to help uncover the processes involved in hormonal regulation and reproduction in *G. pulex*.

Despite being a worthwhile attempt at candidate gene identification, the limited genetic information available for *G. pulex* and similar species has prevented a more stringent experimental design consequently preventing a more positive outcome. Until

this knowledge has been expanded by genome sequencing projects, this methodology should be limited to more conserved genes. Although the candidate gene approach could have provided sequence information on a few key genes within the endocrine system of *G. pulex*, the genomics approach utilising microarray technology would allow the elucidation of not only key genes but information on entire pathways involved in hormone regulation.

CHAPTER 5

ANALYSIS OF MOLECULAR BIOMARKERS

5.1 INTRODUCTION

Disruption of the hormonal systems in a variety of wildlife by compounds derived from anthropogenic sources is a major concern. These endocrine disrupting chemicals (EDCs) impede normal processes involving endogenous hormones and interfere with their actions affecting many biological systems including reproduction and development. Disruption can be caused through a variety of mechanisms: mimicry, antagonistic effects, altering hormone receptor levels and modification of hormone synthesis and metabolism; all result in detrimental changes in reproduction, development and related processes.

The effects of EDCs on organisms at the molecular level have been widely studied in fish over the last 12 years (Sumpter and Jobling 1995; Tyler *et al.* 1999; Bowman *et al.* 2000; Sole *et al.* 2000; Cheek *et al.* 2001; Denslow *et al.* 2001a; Denslow *et al.* 2001b; Fossi *et al.* 2002; Larkin *et al.* 2002; Sole *et al.* 2003; Garcia-Reyero *et al.* 2004; Kirby *et al.* 2004; Brian *et al.* 2005; Eidem *et al.* 2006; Jobling *et al.* 2006; Muncke and Eggen 2006), and to a lesser extent, invertebrates such as *Gammarus* sp., *Daphnia* sp., barnacles, mussels and shrimps (Billinghurst *et al.* 2000; Bultelle *et al.* 2002; Quinn *et al.* 2004; Gagne *et al.* 2005; Heckmann *et al.* 2005; Zapata-Perez *et al.* 2005; Ghekiere *et al.* 2006; Puinean *et al.* 2006; Puinean and Rotchell 2006; Soetaert *et al.* 2007).

The most commonly used biomarker of endocrine disruption in the aquatic environment is the yolk protein, vitellogenin. EDCs were found to induce expression of vitellogenin (usually triggered by the presence of 17β-oestradiol) in male fish exposed to sewage treatment works effluent (Sumpter and Jobling 1995). The normal expression of this gene is absent or at very low levels in male fish (Copeland *et al.* 1986). Other molecular markers include cypris major protein (Billinghurst *et al.* 2000), zona radiata proteins and their precursors (Fossi *et al.* 2002; Lee *et al.* 2002), aromatase (Kallivretaki *et al.* 2006) and biotransformation enzymes (Sole *et al.* 2003; Zapata-Perez *et al.* 2005).

5.1.1 Vitellogenin

The process of vitellogenesis was first used to detect effects of EDCs on fish using vitellogenin as a biomarker in 1995 (Sumpter and Jobling 1995). Blood samples were taken from fish exposed to sewage treatment effluents and vitellogenin levels determined by radioimmunoassay. The success of vitellogenin as a biomarker of environmental oestrogenic exposure led to its establishment as a key molecular marker of endocrine disruption in the UK (Allen *et al.* 1999; Matthiessen *et al.* 2002). This led to development of an enzyme linked immunosorbent assay (ELISA) to detect changes caused by EDCs during early life stages (Tyler *et al.* 1999). Due to technical advances in the field of molecular biology, QPCR is now commonly used to determine effects of EDCs on vitellogenin by determining *Vtg* expression levels in exposed and non exposed fish (Muncke and Eggen 2006).

The process of vitellogenesis also occurs in crustaceans as well as in fish (Charniaux-Cotton 1985) and is therefore a potential biomarker of endocrine disruption in aquatic crustaceans. The presence of vitellogenin in aquatic crustaceans using molecular techniques has been frequently recorded, the presence of a vitellogenin-like protein has been recorded in *Gammarus* sp. (Gagne *et al.* 2005) and the gene has been identified in prawns (Chen *et al.* 1999; Okumura *et al.* 2007) and shrimps (Tsang *et al.* 2003; Raviv *et al.* 2006).

Vitellogenin has been used as a biomarker of endocrine disruption in invertebrates; expression levels were elucidated using QPCR after exposure of the marine mussel *M. edulis* to 17β-oestradiol (Puinean *et al.* 2006). However, in this particular species there was no significant difference in *Vtg* expression between sexes.

5.1.2 Cuticle protein

Moulting is an essential process undertaken by crustaceans to allow for somatic growth and reproduction. It involves separation of the exoskeleton from the epidermis during premoult, followed by shedding of the outer cuticle, a process known as ecdysis. During postmoult and intermoult a new cuticle is formed. Due to these extreme habitual physiological changes, alterations in gene expression throughout this process are a good indicator of moult stage. As well as a physiological marker,

changes in the exoskeleton and moulting process are known to be caused by exposure to pollutants (Zou and Fingerman 1997a; Zou and Fingerman 1997b; Zou and Bonvillain 2004; Gagne *et al.* 2005; Zou 2005).

5.1.3 Quantitative PCR (QPCR)

A universally used method for measuring changes in individual gene expression is quantitative PCR (Bustin 2000). QPCR has been used in the toxicology field to assess effects of ibuprofen on candidate genes in *Daphnia magna* (Heckmann *et al.* 2006) after their identification using gene expression profiling. QPCR technology has also been used in the crab, *Carcinus maenas*, to investigate possible moult cycle-related changes in neuropeptide biosynthesis (Chung and Webster 2003). QPCR has also been exploited in the field of ecotoxicology. Inter-individual variations of aromatase gene (*cyp19a2*) levels in brains of adult male zebrafish were ascertained after exposure to bisphenol A, 17α-ethinyloestradiol or 17β-oestradiol (Kallivretaki *et al.* 2006). In medaka, expression levels of an EDC biomarker, choriogenin, were determined after exposure to bisphenol A, nonylphenol or 17α-ethinyloestradiol (Lee *et al.* 2002). More relevantly, in an invertebrate, vitellogenin gene expression has been quantified using QPCR in the marine mussel *Mytilus edulis,* after exposure to 17β-oestradiol (Puinean *et al.* 2006).

Two candidate genes for the detection of endocrine disrupters have previously been identified in *Gammarus pulex*; female yolk protein, vitellogenin (*vtg*) and exoskeletal cuticle protein (Dr. Judith Richards, unpublished data). Both of these genes were identified from sequences generated from mixed and female subtractive cDNA libraries respectively by sequence homology (BLASTX against GenBank database) to vitellogenin from the nematode, *Caenorhabditis elegans* and cuticle protein from the edible crab, *Cancer pagurus*.

After identification of these potential biomarkers and determination of their gene sequence, further analysis was performed in order to assess the viability of these genes as molecular biomarkers with the goal of gaining insight into the possible mechanisms of endocrine disruption.

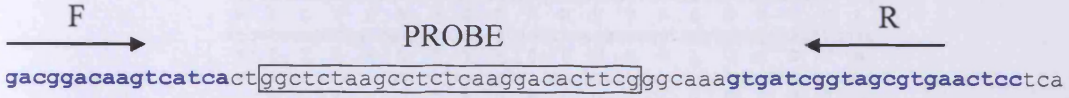
5.2 PROBE DESIGN AND DEVELOPMENT

5.2.1 Vitellogenin

A section of the gene encoding vitellogenin was identified from the *G. pulex* female subtractive library (Gp_feSA_01B01) and a probe and flanking primers, specific to the gene, were designed for QPCR analysis (Dr. Judith Richards, unpublished data; Figure 5.1). The 3′ end of the gene was identified through RACE PCR (Judith Richards, unpublished data) and subsequent sequencing of *G. pulex* cDNA libraries at a later date (Chapter 3), revealed a larger section of the vitellogenin gene. The overlapping ESTs (Gp_mxAA_94A11, Gp_mxAB_05G06 and Gp_mxAA_30G12) were aligned and a 1411bp consensus sequence generated, corresponding to a 343aa fragment of uninterrupted amino acid sequence (Figure 5.2).

Sequence homology was determined using BLASTX analysis on the DNA sequence and the highest match was to the vitellogenin protein (A/N: BAA22791) from sawfly, *Athalia rosae* (*A. rosae*; Hymenoptera), with an Expect value of $5e^{-12}$. Peptide sequence alignment (BLOSUM62) resulted in a 19% homology over the aligned section of the 1880aa *A. rosae* vitellogenin protein at the C-terminal end (Figure 5.3).

```
LOCUS       DV631405                449 bp    mRNA    linear   EST 03-NOV-2005
DEFINITION  Gp_feSA_01B01_M13F feS Gammarus pulex cDNA clone Gp_feSA_01B01 3',
            mRNA sequence.
ACCESSION   DV631405
VERSION     DV631405.1  GI:78771829
KEYWORDS    EST.
SOURCE      Gammarus pulex
  ORGANISM  Gammarus pulex
            Eukaryota; Metazoa; Arthropoda; Crustacea; Malacostraca;
            Eumalacostraca; Peracarida; Amphipoda; Gammaridea; Gammaroidea;
            Gammaridae; Gammarus.
REFERENCE   1  (bases 1 to 449)
  AUTHORS   Richards,J., Weeks,J., Johnson,I., Sambles,C. and Kille,P.
  TITLE     Gammarus pulex response to oestrogen and testosterone
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Christine Sambles
            PK Laboratory
            Cardiff University
            BIOSI 1, School of Biosciences, Cardiff University, Museum Avenue,
            Cardiff, CF10 3TL
            Tel: (+44)02920876680
            Email: samblescm@cardiff.ac.uk
            PCR PRimers
            FORWARD: M13F(GGTTTCCCAGTCACGACGTT)
            BACKWARD: M13R(CAGGAAACAGCTATGACCATG)
            Plate: 01  row: B  column: 01
            Seq primer: M13F(GGTTTCCCAGTCACGACGTT)
            High quality sequence stop: 449.
FEATURES             Location/Qualifiers
     source          1..449
                     /organism="Gammarus pulex"
                     /mol_type="mRNA"
                     /db_xref="taxon:52641"
                     /clone="Gp_feSA_01B01"
                     /sex="Female"
                     /dev_stage="Adult"
                     /clone_lib="feS"
                     /note="Vector: pGEM-T; Gammarus pulex female subtractive
                     library "
ORIGIN
        1 acaacctcaa agatcgttcc ggtcgacgag ccctccaggc agcgggcggc aacaagtgcc
       61 tgataaccgc acagcaagtg gtcactttcg atggcgtgtc ttttaagaag actcccacaa
      121 cctgctggag agttgttact caagttcagc accaacagga taccttcgct gtcatgtcca
      181 gacaggacaa atcttcagca cagagtgaaa aagaagtgcg gatcctcatt ccaggagttg
      241 gcaaaattga aatactcaaa ggacaacaag ttaaagtgga cggacaagtc atcactggct
      301 ctaagcctct caaggacact tcggcaaag tgatcggtag cgtgaactcc tcacccgacc
      361 gcattgtcgt ctcggtgccc agtaagttgg aagtcgttct tcaggacaag acagacttga
      421 gcatcacgct tgccgagagg tacctgccc
//
```



gacggacaagtcatcactggctctaagcctctcaaggacacttcggcaaagtgatcggtagcgtgaactcctca

Gp_Vtg_F: aaa gtg gac gga caa gtc atc a

Vtg probe: ggc tct aag cct ctc aag gac act tcg (dual labelled with 3' TAMRA and 5' 6-FAM)

Gp_Vtg_R: gga gtt cac gct acc gat cac

Figure 5.1: Vitellogenin primer and probe sequence
F = forward; R = reverse

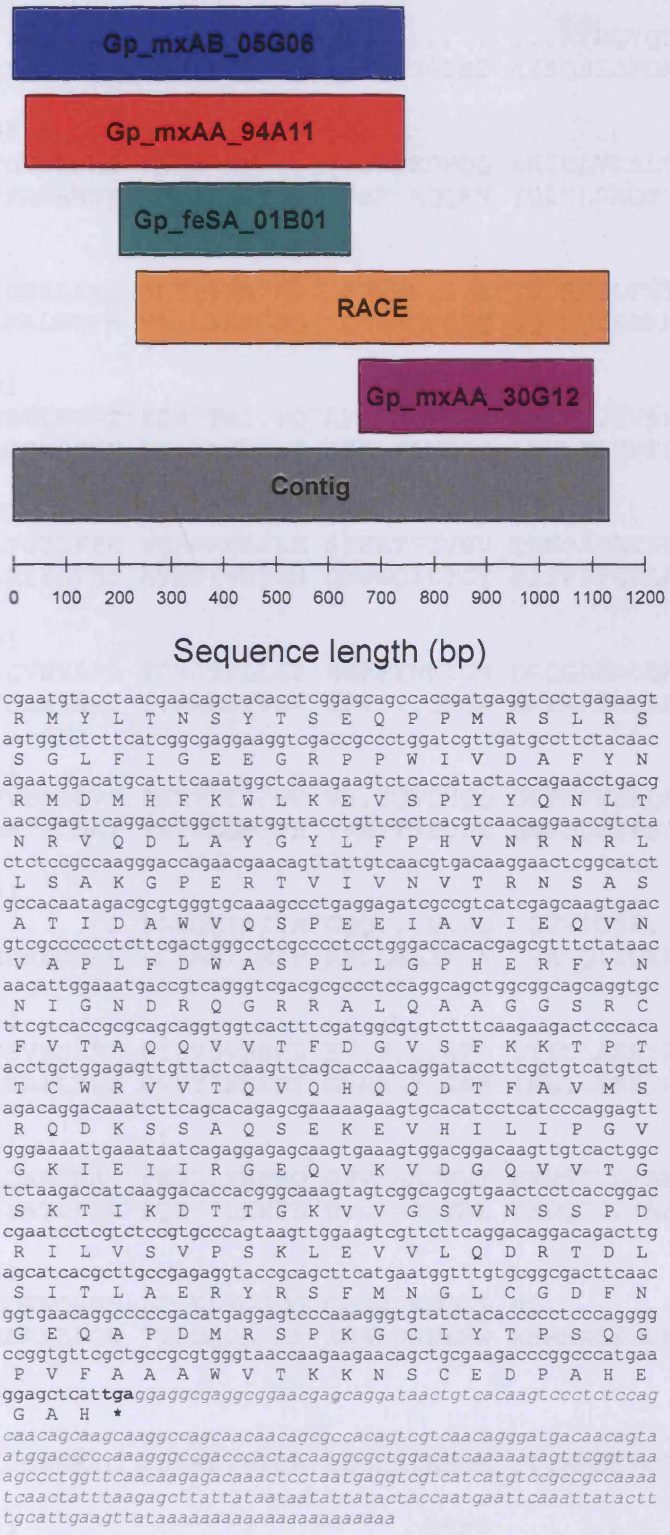


Figure 5.2: Schematic diagram showing the contribution of the different Vtg sequences

Original female library product aligned with overlapping ESTs. Underneath the schematic overlapping diagram is the consensus nucleotide sequence with translation shown below.

```
          1251                                                1300
G.pulex   ..........  ..........  ..........  ...RTRQYQL ALTGTNITKP
A.rosae   VPAFQPSSQK PDSESRQQEL LQKAASGIKD AYSQSIDFGA VFQGKNKAEY

          1301                                                1350
G.pulex   NPTGYEQNMI LNIMSM....  ..GRPEDPQQ KKICLVTSLS EKLTPEMD..
A.rosae   AATFAMANSP VDPKSRVVFY YGSNPAQQKN YQICLSAQSS RPKVPVMDFI

          1351                                                1400
G.pulex   .QKDSELYTL AATDIYAGKD CSVNEKVG.E TRLRMYLTNS YTSEQPPMRS
A.rosae   KALKADPTTK FAADLKFGEQ CQSGAQVQIQ GQLEQTEGRR EYLRRTPTGQ

          1401                                                1450
G.pulex   LRSSGLFIGE EGRPPWI.VD AFYNRMDMHF KWLK..EVSP YYQNLTNRVQ
A.rosae   QCAQQMQEGN KILPACRNVT ARANFLNEGS LTVKFSHVTD YMKNVSARAY

          1451                                                1500
G.pulex   DLAYGYLFPH VNRNRLLSAK GPERTVIVNV TRNSASATID AWVQSPEEIA
A.rosae   GLARYFGYHS AYENIVNPGN QKNGQIDFGY EISPDFQRLN VSMSAPGLDA

          1501                                                1550
G.pulex   VIEQVNVAPL FDWASPLLGP HERFYNNIGN DRQGRRALQA AGGSRCFVTA
A.rosae   NITDLEIHP. ..WARSIVAV DPV...QSGA RRFAREALAG QYNPTCVIDN

          1551                                                1600
G.pulex   QQVVTFDGVS FKKTPTTCWR VV.TQVQHQQ DTFAVMSRQD KSSAQSEKEV
A.rosae   NQASTFDNKT YPIELGNCWH VVMTTIPDDE QNGDDEIPED MRVSVVCRDI

          1601                                                1650
G.pulex   QI......LI PGVGKIEIIR GQQ...VKVD GQVVTGSK.. TIKDTT..GK
A.rosae   NEHQRECRLV LGEAEIEMYP NNDQAAIKVN GRAVQLSQEE TYRYTSQNGE

          1651                                                1700
G.pulex   VVGSVNSSPD RILVSVPSKL EV.VLQDKTD LTITLAEKYR SFMNGLCGDF
A.rosae   IQFEAYLKPS GVVYIHSEEY GVRALYDGER IKLILNNEYR DQIRGLCGTF

          1701                                                1750
G.pulex   NGEQAADMRS PKGCIYNPSQ GPVFAASWVT KKNSCEDPAM KEL..LEEAR
A.rosae   NGEPATDFTA PQNCILKNPE H..FAASYAL TKDQCEGPAK ENARRAQQAH

          1751                                                1800
G.pulex   RNEQENCHKS LSSNSKQGQQ QQRHSRQQGR HQ........ ..........
A.rosae   CYREAVLFGN VVSDQDAGRS KNKDSKWENN KSRKSGSGGK KSGSCTKFAT

          1801                                                1850
G.pulex   ..........  ..........  ..........  ..........  ..........
A.rosae   ATIEENGKTC FSLRPVATCA AGCKATRKQE TTVKMHCVQS SSSSQQLVHR

          1851                      1880
G.pulex   ..........  ..........  ..........
A.rosae   IKQGANPDFS QKSHDKSSRI EVPESCVAQQ
```

Figure 5.3: Alignment of *G. pulex* vitellogenin amino acid sequence with *A. rosae*
Red residues are exact amino acid matches.

Although *A. rosae* and *G. pulex* vitellogenin showed a relatively low amino acid sequence homology, the small section of *G. pulex* vitellogenin isolated to date contained a von Willebrand factor type D domain (VWD; pfam00094) (Figure 5.4); a motif commonly found in invertebrate vitellogenin proteins in this C-terminal position (Baker 1988). Due to the evolutionary amino acid sequence similarity of vitellogenin to crustacean clottable proteins (Yeh *et al.* 1999), there is a possibility that the vitellogenin and crustacean clottable protein gene DNA sequences may also be similar enough in homology for probes designed to vitellogenin gene sequences to bind non-specifically to crustacean clottable protein gene sequences. Therefore, female-specific expression of the isolated gene would suggest that the designed probe is specific for vitellogenin gene expression and is not cross-reacting with crustacean clottable protein genes. This was confirmed by QPCR analysis before exposed samples were analysed (Section 5.3.1).

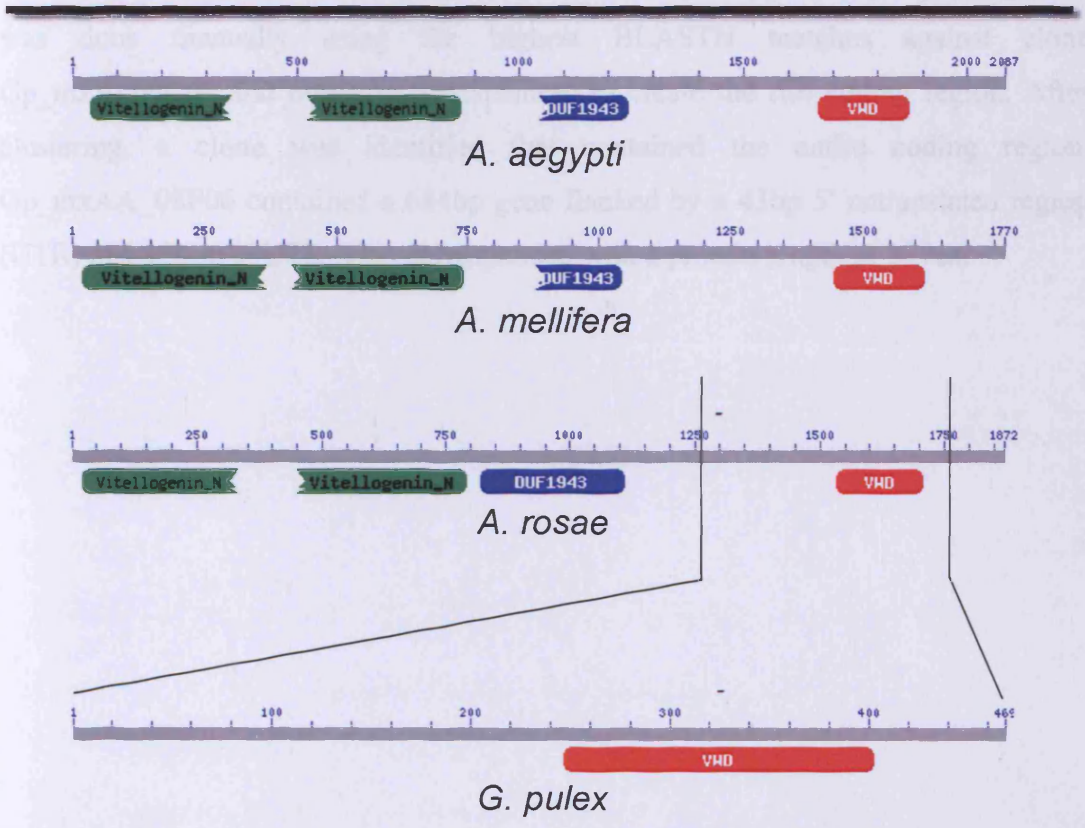


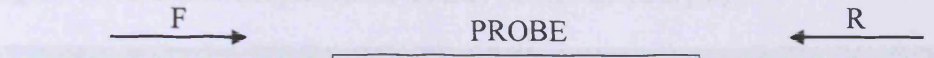
Figure 5.4: Conserved domains of vitellogenin in selected arthropods
The yellow fever mosquito, *Aedes aegypti*; honey bee, *Apis mellifera* and sawfly, *Athalia rosae* all contain the same lipoprotein amino terminal region (Vitellogenin_N), a domain of unknown function (DUF1943) and a von Willebrand factor type D domain (VWD). To date only sequences containing a C-terminal VWD domain in *G. pulex* has been identified.

5.2.2 Cuticle protein

One potential marker of moult stage is a cuticular protein that forms part of the protein matrix of the exoskeleton. After analysing mixed cDNA library sequences by BLASTX, a *G. pulex* sequence (Gp_mxA_04C09) showing high homology to a cuticle protein (CpCP11.58, A/N: P81580) from the edible crab, *Cancer pagurus* (Andersen 1999), was initially identified. A QPCR probe and primers flanking the probe were designed using this clone sequence. (Dr. Judith Richards, unpublished data; Figure 5.5).

Only 440bp of DNA coding region was available for this clone, corresponding to a peptide sequence of 146aa. Although the length of *G. pulex* cuticle protein was not known, absence of a suitable start methionine or a stop codon indicated that the full length sequence was not contained within this clone. EST sequencing of cDNA libraries created a resource for identifying the full length sequence by clustering. This was done manually using the highest BLASTN matches against clone Gp_mxA_04C09 and overlapping sequences to create the full coding region. After clustering, a clone was identified that contained the entire coding region. Gp_mxAA_08F06 contained a 684bp gene flanked by a 43bp 5′ untranslated region (UTR) and 126bp 3′ UTR. This corresponded with a protein length of 227aa.

```
LOCUS       CN211258               440 bp    mRNA    linear   EST 03-NOV-2004
DEFINITION  Gp_mxA_04C09_M13R mxA Gammarus pulex cDNA clone Gp_mxA_04C09 5',
            mRNA sequence.
ACCESSION   CN211258
VERSION     CN211258.1  GI:55254174
KEYWORDS    EST.
SOURCE      Gammarus pulex
  ORGANISM  Gammarus pulex
            Eukaryota; Metazoa; Arthropoda; Crustacea; Malacostraca;
            Eumalacostraca; Peracarida; Amphipoda; Gammaridea; Gammaroidea;
            Gammaridae; Gammarus.
REFERENCE   1  (bases 1 to 440)
  AUTHORS   Richards,J., Weeks,J., Johnson,I., Sambles,C. and Kille,P.
  TITLE     Gammarus pulex response to oestrogen and testosterone
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Judith Richards
            Pete Kille's Lab
            Cardiff University
            Main Building, Park Place, Cardiff, CF10 3TL
            Tel: 029 20874507
            Email: kille@cf.ac.uk
            PCR PRimers
            FORWARD: M13F
            BACKWARD: M13R
            Plate: 04  row: C  column: 09
            Seq primer: M13R
            High quality sequence start: 5
            High quality sequence stop: 324.
FEATURES             Location/Qualifiers
     source          1..440
                     /organism="Gammarus pulex"
                     /mol_type="mRNA"
                     /db_xref="taxon:52641"
                     /clone="Gp_mxA_04C09"
                     /sex="Mixed"
                     /clone_lib="mxA"
                     /note="Vector: pBluescript II SK+; Site_1: XhoI; Site_2:
                     EcoRI"
ORIGIN
        1 actaaaggga acaaaagctg gagctccacc gcggtggcgg ccgctctaga actagtggat
       61 cccccgggct gcaggaattc ggcacgaggc agaggagttc tattggagca ctatacccaa
      121 ggcatcatga agattctggt agctctctgc tttttggccg tcgtcgccaa cgcgcaggtt
      181 ggaccttcag gaatcatcga ccccagtgga aagaacgtcc agttctccca cgatttcgct
      241 tccaacattg tactgattgg accgtcgggc atcgtcacta gagatggacg caaccttcag
      301 cttacccacg gacaggccac actgaacgct gccgcaccca caccggtcac cgtacctcaa
      361 tatgtgtcca acaatcctgc tgtcggacct tcaggaattg tgcgtcccga tggcaggaat
      421 gttcagttca ctcaggactt
//
```



caccggtcaccgtacctcaatatgtgtcca acaatcctgc tgtcggaccttcaggaattgtgcgtcccga

Gp_Cut_F: acc ggt cac cgt acc tca ata t

CUT probe: tgt cca aca atc ctg ctg tcg gac c (dual labelled with 3' TAMRA and 5' 6-FAM)

Gp_Cut_R: tcg gga cgc aca att cct

Figure 5.5: Cuticle primer and probe sequence
F = forward; R = reverse

The amino acid sequence was analysed for a Rebers-Riddiford motif commonly found in flexible arthrodial membrane cuticle proteins and for 18-residue motifs found in cuticle proteins from solid calcified regions (Andersen 1999). The Rebers-Riddiford consensus sequence is $Gx_7[DEN]Gx_6[FY]xA[DGN]x_{2,3}G[FY]x[APV]$ (ExPASy PROSITE PS00233) and is found in approximately half the arthropod cuticle protein sequences known to date (Rebers and Riddiford 1988; Rebers and Willis 2001).

The *G. pulex* cuticle protein did not contain a Rebers-Riddiford motif and was only found to contain 18-residue repeat motifs identified in *C. pagurus* and American lobster *Homarus americanus* (Pfam ID:PF08140; InterPro ID:IPR012539). Cuticle proteins containing the 18-residue motifs have been extracted from calcified regions of these species, and so far these motifs have only been found in crustacean calcified exoskeletons (Andersen 1999). This indicated that the cuticle protein gene isolated from *G. pulex* is a calcified cuticle protein. This putative cuticle protein sequence along with the motifs identified is shown in Figure 5.6.

>*Gammarus pulex* cuticle protein
MKILVALCILAVGANAQVGPSGIIDPSGKNVQFSHDFASNVVLIGPSGIVTRDGRNLQLTHGQATLNA
AAPTPVSVSQYVSKNPAVGPSGIVRPDGRNVQFTQDFADNIVLIGPSGIVTRDGRNMQLRKKRAAPLV
GPSGIITPDGQLIQLEAGVTVVNAGPSGAVLSNGRSIQYRGKRAVPEAIVGPSGIITPDGQLIQLPAG
VTVVNAGPSGAVLSNGRSIQYKL*

Figure 5.6: Amino acid sequence for *G. pulex* cuticle protein.
Sequence identified from whole animal mixed cDNA library (mxA). Boxed sequence indicates similar 18-residue motifs as found in *C .pagurus* and *H. americanus* (Andersen 1999). Conserved residues are highlighted in red, conserved hydrophobic residues are highlighted in purple.

The *C. pagurus* protein that gave the original BLASTX hit was only 114aa in length, in which 77aa were an exact match to that of the *G. pulex* cuticle protein, giving a 68% homology on overlapping sequences. The protein alignment can be seen in Figure 5.7.

After the full length sequence was identified, a further BLAST search on both nucleotide and protein sequences were performed. This produced a top BLAST hit for the protein sequence to a 233aa sand crab, *Portunus pelagicus*, cuticle protein sequence (CUT1). This protein also contained four occurrences of the 18-residue crustacean cuticle protein repeat pairs. With an expect value of $3e^{-49}$, 120aa were an exact match with *G. pulex* cuticle protein giving a homology of 53%. Although the percentage homology is lower than that of *C. pagurus*, the sequence length is closer to that of *G. pulex* and many of the amino acids that are not an exact match are considered a neutral substitution not affecting the overall function of the protein. This first *G. pulex* cuticle protein to be identified was named 'cuticle protein A' as further studies of cDNA library sequences revealed a further two isoforms, named cuticle protein B and C. An alignment of the three isoforms is shown in Figure 5.8.

```
           1                                                          60
   GpCUTA  MKILVALCIL AVGANAQVGP SGIIDPSGKN VQFSHDFASN VVLIGPSGIV TRDGRNLQLT
CpCP11.58                  QVGY SGIVSPDGNN IQFTHDFAHS IVLKGPSGIV TSDGKNLQLT
Consensus  .......... ......QVGp SGI!dPdGnN !QFsHDFAhn !VLiGPSGIV TrDGrNLQLT

           61                                                         120
   GpCUTA  HGQATLNAAA PTP-VSVSQY V-SKNPAVGP SGIVRPDGRN VQFTQDFADN IVLIGPSGIV
CpCP11.58  AGQASLQAAA PAPPLPVSHY VASQQSVVGP SGIVSPSG-N VQFSHEFADN VVLVGPSGIV
Consensus  aGQAsL#AAA PaP.lpVSqY V.Sq#paVGP SGIVrPdG.N VQFsq#FADN !VL!GPSGIV

           121                                                        180
   GpCUTA  TRDGRNMQLR KKRAAPLVGP SGIITPDGQL IQLEAGVTVV NAGPSGAVLS NGRSIQYRGK
CpCP11.58  TKDGNNLQLR A
Consensus  TrDGrN$QLR a......... .......... .......... .......... ..........

           181                                                 229
   GpCUTA  RAVPEAIVGP SGIITPDGQL IQLPAGVTVV NAGPSGAVLS NGRSIQYKL
CpCP11.58
Consensus  .......... .......... .......... .......... .........
```

Figure 5.7: Amino acid alignment of *G. pulex* (CUTA) and *C. pagurus* cuticle protein.

Alignment shows a 68% homology between the cuticle protein of *G. pulex* and *C. pagurus* on the overlapping sequence. Exact matches are highlighted in red.

```
           1                                                          60
     CUTA  MKILVALCIL AVGANAQVGP SGIIDPSGKN VQFSHDFASN VVLIGPSGIV TRDGRNLQLT
     CUTB  MKILVALCIL AVGANAQVGP SGIIDPSGKN VQFSHDFASN VVLIGPSGIV TRDGRNLQLT
     CUTC  MKILVALCIL AVGANAQVGP SGIIDPSGKN VQFSHDFASN VVLIGPSGIV TRDGRNLQLT
Consensus  MKILVALCIL AVGANAQVGP SGIIDPSGKN VQFSHDFASN VVLIGPSGIV TRDGRNLQLT

           61                                                         120
     CUTA  HGQATLNAAA PTPVSVSQYV SKNPAVGPSG IVRPDGRNVQ FTQDFADNIV LIGPSGIVTR
     CUTB  HGQATLNAAA PTPVSVPQYV SNNPAVGPSG IVRPDGRNVQ FTQDFADNIV LIGPSGIVTR
     CUTC  HGQATLNAAA PTPVSVPQYV SNNPAVGPSG IVRPDGRNVQ FTQDFADNIV LIGPSGIVTR
Consensus  HGQATLNAAA PTPVSVpQYV SnNPAVGPSG IVRPDGRNVQ FTQDFADNIV LIGPSGIVTR

           121                                                        180
     CUTA  DGRNMQLRKK RAAPLVGPSG IITPDGQLIQ LEAGVTVVNA GPSGAVLSNG RSIQYRGKRA
     CUTB  DGRNMQLRKK RAAPLVGPSG IITPDGQLIQ LEAGVTVVNA GPSGAVLSNG KSIQYRGKRA
     CUTC  DGRNMQLRKK RAAPLVGPSG IITPDGQLIQ LEAGVTVVNA GPSGAVLSNG RSIQYRGEQT
Consensus  DGRNMQLRKK RAAPLVGPSG IITPDGQLIQ LEAGVTVVNA GPSGAVLSNG rSIQYRGkra

           181                                                 227
     CUTA  VPEAIVGPSG IITPDGQLIQ LPAGVTVVNA GPSGAVLSNG RSIQYKL
     CUTB  LPDAIVGPSG IITPDGQLIQ LEAGVTVVNA GPSGAVLSNG RSIQYKL
     CUTC  --DALLGPSG IITPDGQLIQ LPAGVTVVNA GPSGAVLSNG RSIQYKL
Consensus  .p#AivGPSG IITPDGQLIQ LpAGVTVVNA GPSGAVLSNG RSIQYKL
```

Figure 5.8: Amino acid alignment of three cuticle protein isoforms identified in *G. pulex*

Coloured blocks represent 18-motif pairs commonly found in crustacean calcified exoskeletons; green residues are hydrophilic substitutions, purple residues are hydrophobic substitutions, grey residues are neutral substitutions and black residues are possible change of function mutations.

5.2.3 Glyceraldehyde 3-phosphate dehydrogenase (GAPDH)

Although QPCR is a widely accepted method of mRNA quantification; the process of normalisation is a critical procedure to ensure valid results (Huggett *et al.* 2005). To reduce variation, equal quantities of starting RNA were used in reverse transcription reactions. However, efficiency of reverse transcription reactions can vary between samples so an internal reference gene is required; a housekeeping gene with consistent expression levels. To take into account the starting quantity of cDNA used in QPCR analysis, a reference gene probe and primer set is required to normalise results. Common control genes include β-actin, β_2-microglobulin, glyceraldehyde 3-phosphate dehydrogenase (GAPDH) and 18S ribosomal RNA.

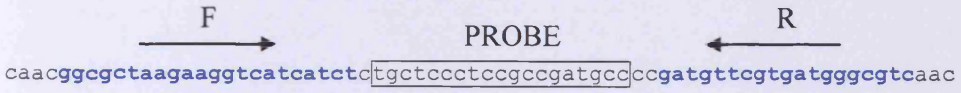
A section of the gene encoding GAPDH was identified in *G. pulex* by PCR using primers specific to *Daphnia pulex* GAPDH gene. A BLASTX search of *G. pulex* GAPDH showed highest consensus with red swamp crayfish GAPDH gene (Figure 5.9). A probe and flanking primers, specific to the gene, were designed for QPCR analysis (Dr. Judith Richards, unpublished data; Figure 5.10).

```
                  1                                                  50
        G. pulex
      P. clarkii MSKIGINGFG RIGRLVLRAA LQNGAQVVAV NDPFIALDYM VYMFKYDSTH
C. quadricarinatus MSKIGINGFG RIGRLVLRAA LQNGAEVVAV NDPFIALDYM VYMFKYDSTH
      C. sapidus              RLVLRAA LLKGAEVVAV NDPFIALDYM VYMFKYDSTH

                  51                                                 100
        G. pulex                                                   ESTGVFT
      P. clarkii GVFKGEVKAE DGALVVDGNK IVVYNEMKPE NIPWSKAGAE YIVESTGVFT
C. quadricarinatus GRFKGEVKEE DGALVVNGHN IQVFNEMKPE NIPWSKAGAE YIVKSTGVFT
      C. sapidus GVYKGEVKAE DGALVVDGHK ITVYNEMKPE NIPWSKAGAE YVVESTGVFT

                 101                                                 150
        G. pulex TLDKAKVHFS NGAKKVIISA PSADAPMFVM GVNEEKYTKD MTIVSNASCT
      P. clarkii TIEKASAHFT GGAKKVVISA PSADAPMFVC GVNLEKYSKD MTVVSNASCT
C. quadricarinatus TIDKAQAHID GGAKKVIISA PSADAPMFVV GVNLEAYDPS MKIVSNASCT
      C. sapidus TIEKASAHFT GGAKKVIISA PSADAPMFVC GVNLEKYSKD MKVVSNASCT

                 151                                                 200
        G. pulex TNCLAPIAKV LNDNFGIEEG LMTTVHAVTA TQKTVDGPSA KDWRGGRGAG
      P. clarkii TNCLAPVAKV LHENFEIVEG LMTTVHAVTA TQKTVDGPSA KDWRGGRGAA
C. quadricarinatus TNCLAPLAKV INDNLEIVEG LMTTVHAVTA TQKTVDGPSA KDWRGGRGAA
      C. sapidus TNCLAPVAKV LHDNFEIVQG LMTTVHAVTA TQKTVDGPSA KDWRGGRGAA

                 201                                                 250
        G. pulex QNIIPSSTGA AKAVGKVIPE LNGKLTGMAF RVPTPDVSVV DLTVRLAKEC
      P. clarkii QNIIPSSTGA AKAVGKVIPE LNGKLTGMAF RVPTPDVSVV DLTVRLGKEC
C. quadricarinatus QNIIPSSTGA AKAVGKVIPE LNGKLTGMAF RVPTPDVSVV DLTVRLGNQC
      C. sapidus QNIIPSSTGA AKAVGKVIPE LNGKLTGMAF RVPTPDVSVV DLTCILGKDC

                 251                                                 300
        G. pulex TYDDIKAAMK AASEAGPLKG YLGYTEDDVV SCDFTGDSRS SIFDAKAGIQ
      P. clarkii SYDDIKAAMK AASE-GPLKG ILGYTEDDVV SCDFTGDIRS SIFDAKAGIQ
C. quadricarinatus SYDDIKAAMK AASQ-GPLKG VLGYTEDDVV STDFTGDVRS SIFDAKAGIQ
      C. sapidus SYDDIKAAMK AAAE-GPLKG VLGYTEDDVV SCDFTGDERS SIFDAKAGIQ

                 301                                335
        G. pulex LSKTFVKVVS WYDNE
      P. clarkii LSKTFVKVVS WYDNEFGYST RVIDLLKHMQ KVDSA
C. quadricarinatus LSKTFVKVVS WYDNEFGYST RVIDLLKHMQ KVDGA
      C. sapidus LSKTFVKVVS
```

Figure 5.9: Alignment of *G. pulex* GAPDH protein with similar proteins

G. pulex GAPDH is compared with the red swamp crayfish, *Procambarus clarkia* (*P. clarkia*), the red claw crayfish, *Cherax quadricarinatus* (*C. quadricarinatus*) and the blue crab, *Callinectes sapidus* (*C. sapidus*).

```
LOCUS       DQ272519                666 bp    mRNA    linear   INV 22-NOV-2005
DEFINITION  Gammarus pulex glyceraldehyde 3-phosphate dehydrogenase (GAPDH)
            partial sequence.
ACCESSION   DQ272519
VERSION     DQ272519
KEYWORDS    .
SOURCE      Gammarus pulex
  ORGANISM  Gammarus pulex
            Eukaryota; Metazoa; Arthropoda; Crustacea; Malacostraca;
            Eumalacostraca; Peracarida; Amphipoda; Gammaridea; Gammaroidea;
            Gammaridae; Gammarus.
REFERENCE   1  (bases 1 to 666)
  AUTHORS   Sambles,C., Richards,J., Johnson,I. and Kille,P.
  TITLE     Gammarus pulex response to estrogen and testosterone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 666)
  AUTHORS   Sambles,C., Richards,J., Johnson,I. and Kille,P.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-NOV-2005) Biosciences, Cardiff University, BIOSI 1,
            School of Biosciences, Cardiff University, Museum Avenue, Cardiff
            CF10 3TL, Wales
FEATURES             Location/Qualifiers
     source          1..666
                     /organism="Gammarus pulex"
                     /mol_type="mRNA"
                     /db_xref="taxon:52641"
     CDS             <1..>666
                     /note="GAPDH"
                     /codon_start=1
                     /product="glyceraldehyde 3-phosphate dehydrogenase"
                     /protein_id="ABB86522"
                     /translation="ESTGVFTTLDKAKVHFSNGAKKVIISAPSADAPMFVMGVNEEKY
                     TKDMTIVSNASCTTNCLAPIAKVLNDNFGIEEGLMTTVHAVTATQKTVDGPSAKDWRG
                     GRGAGQNIIPSSTGAAKAVGKVIPELNGKLTGMAFRVPTPDVSVVDLTVRLAKECTYD
                     DIKAAMKAASEAGPLKGYLGYTEDDVVSCDFTGDSRSSIFDAKAGIQLSKTFVKVVSW
                     YDNE"
ORIGIN
        1  gagtccactg gggttttcac aaccctcgac aaggccaagg ttcacttcag
       51  caacggcgct aagaaggtca tcatctctgc tccctccgcc gatgcccgga
      101  tgttcgtgat gggcgtcaac gaagagaagt acacaaagga catgaccatc
      151  gtgtctaacg cctcctgcac caccaactgc ctcgcaccca tcgctaaggt
      201  cctcaacgac aatttcggta tcgaggaggg cctcatgacc accgtgcacg
      251  cagtcaccgc cacccagaag accgtggacg ggccctcagc caaggattgg
      301  cgtggtggcc gtggcgctgg ccagaacatc attccatcct ccaccggagc
      351  cgctaaggcc gtcggtaaag tcatcccaga actgaacggc aagctcaccg
      401  gcatggcgtt ccgtgtgccc acccccgacg tctccgtggt cgacttgacc
      451  gtccgtctcg caaaggagtg cacttacgac gacatcaagg ccgccatgaa
      501  ggcagcctca gaggctggac ccctcaaggg ctaccttgga tacaccgagg
      551  acgatgtcgt gtcctgcgac ttcaccggcg acagccgatc ctctatcttt
      601  gacgcaaaag ccggcatcca gctgagcaag actttcgtca aggtggtgtc
      651  ctggtacgac aatgag
//
```



Gp_GAPDH_F: ggc gct aag aag gtc atc atc t

GAPDH probe: tgc tcc ctc cgc cga tgc c (dual labelled with 3' TAMRA and 5' 6-FAM)

Gp_GAPDH_R: gac gcc cat cac gaa cat c

Figure 5.10: GAPDH probe and primer sequences
F = forward; R = reverse

5.3 NORMAL PHYSIOLOGICAL GENE EXPRESSION

In order to test whether vitellogenin had the potential to be a suitable indicator of endocrine disruption, normal physiological responses were identified.

5.3.1 Vitellogenin

Total RNA was extracted (Chapter 2, Section 2.11.3) from male and female *G. pulex*, which had been sexed by their precopulatory pairing position (see Chapter 1, Section 1.1.4) before preservation in RNA*later*®. This was performed by Dr. Jason Weeks and Dr. Ian Johnson at WRc-NSF at the same time as chemical exposures described in Section 5.4. After reverse transcription of RNA to produce cDNA (Chapter 2, Section 2.11.6), 14 male and 16 female cDNA samples were analysed for differential vitellogenin expression levels using QPCR analysis (Chapter 2, Section 2.11.13) using the vitellogenin probe and primer set (Section 5.1.1) with GAPDH as an internal reference gene (Section 5.2.3).

Amplification efficiency was calculated from the gradient of the standard curve ($E = 10^{-1/gradient}$) and average efficiency (2.11) used to determine normalised expression fold change. Data were normalised with the reference gene, GAPDH, using the ΔΔCt method (Livak and Schmittgen 2001) to give normalised expression fold change ($E^{-\Delta\Delta Ct}$). Male samples were calibrated against female and *vice versa*. The results of this are shown in Figure 5.11. Probability of a statistical difference between male and female population was determined using ΔCt values in a Mann-Whitney test.

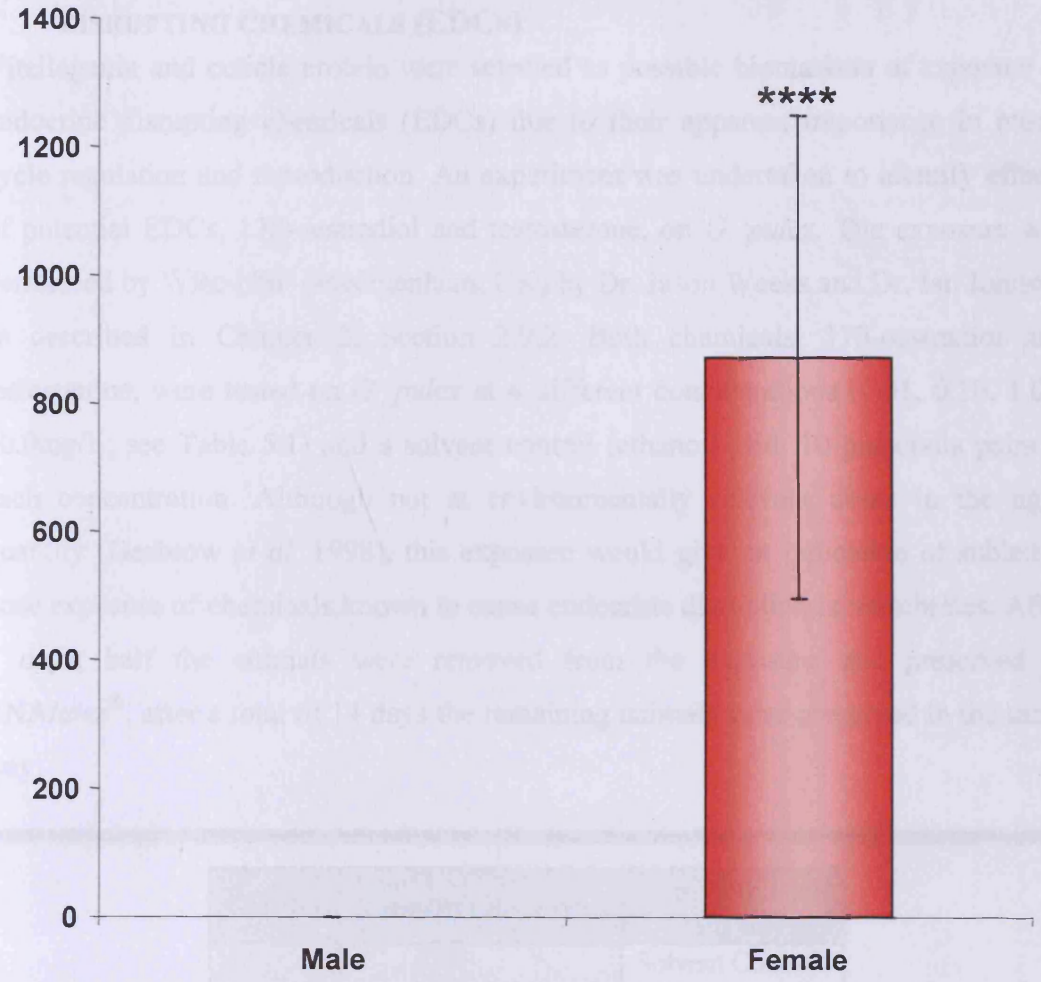


Figure 5.11: Vitellogenin expression levels in *G. pulex*

Levels of vitellogenin in males and females were analysed by QPCR. Females showed significantly higher expression (~870x) than males (mean = 0.053) confirming the suitability of this gene as a sex determination molecular marker. Probability was calculated using the non-parametric Mann-Whitney test; **** = p is significant at ≤ 0.0001 compared to male normalised expression fold change.

5.3.2 Cuticle protein

In order to determine whether or not cuticle protein can be utilised as a marker of moult stage, QPCR analysis would need to be performed on moult staged cDNA populations. Due to the expert knowledge required to identify the moult stage of *G. pulex* by microscopy, moult staging was not available at this stage of the study. However, previous preliminary tests had been performed successfully by Dr. Judith Richards and Mr. John Morgan in order to test the efficiency of the probe and primer set for cuticle gene expression in *G. pulex*.

5.4 RESPONSE OF GENE EXPRESSION AFTER EXPOSURE TO ENDOCRINE DISRUPTING CHEMICALS (EDCs)

Vitellogenin and cuticle protein were selected as possible biomarkers of exposure to endocrine disrupting chemicals (EDCs) due to their apparent importance in moult cycle regulation and reproduction. An experiment was undertaken to identify effects of potential EDCs, 17β-oestradiol and testosterone, on *G. pulex*. The exposure was performed by WRc-NSF (Medmenham, UK) by Dr. Jason Weeks and Dr. Ian Johnson as described in Chapter 2, Section 2.9.2. Both chemicals, 17β-oestradiol and testosterone, were tested on *G. pulex* at 4 different concentrations (0.01, 0.10, 1.00, 10.0mg/L; see Table 5.1) and a solvent control (ethanol) with 10 precopula pairs at each concentration. Although not at environmentally relevant doses in the ng/L quantity (Desbrow *et al.* 1998), this exposure would give an indication of sublethal dose exposure of chemicals known to cause endocrine disruption in vertebrates. After 7 days, half the animals were removed from the exposure and preserved in RNA*later*®, after a total of 14 days the remaining animals were preserved in the same way.

Gender	Exposure chemical	Concentration mg/L
Female	17β-Oestradiol	Solvent Control 0.01 0.1 1 10
	Testosterone	Solvent Control 0.01 0.1 1 10
Male	17β-Oestradiol	Solvent Control 0.01 0.1 1 10
	Testosterone	Solvent Control 0.01 0.1 1 10

Table 5.1: *G. pulex* samples analysed by QPCR

Total RNA was extracted from 159 surviving exposed and control animals (Table 5.2; preserved in RNA*later*®) as described in Chapter 2, Section 2.11.3. First strand cDNA was synthesised from 129 successfully extracted total RNA samples (Chapter 2, Section 2.11.6) and used as a template in QPCR (Chapter 2, Section 2.11.13).

Gender	Exposure chemical	Concentration mg/L	Successful RNA extractions	
			7 days	14 days
Female	17β-Oestradiol	Solvent Control	3	2
		0.01	3	4
		0.1	2	2
		1	3	3
		10	2	1
	Testosterone	Solvent Control	3	4
		0.01	3	4
		0.1	2	4
		1	5	3
		10	5	4
Male	17β-Oestradiol	Solvent Control	4	5
		0.01	3	3
		0.1	3	3
		1	4	4
		10	4	2
	Testosterone	Solvent Control	4	4
		0.01	1	4
		0.1	2	3
		1	4	5
		10	1	4

Gender	Exposure chemical	Concentration	Surviving animals after 14 days (%)
All	17β-Oestradiol	Solvent Control	70
		0.01	80
		0.1	55
		1	70
		10	50
	Testosterone	Solvent Control	95
		0.01	95
		0.1	80
		1	90
		10	90

Table 5.2: Animals surviving after EDC exposures and successfully extracted RNA samples

5.4.1 Vitellogenin expression analysis

QPCR was performed in duplicate using the vitellogenin probe and primer set (Section 5.1.1) and the GAPDH probe and primer set (Section 5.2.3) as an internal control. Individual *G. pulex* cDNA was analysed by QPCR using the same method. QPCR results were separated into male 17β-oestradiol-exposed, female 17β-oestradiol-exposed, male testosterone-exposed and female testosterone-exposed in separate exposure concentrations. Data from organisms exposed for 7 and 14 days were separated in order to see effects of a longer exposure time. Amplification efficiency was calculated from the gradient of the standard curve ($E = 10^{-1/gradient}$) and average efficiency (2.11) used to determine normalised expression fold change. For each sample, data was normalised with the reference gene, GAPDH, using the ΔΔCt method (Livak and Schmittgen 2001) to give normalised expression fold change ($E^{-\Delta\Delta Ct}$) calibrated against the solvent control. A graph of these results is shown in Figure 5.12. Probability of a statistical difference between exposed and solvent control animals was determined using ΔCt values in a Mann-Whitney test.

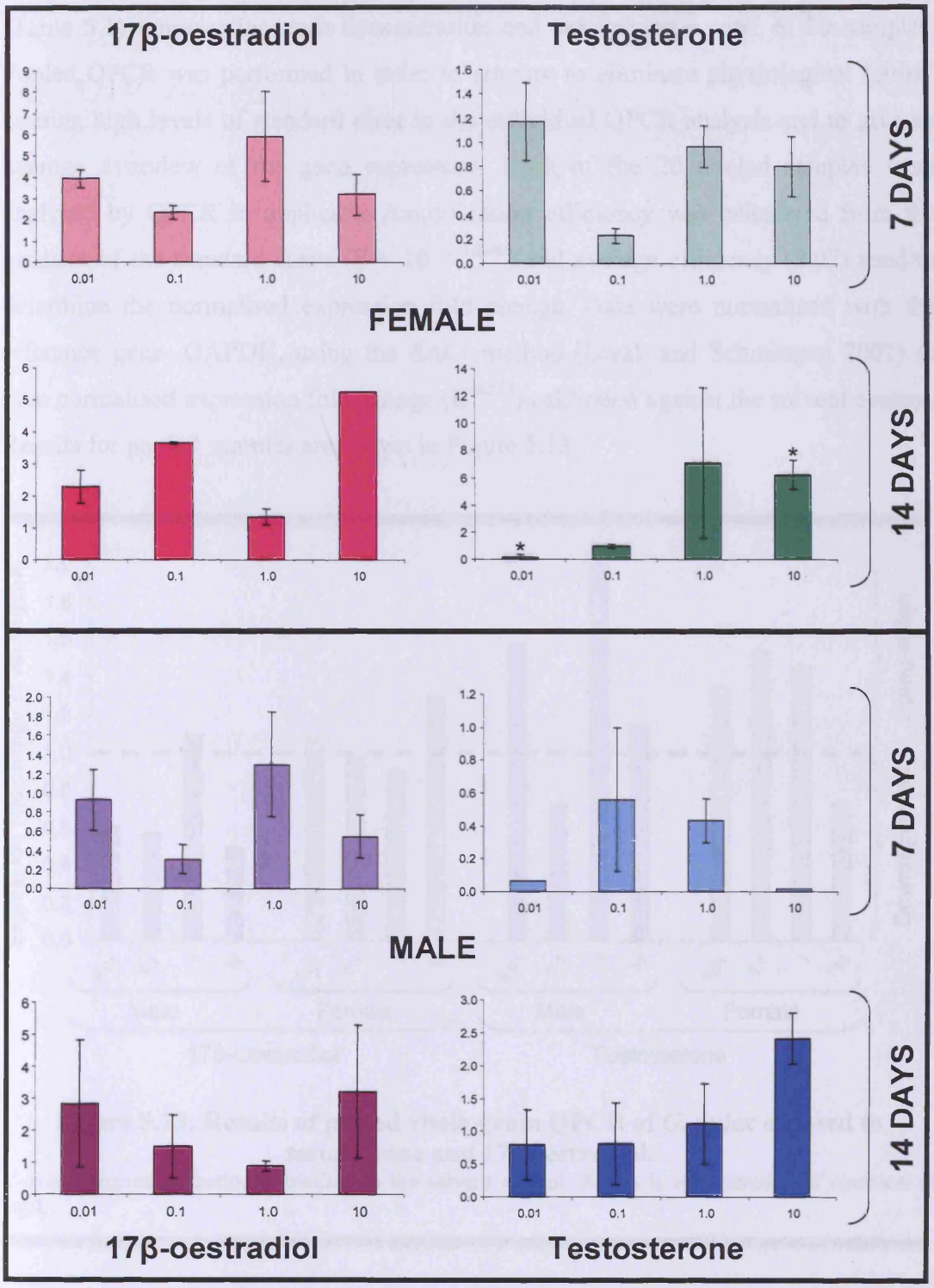


Figure 5.12: Results of vitellogenin QPCR of individual *G. pulex* exposed to testosterone and 17β-oestradiol.

Y-axis is expression ratio, normalised to the solvent control. X-axis is concentration of chemical in mg/L. Probability was calculated using the non-parametric Mann-Whitney test; * = p is significant at ≤ 0.05 compared to solvent control. Low number of statistical outcomes are due to high levels of error in control samples.

An equal quantity of cDNA from each individual sample was pooled into groups (Table 5.1) representing each concentration and sex, giving a total of 20 samples. Pooled QPCR was performed in order to attempt to eliminate physiological 'noise' causing high levels of standard error in the individual QPCR analysis and to give an average overview of the gene expression. Each of the 20 pooled samples were analysed by QPCR in duplicate. Amplification efficiency was calculated from the gradient of the standard curve ($E = 10^{-1/gradient}$) and average efficiency (2.07) used to determine the normalised expression fold change. Data were normalised with the reference gene, GAPDH, using the ΔΔCt method (Livak and Schmittgen 2001) to give normalised expression fold change ($E^{-\Delta\Delta Ct}$) calibrated against the solvent control. Results for pooled samples are shown in Figure 5.13.

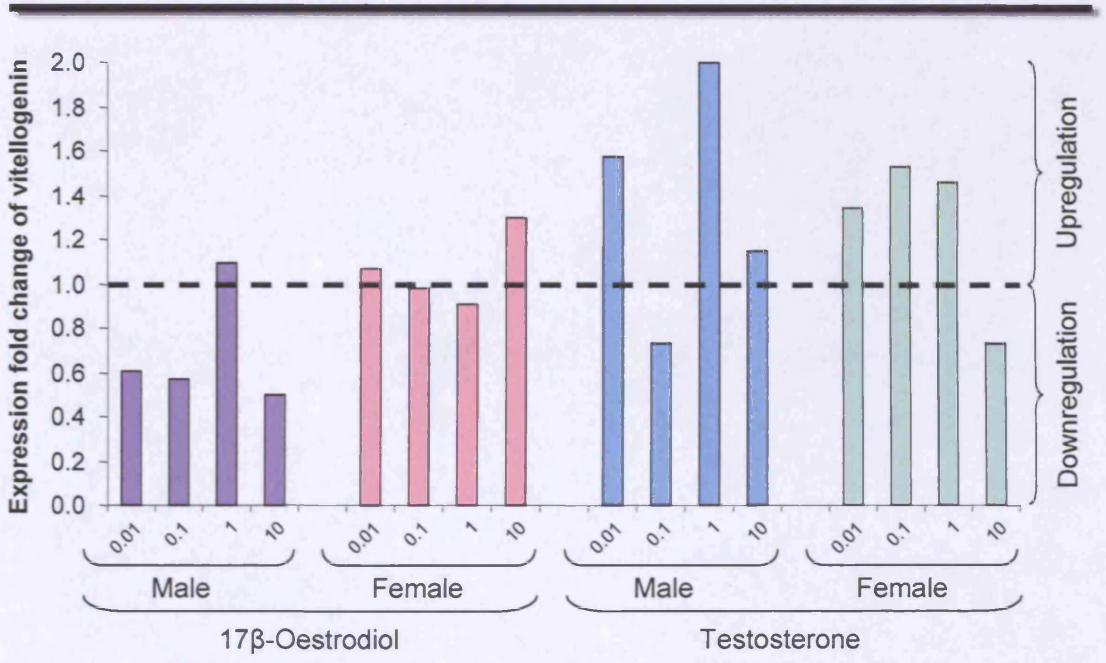


Figure 5.13: Results of pooled vitellogenin QPCR of *G. pulex* exposed to testosterone and 17β-oestradiol.

Y-axis is expression ratio, normalised to the solvent control. X-axis is concentration of chemical in mg/L.

Due to the small number of animals that survived exposure, pooled samples were used to give an indication of any potential changes in vitellogenin expression after exposure to testosterone and 17β-oestradiol. As these were pooled samples and not biological replicates, no statistical tests can be performed; each pooled sample was tested in duplicate to account for technical errors.

5.4.2 Cuticle protein expression analysis

QPCR was performed using a probe and primer set recognising a section of cuticle protein gene (Section 5.2.2). A *G. pulex* cuticle protein gene had been identified previously from a cDNA library (Dr. Judith Richards, unpublished data). As an internal control, the GAPDH probe and primer set were used (Section 5.2.3).

In the same way that both individual and pooled samples were analysed by QPCR for vitellogenin expression (Section 5.4.1), cuticle protein expression was determined (for sample numbers see Table 5.2). Amplification efficiency was calculated from the gradient of the standard curve ($E = 10^{-1/gradient}$) and average efficiency (2.125) used to determine normalised expression fold change. Data was normalised with the reference gene, GAPDH, using the ΔΔCt method (Livak and Schmittgen 2001) to give normalised expression fold change ($E^{-\Delta\Delta Ct}$) calibrated against the solvent control.

Results for individual QPCR are shown in Figure 5.14 and results for pooled results are shown in. Figure 5.15. For individual QPCR samples, probability of a statistical difference between exposed and solvent control animals was determined using ΔCt values in a Mann-Whitney test. As the pooled samples were not biological replicates, no statistical tests were performed; each pooled sample was tested in duplicate to account for technical errors.

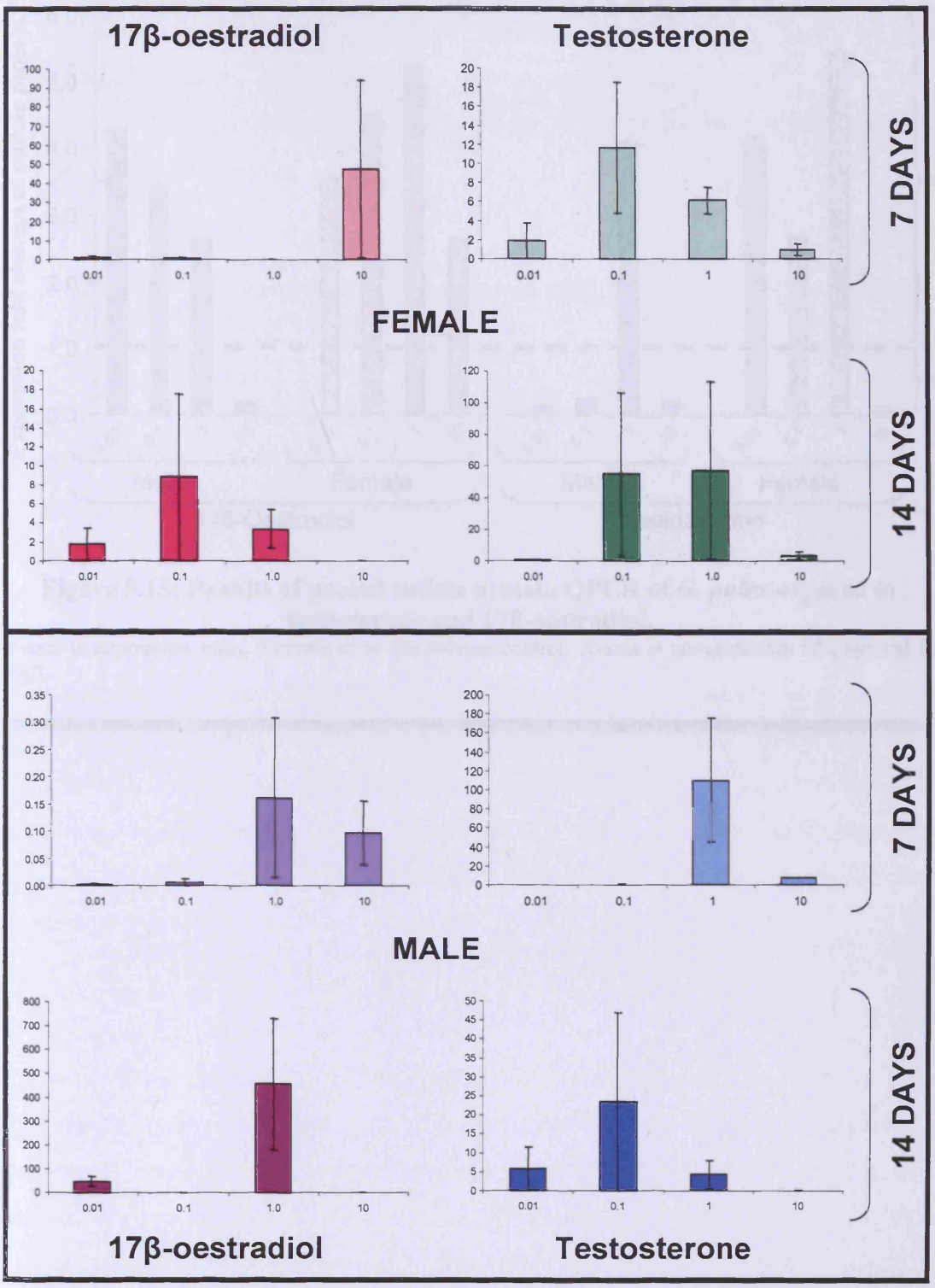


Figure 5.14: Results of individual cuticle protein QPCR of *G. pulex* exposed to testosterone and 17β-oestradiol.

Y-axis is expression ratio, normalised to the solvent control. X-axis is concentration of chemical in mg/L. Probability was calculated using the non-parametric Mann-Whitney test; no significant differences compared to solvent control

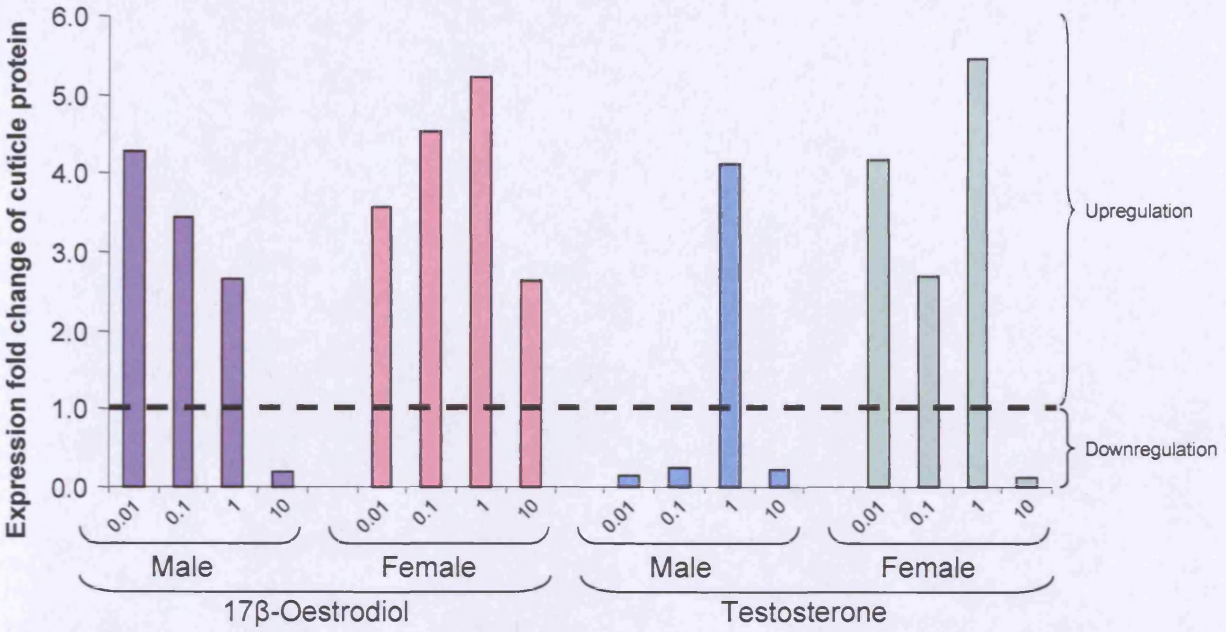


Figure 5.15: Results of pooled cuticle protein QPCR of *G. pulex* exposed to testosterone and 17β-oestradiol.

Y-axis is expression ratio, normalised to the solvent control. X-axis is concentration of chemical in mg/L.

5.5 DISCUSSION

From these experiments, vitellogenin was identified as a sex determination marker, with gene expression up-regulated by ~870x in females than in males when both genders were identified morphologically. It can therefore be used as a female sex specific marker in *G. pulex*.

5.5.1 Changes in vitellogenin gene expression after exposure to 17β-oestradiol in male and female *G. pulex*

5.5.1.1 Female response

After a 7 day exposure to 17β-oestradiol, no clear trend in vitellogenin gene expression was observed across chemical concentrations, however all doses showed an up-regulation of vitellogenin compared to control animals. The exact opposite dose response pattern was observed in females after 14 days, but similarly all doses showed an up-regulation of vitellogenin gene expression. Due to only 1 sample being available for analysis for females exposed to 10mg/L 17β-oestradiol for 14 days, this particular value is not reliable. Pooled samples revealed a U-shaped dose response curve with a low dose reduction in gene expression contrasting with a high dose enhancement of vitellogenin gene expression. The removal of certain biological variables including variation in expression levels due to different moult stage was achieved by pooling samples. This appears to have revealed a possible U-shaped dose response trend.

5.5.1.2 Male response

Male *G. pulex* exposed to 17β-oestradiol for 7 days showed no clear trend in vitellogenin gene expression in relation to dose showing changes in expression between 0.3x and 1.3x the expression levels of the control animals. However after a total of 14 days exposed to 17β-oestradiol, an up-regulation was evident in all doses except 1.0mg/L which showed expression levels comparable with control animals. Pooled samples showed down-regulation of vitellogenin compared to the gene expression levels of control animals apart from at a dose of 1.0mg/L where a small up-regulation of 1.1x was evident. Therefore, in pooled samples for male *G. pulex*

expose to 17β-oestradiol, any biological variation appears to have been removed showing almost no alteration in vitellogenin gene expression irrespective of dose.

5.5.2 Changes in vitellogenin gene expression after exposure to testosterone in male and female *G. pulex*

5.5.2.1 Female response

The individual female vitellogenin response to testosterone shows a possible dose-response trend after a 14 day exposure, with gene expression increasing with concentration, peaking at 1mg/L with expression levels at 7x higher than in control amphipods. Gene expression started to decrease at 10mg/L which may be the start of a fall in expression with increased concentrations. It is possible that by studying the response to vitellogenin expression in females at testosterone concentrations above 10mg/L, an inverted U-shaped dose response curve may have been evident. The 14 day exposure showed a more prominent vitellogenin gene expression response; exposure to testosterone for 7 days only elicited changes in expression of between 0.2x to 1.2x showing no observable difference to control animals, compared to a peak of 7x up-regulation after a longer exposure. Changes in vitellogenin gene levels in pooled female samples showed an inverted U-shaped hormetic response to testosterone exposure showing stimulation in vitellogenin gene expression at lower doses and inhibition of expression at the highest dose. Samples were pooled to eliminate individual biological variations such as moult stage or age. Due to small sample numbers, both 7 day and 14 day exposure samples were combined to give an overall response to acute testosterone exposure.

5.5.2.2 Male response

Exposure to males to testosterone over a 7 day period showed a reduction in vitellogenin expression of ~0.1x to ~0.6x compared to control animals. After a total of 14 days exposed to testosterone, a 2.5x up-regulation in vitellogenin was observed at the highest concentration of 10mg/L, with the lower doses showing expression levels comparable to control animals. Pooled samples changed the pattern of vitellogenin gene expression and no clear trend was evident. This could be due to the opposite response shown in the shorter exposure where down-regulation was observed,

compared to the full 14 day exposure where there was no change apart from an up-regulation at the highest dose.

5.5.3 Changes in cuticle protein gene expression after exposure to 17β-oestradiol in male and female *G. pulex*

5.5.3.1 Female response

An increase in cuticle protein gene expression was evident in female *G. pulex* exposed to 10mg/L 17β-oestradiol for 7 days. Expression peaked at 50x that of control animals; down-regulation of cuticle protein gene expression was observed at all other doses. Females exposed to 17β-oestradiol for 14 days revealed a different expression pattern. Peak expression was shifted towards the lower doses showing an 11x up-regulation at 0.1mg/L compared to control animals, the doses immediately lower and higher (0.01mg/L and 1.0mg/L) showed similar levels of up-regulation of gene expression at 2x to 3x that of control animals. The highest dose of 10mg/L displayed a down-regulation of cuticle protein gene expression contrasting with the observations recorded in females after 7 days of exposure to 17β-oestradiol. Pooled samples reveal a partial inverted U-shaped dose response pattern, but with no inhibitory effect observed at higher concentration, this may be evident if the concentration range had been expanded to levels greater than 10mg/L. All doses showed an up-regulation after pooling samples at a magnitude of 3x to 5x cuticle protein gene expression levels compared to control animals.

5.5.3.2 Male response

A 7 day exposure of male *G. pulex* to 17β-oestradiol showed down-regulation of cuticle protein gene expression compared to control animals at all doses. In contrast, after a total of 14 days, exposure of males to 17β-oestradiol revealed an increase in cuticle protein gene expression at a dose of 1.0mg/L 17β-oestradiol; up-regulation of expression was also observed at the lowest dose of 0.01mg/L at a magnitude of ~40x higher than expression levels recorded in control animals. At both 0.1mg/L and 10mg/L 17β-oestradiol, there was no change in cuticle protein gene expression compared to control animals. Pooled samples showed a distinct inverted U-shaped hormetic dose response to 17β-oestradiol, with stimulatory effects on cuticle protein

gene expression levels at lower doses and an inhibitory effect at the highest concentration of 10mg/L.

5.5.4 Changes in cuticle protein gene expression after exposure to testosterone in male and female *G. pulex*

5.5.4.1 Female response

Female *G. pulex* exposed to testosterone for 7 days show a distinct inverted U-shaped dose response, peaking at 11x increase in cuticle gene expression at 0.1mg/ml compared to control animals. All doses showed an up-regulation in cuticle protein gene expression except for those exposed to the maximum does of 10mg/L which were comparable to the expression levels in control animals. After a total of 14 days, females exposed to testosterone continued to show an inverted U-shaped dose response, but at augmented levels. The peak expression level was shifted toward the higher doses with the greatest level of cuticle gene expression at both 0.1mg/L and 1.0mg/L at ~50x that observed in control animals or the other doses. The highest and lowest levels show an inhibitory response to testosterone, revealing there may be optimal dose required for stimulation of cuticle gene expression. Pooled samples do not show any clear trends through concentration, highlighting a down-regulation of cuticle protein gene expression at the highest dose of 10mg/L contrasting with an up-regulation of between 3x and 6x at all other doses. It can be hypothesised, therefore, that lower doses induce cuticle gene expression up to an optimal value after which expression is inhibited.

5.5.4.2 Male response

Changes in cuticle protein gene expression levels in male *G. pulex* after a 7 day exposure to testosterone reveal an increase at 1.0mg/L of 110x that of control animals. Lower doses show expression levels down-regulated when compared with control animals, with a small up-regulation of ~5x evident at the highest dose of 10mg/L. When animals were exposed to testosterone for a longer period of 14 days, peak cuticle protein gene expression levels shifted toward the lower doses of testosterone, peaking at 0.1mg/L showing a 25x increase compared to control animals. Pooled samples showed a similar response to the 7 day exposure pattern indicating an optimal stimulatory dose of 1mg/L and inhibitory responses at all other doses. The process of

pooling samples masked the stimulatory effect observed in males exposed to 0.1mg/L testosterone for 14 days, highlighting the risk of eliminating evidence of biological responses using pooled samples.

5.5.5 Conclusions

Despite the apparent trends discussed, there were high levels of errors meaning that the majority of results were not statistically different. The only statistically relevant effect was observed in vitellogenin expression in females exposed to testosterone for 14 days with an inverted U-shaped hormetic dose response; an inhibitory effect was evident at lower concentrations with a stimulatory effect observed at the highest dose (Murado and Vazquez 2006). A hormetic response of vitellogenin protein concentration in response to EDC exposure has already been observed in fathead minnows exposed to 4-nonylphenol (Giesy *et al.* 2000). Errors may be either technical or biological. The effects of technical variation were reduced by replicating pooled samples and standards. Individual samples were not replicated due to insufficient sample volume and this may account for high levels of variation. Results from the pooled sample analysis showed that the highest technical variation recorded in the duplicate samples was 26.6% (coefficient of variation (CV)). Despite the substantial up- or down- regulation observed in some of the experiments, the variation in the control samples was one of the factors preventing the changes in expression from being statistically significant. Biological variation, not including changes induced by chemical exposure, was assessed by pooling samples; however the imbalance of sample numbers per exposure group may have led to a skewed representation of dose response, especially due to the pooling of animals exposed for both 7 days and 14 days. Biological variation is likely to be due to organisms varying in cuticle protein expression throughout their moult cycle, and in this exposure, animals in a variety of moult stages were used. Vitellogenin gene expression has also been observed at different levels in different moult cycle stages in the Pacific white shrimp, *Litopenaeus vannamei* where vitellogenin gene expression peaked in intermoult and early premoult stages and subsequently down-regulated immediately before and after ecdysis (Raviv *et al.* 2006).

The major contributor to experimental error was sample size. Assuming a 50% mortality rate as observed in *G. pulex* exposed to the highest 17β-oestradiol dose and a RNA extraction success rate of 80% as observed in this study, the number of animals used in the exposure should be 40 at each dose (20 females and 20 males) if 8 biological replicates are required. Ideally animals should be moult-staged and grouped accordingly to reduce biological variability. Mortality information received does not differentiate whether survival was equally distributed between gender and whether males or females were more susceptible to toxic effects. Therefore only half the numbers of animals were used in the exposures than were required, to produce at least 8 replicates for each chemical, dose and gender.

In these exposures, any vitellogenin or cuticle protein expression trends or responses to 17β-oestradiol or testosterone have been masked by error levels resulting from the experimental setup of this exposure. There were many variables involved which resulted in a small number of organisms within each exposure, dose and gender group. Variables included, sex, age/size, pairing status, exposure time, exposure chemical and moult stage.

CHAPTER 6

OPTIMISATION OF *GAMMARUS PULEX* MICROARRAYS

6.1 INTRODUCTION

Microarrays allow changes in expression of thousands of genes to be analysed simultaneously, providing an excellent tool for global gene expression profiling. This technology is the preferred method for the assessment of quantitative differences in gene expression and is not dependent on the availability of sequence information for the species used (Kuhn 2001). Alternative methodologies available are macroarrays or serial analysis of gene expression (SAGE). Macroarrays, although more cost effective, do not provide as much information on differential gene expression as microarrays due to limitations in reporter gene numbers; there is a high possibility that a large proportion of genes will not be represented. SAGE allows the analysis of multiple transcripts simultaneously, but requires subsequent 'tag' primer driven PCR to identify every gene of interest. It is therefore an anonymous gene expression profiling technique with the potential to provide selected information.

Microarray technology can utilise 'genomics' to assess the effects of different conditions on biological processes. By comparing samples from different tissues, from animals with a different physiological status or from animals exposed to different doses of chemicals, the molecular mechanisms underlying a biological question can be elucidated.

Microarray technology has been used successfully in the field of ecotoxicogenomics, a combination of ecotoxicology, genomics and bioinformatics (Chapter 1, Section 1.4.3). Expression profiling of *Daphnia magna* in response to copper, cadmium and zinc exposure, using microarray technology, was used to discover potential modes of action and potential biomarkers of exposure (Poynton *et al.* 2007). Effects of endocrine disrupting chemicals (EDCs) on gene expression were determined in zebrafish (*Danio rerio*). Exposure to 17α-ethinyloestradiol at environmentally relevant concentrations affected the reproductive health of breeding zebrafish by decreasing the number and quality of gametes produced in both genders. The molecular mechanisms involved were identified, including processes associated with

gametogenesis in males and females (Santos *et al.* 2007). cDNA microarrays were employed to compare gene expression profiles of the aquatic invertebrate, *Ciona intestinalis*, after exposure to tributyltin (TBT) (Azumi *et al.* 2004). These successful studies reinforced the feasibility of ecotoxicogenomics which is increasingly being utilised in aquatic species (Ju *et al.* 2007).

Further gene expression profiling studies in crustaceans using microarray technology are underway. In *D. magna*, a 'microarray system' is being created to identify transcriptional profiles during chemical toxicity, specifically EDCs (Iguchi *et al.* 2007). In *Cancer magister*, gene expression profiling will be used to identify molecular mechanisms of development and moulting (Terwilliger *et al.* 2006).

cDNA microarrays allow changes in gene expression to be identified in species where there is little genetic data available. As there is little genetic data available for amphipods, in particular *Gammarus* sp., to analyse gene expression, a cDNA microarray was constructed. An alternative method to cDNA arrays often employed is spotted oligonucleotide arrays (Lockhart *et al.* 1996). These are similar to cDNA arrays except the printed target DNA is 50 – 60mers of synthesised oligonucleotide specific to the gene of interest, as opposed to DNA complementary to mRNA (cDNA). The reason this approach cannot be utilised in *G. pulex* is that oligonucleotide arrays require extensive genetic knowledge in order to synthesise oligonucleotide targets.

For each gene profiling experiment using microarrays, a certain amount of optimisation is required to ensure the vast quantities of data generated are reproducible and accurate. The protocol needs to be robust, allowing a large number of samples to be processed as consistently and similarly as possible, as any deviance from the standard methods can ultimately effect the transcriptomic profile. Although the best approach for removing error caused by technical variation is randomisation of samples, further steps must be taken to reduce error to an absolute minimum. Certain stages of this complex technique need to be modified to take into account such variables as different species, quantity of starting tissue and RNA available. Optimisation steps are often taken when assessing the use of CyDye incorporated cDNA synthesised from RNA, as a probe for microarray hybridisation (Naderi *et al.*

2004). Commercially available labelling methods have been compared (Badiee *et al.* 2003) but these are not the most cost effective, or efficient way to label a large number of samples. In addition, optimisation of spotting solutions has been performed using CMT-GAPS slides (Corning, New York, USA), which are coated with γ-amino propyl silane (Rickman *et al.* 2003). This process of testing spotting solutions is required for different slide types as the effectiveness of a spotting solution varies with subtle changes in the slide coating. An approach such as the protocol development undertaken by The Institute For Genomic Research (TIGR; Rockville, Massachusetts, USA (Hegde *et al.* 2000)) would allow for a stable and consistent procedure for the production of high quality differential gene expression profiles. Another study has suggested a 'Design Of Experiments' (DOE) approach to optimising microarrays (Wrobel *et al.* 2003). By developing a protocol with a number of the critical steps tested, a method specific to *G. pulex* was created using a DOE approach. The outline of this process is shown in Figure 6.1.

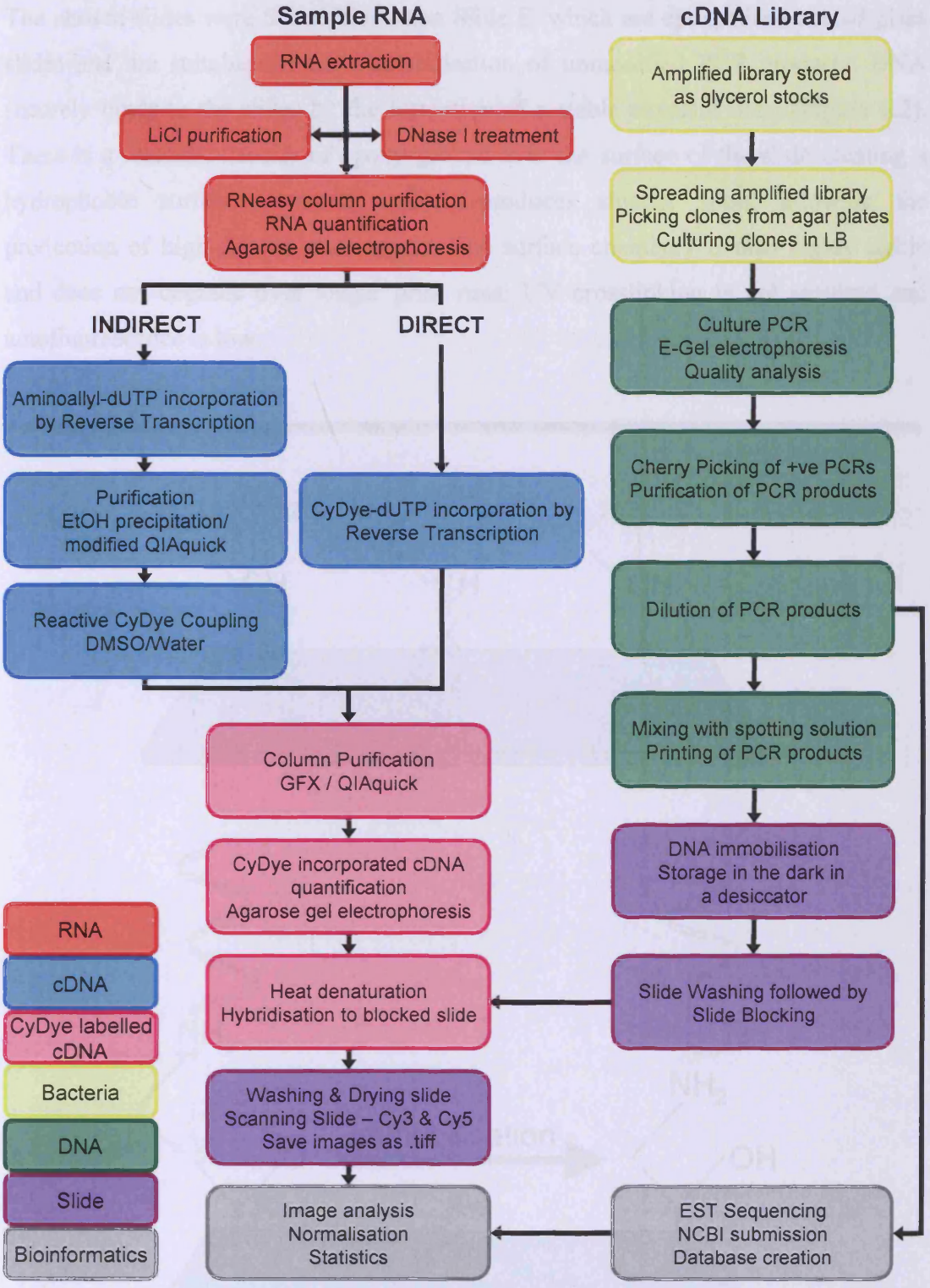


Figure 6.1: Flow diagram showing the process of genomic analysis using microarray technology.

6.1.1 Microarray slide selection

The chosen slides were Schott Nexterion Slide E, which are epoxysilane coated glass slides and are suitable for the immobilisation of unmodified PCR products. DNA securely binds to the slides by the formation of a stable covalent bond (Figure 6.2). There is a constant density of epoxy groups over the surface of the slide creating a hydrophobic surface chemistry, which produces smaller spots, allowing the production of high-density microarrays. The surface chemistry is also highly stable and does not degrade over longer print runs; UV crosslinking is not required and autofluorescence is low.

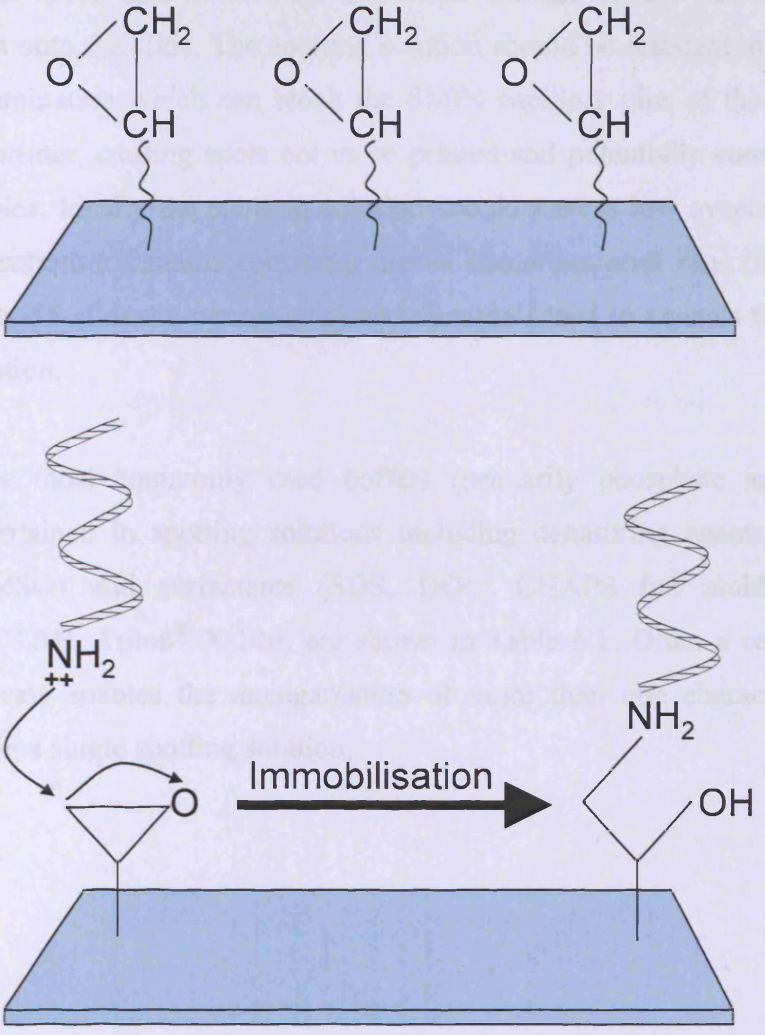


Figure 6.2: Slide chemistry and DNA immobilisation of Schott Nexterion Slide E

6.1.2 Spotting Solutions

The spotting solution used to dilute the PCR products for microarray printing is a fundamental factor governing success of the procedure. The solution used plays a significant part in the size, morphology and brightness of resulting spots and is crucial to deliver the largest amount of data of the best quality and to allow better reproducibility of experiments.

The ideal spotting solution would produce a distinct 'spot' of cDNA on the slide surface, which, after hybridisation, washing and drying retained a perfect circular morphology with equal fluorescent intensity across the spot area. Printed cDNA spots should be large enough in diameter to retrieve useful information (the more pixels available, the more data available), but small enough so that all the spots can physically fit onto the slide. The spotting solution should be resistant to bacterial and fungal contamination which can block the SMP4 capillary pins of the SpotArray72 microarray printer, causing spots not to be printed and potentially contaminating all cDNA samples. Ideally the spotting solution should have a low evaporation rate so cDNA concentration remains consistent across numerous print runs (the process of printing up to 55 slides in one session) and shouldn't lead to crystals forming in the spotting solution.

Some of the most commonly used buffers (primarily phosphate and SSC) and additives contained in spotting solutions including denaturing agents (formamide, betaine, DMSO) and surfactants (SDS, DOC, CHAPS (an amidosulfobetaine detergent), CTAB, Triton® X-100) are shown in Table 6.1. Often a combination of these chemicals enables the amalgamation of more than one characteristic to be represented in a single spotting solution.

Denaturing agents		
Common name	**Chemical name**	**Charge**
Formamide	Formamide	protic solvent
Betaine	2-trimethylammonioacetate	zwitterionic
DMSO	**Di**Methyl**Sulph**Oxide	aprotic solvent
Surfactants		
Common name	**Chemical name**	**Charge**
DOC	**De**Oxy**C**holate	anionic
Tween 20	Poly(oxyethylene)$_n$ sorbitane monolaurate	non-ionic
SDS	**S**odium **D**odecyl **S**ulphate	zwitterionic
CHAPS	3-[(3-Cholamidopropyl)dimethylammonio]-1-propanesulfonate	zwitterionic
CTAB	**C**etyl**T**rimethyl **A**mmonium **B**romide	cationic
Triton® X-100	Octylphenolpoly(ethyleneglycolether)$_n$	non-ionic
Buffers		
Common name	**Chemical name**	**Charge**
SSC	**S**aline **S**odium **C**itrate	N/A
Phosphate	N/A	N/A

Table 6.1: Commonly used buffers and additives used in microarray spotting solutions (discussed in Rickman *et al.* 2003).

A suitable selection of spotting solutions was selected taking into account ease of use, contamination risk, evaporation rate, expense, chemical availability and slide compatibility.

6.2 SPOTTING SOLUTION OPTIMISATION

The advantages and disadvantages of each spotting solution (Table 6.2), surfactant and denaturant were considered.

Denaturing agents		
Common name	**Advantages**	**Disadvantages**
Formamide	Increases spot intensity	Biological hazard
Betaine	Reduces rate of evaporation Increases spot intensity	Crystallises Require rehydration
DMSO	Denatures DNA Low evaporation rate Uniform spots on UltraGAPS (Corning) Most reliably increases spot intensity Reduces rate of evaporation	Strong irritant Increases spot diameter Causes aggregation of DNA at concentrations over 70% Not thought to be compatible with Nexterion Slide E
Surfactants		
Common name	**Advantages**	**Disadvantages**
DOC	Increases spot diameter Decreases surface tension	Crystallises
Tween 20	Increases spot diameter Decreases surface tension	Crystallises
SDS	Increases spot diameter Decreases surface tension	Crystallises
CHAPS	Increases spot diameter Decreases surface tension	Crystallises
CTAB	Increases spot diameter Decreases surface tension	Crystallises
Triton® X-100	Increases spot diameter Decreases surface tension	Crystallises
Buffers		
Common name	**Advantages**	**Disadvantages**
SSC	Commonly used aqueous solvent Produces small diameter spots	Does not denature DNA High evaporation rate
Phosphate	Commonly used aqueous solvent	Prone to bacterial contamination
Proprietary spotting solutions		
Common name	**Advantages**	**Disadvantages**
Nexterion Spot	Readymade	Phosphate based therefore prone to bacterial / fungal contamination
Nexterion Spot Modified	Readymade	Phosphate based therefore prone to bacterial / fungal contamination
Nexterion Spot LE	Readymade Low evaporation rate	Requires evaporation of PCR products
Preservatives		
Common name	**Advantages**	**Disadvantages**
Sodium azide	Reduces the risk of bacterial / fungal contamination	Crystallises

Table 6.2: Advantages and disadvantages of common spotting solution components

A selection of 14 spotting solutions were identified for further testing (Table 6.3).

PRIMARY BUFFER	ADDITIVE	ABBREVIATION
Nexterion Spot Modified *Schott proprietary*	None	NXT
Nexterion Spot *Schott proprietary*	None	NSB
Nexterion Spot *Schott proprietary*	0.1% azide	NSA
Nexterion Low Evaporation *Schott proprietary*	None	NLE
25% DMSO	1.5X SSC	25D15
50% DMSO	None	DM50
25% DMSO	None	DM25
50% DMSO	3X SSC	50D3
50% DMSO	1.5X SSC	50D15
25% DMSO	3X SSC	25D3
3X SSC	None	SSC
3X SSC	0.004% SDS	SDS1
3X SSC	0.001% SDS	SDS2
3X SSC	0.0005% SDS	SDS3.

Table 6.3: Primary buffers, additives and abbreviations for spotting solutions

Three solutions were eliminated due to precipitation; they were 50D3 (precipitated at RT), 50D15 (precipitated at RT) and 25D3 (precipitated slightly when temperature dropped below 20°C).

6.2.1 Contamination

In order to assess the risk of bacterial contamination of the spotting solutions, 200µl of each potential spotting buffer was transferred to a separate well of a 96-well plate and each spotting solution spiked with two different quantities of *E. coli* broth (1µl and 10µl of DH5α; Invitrogen Ltd., Paisley, UK) and with sterile water as a control. The plate was then incubated at 37°C with the lid off (to replicate the possible exposure to air during a print run) for 24 hours and the absorbance recorded before and after incubation. This test was performed on the following 12 potential spotting buffers: NXT, NSB, NSA, NLE, 25D15, DM50, DM25, 50D3, 50D15, 25D3, SSC,

SDS1, SDS2 and SDS3. In addition, water (H_2O) and Luria broth (LB) were also tested as suitable comparisons. The changes in absorbance at 595nm (at 22.2°C) are shown in Figure 6.3.

6.2.2 Evaporation

Each buffer (200µl) was transferred to a separate well of a 96-well plate, which was placed, without the lid on, at room temperature to test how prone the different spotting solutions are to evaporation. The following solutions were tested: NXT, NSB, H2O, NSA, SSC, SDS1, SDS2, SDS3, DM25, DM50, DM25, 25D15, NLE, and Luria Broth (LB). The only buffers that didn't evaporate after 10 days were the DMSO containing solutions, DM25, DM50, 25DM3, 25DM15, also NLE and LB. The other Nexterion solutions (NLA, NLB and NXT), the SSC and SDS based solutions and the water evaporated. In order to test a range of spotting solutions some buffers from both evaporated and non-evaporated groups were tested further. A more accurate evaporation test was undertaken at a later date to assess the effects of evaporation during a microarray print run.

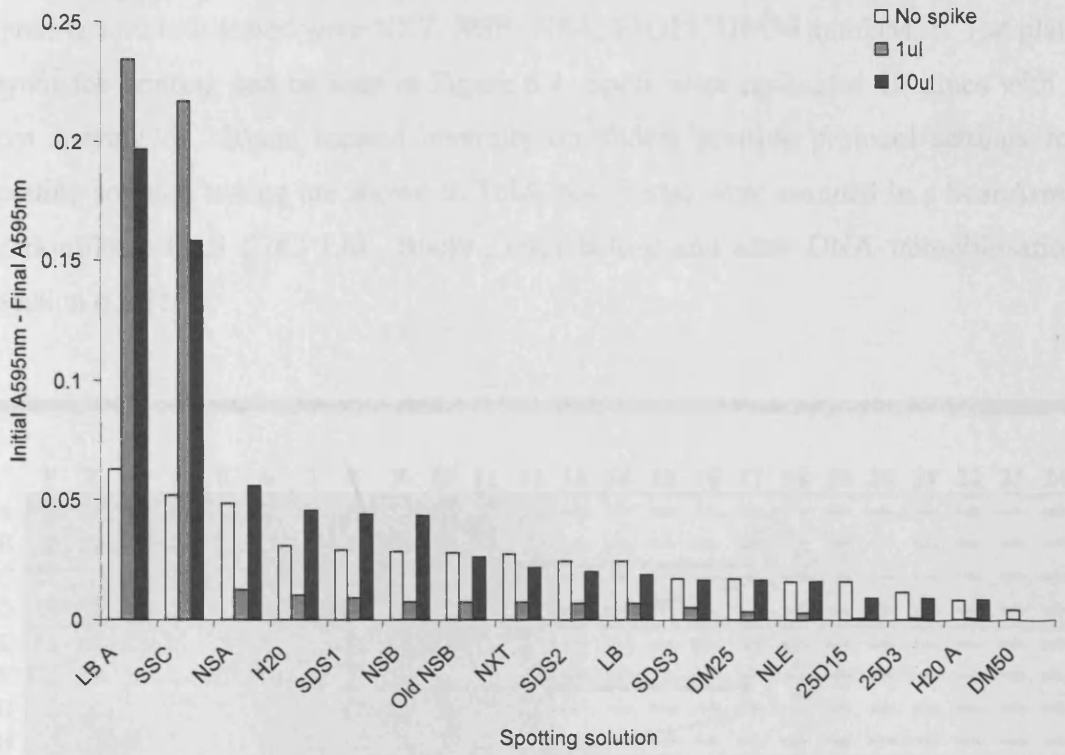


Figure 6.3: The susceptibility of various spotting buffers to bacterial contamination and proliferation

6.2.3 Spotting solution test print run

Spotting solutions tested were NXT, NSB, NSA, 25D15, DM50 and DM25. The plate layout for printing can be seen in Figure 6.4. Spots were replicated 10 times with a spot spacing of 250µm, located centrally on slides; printing protocol settings for spotting solution testing are shown in Table 6.4. Slides were scanned in a ScanArray (PerkinElmer LAS (UK) Ltd., Bucks., UK) before and after DNA immobilisation (Section 6.3.2).

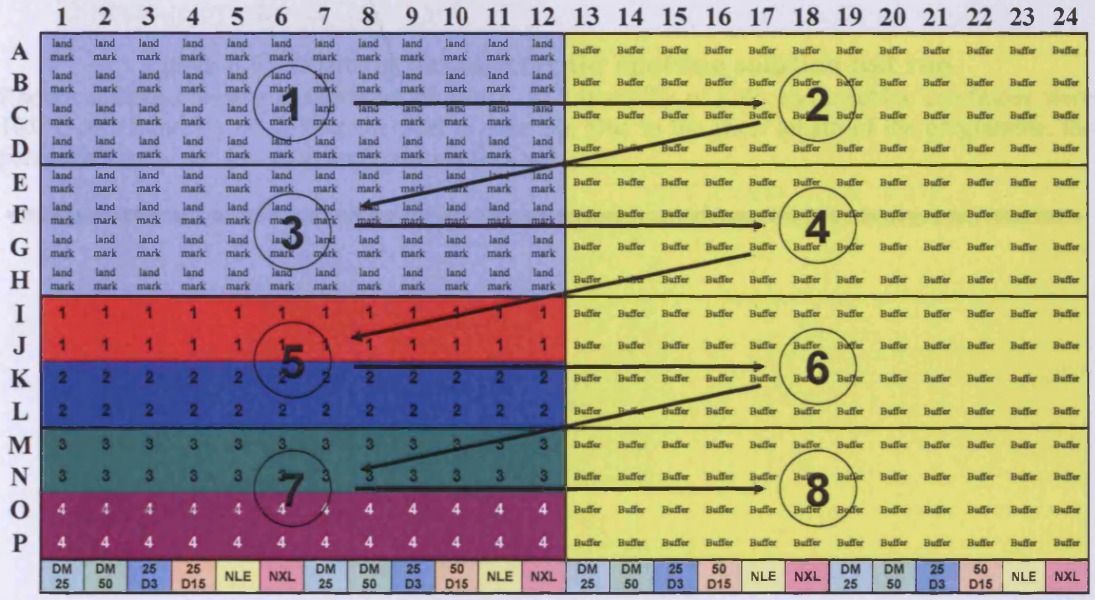


Figure 6.4: Plate layout for Buffer Test

Blocks 1 and 3 contained landmark DNA, prepared from a mixture of CyDye labelled zebrafish (*Danio rerio*) and mouse (*Mus musculus*) PCR products. The production of landmark DNA is described in Section 6.8. Blocks 2, 4, 6 and 8 contained the relevant buffer adjusted to the correct concentration with HPLC grade water. Blocks 5 and 7 contained a combination of 4 different *G. pulex* PCR products which were haemocyanin (maSB_01F03), ferritin (mxAB_02D05), HR3 receptor (mxA_5B12) and an Na^+-coupled citrate transporter gene (maSB_05B2). Arrows and numbers show order of printing.

Buffer Test			
Plate changes	1	Substrate dwell time	200
Pins used	48	Sample overtravel	500
Pre-blots	10	Sample dwell time	2000
Pre-blot spacing	450	Substrate thickness	1
Spot replicates	10	Maximum spot size	150
Nominal spot size	120	X-Y	MEDIUM
Spot spacing	250	Wash	3
Location on slide	CENTRE	Times (Reps)	3
Space for barcode	X	Length (Dry)	3
Approach velocity	15	Humidity	50-65
Departure velocity	25	Optimal humidity	60
Substrate overtravel	500		

Table 6.4: Printing parameters for spotting solution test run

Room conditions were a temperature of 17.1°C at a humidity of 46% and machine conditions were 18.0°C at a humidity of 60% at the start of printing. Due to the short length of the programme, the parameters remained the same until printing finished.

6.2.3.1 Spot size

Spot size was compared for each spotting solution tested, before immobilisation, after immobilisation and after hybridisation (Figure 6.5).

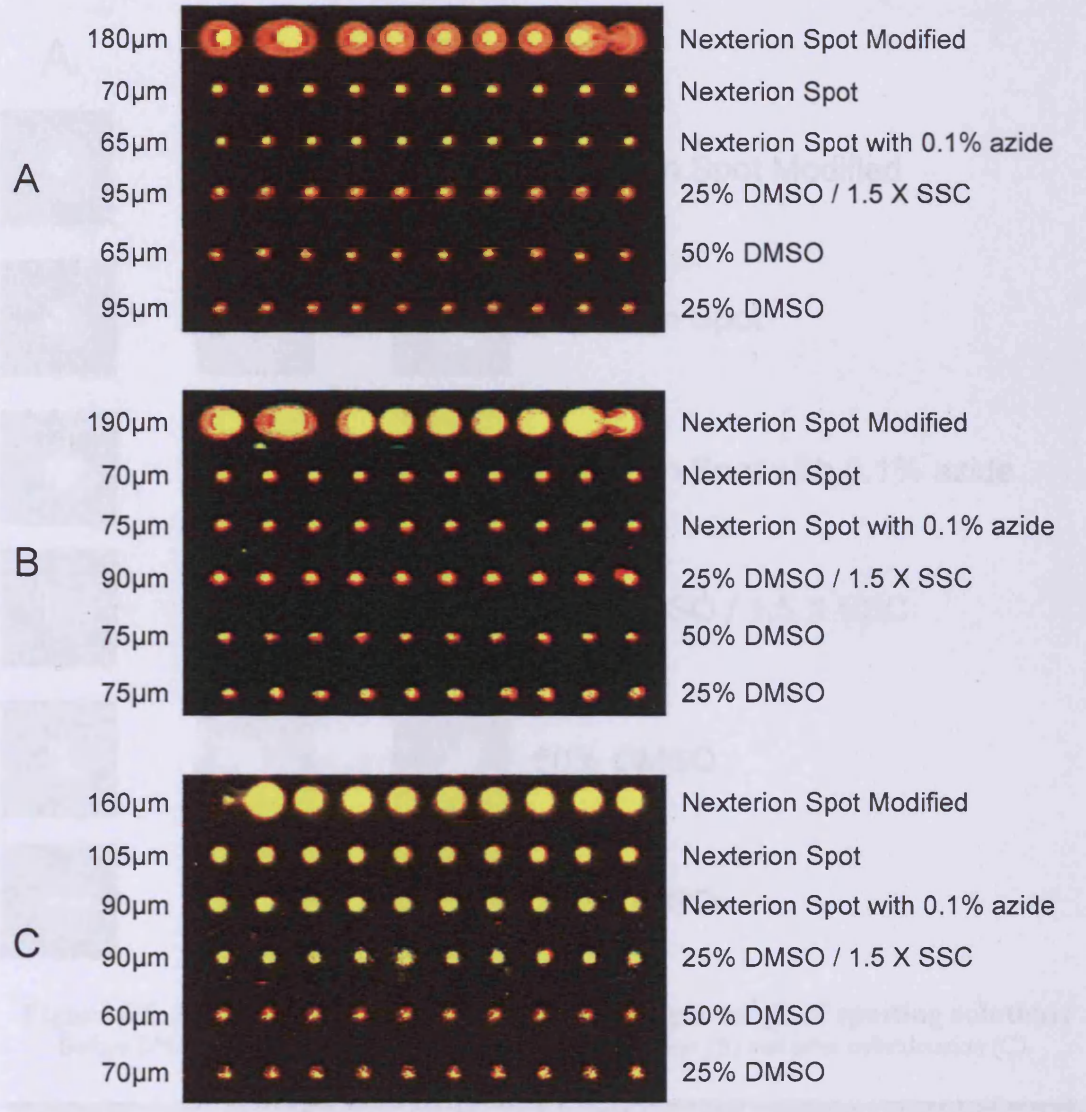


Figure 6.5: Differences in spot size using a range of spotting solutions
Before DNA immobilisation (A), after DNA immobilisation (B) and after hybridisation (C).

6.2.3.2 Spot morphology

Spot morphology was compared for each spotting solution tested, before immobilisation, after immobilisation and after hybridisation (Figure 6.6).

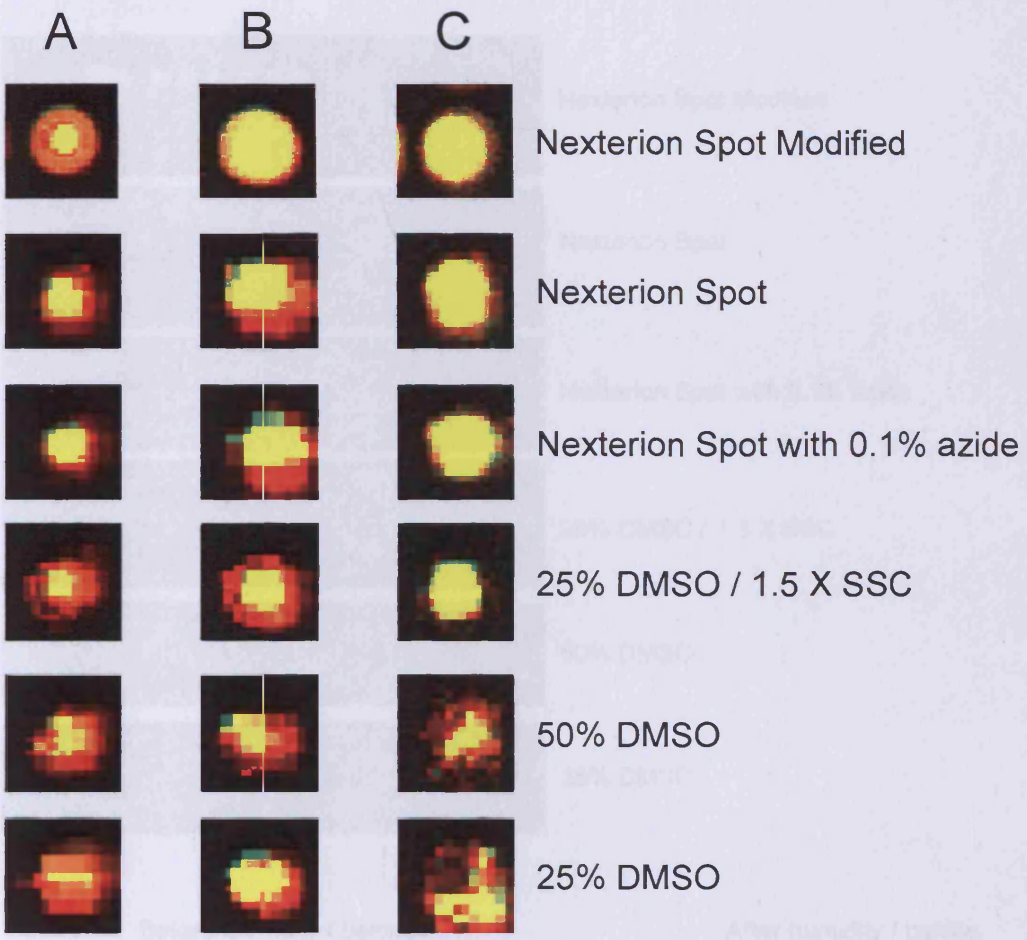


Figure 6.6: Differences in spot morphology using a range of spotting solutions
Before DNA immobilisation (A), after DNA immobilisation (B) and after hybridisation (C).

6.2.3.3 Spotting solution autofluorescence

Before and after DNA immobilisation the level of buffer autofluorescence was determined (Figure 6.7) highlighting the importance of washing slides before hybridisation to prevent signal interference.

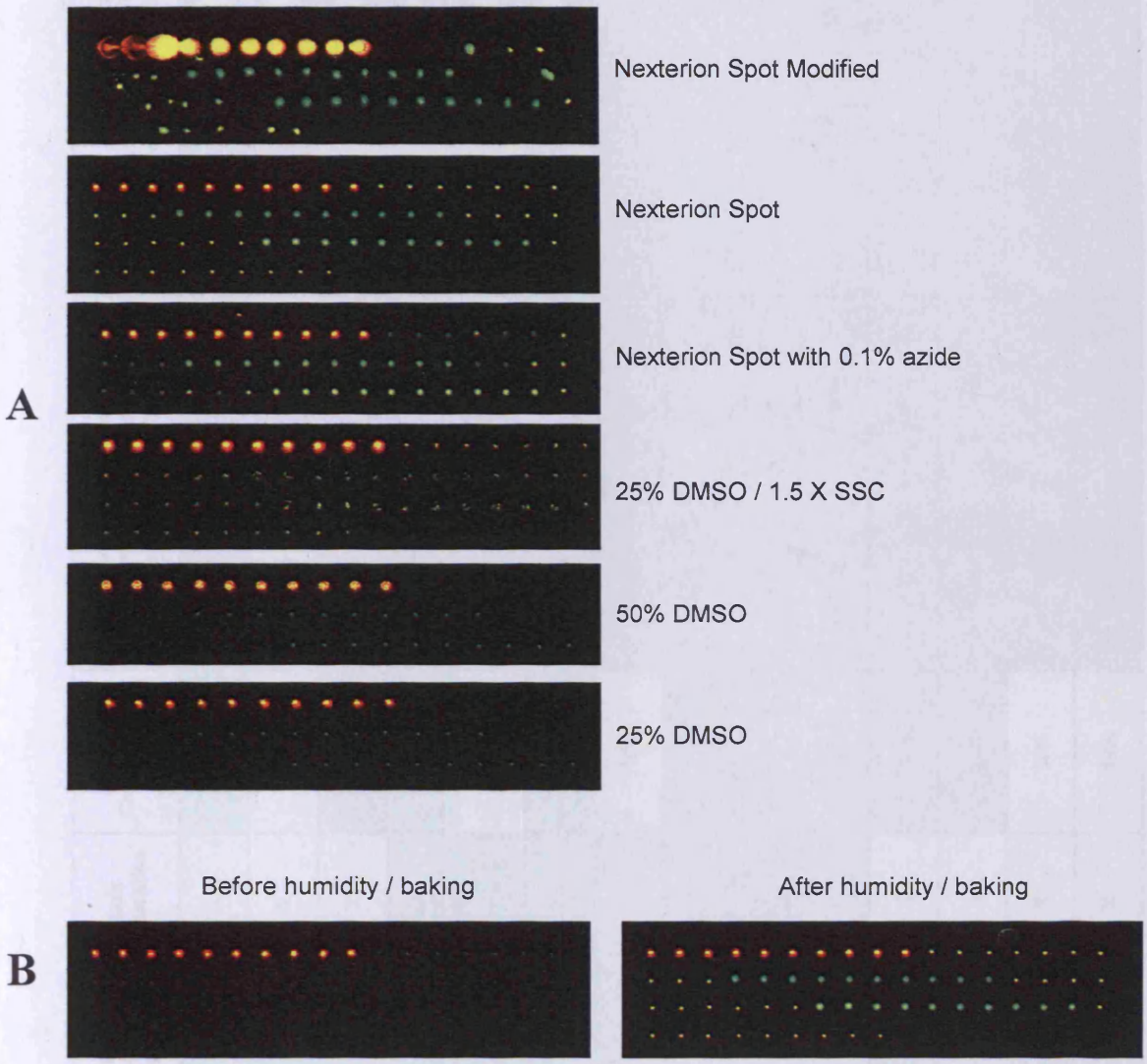


Figure 6.7: Spotting solution autofluorescence
A: Autofluorescence of spotting solution after DNA immobilisation.
B: Autofluorescence of Nexterion Spot before and after DNA immobilisation

6.2.4 Spotting solution comparison

Spotting solutions were compared in order to select the most viable for printing cDNA onto Nexterion Slide E (Table 6.5).

Spotting Solutions		Precipitation at working temperature	Basic evaporation	Contamination	Spot size (before DNA immobilisation)	Spot morphology (before DNA immobilisation)	Spot size (after DNA immobilisation)	Spot morphology (after DNA immobilisation)
NXT	Nexterion Spot Modified	✓	X	81%	180µm		190µm	
NSB	Nexterion Spot	✓	X	50%	70µm		70µm	
NSA	Nexterion Spot +Azide	✓	X	28%	65µm		75µm	
DM50	DMSO 50%	✓	✓	89%	65µm		75µm	
DM25	DMSO 25%	✓	✓	42%	95µm		75µm	
25D15	DMSO 25% SSC 1.5X	✓	✓	72%	95µm		90µm	
NLE	Nexterion Spot Low Evaporation	✓	✓	44%	N/A	N/A	N/A	N/A
50D3	DMSO 50% SSC 3X	X	N/A	N\A	N/A	N/A	N/A	N/A
50D15	DMSO 50% SSC 1.5X	X	N/A	N\A	N/A	N/A	N/A	N/A
25D3	DMSO 25% SSC 3X	X	✓	75%	N/A	N/A	N/A	N/A
SSC	SSC 3X	✓	X	28%	N/A	N/A	N/A	N/A
SDS1	3X SSC with 0.004% SDS	✓	X	28%	N/A	N/A	N/A	N/A
SDS2	3X SSC with 0.001% SDS	✓	X	56%	N/A	N/A	N/A	N/A
SDS3	3X SSC with 0.0005% SDS	✓	X	39%	N/A	N/A	N/A	N/A

Table 6.5: Comparison of spotting solutions

Red highlights the worst solutions tested; green represents the best; yellow indicates those that were relatively successful. Contamination was graded by order of success, with the buffers showing a low absorbance increase awarded the highest points. This ranking was repeated for all 3 contamination tests (1µl, 10µl and no spike) and the sum of the scores converted into percentages compared to the maximum possible score.

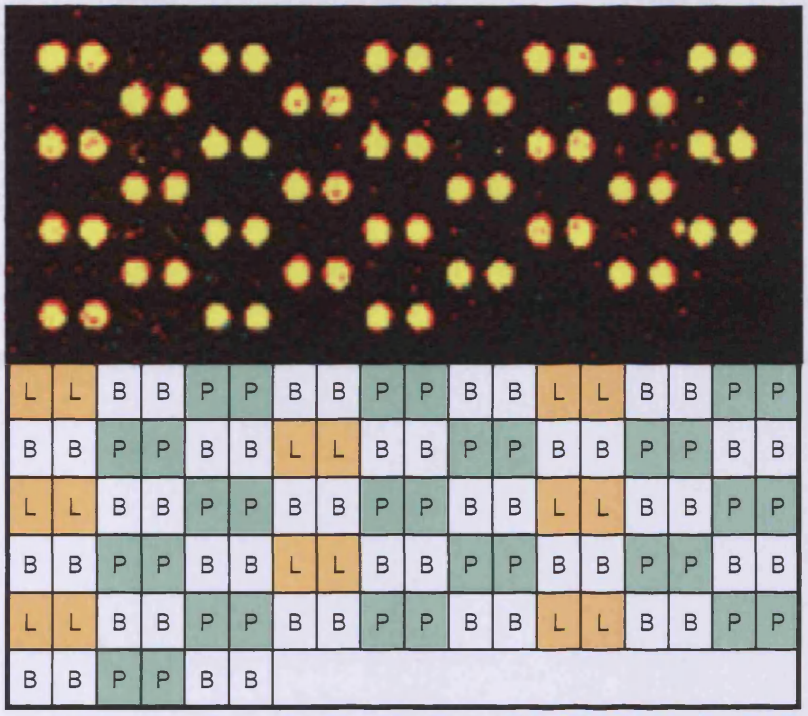
6.2.5 Nexterion Modified Spotting Solution

It was decided that the most suitable spotting buffer, based on spot size and limited proliferation of bacteria after contamination, was Nexterion Spot Modified (NXT) spotting solution. The issue of spots dragging together, caused by overtravel during spot replication, should not arise in a normal print run. A mock print run was performed with pins touching down onto slides only once, as would occur during a standard print run. The print run was performed using the same plate as in the previous buffer test print run (Figure 6.4). Only columns containing the NXT buffer (columns 1, 7, 13 and 19) were used by arranging 4 pins in a single row formation at the top of the print head. Ten artificial plate changes were used to replicate a short print run. This printed in the pattern shown in Figure 6.8, which allowed spots running into each other to be identified. By considering the total number of spots (13,728) to be printed onto the *G. pulex* microarray, optimal spot spacing was calculated to be 235µm. The ideal spotting solution was required to be able to print using this spacing, if a nominal spot size of 100µm was assumed. Printing parameters are shown in Table 6.6. Humidity was kept between 52% and 59% at all times and the temperature fluctuated between 18.0°C and 20.0°C inside the machine. In the room, humidity was maintained consistently between 45% and 48% and the temperature between 16.4°C and 17.6°C.

NXT Test			
Plate changes	10	Substrate dwell time	200
Pins used	4 in a row	sample overtravel	500
Pre-blots	10	Sample dwell time	2000
Pre-blot spacing	450	Substrate thickness	1
Spot replicates	1	Maximum spot size	150
Nominal spot size	100	X-Y	MEDIUM
Spot spacing	235	Wash	3
Location on slide	CENTRE	Times (Reps)	3
Space for barcode	X	Length (Dry)	3
Approach velocity	15	Humidity	50-65
Departure velocity	25	Optimal humidity	60
Substrate overtravel	500		

Table 6.6: Printing parameters for the NXT spotting solution test

After DNA immobilisation (Initial method – Section 6.3.2.1), and blocking (Chapter 2, Section 2.13.10.2.1), hybridisation of labelled PCR products (the same products as printed on the slides) was performed. The PCR products were directly labelled with Cy3-dUTP and Cy5-dUTP as described in Chapter 2, Section 2.13.2.1 using appropriate primers as described in Section 6.2.3. Each of the labelled PCR products (1μl) was combined with M13, LibN, SP6 and T7 primers (in order to block the primer sites of the cDNA making up the spots) and then combined with hybridisation solution and hybridisation performed as described in Chapter 2, Section 2.13.2.4. The slide was incubated at 55°C overnight; a high temperature was used due to the high specificity of the labelled probes. Slides were washed, dried and scanned as described in Chapter 2, Sections 2.13.10.3.1 and 2.13.11. The resulting image can be seen in Figure 6.8.



L	L	B	B	P	P	B	B	P	P	B	B	L	L	B	B	P	P
B	B	P	P	B	B	L	L	B	B	P	P	B	B	P	P	B	B
L	L	B	B	P	P	B	B	P	P	B	B	L	L	B	B	P	P
B	B	P	P	B	B	L	L	B	B	P	P	B	B	P	P	B	B
L	L	B	B	P	P	B	B	P	P	B	B	L	L	B	B	P	P
B	B	P	P	B	B												

Figure 6.8: Hybridisation of labelled PCR products using NXT spotting solution
Top panel shows PCR products hybridised to the same PCR products printed on Nexterion Slide E with Nexterion Spot Modified Spotting Solution. Lower panel describes what was printed onto the slides and their location. L = Landmark; P = PCR product, B = Buffer only.

6.2.5.1 Evaporation Test

In order to assess the effect and rate of evaporation whilst printing with NXT spotting solution, a fresh printing plate was prepared and a 7 hour print run undertaken to mimic the length of a *G. pulex* print run. The plate layout used in this test is shown in Figure 6.9.

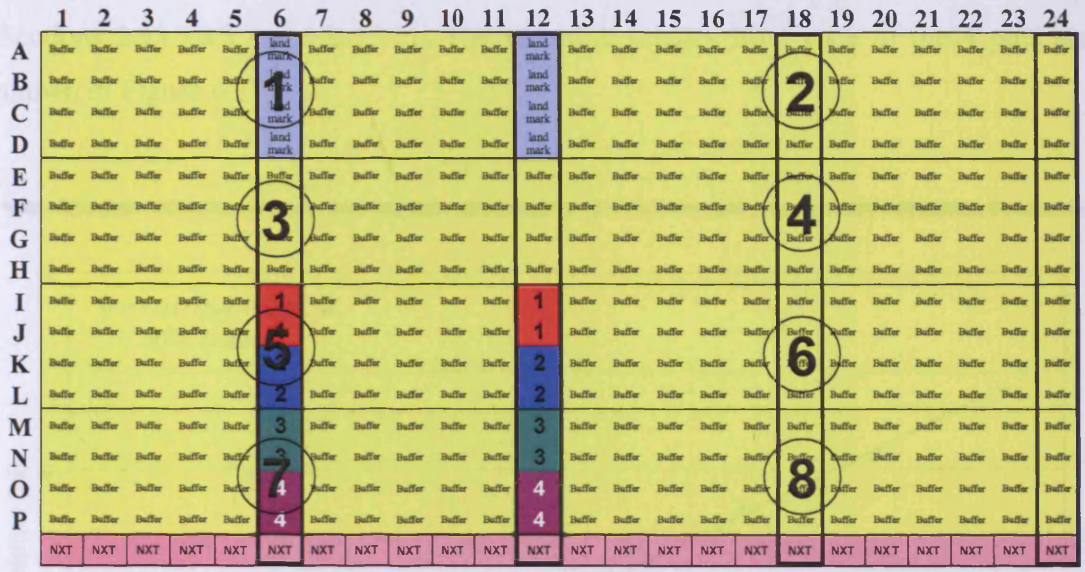


Figure 6.9: Plate layout for the evaporation test using Nexterion Spot Modified
Columns 6, 12, 18 and 24 (with border) were printed onto Nexterion Slide E

NXT buffer was in the same wells as for the buffer test; columns 6, 12, 18 and 24. The remaining wells were filled with buffer to recreate the realistic structure of a print plate and the temperature and insulation effects of the surrounding wells. The plate was printed from each hour and, in between printing, left on the printer bed under normal print run environmental conditions. After each print, the volume of a well not being used by the pins, but containing 1x printing buffer, was recorded to assess the evaporation rate. A graph showing the rate of evaporation over the print run is shown in Figure 6.10. Slides were washed, dried and scanned as described in Chapter 2, Sections 2.13.10.3.1 and 2.13.11. The resulting images from each of the 4 pins are shown in Figure 6.11.

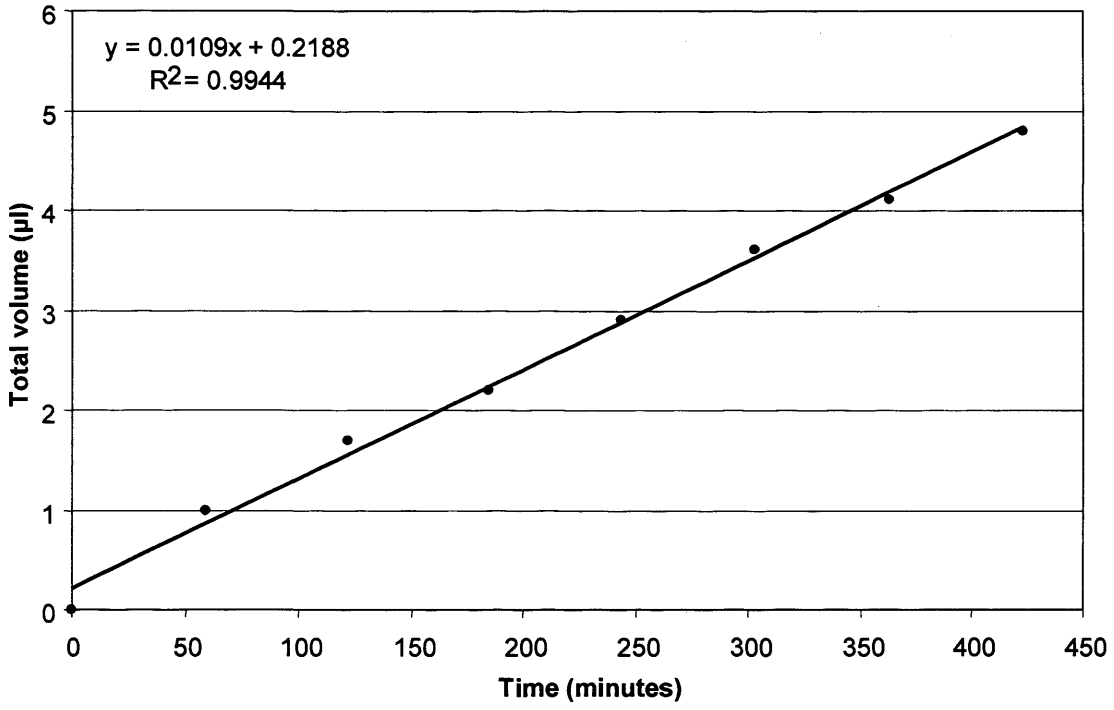


Figure 6.10: Rate of evaporation of NXT spotting solution during a 7 hour print run

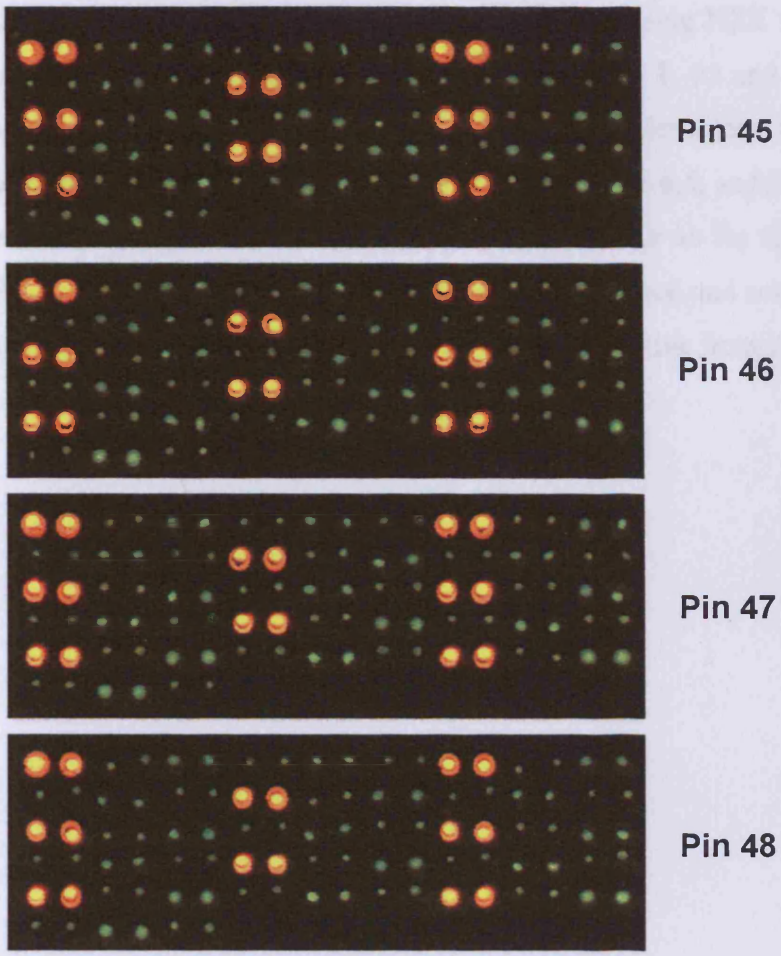
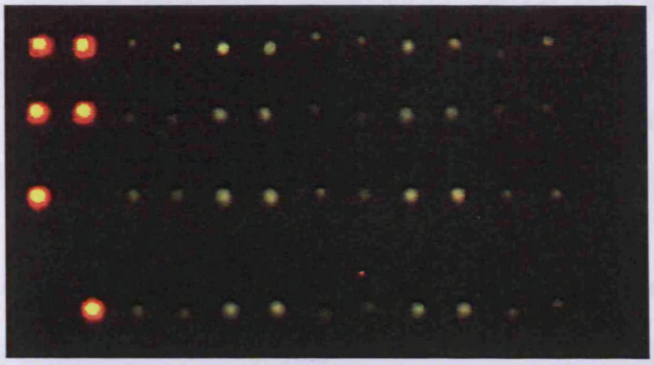


Figure 6.11: Effects of evaporation on landmark printing using NXT spotting solution – After DNA immobilisation

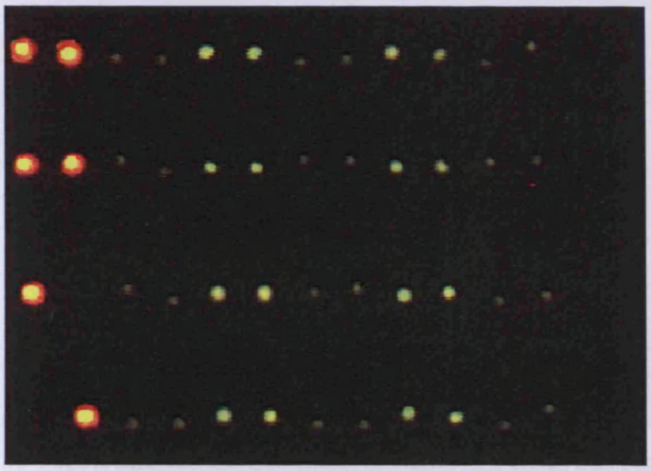
Starting from the top, from left to right, each pair of landmarks were printed an hour after the previous pair. Below the slide images, the printing pattern for each patch is shown. L = landmark; P = PCR product; B = Buffer only. By printing buffer in between, the likelihood of carryover is also tested.

6.2.5.2 Volume test

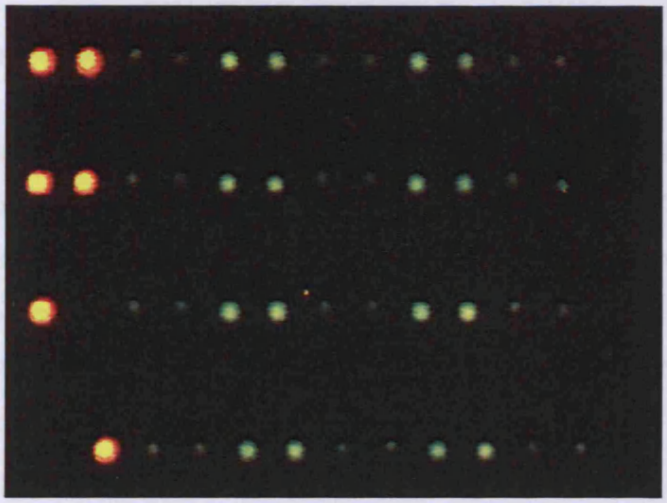
In order to test the ability to print a full 50 slides in one print run using NXT spotting solution, Nexterion Slide E were placed on the printer at position 1, 25 and 50 (for positions, see Chapter 2, Figure 2.6) with normal, uncoated glass microscope slides in all positions in between. The plate layout was as described in Figure 6.4, and four pins in a row were used in the NXT containing wells, in the same way as for the mock NXT print run (Section 6.2.5). Slides 1, 25 and 50 were washed, dried and scanned as described in Chapter 2, Sections 2.13.10.3.1 and 2.13.11. The resulting images can be seen in Figure 6.12.



SLIDE 1



SLIDE 25



SLIDE 50

Figure 6.12: The consistency of printing landmarks on 50 slides using NXT spotting solution

6.3 PRINTING

PCR products in Nexterion Modified Buffer were printed onto Nexterion Slide E using the methods described in Chapter 2, Section 2.13.3. In total, 3 batches of slides were printed on separate day resulting in 153 slides. The environmental parameters and slide numbers are describe in Table 6.7.

BATCH	NO. OF SLIDES	ROOM		MACHINE	
		TEMP °C	HUMIDITY %	TEMP °C	HUMIDITY %
1	50	19.2±1.5	51.5±5.5	21.5±0.5	59.5±2.5
2	50	17.9±1.3	57.0±3.0	20.5±0.5	60.0±1.0
3	53	17.6±1.2	56.5±2.5	20.5±0.5	60.0±2.0
TOTAL	**153**	**18.6±2.2**	**53.0±7.0**	**21.0±1.0**	**59.5±2.5**

Table 6.7: Environmental parameters during microarray printing
During the *G. pulex* microarray printing, the room parameter varied and the machine parameters remained tighter.

After each print run, once the DNA had been immobilised by the humidity and baking step, reflective imaging analysis was performed using a 'RedReflect' scan (excitation/emission: 633nm) of each slide to check that the spots had been successfully printed as this scan picks up residual salt on the slide. An example of a slide scanned with 'RedReflect' is shown in Figure 6.13. The print bed layout for print run one can be seen in Chapter 2, Figure 2.6.

It was noted that for the first 5 slides printed, some of the spots were poor and ran into each other. Therefore, the number of blots was increased. This could not be done by the increasing of the pre-blots number as there are only 4 pre-blot slide spaces on the print bed. Therefore 5 fake slides (blotting slides) were used at the beginning of all subsequent print runs.

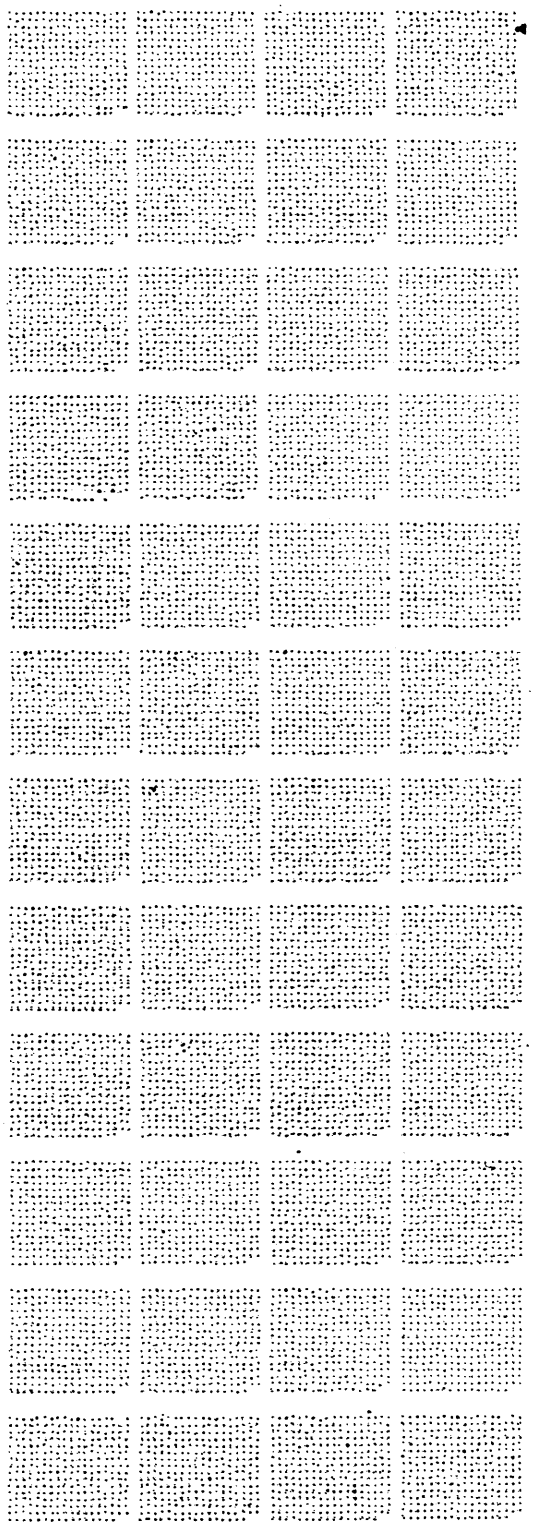


Figure 6.13: Reflection image acquired by 'RedReflect' scanning
PCR products printed using Nexterion Modified Buffer on Nexterion Slide E. Slide was scanned after DNA immobilisation.

6.3.1 Carryover

Before the start of each print run, a carryover test was performed as described in Chapter 2, Section 2.13.3.1. An example of a successful carryover plate (i.e. no blockages or carryover) can be seen in Figure 6.14.

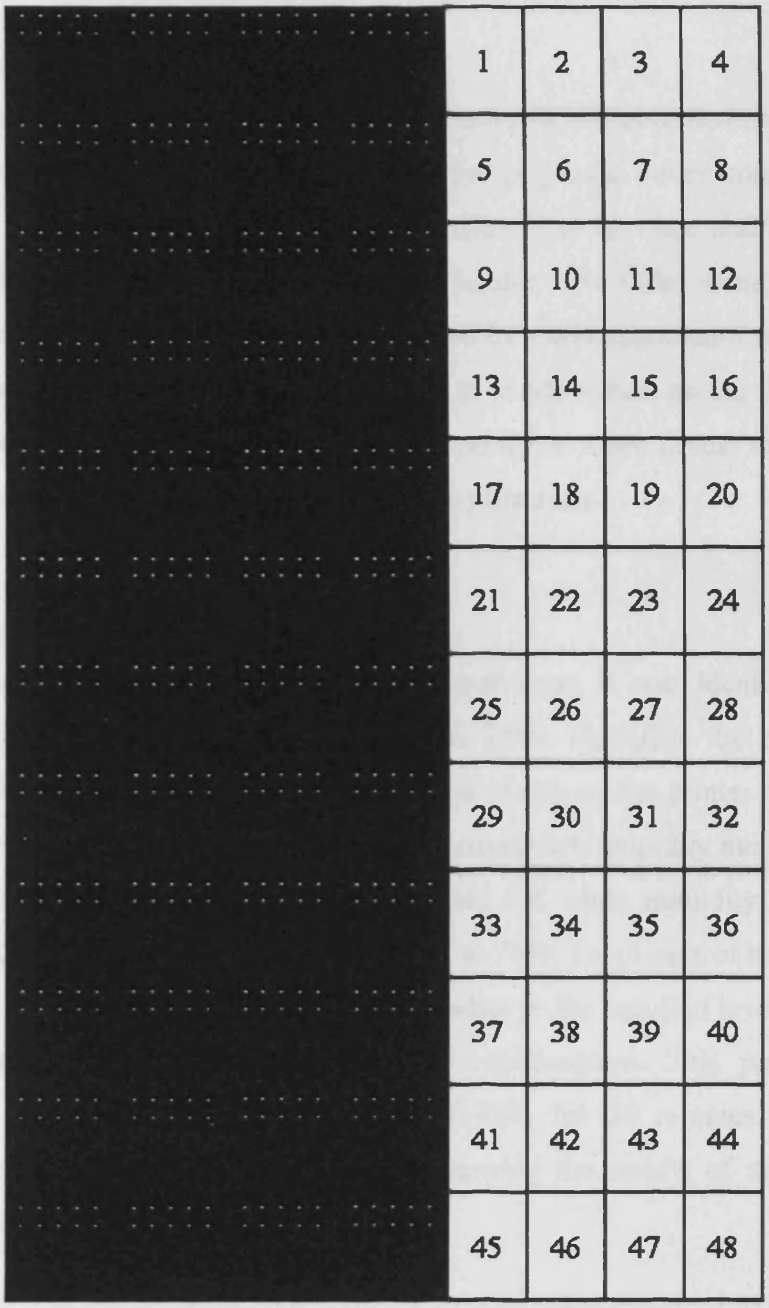


Figure 6.14: Carryover test slide scan
Numbers represent pin number used to print respective 'patches' on the slide

6.3.2 DNA immobilisation

Immobilisation of printed cDNA is required to allow the covalent binding of the amine group of amino acids with the epoxy group on the slide coating. The process of immobilisation is therefore dependent on slide coating and not spotting solution and guidelines are given by the slide manufacturer, in this case, Schott.

6.3.2.1 Initial method

After printing, DNA was immobilised onto the slides by a humidification and baking procedure. Initially, the slides were incubated flat in plastic boxes containing 50°C water for 30 minutes. The slides were then transferred to an oven and baked for 30 minutes at 120°C on their side in a metal slide holder. The slides were stored inside lightproof slide boxes, wrapped in foil and stored in a desiccator until required. This process of immobilisation caused problems due to condensation on the slides causing the spots to run into each other (Figure 6.15 Initial). A more robust and consistent method of humidification was required for future print runs.

6.3.2.2 Final method

After communication with Schott technical department, it was identified that the optimal humidity for DNA immobilisation was 75%. Therefore the most suitable method for humidifying the slides appeared to be *in situ* on the printer bed using the humidification feature of the SpotArray72. The maximum humidity attainable on the Spotarray72 is 70%, so a method recommended for when humidity control fails during a print run, was used to increase the level to 75%. Small pots of hot water were placed at the front of the print bed to boost humidity to the required level, far enough from the slides to prevent the formation of condensation. This procedure was sufficient to maintain a constant humidity of 75% for 30 minutes. The baking procedure remained the same. Patch images showing the results of this change in procedure are shown in Figure 6.15.

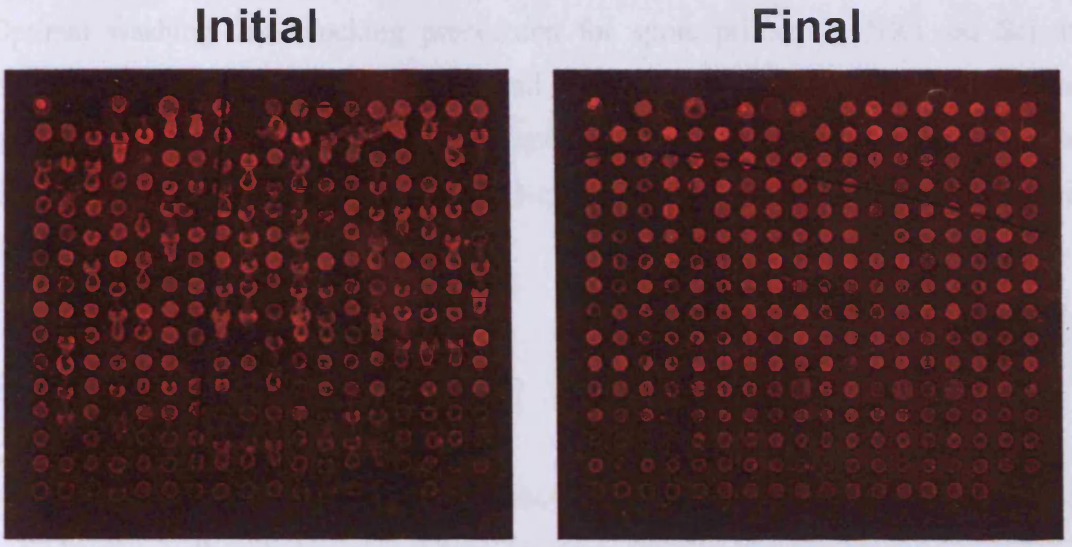


Figure 6.15: Differences in spot morphology using the initial and final method of humidity DNA immobilisation

6.4 SLIDE PREPARATION FOR HYBRIDISATION

Optimal washing and blocking procedures for spots printed in NXT on Schott Nexterion Slide E are described in detail in Chapter 2, Section 2.13.10.2.1. The blocking procedure had been previously optimised by colleagues who compared this approach to one-step BSA and one-step Nexterion Block E; the dual step procedure resulted in the lowest background.

6.5 TARGET PROBE LABELLING

6.5.1 Direct vs. Indirect Labelling

There are two main approaches to cDNA labelling: direct and indirect. During optimisation, both techniques were tested and are described in detail below:

Four samples of male *Echinogammarus marinus* RNA were used to test both direct and indirect labelling. Two samples (GM2, GM3) were purified using mRNA Oligotex columns (Oligotex mRNA Spin-Column Protocol) as per the manufacturer's instructions and the other 2 (GM8, GM1) by lithium chloride precipitation (Chapter 2 Section 2.16.6.1). The Oligotex eluate was reduced in volume to ~20µl per sample (from 40µl) in a SpeedVac and made up to 25.8µl with RNase-free water.

From the 4 samples (2 LiCl, 2 Oligotex purified), 2 (GM3, GM8) were subject to fluorophore labelling directly (SuperScript™ Plus Direct cDNA Labeling System) and the other 2, indirectly (GM2, GM1). Alongside the samples, HeLa RNA, supplied with the direct labelling kit was used as a positive control in both cases. Samples used in these procedures are shown in Table 6.8.

Indirect labelling was performed as previously described, using 10µg of total RNA or ~0.15µg mRNA (based on 1.5% yield from 10µg total RNA after Oligotex purification). Direct labelling was performed as described in the manufacturer's instructions using 10µg of total RNA or ~0.15µg mRNA.

	LABELLING PROTOCOL	
RNA Purification Method	**DIRECT**	**INDIRECT**
Oligotex column	GM3	GM2
Lithium chloride precipitation	GM8	GM1
Control HeLa RNA	Control	Control

Table 6.8: Samples used in Indirect vs. Direct CyDye incorporation into cDNA synthesised from RNA purified using 2 different methods.

All 10 samples were analysed by spectrophotometry, using a spectral scan, recording measurements at 230nm, 260nm, 280nm, 320nm, 550nm and 650nm. The amount of incorporated dye was calculated using the following equation:

$$\text{Incorporated dye (pmols)} = A_{550} \times \frac{\text{Volume (µl)}}{0.15}$$

The frequency of incorporation (FOI) was calculated using the following equation:

$$\text{FOI} = \frac{\text{incorporated dye (pmols)} \times 324.5}{\text{Amount of cDNA (ng)}}$$

The results of the comparison are shown in Table 6.9. The indirect labelled samples were run on a minigel (Chapter 2, Section 2.13.9) shown in Figure 6.16.

PURIFICATION PROCEDURE	SAMPLE	CDNA (NG)	INCORPORATED DYE (PMOLS)	FOI
DIRECT				
Oligotex	GM3	421.8	16.0	12.3
LiCl	GM8	321.9	18.0	18.1
CONTROL		390.7	15.2	12.6
INDIRECT				
Oligotex	GM2	712.6	81.6	37.2
LiCl	GM1	2357.6	298.8	41.1
CONTROL		2488.6	246.0	32.1

Table 6.9: Data recorded following Indirect and Direct CyDye incorporation into cDNA synthesised from RNA purified using 2 different methods.

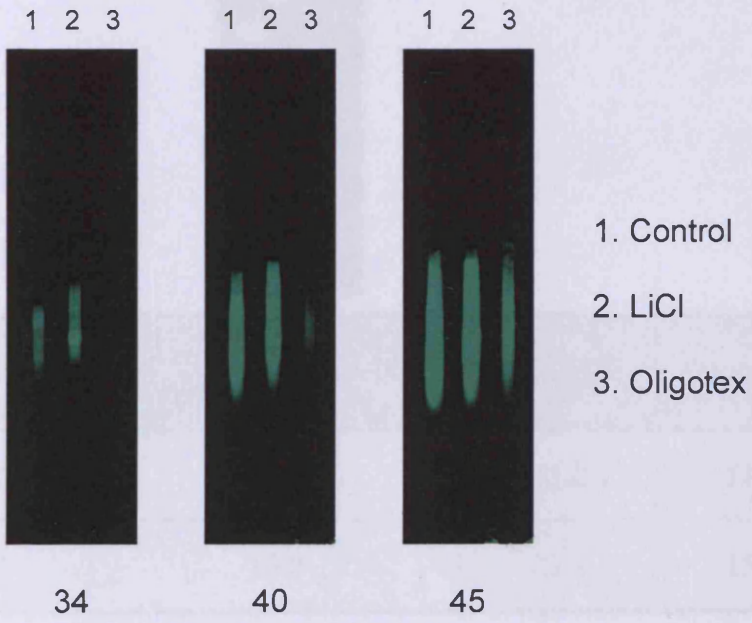


Figure 6.16: Minigel electrophoresis of cDNA containing indirectly incorporated Cy3 synthesised from RNA purified using 2 different methods.
Scans were performed at 3 different gain settings, 35, 40 and 45.

6.5.2 Reverse transcription

6.5.2.1 Superscript III vs. MMLV

Two reverse transcriptase enzymes, Superscript™ III and MMLV were compared side-by-side. Indirect labelling was performed on a sample of *G. pulex* RNA that had been divided into two equal aliquots. Reverse transcription (RT) was performed using with MMLV RT or Superscript™ III RT with appropriate buffers. After labelling the probe was analysed quantitatively in a spectrophotometer and qualitatively on a minigel (Figure 6.17).

Experiment	Incorporated Dye (pmols)	FOI	cDNA (ng)
Superscript III @ 42°C	14.0	38.2	118.8
MMLV @ 42°C	10.8	22.4	156.4

Figure 6.17: Superscript III vs. MMLV Reverse transcriptase
A: Superscript™ III; B: MMLV

6.5.2.2 Random Hexamers and OligodT

A comparison of labelling efficiency was carried out using either random hexamers and oligodT as primers or oligodT only. Both indirect labelling reactions were performed using the same *G. pulex* RNA sample divided into 2 equal aliquots; RT was undertaken using MMLV reverse transcriptase. The labelled cDNA was analysed by spectrophotometry and minigel electrophoresis (Figure 6.18).

Experiment	Incorporated Dye (pmols)	FOI	cDNA (ng)
Random Hexamers + anchored oligodT	51.2	32.5	510.8
Anchored oligodT only	17.6	32.5	176.2

Figure 6.18: Random Hexamers + OligodT vs. OligodT only
A: Random Hexamers + OligodT; B: OligodT only

6.5.2.3 *dNTP Ratio & Molarity Comparison*

By comparing other research group's protocols, dNTP ratio and molarities / volumes were selected (Table 6.10).

	Ratio			Volumes		Molarity		
Protocol	**ACG**	**T**	**aa-U**	**Volume Used uL**	**Total Volume uL**	**ACG (M)**	**T (M)**	**aa-U (M)**
1	5	3	2	0.6	30	0.33	75	65
2	5	3	2	0.6	30	0.17	60	65
3	5	3	2	1.2	30	-	-	65
4	5	1	4	0.6	30	0.33	75	65
5	5	3	1	0.6	30	0.50	-	65
6	-	-	-	1.5	30	0.20	75	70
7	5	3	2	0.6	30	0.33	75	65
8	5	4	4	1	20	0.67	30	37
9	5	3	2	0.75	30	0.45	61	65
10	5	1	1	0.6 ACG 1.5 T 1.5 aa-U	30	0.25	40	37
Current	5	2	3	2	20	0.67	32.5	37

Table 6.10: dNTP Ratio and molarities

1: Aminoallyl labelling of RNA for microarrays, SOP, St. Jude Children's Research Hospital, a modification from The Institute for Genomic Research.
2: Amino allyl labelling of RNA for Microarrays, Jeremy Hasseman, The Institute for Genomics Research SOP.
3: High Throughput Aminoallyl labelling of RNA for microarrays, Molly G. Friedman, The Institute for Genomics Research SOP.
4: Amino-allyl reverse transcription labelling protocol for oligo arrays, Version 1.0, Fred Hutchinson Cancer Research Centre.
5: Amino-allyl Dye Coupling Protocol, Joseph DeRisi, University of California, San Francisco.
6: Amino Allyl Labeling and Hybridization Protocol, Andrews, J. et al, Indiana University, February 28, 2005.
7: RNA Reverse Transcriptase Labeling with Aminoallyl-dUTP, Marine Genomics Laboratory, University of Miami.
8. Protocol for Amino-allyl Reverse Transcription and NHS-Cy dye labelling, 18 March 2000, Department of Microbiology, University of Washington School of Medicine.
9: Fluorescent Probe Preparation, Chris Seidel, University of California, Berkeley.
10: Amino allyl labelling of RNA for Microarrays, Oregon Health & Science University.
Current: Cardiff University Original Labelling Protocol

6.5.2.4 dUTP:dTTP ratio

Five different dUTP:dTTP ratios (1:4, 2:3, 1:1, 3:2, 4:1) were compared in three different species using the indirect labelling procedure (Chapter 2, Section 2.13.8). The species were *E. marinus*, *G. pulex* and the earthworm, *Lumbricus rubellus*. The ratios 0:5 and 5:0 were also used as positive and negative controls. The quality of the resulting probes were analysed by minigel electrophoresis (Figure 6.19).

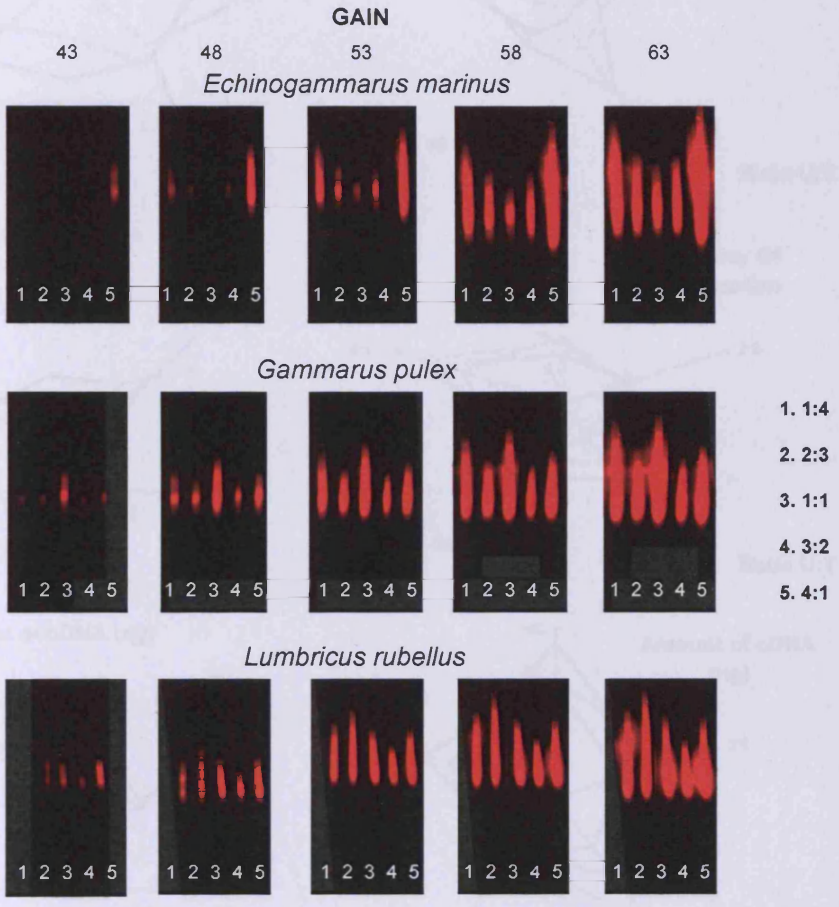


Figure 6.19: Minigel electrophoresis of labelled cDNA from 3 different invertebrates using 5 different ratios of dNTPs at 5 different scan gains.

The quantitative parameters were compared for each species using line and radar charts (Figure 6.20). The parameters measured were incorporated dye, frequency of incorporation and quantity of cDNA.

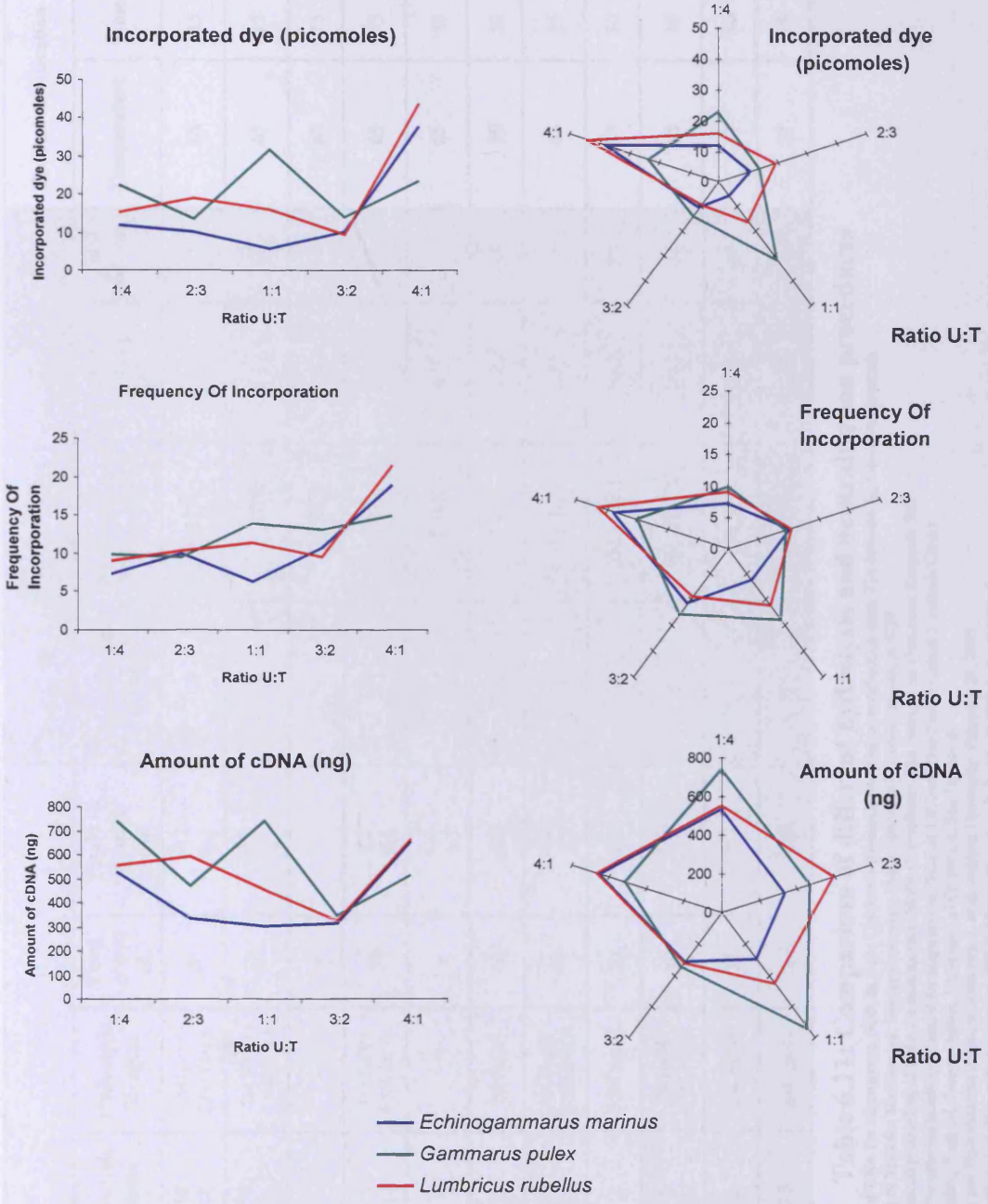


Figure 6.20: Comparing U:T ratio effects on aa-cDNA synthesis using RNA from 3 different invertebrates

6.5.3 RNA Hydrolysis and Neutralisation

By comparing other research group's protocols, the RNA hydrolysis and neutralisation methods were selected (Table 6.11).

	Hydrolysis				Neutralisation				Enzyme Inactivation	
Protocol	Hydrolysis Volume	Hydrolysis Reagent	Total volume uL	Final HR Molarity (M)	Neutralisation Volume uL	Neutralisation Solution	Final NS Molarity (M)	Total volume uL	Temperature	Time
1	10 10	1M NaOH 0.5M EDTA	50	0.2 0.1	25	1M HEPES	0.33	75	65	15
2	10 10	1M NaOH 0.5M EDTA	50	0.2 0.1	10	1M HCl	0.17	60	65	15
3	-	-	-	-	-	1M HCl	-	-	65	15
4	10 10	1M NaOH 0.5M EDTA	50	0.2 0.1	25	1M Tris	0.33	75	65	15
5	-	-	-	0.1 0.1	-	HEPES	0.50	-	65	10
6	15	1M NaOH	60	0.25	15	1M HCl	0.20	75	70	10
7	10 10	1M NaOH 0.5M EDTA	50	0.2 0.1	25	1M Tris	0.33	75	65	15
8	2	2.5M NaOH	20	0.25	10	2M MOPS	0.67	30	37	10
9	5 1	1M NaOH 0.5M EDTA	36	0.14 0.14	25	1M HEPES	0.45	61	65	10
10	5	2M NaOH	35	0.3	5	2M Tris	0.25	40	37	10
Current	2.5	2M NaOH	22.5	0.28	10	2M HEPES	0.67	32.5	37	15

Table 6.11: Comparison of different hydrolysis and neutralisation procedures

1: Aminoallyl labelling of RNA for microarrays, SOP, St. Jude Children's Research Hospital, a modification from The Institute for Genomic Research.
2: Amino allyl Labelling of RNA for Microarrays, Jeremy Hasseman, The Institute for Genomics Research SOP.
3: High Throughput Aminoallyl labelling of RNA for microarrays, Molly G. Friedman, The Institute for Genomics Research SOP.
4: Amino-allyl reverse transcription labelling protocol for oligo arrays, Version 1.0, Fred Hutchinson Cancer Research Centre.
5: Amino-allyl Dye Coupling Protocol, Joseph DeRisi, University of California, San Francisco.
6: Amino Allyl Labeling and Hybridization Protocol, Andrews, J. et al, Indiana University, February 28, 2005.
7: RNA Reverse Transcriptase Labeling with Aminoallyl-dUTP, Marine Genomics Laboratory, University of Miami.
8. Protocol for Amino-allyl Reverse Transcription and NHS-Cy dye labelling, 18 March 2000, Department of Microbiology, University of Washington School of Medicine.
9: Fluorescent Probe Preparation, Chris Seidel, University of California, Berkeley.
10: Amino allyl labelling of RNA for Microarrays, Oregon Health & Science University.
Current: Cardiff University Original Labelling Protocol

6.5.4 cDNA purification and CyDye coupling

Two alternative purification methods were tested, ethanol precipitation and modified QIAGEN QIAquick PCR purification (Hegde *et al.* 2000). Different samples of *G. pulex* RNA were indirectly labelled, cleaning the cDNA either by ethanol precipitation or using a QIAGEN QIAquick PCR purification kit (data not shown). CyDye was resuspended in either RNase-free water or DMSO before being combined with carbonate buffer (data not shown).

6.5.5 Probe purification and analysis

Two alternative purification methods were tested: GFX column purification and standard QIAGEN QIAquick PCR purification. In general, QIAquick columns resulted in an increased compared with GFX columns (data not shown). After purification, labelled cDNA was assessed for quantity, purity and incorporation efficiency, using stringent specifications to assess its suitability for hybridisation as described in Chapter 2, Section 2.13.9.

6.6 HYBRIDISATION

6.6.1 Oligo test

An oligonucleotide mix specific to the library vectors was used labelled with Cy5. The location of the oligonucleotides within the vector is shown in Figure 6.21. Different concentrations of oligo were hybridised to a *G. pulex* microarray to ascertain the most appropriate quantity of oligonucleotide mix to use for microarray hybridisation. The scan images are shown in Figure 6.22.

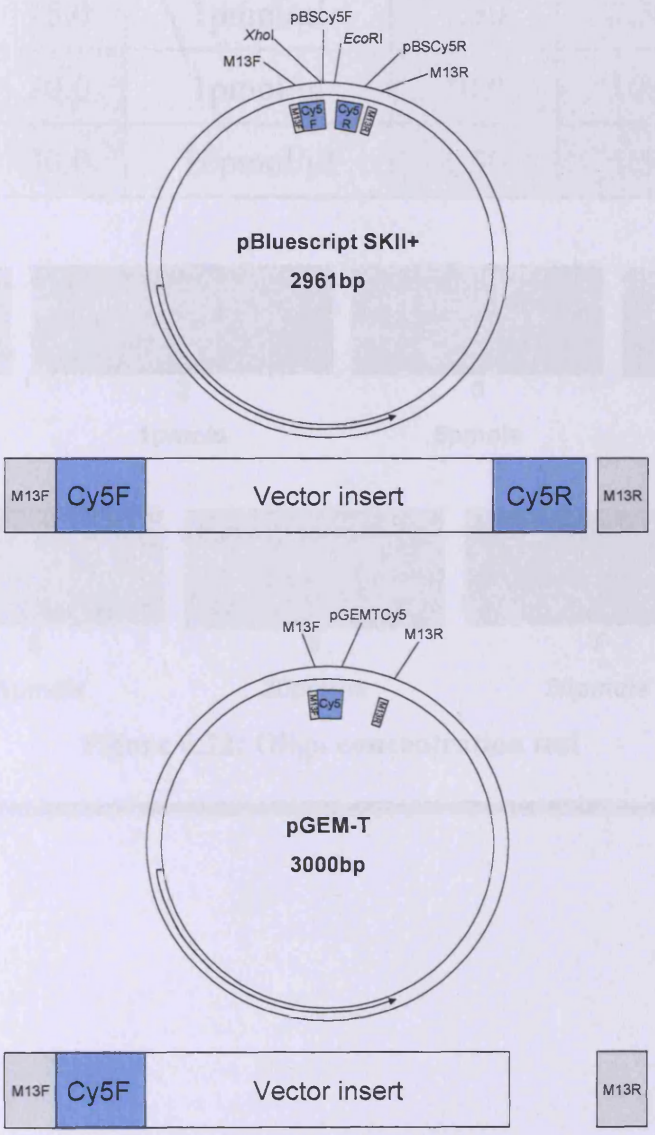


Figure 6.21: pBluescript SKII+ and pGEM-T Cy5 labelled oligonucleotides

Experiment Number	Total pmols	Stock oligo concentration	Volume pBS (µl)	Volume pGEM (µl)	Volume Water (µl)
1	0.5	1pmol/µl	0.25	0.25	19.5
2	1.0	1pmol/µl	0.50	0.50	19.0
3	5.0	1pmol/µl	2.50	2.50	15.0
4	10.0	1pmol/µl	5.00	5.00	10.0
5	15.0	1pmol/µl	7.50	7.50	5.00
6	20.0	1pmol/µl	10.0	10.0	0.00
7	30.0	10pmol/µl	1.50	1.50	17.0

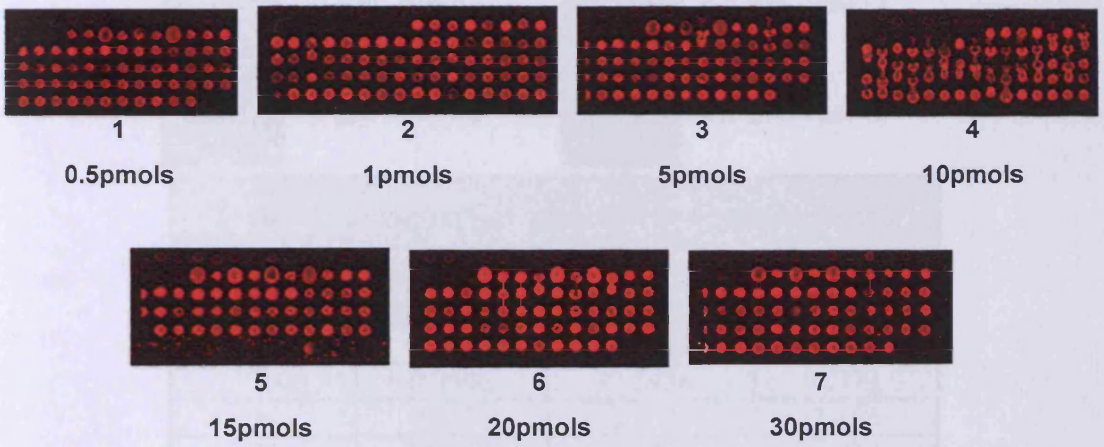
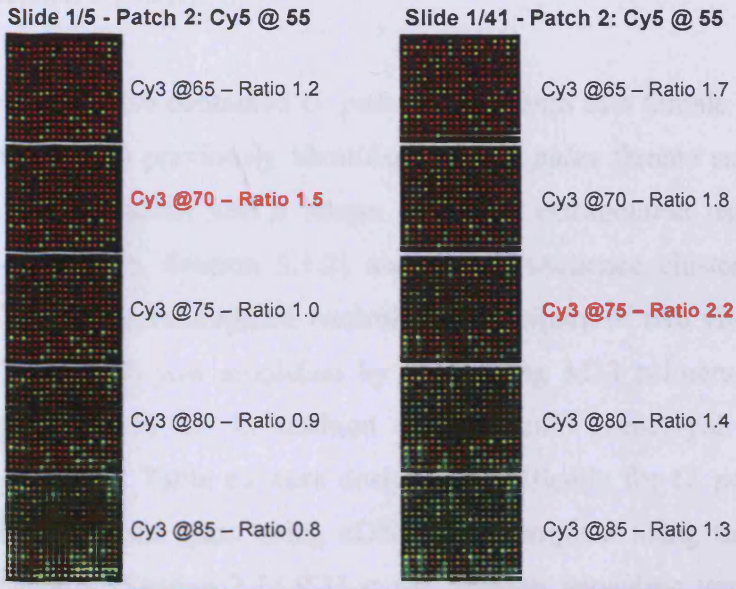


Figure 6.22: Oligo concentration test

6.7 SCANNING

Slides were scanned using a PerkinElmer ScanArray® Express HT microarray scanner. Resulting images were stored as .tiff files for subsequent image analysis. Using the aforementioned labelling and hybridisation procedures, a test slide was scanned at increasing laser intensities from 65% to 80% (Figure 6.23).



Signal : noise ratios at different gain settings			
Slide 1/41 Cy3 Gain	*Signal* **I / µm**	*Background* **I / µm**	**Ratio**
65	0.2028	0.1211	1.675
70	0.2546	0.1438	1.771
75	**0.3331**	**0.1524**	**2.186**
80	0.4351	0.3076	1.414
85	0.5565	0.4417	1.260
Slide 1/5 CY3 Gain	*Signal* **I / µm**	*Background* **I / µm**	**Ratio**
65	0.1755	0.1471	1.193
70	**0.2103**	**0.1384**	**1.520**
75	0.2746	0.2878	0.954
80	0.3601	0.4206	0.856
85	0.4645	0.5986	0.776

Figure 6.23: Background to Signal ratio at different laser intensities

PMT (PhotoMultipler Tube) was fixed at 90 and laser intensity varied. The signal:background ratio was determined using image information provided by the ScanArray software. The same area of slide adjacent to patch 2 was analysed for brightness, as was the patch itself. Results highlighted in red indicate the parameters with the optimal signal:noise ratio.

6.8 LANDMARKS AND PLATE LAYOUTS

Landmarks were made from CyDye (GE Healthcare UK Ltd., Bucks., UK) labelled mouse and zebrafish PCR products using a direct labelling method. A zebrafish metallothioneins gene was amplified and directly labelled using cDNA as a template. Mouse genes were amplified and directly labelled using a PCR product purified by agarose gel extraction, as a template. The labelling procedure is described in detail Chapter 2, Section 2.13.2.1.

The scorecard plate also contained *G. pulex* vitellogenin as a female positive control. Vitellogenin had been previously identified in a *G. pulex* female subtractive library (feS; Dr. Judith Richards) and a longer sequence extrapolated using both RACE technology (Chapter 5, Section 5.1.2) and library sequence clustering (Chapter 5, Section 5.1.2). For the vitellogenin controls, the miniprep of two vitellogenin RACE fragments (13 and 17) was amplified by PCR using M13 primers as described in Chapter 2, Section 2.11.9.3. In addition a vitellogenin primer pair (Gp_Vg_F and Gp_Vg_R; Appendix, Table 6) were designed specifically for *G. pulex* vitellogenin and used to amplify the gene using cDNA as a template using the standard PCR protocol (Chapter 2, Section 2.11.9.3) at the relevant annealing temperature for the Gp_Vg primer pair (Appendix, Table 6). The fragment amplified from the cDNA was 206bp in size, the RACE fragments were 1119bp. The PCR products were cleaned up using single sample Montage PCR cleanup tubes as per manufacturer's instructions. The PCR and the RACE17 products were included on the landmark / scorecard plate (Chapter 2, Figure 2.6) for printing, by dilution with NXT buffer 1:1 into wells B21 and B22 respectively.

Each time the landmark / scorecard plate (Chapter 2, Figure 2.6) was printed on, the printer was paused and the plate removed, the lid replaced and the plate stored at 4°C until required at the next plate change. This prevented evaporation and decreased the risk of contamination. A blank plate was put in its place so that the printer did not stall when the print run was resumed after pausing.

6.9 DISCUSSION

Successful optimisation of microarray slide printing and sample probe hybridisation was essential to obtain accurate and repeatable data from generated images. Additionally, a robust method was created so that there is sample processing consistency and as little variation between printed slides as possible.

6.9.1 Spotting solution test print run

Proprietary Nexterion spotting solutions are designed to be printed specifically onto Nexterion Slide E slide surface and were chosen for that reason. The Nexterion Spot LE (NLE) was not chosen for testing as it would mean drying down all the PCR products and resuspending them in 1x buffer. This introduced an additional step to the process which would increase the risk of contamination. One of the key concerns in using standard NSB is that it is a phosphate based solution which is prone to bacterial contamination. For this reason this buffer was also tested with added sodium azide to reduce this risk. NXT provided an alternative spotting solution with added detergent; this is an advantage on these highly hydrophobic slides as it works by decreasing the surface tension of the spots and increasing the spot size. Also, the addition of the detergent Triton® X-100 to the phosphate based solution decreased the risk of contamination. The main disadvantage of proprietary spotting solutions is the cost and lack of control over buffer component concentrations. No knowledge of the exact chemical composition is disadvantageous if any trouble-shooting is required at a later stage. Therefore three 'homemade' spotting solutions were also tested, 25D15, DM50 and DM25. Some alterative buffers had also been ruled out including 50D3 (precipitated at RT), 50D15 (precipitated at RT), 25D3 (precipitated slightly when temperature dropped below 20°C), SSC, SDS1, SDS2 and SDS3. The SSC and SDS solutions were disregarded due to previous optimisation attempts by colleagues had yielded substandard results. No spotting solution was tested that contains Tris as this competes with PCR products for attachment sites on Nexterion Slide E. These subsequent trials were done with the hope that a much more consistent and superior spot size and morphology could be achieved with different buffers.

6.9.2 Spotting solution autofluorescence

Many of the buffers, in particular the salt containing buffers, show autofluorescence after the DNA immobilisation step. This problem should be resolved when the washing / blocking step removes the salt deposits from the slides. Autofluorescence can also occur from the slide chemistry itself, although Nexterion Slide E has a very low autofluorescence. As the fluorescence was more obvious after baking, it was likely to be autofluorescence of the salt contained in the spotting solution. This is intensified when dried but was easily removed by washing slides as described by the slide manufacturer.

6.9.3 Nexterion Modified Spotting Solution

The 4 PCR products for print and hybridisation testing had been identified from the first sequences from the *G. pulex* mixed library sequenced at Cardiff University. When the mixed *G. pulex* library was originally made, ~150 clones were sequenced to check the library; clones containing a prefix of the three letters mxA only (as opposed to mxAA or mxAB) are from these original clone stocks. For haemocyanin and the transporter gene insert, a combination of primers SP6 / T7 were used to get the best amplicon, for HR3, the M13 primer pair was used to amplify the product and for ferritin, the LibN primer pair was used, the details of these primers can be seen in the Appendix A, Table 6. After initial tests, it was decided that the most suitable spotting buffer, based on spot size and limited proliferation of bacteria after contamination, was Nexterion Spot Modified spotting solution, despite the spot distortion. The unwanted effect of the spots running into each other was thought to potentially be due to the way the SpotArray72 duplicates spots. The pin overtravel height during replication was low enough to drag each spot into each other. Although this is undesirable, during a normal *G. pulex* microarray print run, replication of adjacent spots will not be required and individual pin contact with each slide is limited to once per sample. Therefore the issue of spots dragging together, caused by overtravel during spot replication, should not normally arise. In order to test this hypothesis, a mock run was performed with each pin touching down onto the slide only once, as would occur during a standard print run. In a normal print run, one spot is printed and the print head fully lifted in between the next spot. By reproducing this, it may prevent the spot drag problem and allow the use of this buffer to be confirmed. It was

decided to use 4 pins in a row and 10 fake plate changes keeping the same plate in position to replicate a short print run. The mock run was successful removing the dragging effect caused by pin overtravel. NXT spotting solution was therefore selected and ideal for printing on Nexterion Slide E as it gave the best spot size, brightness and morphology.

6.9.4 Evaporation

Although the basic evaporation test was a crude experiment, it was important to take into consideration evaporation due to the problems this causes during and between print runs. In order to assess the effect and rate of evaporation whilst printing with NXT spotting solution, a 7 hour print run was undertaken to mimic the length of a *G. pulex* print run. The maximum time the sample plates would be on the print bed is 1 hour and the landmark/scorecard plate is potentially 5 hours as this plate is printed 5 times, however it is advisable to remove this plate from the deck and replace it with a "fake plate" so that is only open and able to evaporate whilst it is being used for printing, for approximately 15 minutes of each plate run, totalling only 1 hour 15 minutes. Therefore a run of 7 hours was sufficient to fully test the effect of evaporation on the *G. pulex* print run. The main reason for testing the evaporation rate is to ensure that there is a consistent concentration of cDNA throughout multiple print runs. Once the evaporation rate was determined, it was possible to calculate how much water is required to add to the plates at the end of each print run to account for evaporation, Also, keeping a constant concentration of both cDNA and spotting solution prevented any crystals forming in the solution that may have occurred as a result of evaporation. The presence of crystals in the spotting solution may have obstructed the SMP4 pins preventing successful printing.

6.9.5 Volume

The volume test showed that the pins are taking up a sufficient volume to print over a full 50 slide run. The intensity of the printed landmark had not been significantly reduced (by eye) and there were no obvious differences between the 3 slides. The spot brightness, size and morphology had also not been altered significantly over the 50 slide print run. Although uncoated slides may not truly represent the action of Nexterion Slide E, it is potentially a more rigorous test, as a higher volume will be

printed onto the blank slides due to the lack of a hydrophobic coating. This procedure therefore tested that the volume taken up by the pins is sufficient to print 50 slides. The pins used by the SpotArray72 are SMP4 split pins (TeleChem International, Inc., CA, U.S.A.) and use a capillary action to uptake the liquid, this means the pins take up a volume of 0.25µl and after blotting 10 times, then contact prints 50 slides, theoretically releasing 0.7nl each time it comes into contact with the slide. Therefore in theory, there is a sufficient volume taken up by the pins to print all 50 of the slides (the manufacturer, TeleChem, claim 200 prints are possible from one liquid uptake), however, each spotting buffer has different chemical properties, viscosity, surface tension and ionic charge; these properties can affect the ability of the pins to uptake the spotting solution and successfully deposit it onto the slides. It is therefore important to test each potential spotting buffer as thoroughly as possible before using it to dilute library PCR products.

6.9.6 Carryover

Carryover tests were performed before each print run to check there was no potential sample carryover evident. If the second batch of spots fluoresced, then some of the landmark DNA had not been successfully washed off the pins and had been carried over into the buffer wells. Also, this procedure tested whether or not any of the pins were blocked, which was indicated by missing spots for a certain patch.

6.9.7 DNA immobilisation and Slide surface blocking

The change in procedure for DNA immobilisation using the print bed to produce a humidity controlled atmosphere produced distinct spots and prevented the spots from running into each other. The slide surface was blocked after immobilisation and before hybridisation to reduce the risk of non-specific binding of labelled cDNA to the slide chemistry. In order to prevent cross reactivity of the epoxy groups on the slide coating with the nucleophilic groups of the labelled cDNA, either chemical or protein blocking can be performed. Otherwise, the labelled cDNA will react irreversibly with the epoxy groups to form a covalent bond. In this study both BSA and Nexterion Block E were used to eliminate non-specific binding and decrease background fluorescent levels.

6.9.8 RNA purification

Lithium chloride is used to selectively precipitate out RNA from other contaminants in samples (Meier 1988), it is especially good for removing carbohydrate, DNA and protein and is therefore much more selective than alternative precipitation purification methods. It is a key method in removing inhibitors from RNA samples, but cannot be used in samples with a low concentration of RNA (Naderi *et al.* 2004). LiCl precipitation was followed by purification on RNeasy columns in order to eliminate any lithium molecules which can inhibit subsequent enzymatic applications such as RT and PCR. Oligotex columns are used to isolate poly A^+ mRNA from total RNA using a spin-column. The column eluate produces mRNA free from any contamination, ideal to produce the high level of purity required for CyDye incorporation into cDNA. Comparison of these purification methods revealed both to be successful yielding suitable probes, however LiCl produced the best results and this procedure was used on the more difficult *Echinogammarus marinus* samples described in Chapter 9. DNase I was used in samples showing genomic DNA contamination, this approach was used for all samples within that experiment.

6.9.9 Target probe labelling

In order to visualise the DNA hybridised to the printed spots of cDNA on the microarray slide, fluorescent dye incorporation ('labelling') was required. 'Labelling' is performed at the reverse transcription stage by incorporating modified bases, usually dUTP. There are two main approaches to cDNA labelling: direct and indirect. Direct labelling uses modified dUTP with a fluorescent molecule attached such as a CyDye. Indirect labelling uses dUTP modified with an aminoallyl group; the resulting cDNA is then incubated with a fluorescent probe, such as Cy3 or Cy5 in their NHS ester form. There are advantages and disadvantages to both approaches.

Direct labelling using CyDye modified dUTP (CyDye-dUTP) is a method commonly used to attach fluorophores to DNA strands. CyDye-dUTP can be incorporated into DNA during reverse transcription or PCR, dependent on the application. With respect to RT-based incorporation, the key advantage of direct labelling is the single-step required to achieve labelled DNA followed by purification. In contrast, indirect labelling requires the purification of the cDNA following RT and a further

purification step after labelling. Multiple purifications potentially reduce the final quantity of labelled cDNA. The main disadvantage of direct CyDye incorporation is the relatively low incorporation efficiency by steric hindrance, due to the large CyDye molecule. Successful labelling is therefore dependent on the activity and efficiency of the reverse transcriptase enzyme. Although RT enzymes have been developed with an increased ability to incorporate bulky molecules, they do not completely eliminate the problem. Also the method of direct labelling has a strong bias towards Cy3 incorporation over Cy5 (Cy5 is a larger molecule).

Indirect labelling firstly incorporates an aminoallyl modified dUTP (aa-dUTP), followed by chemical coupling with a reactive CyDye containing an NHS ester. This eliminates the dependence on enzymatic incorporation and therefore is a more efficient method of labelling. It overcomes the problem of incorporating large molecules like CyDye-dUTPs by using smaller aa-dUTP molecules. Both direct and indirect labelling requires the use of a reverse transcriptase enzyme that is able to incorporate modified dUTPs into the synthesised DNA. Although standard RT enzymes such as MMLV (Moloney Murine Leukemia Virus) RT and AMV (Avian Myeloblastosis Virus) are capable of utilising modified bases, they do so less efficiently than a more suitable engineered RT enzyme such as the Superscript™ enzymes (Invitrogen). MMLV is usually preferred over AMV as it does not have any DNA endonuclease activity and is lower in RNaseH activity.

Throughout the microarray community, Superscript™ II, and more recently Superscript™ III have been the enzymes of choice due to their ability to produce probes with a high frequency of modified base incorporation (Hegde *et al.* 2000). They are similar to MMLV in that that are RNaseH- but another significant advantage in using Superscript™ enzymes is their long half life in comparison to MMLV. The most recently adapted Superscript™ III RT claims a half life of 220 minutes (which the manufacturer states is 35X more than other RNaseH- RTs) and is thermostable up to temperature of 50°C. The main disadvantage of the Superscript™ enzymes is the cost effectiveness of their use, at over 3 times the cost of MMLV, the increased efficiency of Superscript™ III over MMLV would have to significantly increase the activity of the probes produced.

After the production of aa-cDNA, a purification step is necessary to remove RT enzyme, salts, RNA fragments and any free aa-dUTP and other nucleotides. Amines are significant inhibitors of CyDye coupling (they compete with the aminoallyl groups to bind CyDye), yet are found in commonly used buffers including Tris and HEPES. Unless removed, these contaminants will effect downstream applications including CyDye coupling and hybridisation. A final purification step is required to remove any free dye that has not been coupled to the aminoallyl groups of aa-cDNA. Free dye causes high background fluorescence on the resulting microarrays, decreasing the signal:background ratio, making it difficult to extract any data. After purification, labelled cDNA was assessed for quantity, purity and incorporation efficiency, using stringent specifications to assess its suitability for hybridisation. Poor labels produce inferior microarray images, preventing the extraction of any useful data. Using QIAquick columns to purify the labelled cDNA was found to be the more consistent method. In general, resuspension in DMSO produced probes with a higher activity.

6.9.10 Oligonucleotide reference channel

The reference channel for the *G. pulex* microarray protocol was an oligo designed to a section of vector contained in each PCR product. A section of pBluescript SKII vector was used to bind to mixed library amplicons and an oligonucleotide designed to pGEM-T vector was used to bind to amplicons from the male subtractive library. A quantity of 0.5pmols (comparable to similar studies (Dudley *et al.* 2002)) was used per slide per oligonucleotide giving a total of 1pmol total oligonucleotides per slide. This is an enhancement of the reference experiment design, allowing all gene expression profile to be compared to each other. The reference channel is used to assess the spot concentration and normalise the gene expression response in the samples channel. Similar methods have been used including creating a reference by pooling part of each amplicon before spotting onto the slides.(Sterrenburg *et al.* 2002) and by creating RNA complementary to the vector DNA represented in each amplicon, this 'universal RNA' reference is then labelled in the same way as the RNA samples and co-hybridised (Khan *et al.* 2006).

6.9.11 Hybridisation and Scanning

Successful hybridisations were affirmed by a low background and high signal intensity with equal hybridisation across the slide. The level of success was increased by thoroughly blocking slides and fastidiously cleaning labels; probe purity and correct environmental conditions were paramount in order to produce valuable data. After successful hybridisation and the removal of excess probe by washing, the slides were scanned with a laser to produce an image allowing the interpretation of differentially expressed genes. Slides were scanned using a PerkinElmer ScanArray® Express HT microarray scanner, a confocal laser scanner with adjustable photomultiplier tube (PMT) and laser controls. It was able to scan at down to a resolution of 5μm increasing the quantity of information extracted from the image.

6.9.12 Conclusion

A procedure to spot *G. pulex* target PCR amplicons onto Schott Nexterion Slide E microarray slides has been developed and optimised using Nexterion Modified Spot Buffer (NXT) as the spotting solution. This optimised procedure produced distinctive spots of circular morphology with evenly distributed levels of fluorescence across each spot after probe hybridisation. Ranges of spotting solutions were tested and the optimal solution identified.

CHAPTER 7

CHANGES IN GENE EXPRESSION IN ADULT *GAMMARUS PULEX*

7.1. INTRODUCTION

The process of moulting is essential to the growth and reproduction of crustaceans. It involves the development of a new cuticle beneath the present exoskeleton, which, when complete is followed by ecdysis, the shedding of the old exoskeleton (Figure 7.1).

Paired female during ecdysis Paired female after ecdysis

Figure 7.1: Paired *G. pulex* with moulting female

The new cuticle is created by secretion from the hypodermis and the readsorption of the medial layers of the old exoskeleton (West 1997). The build up of fluid between the cuticles allow the old exoskeleton to be easily removed at the point of ecdysis (West 1997). The length of the moult cycle and consequently the rate of development is dependent on temperature rather than seasonal rhythm (Sexton 1928).

The moult cycle is longer in males than females (Hynes 1955). It is possible to ascertain the sex of *G. pulex* before full maturity by the presence of the genital papillae in males and oostegites (brood plates) in females. In females the brood plates

develop serrated edges one moult cycle before maturity and developing eggs are clearly visible in the ovaries. When the females are fully mature the serrated edges of the brood plates become long bristles which help hold the fertilised eggs during development, therefore all breeding females contain these bristles (Hynes 1955). Mature females were observed to be lacking these bristles returning to the presence of serrated edges, but still retaining full size brood plates. This occurrence is thought to be due to loss of bristles at ecdysis and may be one cause of breeding cessation during winter months (Hynes 1955). Mature males can sometimes be determined by the presence of curly hairs on the third leg (front pereopod), however not all males have these (Hynes 1955).

One of the biological implications of growth by ecdysis is vast calcium fluctuations in both the haemolymph and hepatopancreas (5.5% of total body weight). The animals lose about 42% of body calcium 2-3 days premoult and a further 54% with the exuviae, leaving only 4% in postmoult animals (Wright 1980). The calcium is then acquired from their environment through food, water and sometimes exuviae consumption; recalcification time is dependent upon calcium concentration. Haemolymph and hepatopancreas concentrations of calcium throughout the moult cycle have been measured and the results shown graphically in Figure 7.2 (Wright 1980).

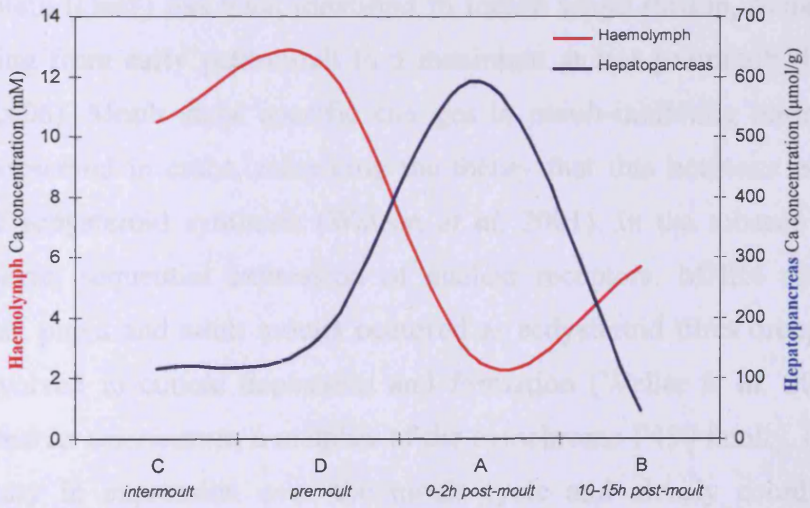


Figure 7.2 Mean haemolymph and hepatopancreas calcium concentrations in intermoult, premoult and postmoult *G. pulex*.
Generated from data published by D.A. Wright (Wright 1980).

This indicates there is likely to be a premoult release of calcium-binding proteins to increase the capacity of the calcium stores in the hepatopancreas immediately before ecdysis, followed by calcium transport to the exoskeleton post-moult. In immature *G. pulex*, the time between moults is much shorter than that of the mature amphipods. After reaching maturity the length of the moult cycle increases and the growth rate slows until the maximum size is achieved (Sexton 1928). These dramatic biochemical changes throughout the moult cycle indicate the major role that ecdysis plays in this species and suggests that there is likely to be significant changes in gene expression. Microarray analysis will allow these mechanisms to be elucidated.

7.1.1. Changes in gene expression over the moult cycle of arthropods

Changes in gene expression over the moult cycle have been observed in various crustaceans, i.e. in *Penaeus vannamei*, dot blot hybridisation revealed alterations in *trypsin* expression during premoult. The highest level occurred during early premoult (D_1), decreasing significantly in late premoult (D_2-D_3) and gradually increasing during and following postmoult (A and B) (Klein *et al.* 1996). A protein identified in *Cancer magister* crab haemolymph, cryptocyanin, was expressed in direct correlation with the changes in moult cycle stage. Cryptocyanin is a member of the haemocyanin family and is thought to be involved in both oxygen binding and moulting in *C. magister* (Terwilliger *et al.* 1999). Stage-specific expression of growth arrest-specific protein (Gas7) has been identified in Indian white shrimp, *Fenneropenaeus indicus*, rising from early post-moult to a maximum at late postmoult (Devaraj and Natarajan 2006). Moult stage specific changes in moult-inhibiting hormone (MIH) have been observed in crabs, reiterating the theory that this hormone is a negative regulator of ecdysteroid synthesis (Watson *et al.* 2001). In the tobacco hornworm, *Manduca sexta*, sequential expression of nuclear receptors, MHR4 and βFTZ-F1 during larval, pupal and adult moults occurred as ecdysteroid titres dropped and are possibly involved in cuticle deposition and formation (Weller *et al.* 2001). In the lobster, *Homarus americanus*, a member of the cytochrome P450 family, *CYP45*, was found to vary in expression over the moult cycle and closely coordinated with ecdysteroid titres. This correlation was tested further by ecdysteroid injections which mirrored an increase in *CYP45* expression (Snyder and Mulder 2001).

7.1.2. Effects of xenobiotics on gene expression during the moult cycle of arthropods

Moult cycling is an essential physiological process that has been linked to pollutant sensitivity and many pesticides are known to inhibit this process. It is though that the structural similarity of oestrogen-mimicking chemicals to crustacean ecdysteroids, cause them to interfere with moulting by binding to ecdysteroid receptor (EcR). This was demonstrated in a *Daphnia magna* population exposed to endosulfan or diethylstilbestrol (DES) which caused moulting to be inhibited, extending the time taken (by ~20-30%) for four moults to be completed (Zou and Fingerman 1997b). It is possible that moult inhibition had been caused by a factor other than EcR binding. Whether or not this inhibitory effect was due to direct competition with ecdysteroids for receptor binding or whether it is a general stress response to exposure has not been ascertained. In addition to normal physiological responses in the lobster, *Homarus americanus* described in Section 7.1.1, *cyp45* expression levels increased after exposure to an insecticide, heptachlor, and an anticonvulsant drug, phenobarbital, suggesting their interference in the endocrine system (Snyder 1998). *cyp45* has since been suggested as a potential biomarker of heptachlor in *H. americanus* larvae (Snyder and Mulder 2001).

The effects of xenobiotics on moulting have also been identified in *D. magna*. Animals were used to create microarray expression profiles after exposure to different doses of a fungicide, fenarimol, a known anti-ecdysteroid, revealing disruption of the moulting cycle including down regulation of cuticle proteins and vitellogenin (Soetaert *et al.* 2007). Moult cycle length is extended in *D. magna* after exposure to *ortho*-chlorinated PCBs: 2,4,5-trichlorobiphenyl (PCB29), Aroclor 1242, and diethyl phthalate (Zou and Fingerman 1997a). The ability of the shore crab, *Carcinus maenas*, to metabolise pyrene is understood to be due to moult-stage specific expression of cytochrome P450 proteins (CYP), which are responsible for pyrene metabolism. Crabs in terminal intermoult (anecdysis) had a higher occurrence of mortality compared to regularly moulting crabs, after exposure to pyrene, suggesting levels of CYP expression is associated with moult-stage sensitivity (Dam *et al.* 2006). The frequency of ecdysis was delayed in daphnids exposed to testosterone, this hormonal interference is believed to be through ecdysteroid receptor antagonism (Mu

and LeBlanc 2002). The insecticide and ecdysteroid agonist, tebufenozide (RH-5992), induces premature ecdysis and prevents the completion of the moult cycle in lepidopteran insects (Sundaram *et al.* 1998; Retnakaran *et al.* 2001). Exposure of *Chironomus riparius* to 17α-ethinylestradiol (EE) and bisphenol A (BPA) delayed moulting at the highest exposure dose of 1mg/L (Watts *et al.* 2003). The possibility that chitinase, a chitinolytic enzyme, has potential as a biomarker of xenobiotic exposure in the sand fiddler crab *Uca pugilator* was identified after exposure to (dichloro-diphenyl-trichloroethane) DDT inhibited epidermal chitinase activity (Zou and Bonvillain 2004).

One study of gammarids collected at intertidal sites exposed to direct sources of pollution showed disturbed gametogenesis and exoskeletal integrity (Gagne *et al.* 2005). Biochemical analysis of the gammarids revealed significant increase in the amount of arthropodin, sclerotin and metallothioneins and an increase in the activity of lipogenic enzymes involved in intermediary glucose metabolism, however the quantity of chitin was reduced.

7.1.3. Gender specific changes in gene expression

The identification of sex-specific markers which are differentially expressed between genders has been extensively studied in vertebrates, but less so in invertebrates. Most recently, gonadal transcriptome profiles were ascertained for male and female zebrafish in order to determine the molecular basis of gender using microarray technology. Sex specific molecular processes were identified and 9 differentially expressed genes were validated by QPCR (Santos *et al.* 2007). Vitellogenin has been identified and confirmed as a female specific marker in both vertebrates and invertebrates, as described in detail in Chapter 5, Section 5.1.2. Research on invertebrates at the molecular level has revealed a male specific insulin-like gene in the red-claw crayfish, *Cherax quadricarinatus*, using an androgenic gland subtractive cDNA library (Manor *et al.* 2007). Male and female specific fragments of genomic DNA have been discovered in the giant tiger shrimp, *P. monodon*, using the DNA fingerprinting technique, AFLP (amplified fragment length polymorphism). Most fragments have not been fully characterised, with only one female specific fragment, similar to NADPH ferrihaemoprotein reductase, identified using BLAST (Khamnamtong *et al.* 2006). This demonstrates the difficulty of identifying markers in non-model invertebrates which have not been fully sequenced.

7.1.4. Effects of xenobiotics on gender-specific gene expression

A review by McClellan-Green (2007) highlighted the importance of the gender-specific effect of toxicants on invertebrates by discussing the differences in endpoint observations in males and females in a range of invertebrate species. It is evident that different pathways regulating the sex hormones are all at risk from exposure to EDCs; these processes differ between genders and therefore need to be assessed separately. Although physiological, metabolic, morphological and behavioural differences in anthropogenic toxicant responses have been observed in invertebrates (Brown *et al.* 1999; Oberdorster *et al.* 2000; Gooding and LeBlanc 2001; reviewed in McClellan-Green *et al.* 2007), no significant or extensive studies have been undertaken to identify gender-specific effects of exposure at the molecular level of gonochoristic invertebrates.

The use of microarrays for transcriptomic profiling offers the potential to enhance the understanding of the molecular mechanisms involved in moulting and gender differentiation in *G. pulex*. By comparing transcriptome profiles throughout the moult cycle and between genders in *G. pulex* our understanding of the differential gene expression in response to major physiological events can be augmented. The basal expression levels of the following moult stages were assessed: intermoult (C_2), late intermoult / early premoult (C_4 / D_0) and premoult (D_1). This allows the gene expression variation throughout this critical process to be determined and assists our understanding of the pathways involved in moulting at the genetic level.

7.2. AMPHIPOD COLLECTION

Amphipods were collected from a site near Cynrig hatchery (Environment Agency site; Figure 7.3) near Llanfrynach, Powys by kick sampling as described in Chapter 2, Section 2.9.1. The catch was roughly sorted on-site in river water, removing any potential predators, and transported to the laboratory. More thorough sorting was performed in the laboratory, transferring only *G. pulex* into a fresh container of river water. Amphipods were placed in an incubator at 17°C overnight to become accustomed to the temperature, with an excess of conditioned alder leaves (Chapter 2, Section 2.9.1). The following day the river water was combined with deionised and dechlorinated tap water, and incubated overnight to allow animals to adapt to the change in water. The water was then replaced entirely with deionised and dechlorinated tap water and the animals left for at least 10 days to acclimatise to the conditions. The incubator was set to a 16 hour light, 8 hour dark cycle.

Figure 7.3: *G. pulex* sampling site at Cynrig Fish Hatchery, near Brecon, Powys
OS grid reference: SO06582641

7.3. SEX DETERMINATION

7.3.1. Paired method

Male and female *G. pulex*, were sexed by their precopulatory pairing position (Figure 7.4). Paired animals were separated by removal from water; the smaller female and the larger male that was grasping her, were placed in different containers before sacrifice.

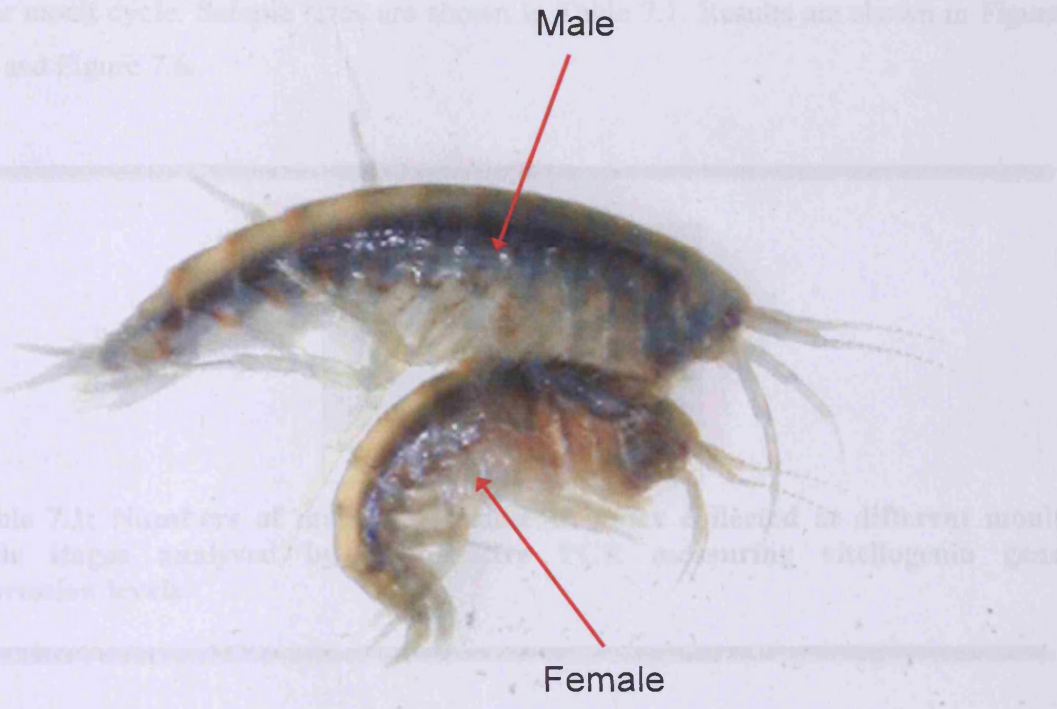


Figure 7.4: Male and female *G. pulex* in the precopulatory pairing position

7.3.2. Vitellogenin gene expression

7.3.2.1. QPCR

Quantitative PCR was performed as described in Chapter 2, Section 2.11.13. After reverse transcription the remaining RNA was hydrolysed, the sample neutralised and cDNA purified as described in Chapter 2, Section 2.13.8.1. Purified cDNA was quantified using PicoGreen® assay (Chapter 2, Section 2.11.8). Amplification efficiency was calculated from the gradient of the standard curve ($E = 10^{-1/gradient}$) and

average efficiency (1.90) used to determine normalised expression fold change. Data were normalised with the reference gene, GAPDH, using the ΔΔCt method (Livak and Schmittgen 2001) to give normalised expression fold change ($E^{-\Delta\Delta Ct}$). The mean and standard deviation of the male and female group were determined and the 'unknown' animals were allocated into groups depending on their vitellogenin expression levels. If expression levels fell within the 95% confidence intervals of the mean, animals were assigned to a group based on gender. Male samples were calibrated against female and *vice versa*. Moult cycle stage was determined by sacrificing animals at ¼ (Stage 1), ½ (Stage 2) and ¾ (Stage 3) of the way through their moult cycle. Sample sizes are shown in Table 7.1. Results are shown in Figure 7.5 and Figure 7.6.

	Female	Male
Stage '1'	6	6
Stage '2'	4	7
Stage '3'	6	3
TOTAL	**16**	**16**

Table 7.1: Numbers of male and female *G. pulex* collected at different moult cycle stages analysed by quantitative PCR measuring vitellogenin gene expression levels

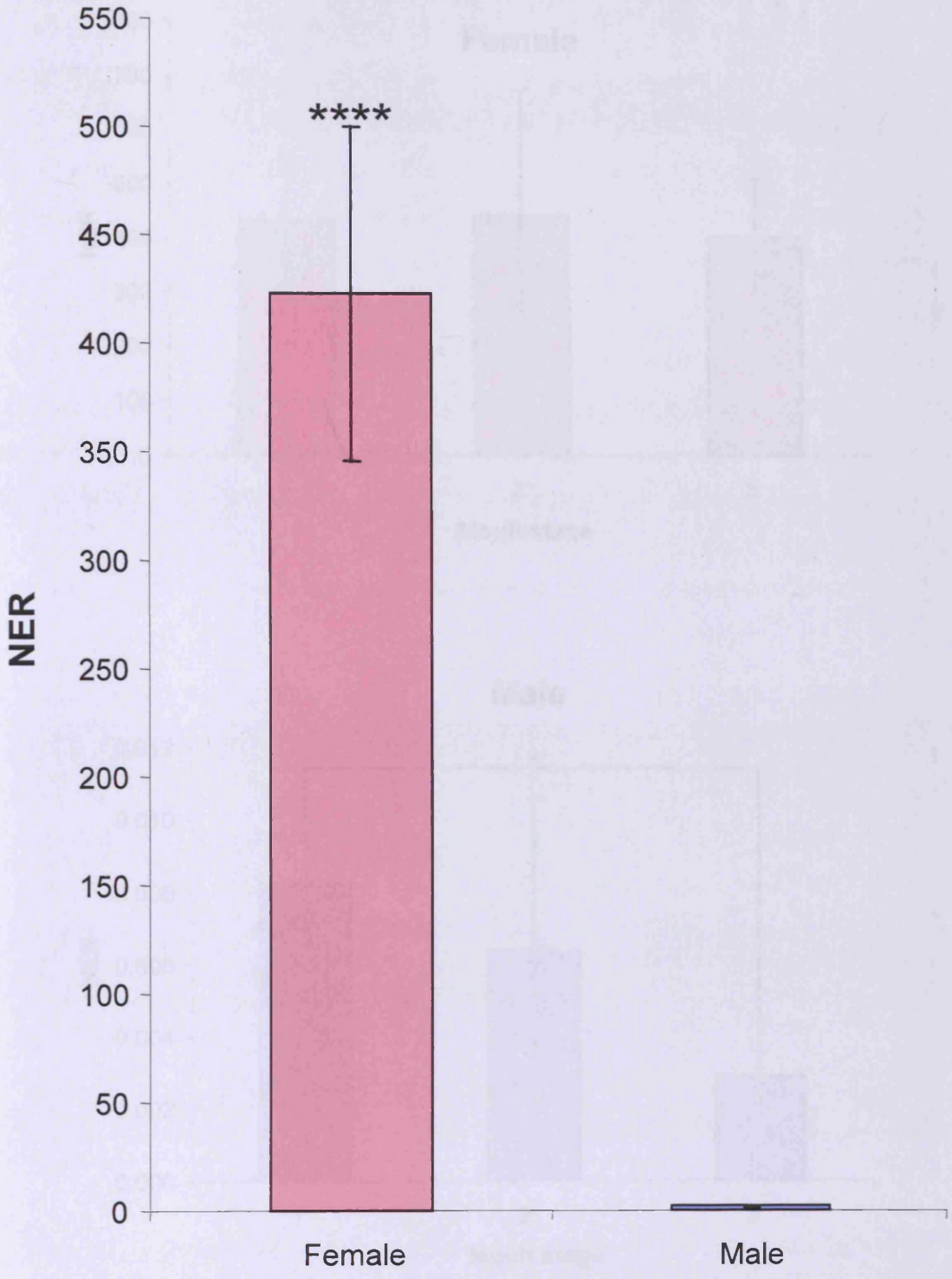
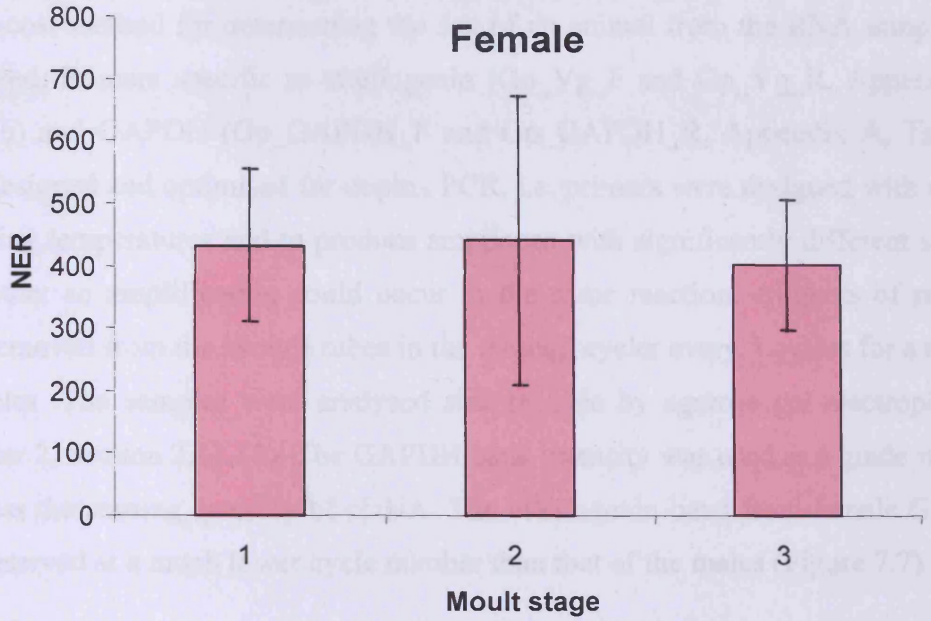


Figure 7.5: Female and male *G. pulex* vitellogenin gene expression levels in animals used for transcriptomic profiling using QPCR

NER = Normalised Expression Ratio. **** = $p < 0.0001$ from non-parametric Mann-Whitney statistical analysis.



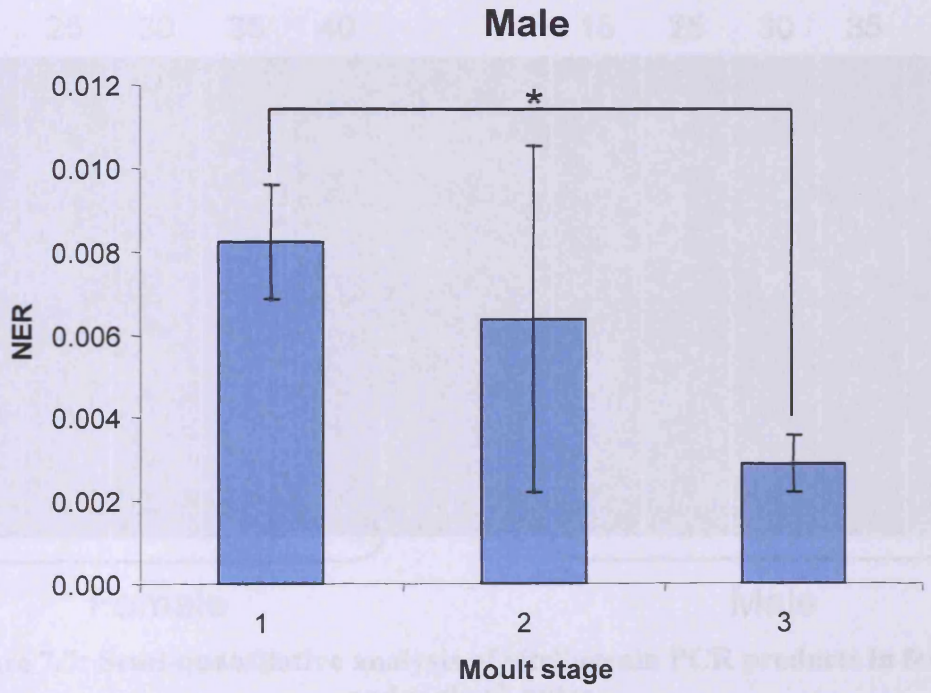


Figure 7.6: Female and male *G. pulex* vitellogenin gene expression over the moult cycle in animals used for transcriptomic profiling using QPCR

NER = Normalised Expression Ratio. * = $p < 0.1$ from non-parametric Mann-Whitney statistical analysis.

7.3.2.2. Semi-Quantitative PCR

A low-cost method for determining the sex of an animal from the RNA sample was optimised. Primers specific to vitellogenin (Gp_Vg_F and Gp_Vg_R, Appendix A, Table 6) and GAPDH (Gp_GAPDH_F and Gp_GAPDH_R, Appendix A, Table 6) were designed and optimised for duplex PCR, i.e. primers were designed with similar annealing temperatures and to produce amplicons with significantly different sizes to each other so amplification could occur in the same reaction. Aliquots of reaction were removed from the sample tubes in the thermal cycler every 5 cycles for a total of 40 cycles. The samples were analysed side by side by agarose gel electrophoresis (Chapter 2, Section 2.11.11). The GAPDH band intensity was used as a guide in order to assess the starting quantity of cDNA. The vitellogenin band from female *G. pulex* was observed at a much lower cycle number than that of the males (Figure 7.7).

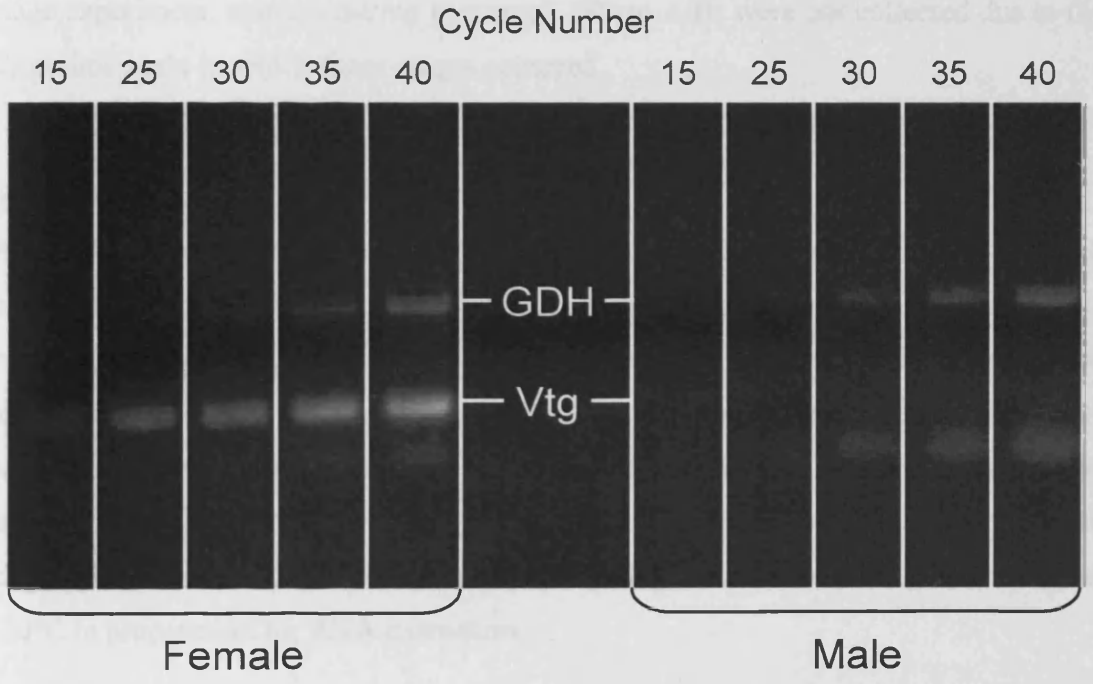


Figure 7.7: Semi-quantitative analysis of vitellogenin PCR products in female and male *G. pulex*

Vtg = vitellogenin; GDH = glyceraldehyde 3-phosphate dehydrogenase (GAPDH)
An aliquot was removed from the PCR sample at cycle number 15, 25, 30, 35 and at the end of the reaction (cycle 40). The bands were separated using agarose gel electrophoresis.

7.4. CYCLE LENGTH AND MOULT STAGE DETERMINATION

After the acclimatisation period, the animals were ready for experimentation. Adults of known or unknown sex (i.e. single or separated pairs, see Section 7.3.1) and approximately equal size (7mm - 12mm) were placed in separate, numbered containers with 80ml of deionised tap water with a 15mm disc of conditioned alder leaf, replaced as required allowing them to feed at a natural rate, but without an excess of food present. Containers were checked everyday for the presence of exuviate, which if present, was immediately removed with a Pasteur pipette. Water was changed every 3 days. The date on which the animal had moulted (within 24 hours) was recorded. After the second moult, the cycle length was calculated and the animal was sacrificed at points ¼, ½ or ¾ of the way through the cycle, corresponding approximately with stages C_2 (¼), late C_4 / early D_0 (½) and mid D_1 (¾). In addition 2 animals were collected at D_2/D_3 as close to predicted ecdysis as possible. These were called Stage 1 (¼), Stage 2 (½), Stage 3 (¾) and Stage 4 (D_2/D_3). In this general moult stage experiment, animals during postmoult (Stage A/B) were not collected due to the short time scale in which these stages occurred.

7.4.1. Sample preparation

Amphipods were placed in a Petri dish of fresh deionised tap water for five minutes, before being transferred to a second dish containing fresh deionised tap water. This process was repeated. The washing procedure removes any detritus from the animal and helps to clear the intestinal tubes from food to minimise carryover into the RNA sample. The amphipod was briefly placed onto a soft tissue to remove excess water and immediately sacrificed in liquid nitrogen. Frozen samples were transferred to pre-chilled RNA*later*®ICE (~0.5ml; Ambion, Huntingdon, Cambs., UK) and stored at -20°C in preparation for RNA extraction.

RNA was extracted, purified and quantified as described in Chapter 2, Sections 2.13.6 and 2.13.7 and stored at -80°C for subsequent reverse transcription and Cy3 labelling (Chapter 2, Section 2.13.8) before being hybridised to a *G. pulex* microarray slide alongside vector-specific Cy5-labelled oligonucleotides.

7.4.2. Expression levels of gene encoding Cuticle protein

7.4.2.1. QPCR

A small volume of RNA used for microarray analysis was reverse transcribed (Chapter 2, Section 2.11.6) and used to investigate cuticle protein gene expression using Quantitative PCR (QPCR) using 10ng cDNA per well. QPCR was performed as described in Chapter 2, Section 2.11.13, using glyceraldehyde 3-phosphate dehydrogenase (GAPDH) as an internal reference gene (see Chapter 5, Section 5.1.4). Samples were analysed in duplicate and controls in triplicate. Amplification efficiency was calculated from the gradient of the standard curve ($E = 10^{-1/gradient}$) and average efficiency (1.96) used to determine normalised expression fold change. Data was normalised with the reference gene, GAPDH, using the ΔΔCt method (Livak and Schmittgen 2001) to give normalised expression fold change ($E^{-\Delta\Delta Ct}$). Expression fold changes of animals of estimated moult stage were used to establish a mean and standard deviation for each stage (C_2, late C_4 / early D_0 and mid D_1). Moult staged groups were compared to each other using a Mann-Whitney U-test. Sample sizes are shown in Table 7.2. Results are shown in Figure 7.8 and Figure 7.9.

	Female	Male
Stage '1'	6	6
Stage '2'	4	7
Stage '3'	6	3
Stage '4'	2	0
TOTAL	**16**	**16**

Table 7.2: Numbers of male and female *G. pulex* collected at different moult cycle stages analysed by quantitative PCR measuring cuticle gene expression levels

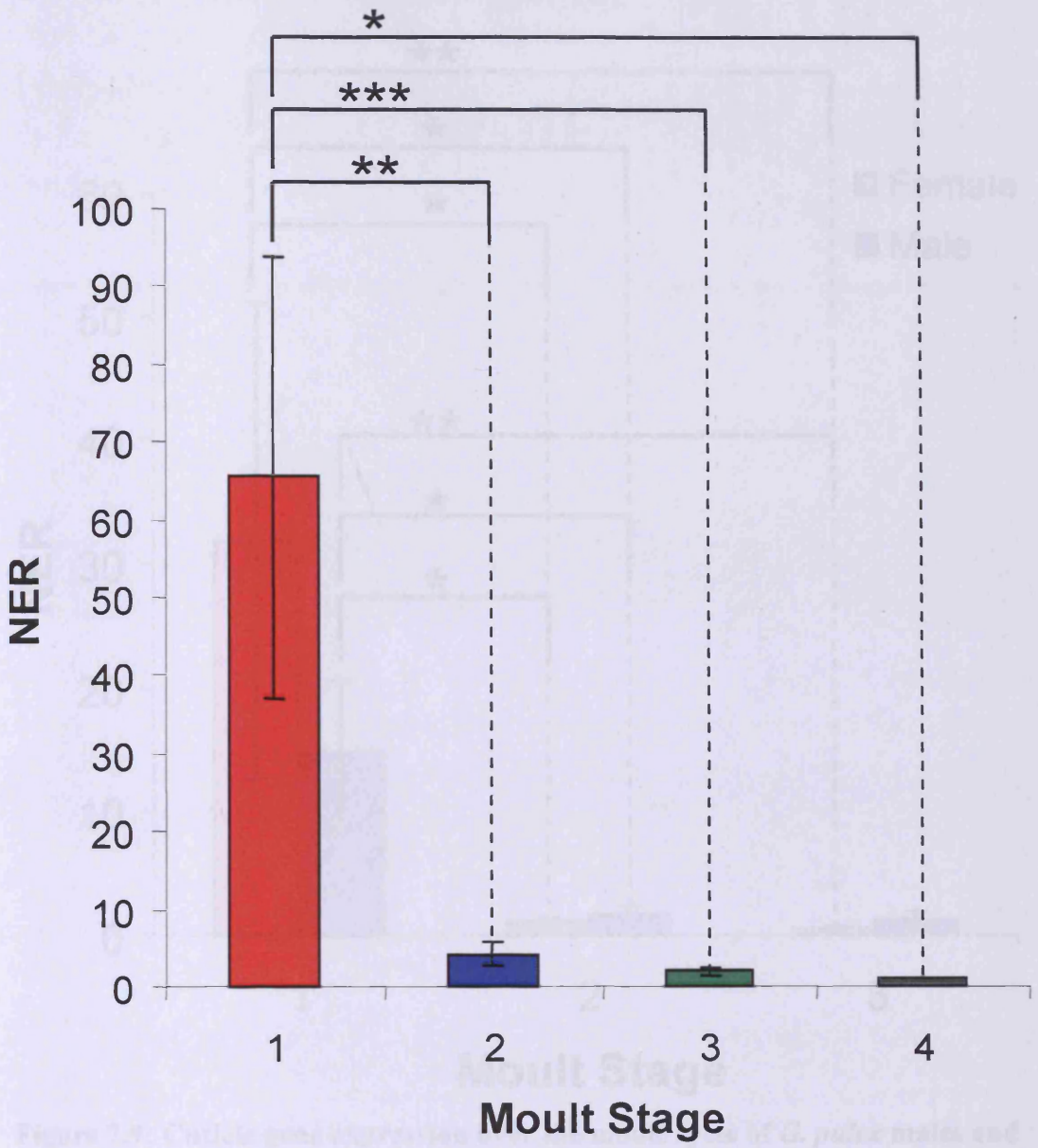


Figure 7.8: Cuticle gene expression over the moult cycle of *G. pulex*

Changes in cuticle gene expression in males and females (data combined) over 4 moult stages. NER = Normalised Expression Ratio. * = $p < 0.1$; ** = $p < 0.01$ and *** = $p < 0.001$ from non-parametric Mann-Whitney statistical analysis.

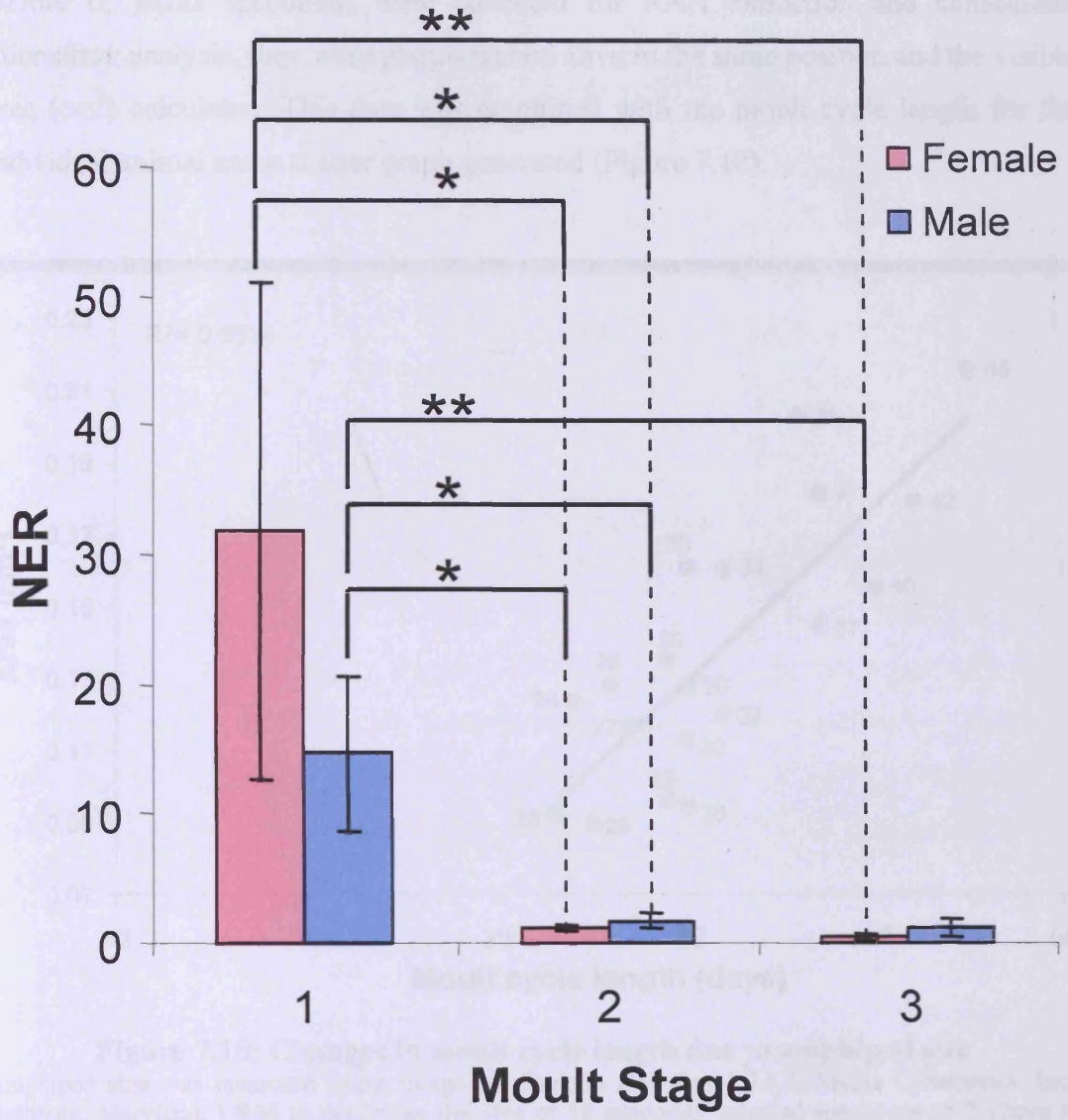


Figure 7.9: Cuticle gene expression over the moult cycle of *G. pulex* males and females

Changes in cuticle gene expression in males or females over 3 moult stages
NER = Normalised Expression Ratio. * = $p < 0.1$ and ** = $p < 0.01$ from non-parametric Mann-Whitney statistical analysis.

7.4.3. Relationship of animal size/area with cycle length

Before *G. pulex* specimens were collected for RNA extraction and consequent microarray analysis, they were photographed alive in the same position and the visible area (cm^2) calculated. This data was combined with the moult cycle length for the individual animal and a scatter graph generated (Figure 7.10).

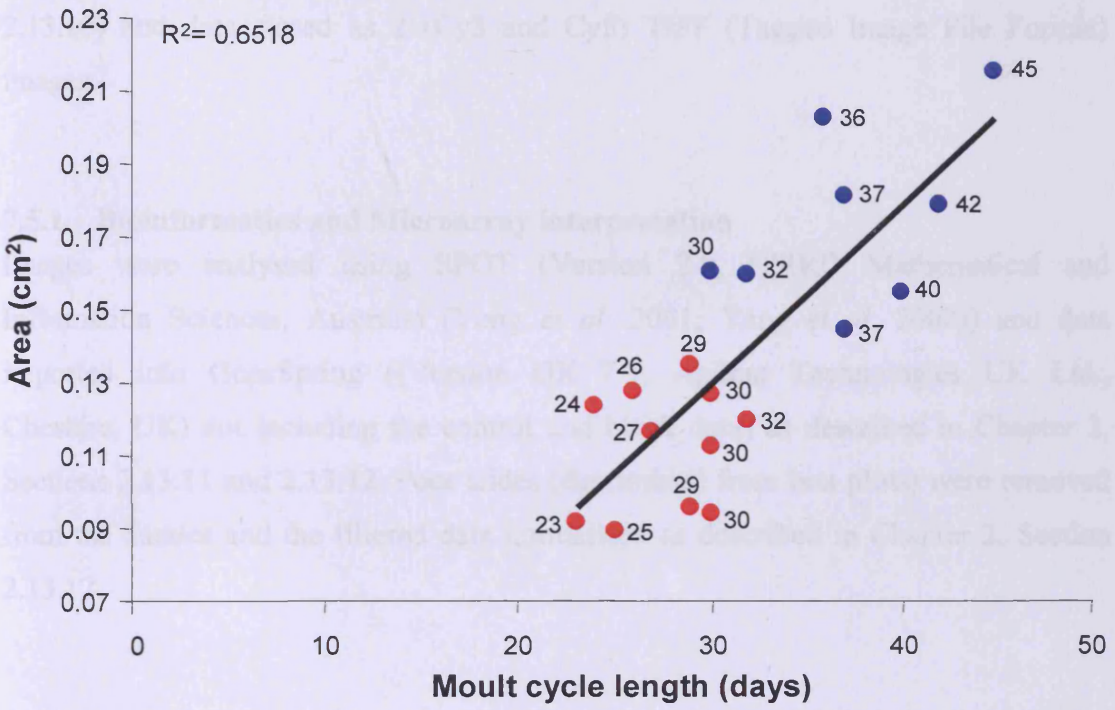


Figure 7.10: Changes in moult cycle length due to amphipod size

Amphipod size was measured using Image-Pro Express (Version 4.5.1.3, Media Cybernetics, Inc., Bethesda, Maryland, USA) to determine the area of 19 randomly selected specimens of 7-12mm in approximate length in the same position, using digital photography. Petri dishes containing the animals were placed onto a paper ruler in order to calibrate the programme and compare images. Red spots = females; Blue spots = males. Mean moult cycle lengths including standard errors were 28±1 days for females and 37±2 days for males at 17°C.

7.5. *G. PULEX* MICROARRAY ANALYSIS

Sample collection and storage is described in 7.4.1. Before labelling, RNA was extracted as described in Chapter 2, Section 2.13.5 and purified (Chapter 2, Section 2.13.6) without the optional LiCl precipitation. Samples passing quality control (Chapter 2, Section 2.13.7) were labelled (Chapter 2, Section 2.13.8) using a 4:1:5 U:T:ACG ratio of dNTPs and hybridised (Chapter 2, Section 2.13.9) onto *G. pulex* microarray slides (Chapter 2, Section 2.13.3). Slides were scanned (Chapter 2, Section 2.13.10) and data stored as 2 (Cy3 and Cy5) TIFF (Tagged Image File Format) images.

7.5.1. Bioinformatics and Microarray interpretation

Images were analysed using SPOT (Version 2.0, CSIRO Mathematical and Information Sciences, Australia (Yang *et al.* 2001; Yang *et al.* 2002)) and data imported into GeneSpring ((Version GX 7.3; Agilent Technologies UK Ltd., Cheshire, UK) not including the control and blank data) as described in Chapter 2, Sections 2.13.11 and 2.13.12. Poor slides (determined from box plots) were removed from the dataset and the filtered data normalised as described in Chapter 2, Section 2.13.12.

7.5.1.1. Microarray quality validation

An MA plot was generated to demonstrate the raw and normalised filtered data of the experimental samples. An MA plot is defined as the log ratio vs. log median intensity where:

$$M = \log_2\left(\frac{Cy5}{Cy3}\right)$$

$$A = \frac{(\log_2 Cy5 + \log_2 Cy3)}{2}$$

The MA of raw data (Figure 7.11A and Figure 7.11C) shows the intensity dependent ratio and demonstrates spot artefacts and systematic variation. The ideal MA plot should have M values that are evenly distributed at 0 across the range of intensities (Yang *et al.* 2002). Raw data was normalised to remove systematic bias from

experimental artefacts as described in Chapter 2, Section 2.13.12. The MA plot demonstrates how the filtering and normalisation process created a data set distributed around M=0 (Figure 7.11B and Figure 7.11D).

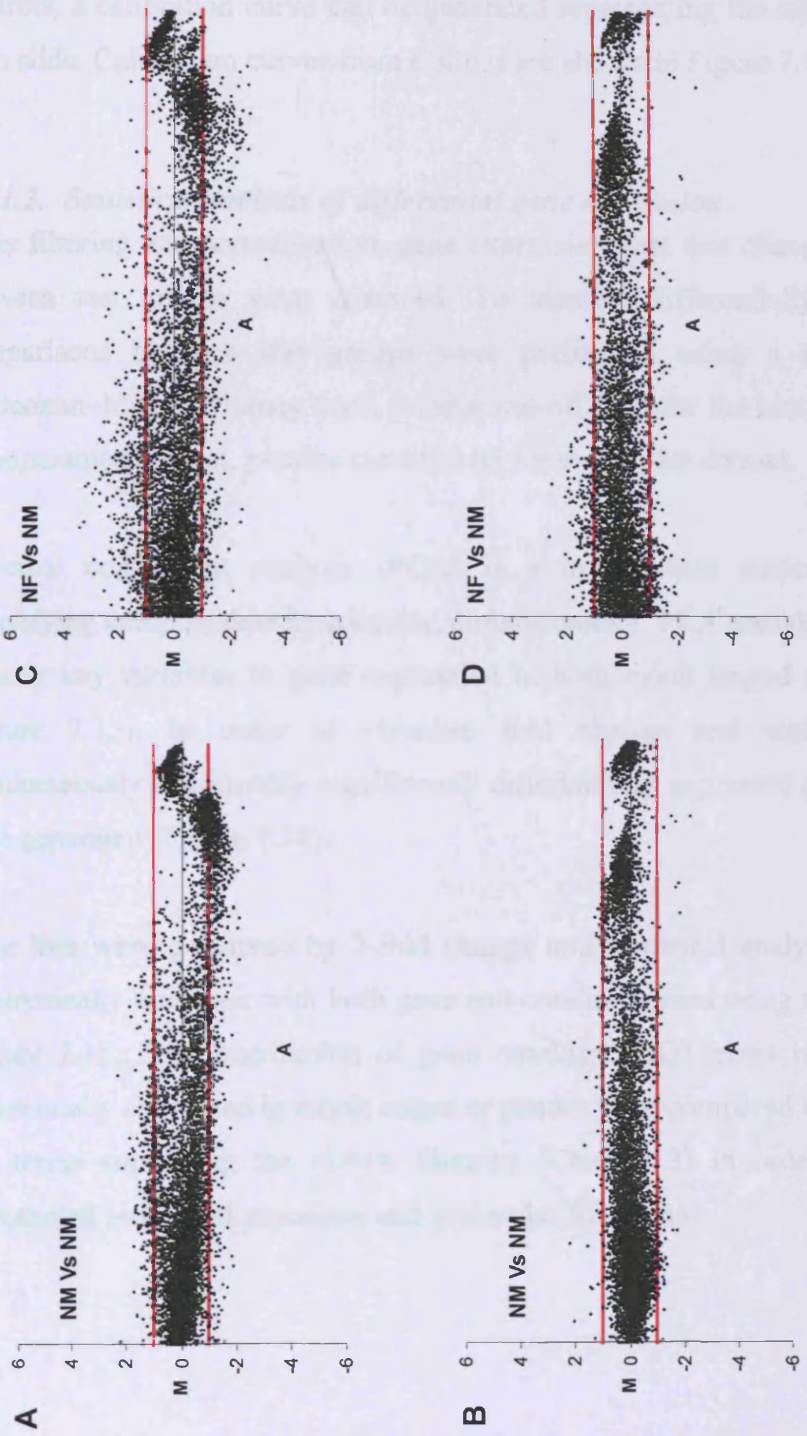


Figure 7.11: MA plots of male and female *G. pulex* microarray data before and after normalisation
A: Male vs. Male before normalisation; B: Male vs. Male after normalisation; C: Female vs. Male before normalisation; D: Female vs. Male after normalisation.

Internal synthetic control mRNA (Universal ScoreCard, GE Healthcare, Little Chalfont, Bucks., UK) was added to each RNA sample and reverse transcribed and indirectly incorporated with Cy3 simultaneously with the RNA. These control probes bind to complementary target sequences printed across the length and width of the microarray slides alongside the *G. pulex* DNA targets. Using the ratio ScoreCard controls, a calibration curve can be generated representing the range of detection for each slide. Calibration curves from 6 slides are shown in Figure 7.12.

7.5.1.2. Statistical analysis of differential gene expression

After filtering and normalisation, gene expression data that changed less than 2-fold between test groups were removed. To identify differentially expressed genes, comparisons between test groups were performed using a non-parametric test (Wilcoxon–Mann–Whitney test), p-value cut-off 0.05 for the moult stage dataset and a nonparametric t-test, p-value cut-off 0.05 for the gender dataset.

Principal component analysis (PCA) is a multivariate statistical technique for simplifying complex data by reducing dimensionality. PCA analysis was performed to identify key variables in gene expression in both moult staged and gender datasets (Figure 7.13). In order to visualise fold change and statistical significance simultaneously and identify significantly differentially expressed genes, volcano plots were generated (Figure 7.14).

Gene lists were generated by 2-fold change and statistical analysis (see above) and hierarchically clustered with both gene and condition trees using a distance algorithm (Figure 7.15). The distribution of gene ontology (GO) terms in the lists of genes differentially expressed in moult stages or gender were compared to the distribution of GO terms annotating the cDNA libraries (Chapter 3) in order to identify over-represented biological processes and molecular functions.

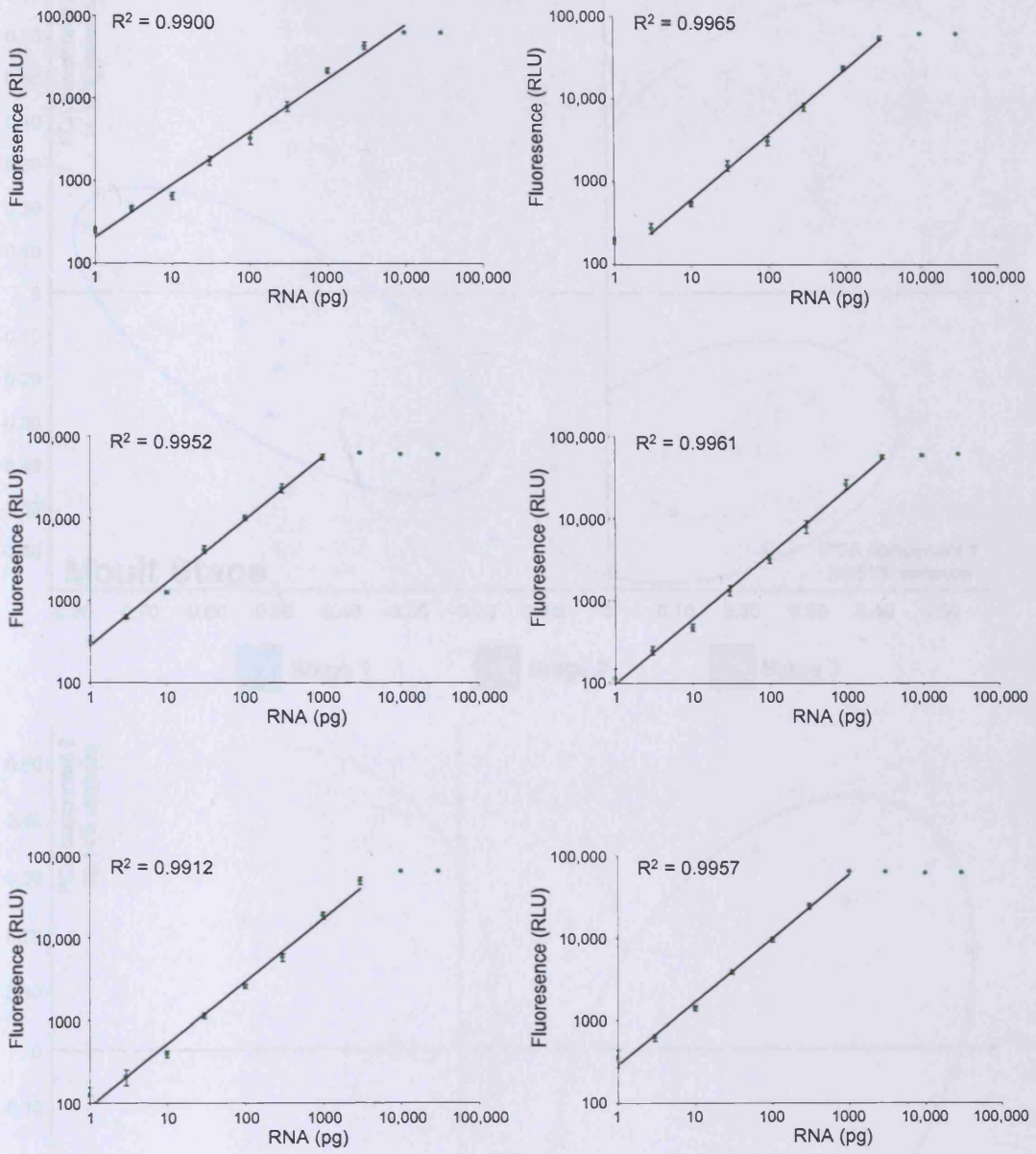


Figure 7.12: Universal ScoreCard calibration of 6 slides used to generate both Moult and Gender datasets

Each ratio control was printed 10 times per slide; the mean signal intensity was used to generate calibration curves which are displayed with standard error bars.

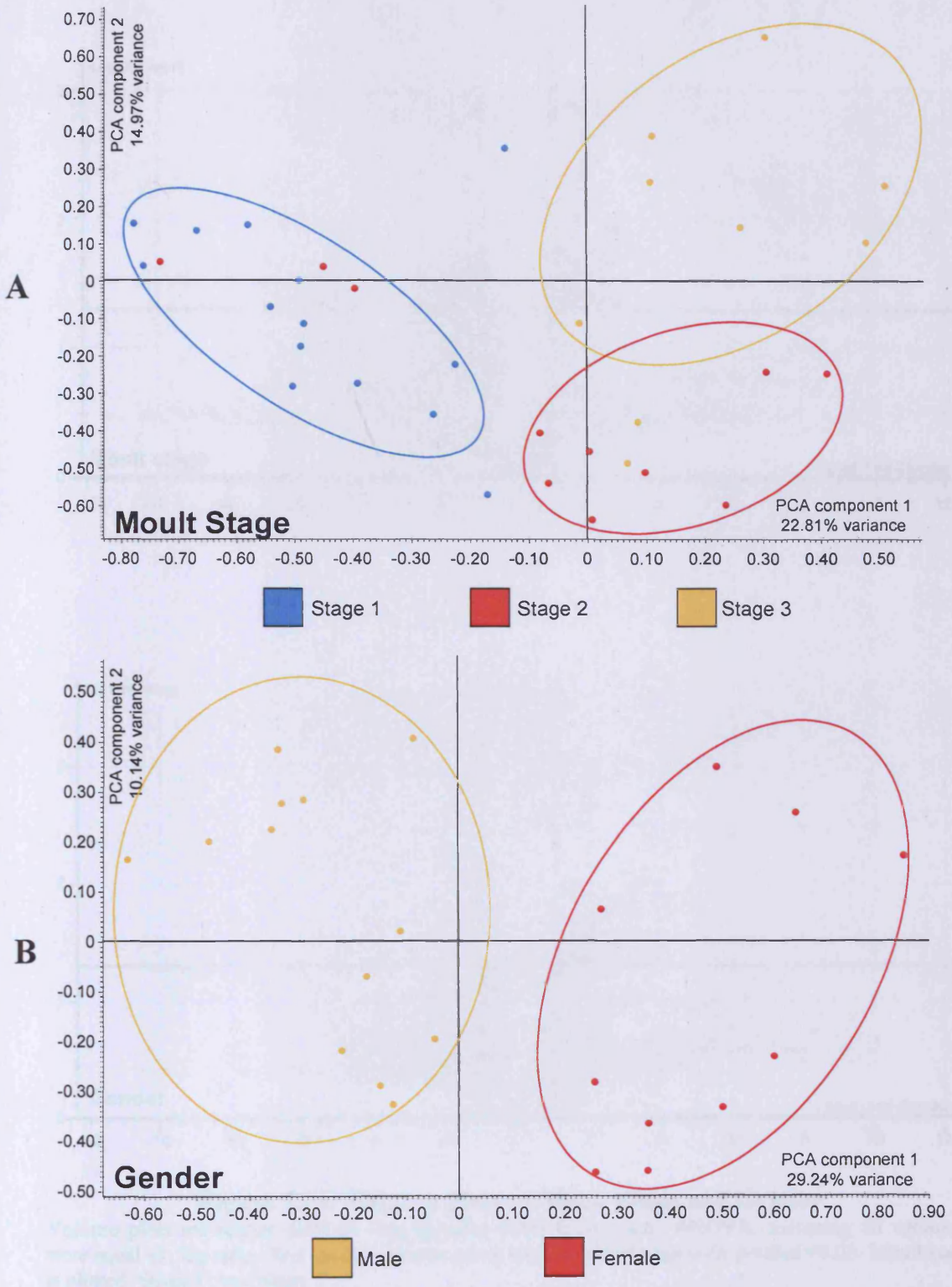


Figure 7.13: Principal component analysis of individual moult stage and gender transcriptomes

Based on all statistically relevant differentially expressed genes from:
A: Mixed adults in moult stages 1 (blue), 2 (red) or 3 (yellow).
B: Males (yellow) and females (red).

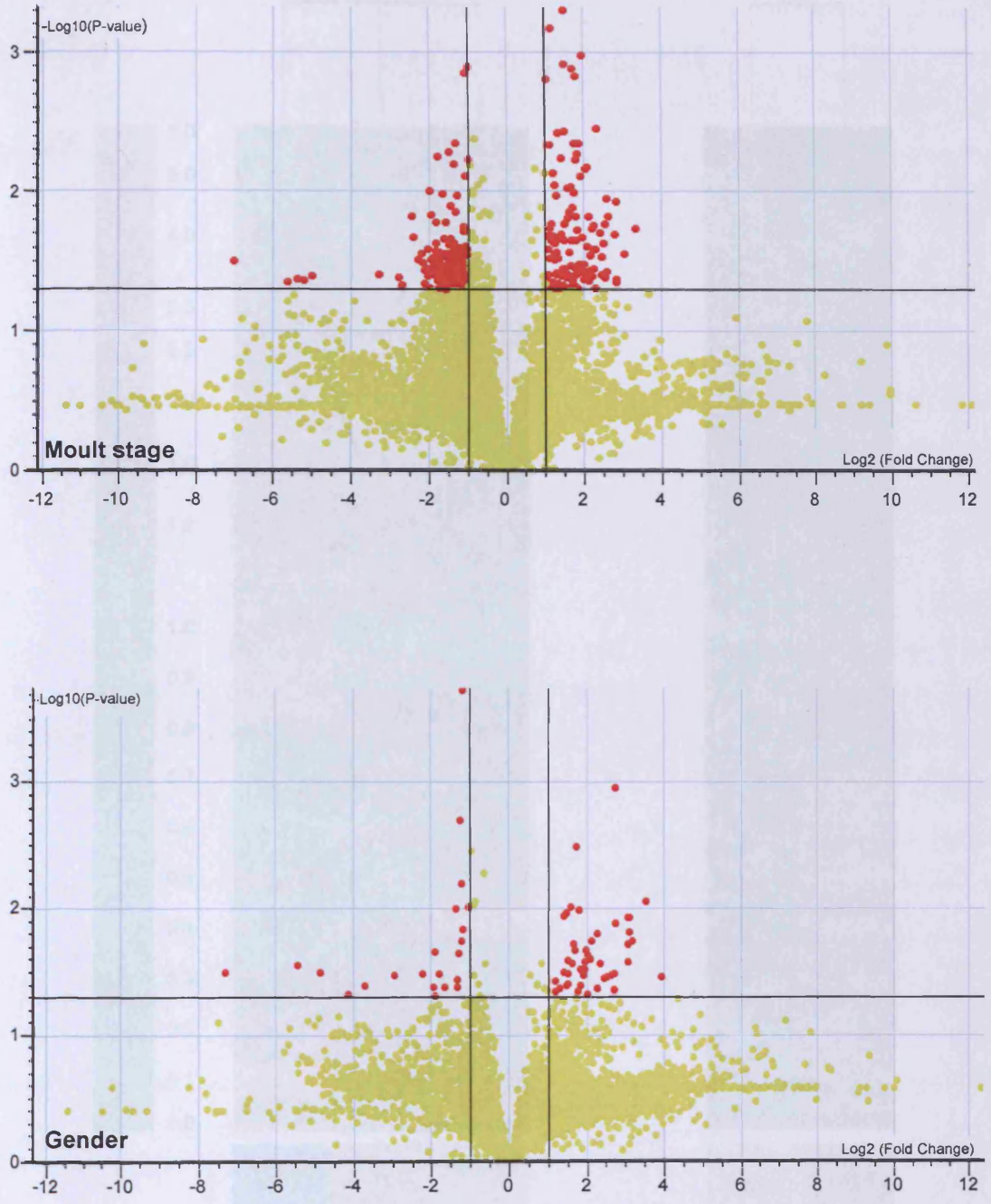


Figure 7.14: Volcano plots of Moult Stage and Gender

Volcano plots are scatter plots of –log (p-value 0.05) from 1-way ANOVA, assuming all variances were equal vs. log ratio. Red spots represent genes with 2-fold change with p-value <0.05. Moult stage is plotted 'Stage 1' vs. 'Stage 3'.

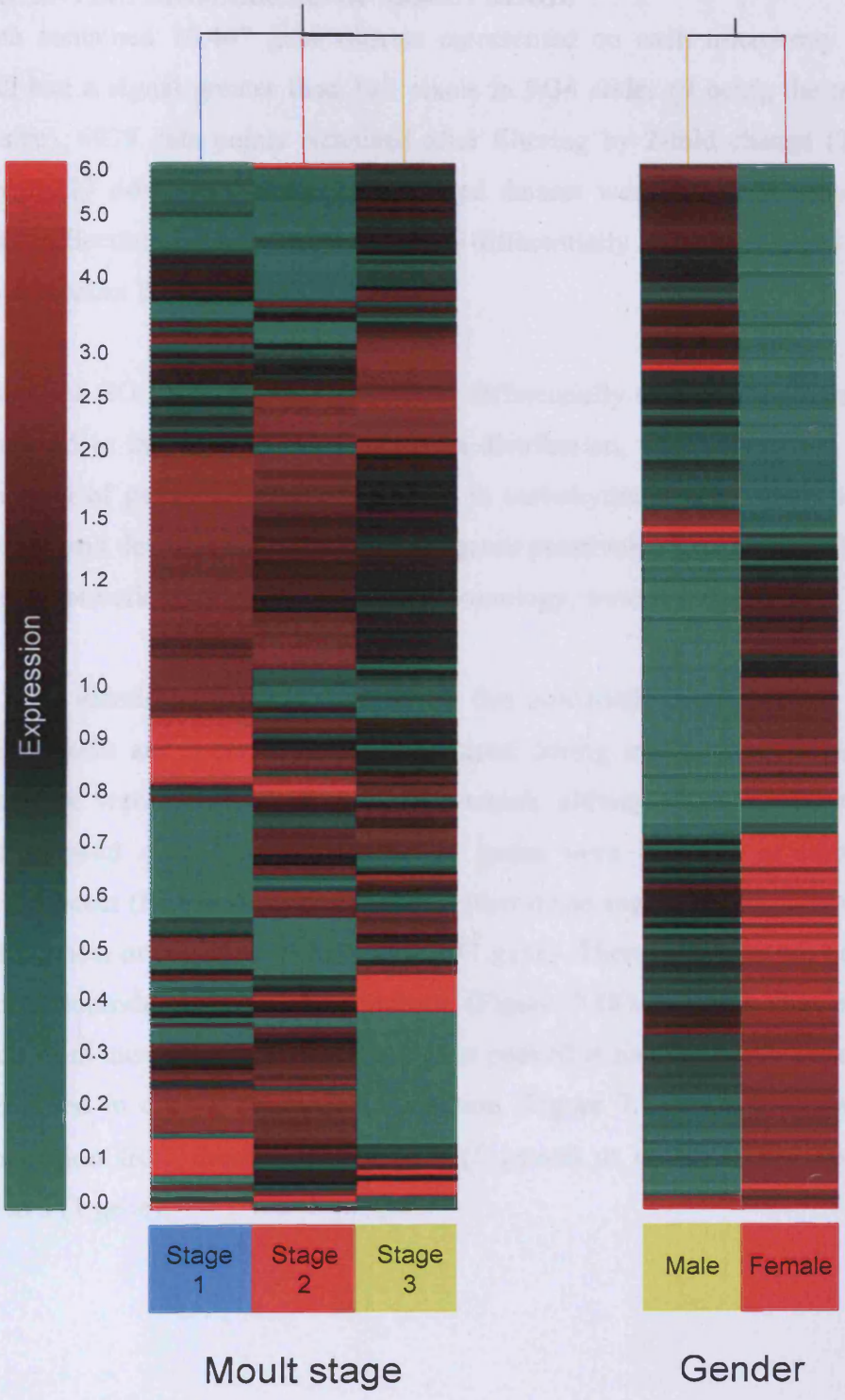


Figure 7.15: Global gene expression profiles for Moult stage and Gender in *G. pulex*

Hierarchical tree of genes and heat map showing differential expression between the 2 genders. The heat map columns left to right are: Normal males and Normal females. Clustering was calculated using 'Distance' algorithm. Expression is average log signal intensity.

7.6. POTENTIAL BIOMARKERS OF MOULT STAGE

Raw data contained 13,447 gene objects represented on each microarray slide of which all had a signal greater than 100 pixels in 9/34 slides (9 being the minimum sample size). 6928 data points remained after filtering by 2-fold change (2236 up-regulated; 5227 down-regulated). This filtered dataset was statistically analysed as described in Section 7.5.1.2. The full list of differentially expressed genes (199) is given in Appendix B, Table 1).

Distribution of GO terms in the list of genes differentially expressed in moult stages was compared to the cDNA library GO term distribution, which identified an over-representation of genes potentially involved in carbohydrate metabolic process and reproduction and development. Additionally genes putatively related to cuticle protein and chorion formation according to BLAST homology, were investigated.

7 genes were identified showing clear trends that potentially could be used as moult stage biomarkers and were significantly altered during moult stage (Figure 7.16) although there were numerous other genes which, although were not significantly different, showed clear trends. In total, 14 genes were involved in carbohydrate metabolic process (Figure 7.17) and showed either down-regulation from moult stages 1 to 3 (13 genes) or peaked at moult stage 2 (1 gene). There were 6 genes potentially involved in reproduction and development (Figure 7.18), 5 genes showed down-regulation from moult stages 1 to 3 and 1 gene peaked at moult stage 2. There were 6 genes involved in cuticle or chorion formation (Figure 7.18) which showed either down-regulation from moult stages 1 to 3 (5 genes) or up-regulation from moult stages 1 to 3 (1 gene).

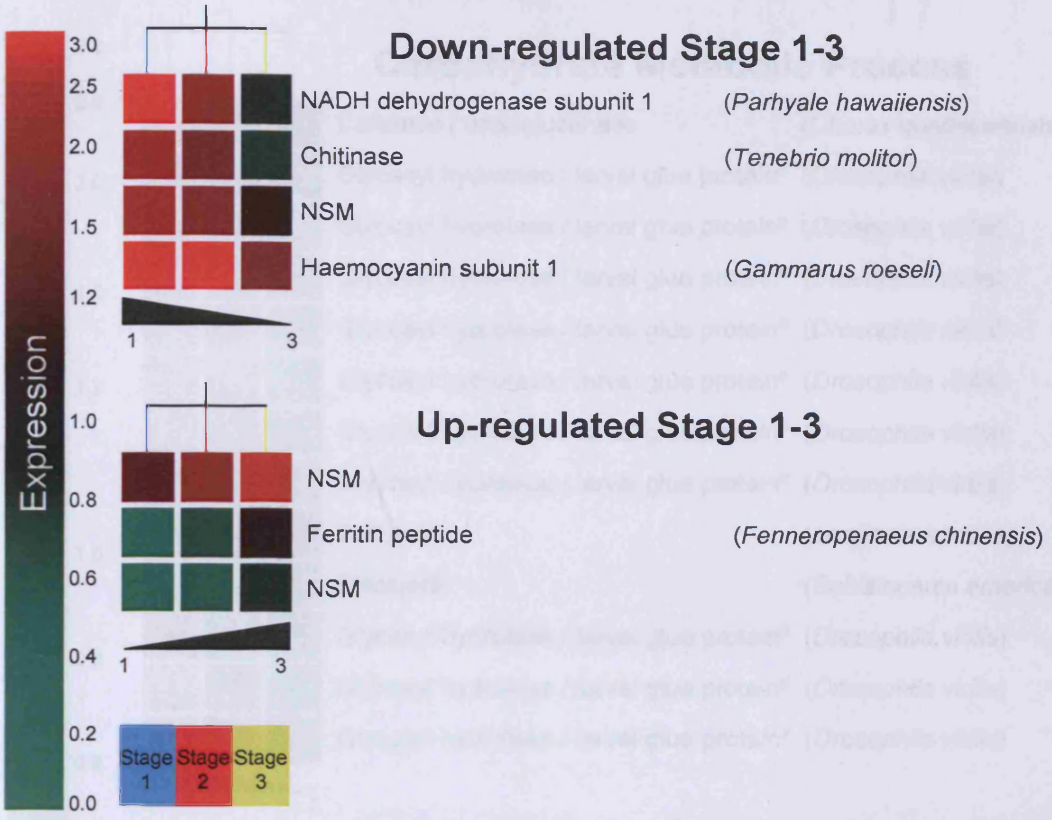


Figure 7.16: Statistically relevant moult stage changes in gene expression
NSM = No Significant Match against BLAST database

Of the 7 genes significantly altered in expression over the moult cycle, only 4 had a significant homology to any known protein according to BLASTX analysis. Of the genes with putative functions based on homology, the fold-changes shown in Table 7.3 were observed:

PUTATIVE FUNCTION	STAGE 1	STAGE 2	STAGE 3
Down-regulated 1 to 3			
NADH dehydrogenase subunit I	3.8	1.8	0.8
Chitinase	2.2	1.5	0.8
Haemocyanin Subunit 1	5.5	3.3	1.6
Up-regulated 1 to 3			
Ferritin Peptide	0.5	0.7	1.0

Table 7.3: Statistically significant fold-changes in expression over the moult cycle of *G. pulex*

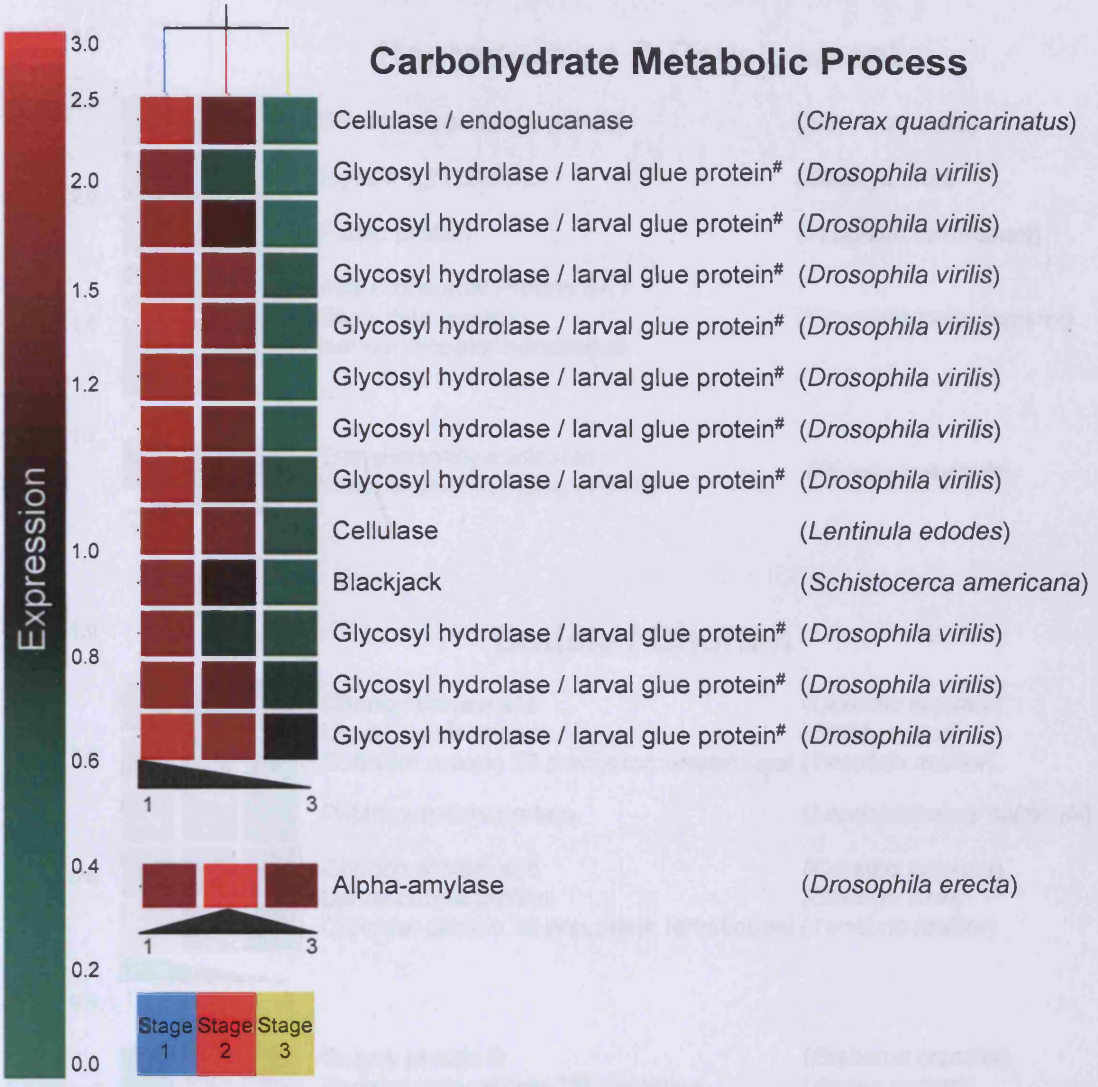


Figure 7.17: Differential expression in genes potentially involved in carbohydrate metabolic processes

These gene fragments are all clustered in contig GAC00436

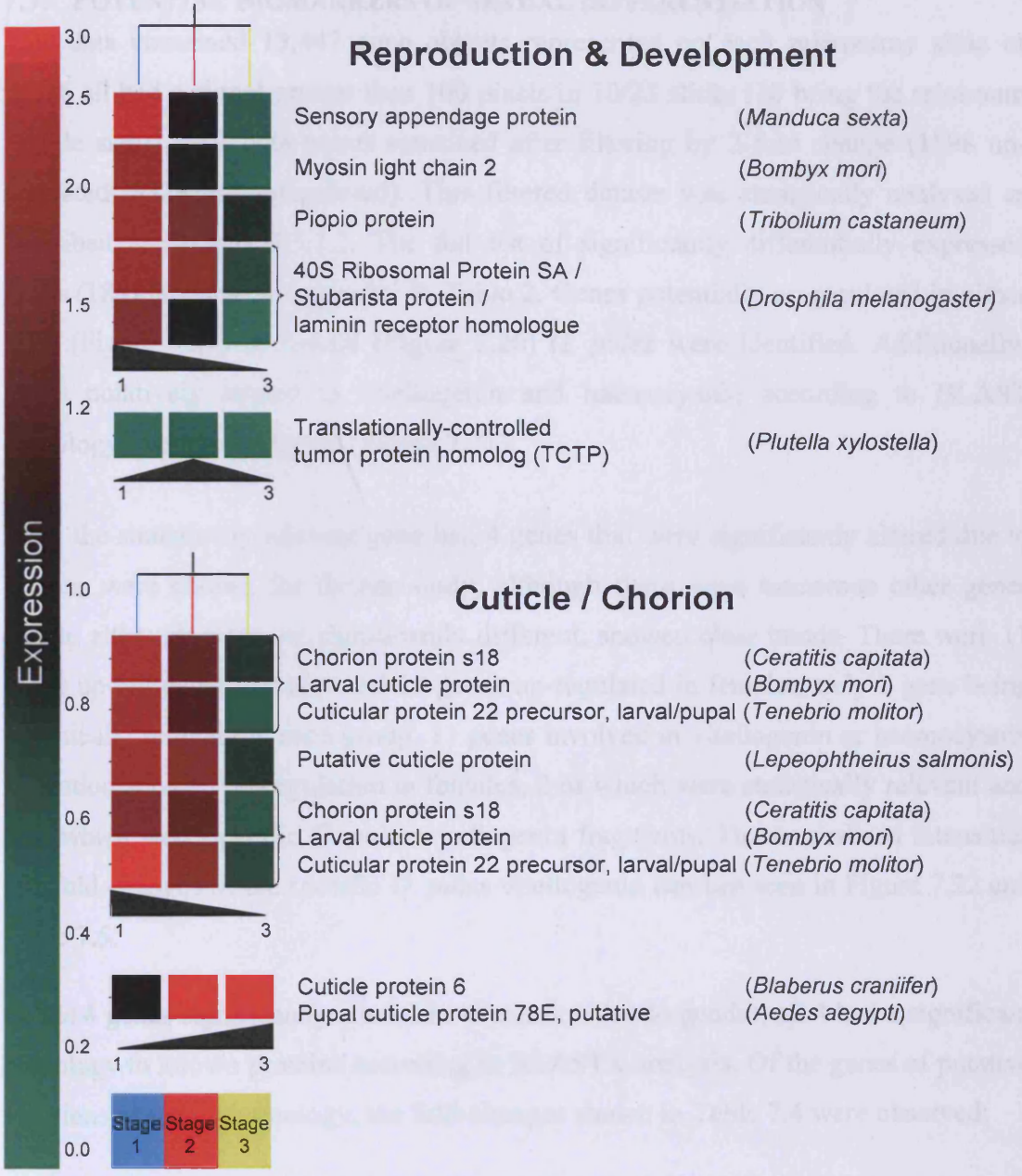


Figure 7.18: Differential expression in genes potentially involved in reproduction, development and cuticle/chorion formation

7.7. POTENTIAL BIOMARKERS OF SEXUAL DIFFERENTIATION

Raw data contained 13,447 gene objects represented on each microarray slide of which all had a signal greater than 100 pixels in 10/23 slides (10 being the minimum sample size). 4982 data points remained after filtering by 2-fold change (1898 up-regulated; 3329 down-regulated). This filtered dataset was statistically analysed as described in Section 7.5.1.2. The full list of significantly differentially expressed genes (188) is given in Appendix B, Table 2. Genes potentially up-regulated in either male (Figure 7.19) or female (Figure 7.20) *G. pulex* were identified. Additionally, genes putatively related to vitellogenin and haemocyanin according to BLAST homology, were investigated (Figure 7.21).

From the statistically relevant gene list, 4 genes that were significantly altered due to gender, were chosen for further study, although there were numerous other genes which, although were not significantly different, showed clear trends. There were 11 genes up-regulated in males and 10 genes up-regulated in females, only 1 gene being statistically relevant in each group. 11 genes involved in vitellogenin or haemocyanin formation showed up-regulation in females, 2 of which were statistically relevant and 5 of which were specific *G. pulex* vitellogenin fragments. The normalised intensities and fold changes of the specific *G. pulex* vitellogenin can bee seen in Figure 7.22 and Table 7.5.

Of the 4 genes significantly altered in expression due to gender, all 4 had a significant homology to known proteins according to BLASTX analysis. Of the genes of putative functions based on homology, the fold-changes shown in Table 7.4 were observed:

PUTATIVE FUNCTION	MALE	FEMALE
Up-regulated in males		
Non-SMC element 1 homologue	2.0	0.5
Up-regulated in females		
Chaoptin	0.4	3.3
Vitellogenin 2	0.5	2.1
Vitellogenin (*A. rosae*)	0.3	4.5

Table 7.4: Selection of statistically significant fold-changes in expression over the moult cycle of *G. pulex*

Fold-changes are calculated against a normalised average expression of 1.0

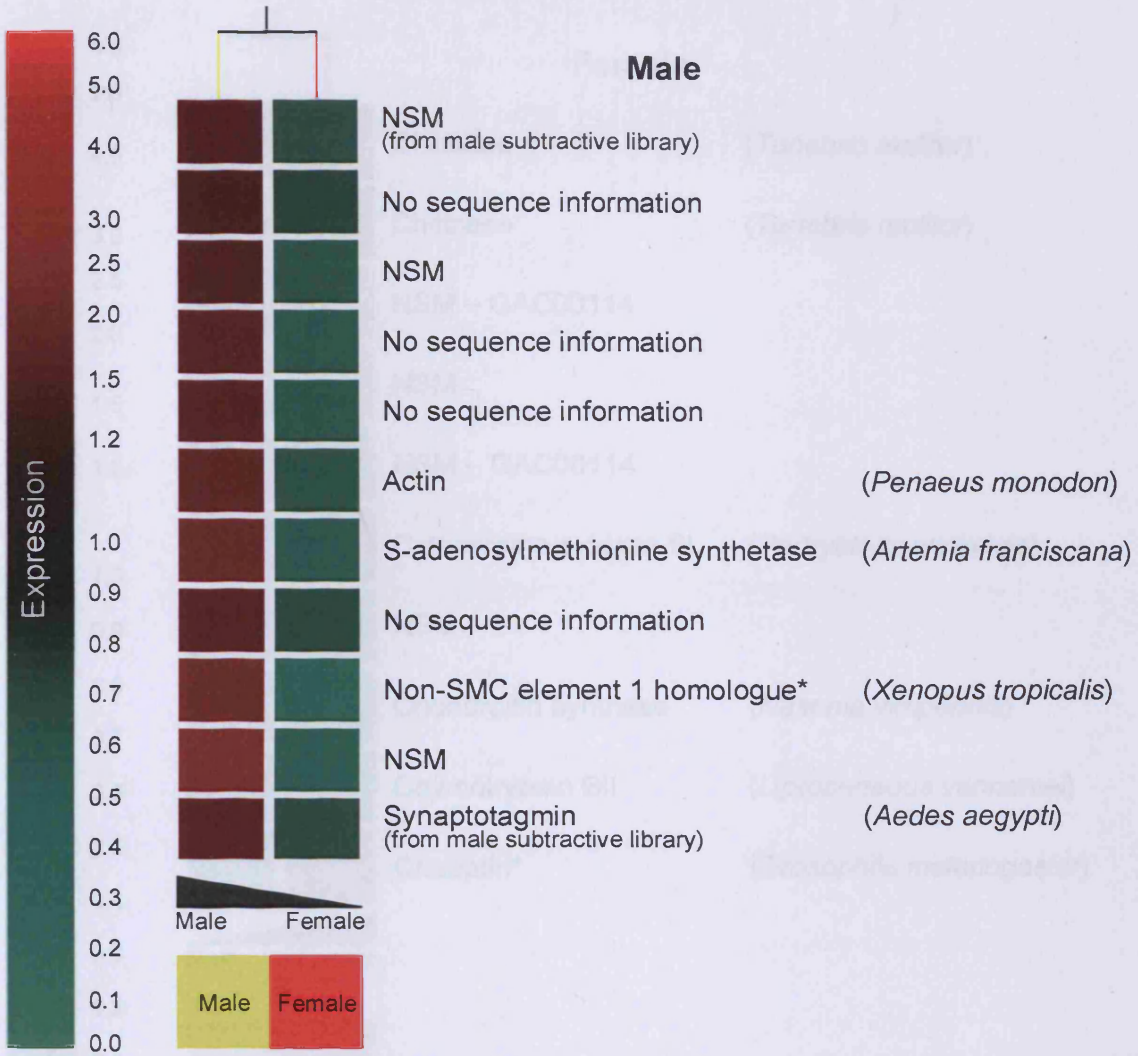


Figure 7.19: Genes potentially up-regulated in male *G. pulex*

* Significantly different gene expression. NSM = No Significant Match against BLAST database.

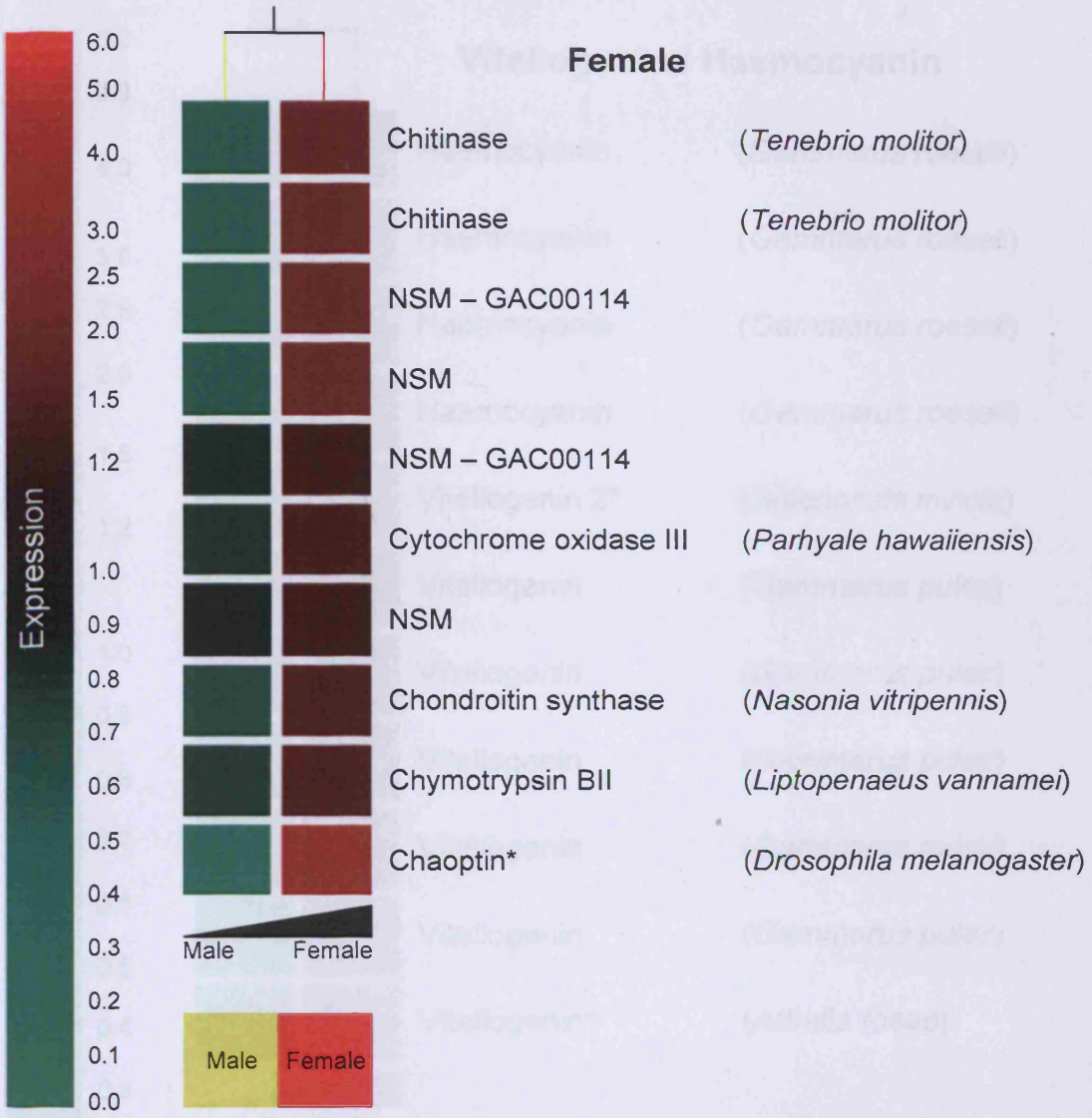


Figure 7.20: Genes potentially up-regulated in female *G. pulex*

* Significantly different gene expression. NSM = No Significant Match against BLAST database.

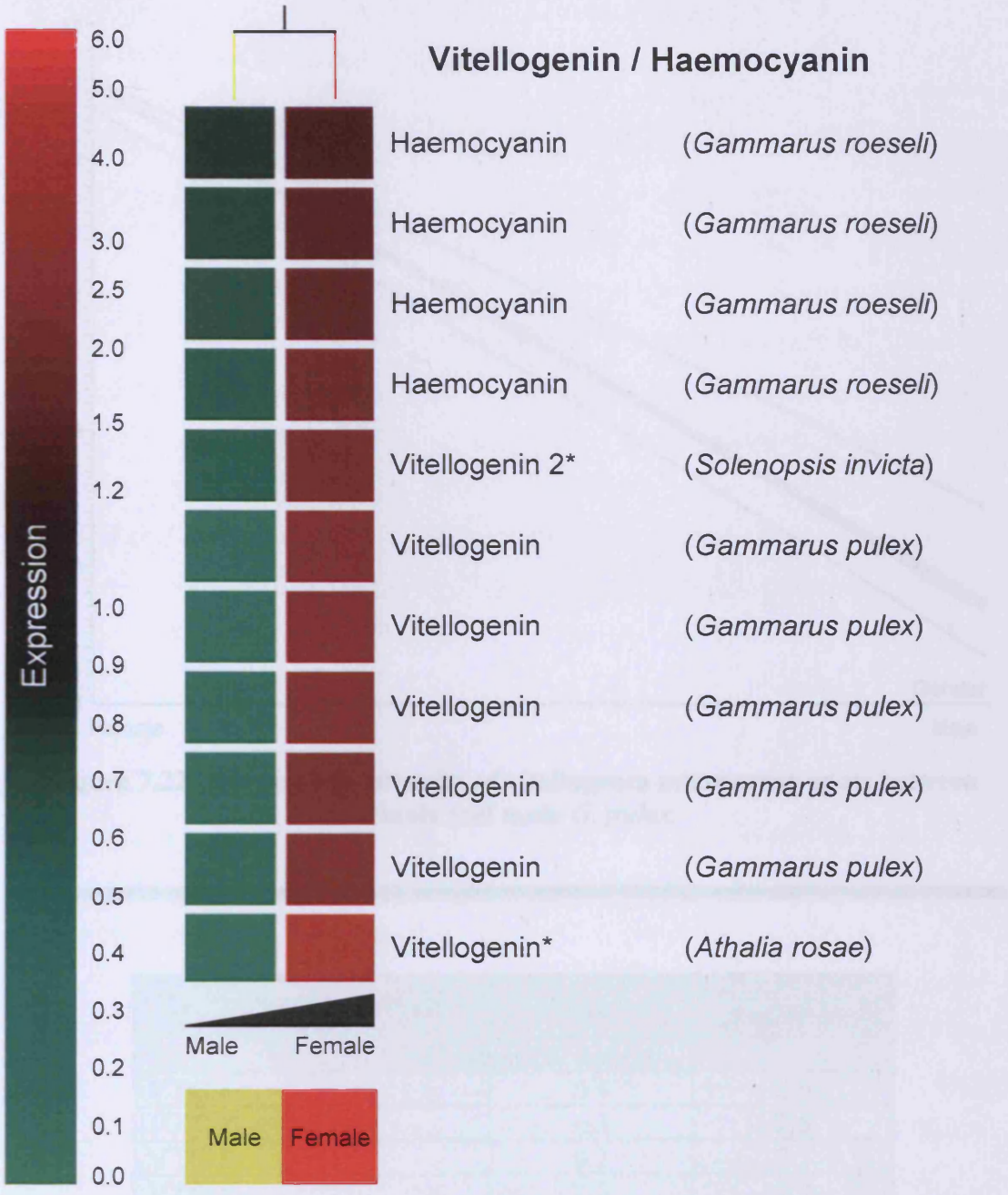


Figure 7.21: Differential expression in genes potentially involved in vitellogenin and haemocyanin formation in male and female *G. pulex*

* Significantly different gene expression

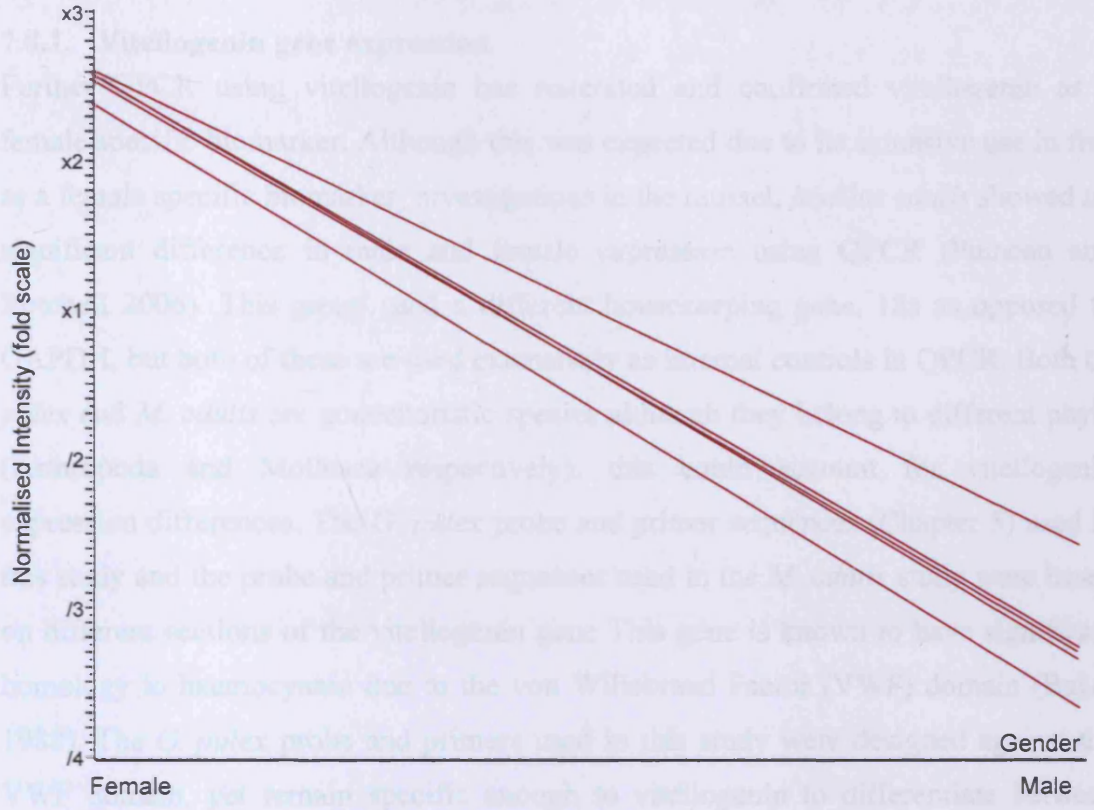


Figure 7.22: Normalised intensity of vitellogenin microarray spots between female and male *G. pulex*

G. PULEX GENE ID	MALE	FEMALE
Vitellogenin (G. pulex)		
VTG 6	0.4	2.6
VTG 7	0.3	2.5
VTG 8	0.3	2.5
VTG 9	0.3	2.3
VTG 10	0.3	2.5

Table 7.5: Fold-changes in *G. pulex* vitellogenin between males and females
Fold-changes are calculated against a normalised average expression of 1.0

7.8. DISCUSSION

7.8.1. Vitellogenin gene expression

Further QPCR using vitellogenin has reiterated and confirmed vitellogenin as a female specific biomarker. Although this was expected due to its extensive use in fish as a female specific biomarker, investigations in the mussel, *Mytilus edulis* showed no significant difference in male and female expression using QPCR (Puinean and Rotchell 2006). This group used a different housekeeping gene, 18s as opposed to GAPDH, but both of these are used extensively as internal controls in QPCR. Both *G. pulex* and *M. edulis* are gonochoristic species although they belong to different phyla (Arthropoda and Mollusca respectively), this could account for vitellogenin expression differences. The *G. pulex* probe and primer sequences (Chapter 5) used in this study and the probe and primer sequences used in the *M. edulis* study were based on different sections of the vitellogenin gene This gene is known to have significant homology to haemocyanin due to the von Willebrand Factor (VWF) domain (Baker 1988). The *G. pulex* probe and primers used in this study were designed against the VWF domain, yet remain specific enough to vitellogenin to differentiate between gender and the *M. edulis* probe and primers were designed to the Vitellogenin_N lipid binding domain, specific to vitellogenin. Despite *G. pulex* vitellogenin QPCR based on the VWD domain, the sequence is specific enough to allow gender specific expression to be identified in female *G. pulex*.

No changes were observed in vitellogenin expression over the moult cycle; this could be due to the possibility of two vitellogenic periods, in intermoult (C) and premoult (D_3/D_4). This was suggested after a study on the isopod *Oniscus asellus*, which was revealed to have different primary vitellogenic periods depending on whether the female was in a first maternal (occurring in premoult) or a subsequent maternal moult cycle (occurring in intermoult). Secondary vitellogenesis occurred in premoult in all maternal moults (Steel 1980). As vitellogenin was assessed across a mixed population, this may have been mis-represented as constant expression across the moult stages, although the females had been in isolation for an entire moult cycle, so they should all have been in a non-breeding state due to the absence of males.

In addition to QPCR, an expensive and laborious approach, semi-quantitative PCR has been optimised for detecting vitellogenin expression levels in *G. pulex*. This allows a visual and rapid identification of gender, although this technique is not ideal for large numbers of animals due to the number of samples generated (5 per animal). These are then separated in individual lanes of an agarose gel by electrophoresis, a time consuming process when large numbers of samples are involved. However, it has the potential to allow the rapid identification of a population bias by using pooled samples, although this approach has not been tested or validated.

7.8.2. Expression levels of gene encoding Cuticle protein

Cuticle gene expression has been confirmed as a marker of moult stage for *G. pulex* in Stage 1 (C_2) where a distinct up-regulation is observed. Animals in late intermoult/early premoult to ecdysis cannot be differentiated between. Additionally there appears to be no significant differences between cuticle expression in males and females throughout the moult cycle. This is not entirely unexpected as the physical processes of ecdysis are similar in males and females and cuticle protein is involved in exoskeletal formation, a process common in both genders.

Moult stage was assessed by recording the moult cycle length and calculating the collection time points as previously described. Alternative methods to identify moult stage are available including using microscopy to assess setae formation in appendages, however expert skills are required to ascertain moult stage efficiently and accurately using this method. Due to the relatively crude method of moult staging, animals were assessed by size and this measurement compared to time between two ecdysis events (moult stage length). If the moult stage length was being correctly recorded, a rough trend of increasing time with increasing size was expected and observed. The animal with the shortest moult cycle length (a female), at 23 days was also the smallest; the largest animal (a male) had the longest moult cycle length at 45 days. Therefore a correlation between amphipod size and moult cycle length was recorded.

7.8.3. *G. pulex* microarray analysis

Microarrays allow changes in expression of thousands of genes to be analysed simultaneously, providing an excellent tool for global gene expression profiling. MA plots and calibration curves have confirmed the success of the profiling process. Volcano plots have shown that differentially expressed genes were identified in all datasets. The vast quantity of data produced by gene expression profiling presents a complex task to identify individual potential biomarkers, so PCA was performed to reduce dimensionality and statistical analysis was used to isolate the best potential biomarkers of moult stage and gender.

7.8.4. *G. pulex* potential biomarkers of Moult Stage

These *G. pulex* studies revealed 6928 potential biomarkers of moult stage that were up- or down-regulated, however statistical analysis reduced this number to 199 genes. Of these genes a selection of 7 were hand picked for further analysis which showed the most consistent expression patterns.

Of particular interest amongst the statistically relevant genes was chitinase (e = 0.16) which was up-regulated in Stage 1 by a 2.2x fold change decreasing to 1.5x in Stage 2 and 0.8x in Stage 3. Homology was to chitinase from the yellow mealworm beetle, *Tenebrio molitor*, in which this particular isoform of chitinase is induced by 2-hydroxyecdysone and is actively involved in cuticle digestion during moulting (Royer *et al.* 2002). It is therefore unexpected that chitinase would be up-regulated in Stage 1, this may be due to the low probability (0.16) associated with the homology, although there was a 50% homology over a 70 amino acid stretch, this up-regulated gene may not be chitinase as characterised in *T. molitor*, but a novel chitinase or similar protein with an alternative function. Due to the up-regulation of chitinase gene expression in *G. pulex* females as well as in Stage 1 mixed adults, the function of this chitinase may be related to oviduct-specific glycoprotein, a member of the human chitinase protein family (Agarwal *et al.* 2002). Further sequence analysis of the original clone is required to further identify this potential Stage 1 biomarker.

The largest gene ontological annotation involved in the moult cycle was carbohydrate metabolic process with 29% of total annotations compared to 6.3% from cluster GO

annotations (Chapter 3, Figure 3.32). The most consistently altered gene, up-regulated in stage 1 animals, had putative glycosyl hydrolase activity and was most homologous to the larval glue protein of *Drosophila virilis* (e = 0.004). In *D. virilis*, this gene contains ecdysone receptor binding sites and transcript expression is observed to decrease in 3rd late instar larvae. It is related to the 68C glue protein family of *D. melanogaster* (Swida 1990). *Drosophila* sp. glue proteins are secreted from the salivary glands and are used to fix the pupal case to a substrate (Fraenkel and Brookes 1953). Glycosyl hydrolases aid in carbohydrate digestion by hydrolysing the glycosidic bonds between carbohydrates. It is therefore possible that this gene may encode for a protein involved in the creation and stabilisation of the new cuticle in the same way that the pupal case is fixed and stabilised to a substrate. Glycosyl hydrolase activity, indicative of cellulases may be linked to the suggestion that *G. pulex* consume the shed exuviate to increase calcium levels (Wright 1980). This may also explain why there are 2 cellulases identified as up-regulated in Stage 1, homologous to cellulases from *Cherax quadricarinatus* and *Lentinula edodes*.

Also involved in carbohydrate metabolism is α-amylase homologous to that of *Drosophila erecta*, and is found to be up-regulated in Stage 2 of the *G. pulex* moult cycle. α-amylase is a digestive enzyme which is involved in starch metabolism. Studies in the Pacific white shrimp, *Litopenaeus vannamei*, revealed specific activity of α-amylase to be highest in Stage C (low salinity) and D_0 (high salinity) (Gaxiola *et al.* 2005) which is also indicated by Stage 2 expression in *G. pulex* corresponding to late C_4 / early D_0.

Reproduction and development accounted for 11% of annotated GOs compared to 3.2% annotations for clustered sequences (Chapter 3, Figure 3.32). Of key interest is the sensory appendage protein homologous to the functional protein of *Manduca sexta* ($e = 2e^{-7}$), up-regulated in Stage 1 of *G. pulex*. The expression of this gene has been identified in the antennae of *M. sexta* moths and is thought to be involved in odourant transport (Robertson *et al.* 1999). The gene is classified under 'reproduction and development' due to its similarity to an insect pheromone-binding protein (InterPro: IPR005055) and GO annotations of gamete generation and embryonic development based on sequence homology (annotated by blast2GO, (Conesa *et al.* 2005)). Also annotated under 'reproduction and development' was piopio from the

red flour beetle, *Tribolium castaneum* up-regulated in Stage 1 of *G. pulex*. In *T. castaneum*, piopio contains a zona pellucida domain; proteins containing this domain are responsible for sperm-adhesion to the zona pellucida, the membrane surrounding oocytes consisting of glycoproteins. Zona pellucida domains can also be found on multidomain transmembrane proteins such as glycoproteins, uromodulin and betaglycan (SMART (Simple Modular Architecture Research Tool) description, EMBL, Heidelberg, Germany). Up-regulation in Stage 1 of the *G. pulex* moult cycle maybe due to the pre-vitellogenic growth of oocytes during brood incubation. It has been observed in the isopod, *Oniscus asellus*, that pre-vitellogenic oocytes are retained in the ovaries until reproduction is triggered, at which point the mature eggs are deposited into the recently formed brood pouch and incubated until the neonates are released (Steel 1980). Egg deposition into the brood pouch immediately after ecdysis, triggers the formation of new pre-vitellogenic oocytes which may require the production of glycoproteins containing a zona pellucida domain to form the egg membrane.

Genes involved in cuticle or chorion formation were mostly up-regulated in Stage 1 in *G. pulex*, a stage associated with tissue growth and calcification (Chan *et al.* 1988), it is therefore expected that there will be an up-regulation in genes involved in cuticle formation, calcium binding and stabilisation. One gene was up-regulated in Stage 3; this was a different sequence to the 5 up-regulated in Stage 1, encoding a putative pupal cuticle protein, which, although not relevant to *G. pulex* adults, may be a different isoform of a cuticle related protein involved in the pre-ecdysis secretion of the epicuticle and exocuticle described in Chapter 1, Section 1.1.3.

7.8.5. *G. pulex* potential biomarkers of gender

These *G. pulex* studies revealed 4982 potential biomarkers of gender that were up- or down-regulated, however statistical analysis reduced this number to 188 genes. Of these genes a selection of 4 were hand picked for further analysis which showed the most consistent expression patterns.

The one statistically relevant gene of interest up-regulated in male *G. pulex* was from the mixed library and homologous to non-SMC element 1 homologue for *Xenopus*

tropicalis ($e = 4e^{-20}$). In *X. tropicalis*, the translated protein contains a domain that is found in subunits of the non-structural maintenance of chromosomes (SMC) complex involved in genomic stability though DNA repair and DNA metabolism (Fujioka *et al.* 2002) and has also been linked to spermatogenesis in medaka fish (Iwai *et al.* 2004).

The most significantly up-regulated gene in females, besides vitellogenin, was homologous to chaoptin from *D. melanogaster* ($e = 4e^{-4}$), a cell surface glycoprotein involved in photoreceptor morphogenesis (Reinke *et al.* 1988). Additionally, as expected, fragments of the vitellogenin gene were up-regulated in females by 2.2x to 4.5x and down-regulated in males by 0.3x to 0.5x in sequences homologous to the red imported fire ant, *Solenopsis invicta* and the sawfly, *Athalia rosae*. Short fragments of confirmed *G. pulex* vitellogenin gene (containing the same section amplified by QPCR) showed similar patterns of expression with female expression up-regulated by ~2.5x and males down-regulated by ~0.3x. As expected, haemocyanin also appeared as up-regulated in females, although to a lesser extent as vitellogenin. This is most likely due to the sequence similarity between the von Willebrand domains contained in both genes. This highlights the problems of false positives using a platform such as DNA microarrays and highlights the importance of further sequence analysis and subsequent QPCR validation.

The interpretation of gene profile data was inhibited by insufficient sequence information for amphipods and the number of uncharacterised proteins, leading to poor GO annotation. This discussion is a brief overview of some genes of interest; however there are a large number of genes identified that may be potential biomarkers of gender and moult stage. Further analysis of the remaining genes of interest shown in Appendix B, Tables 1 and 2 may lead to further understanding of the molecular mechanisms involved in sexual differentiation and moulting.

CHAPTER 8

GENE EXPRESSION CHANGES IN DIFFERENT DEVELOPMENTAL STAGES OF *GAMMARUS PULEX*

8.1. INTRODUCTION

In the amphipod, *Gammarus pulex*, fertilisation of eggs occurs shortly after female moulting following a period of precopulatory pairing with a male. The fertilised eggs are retained in the brood pouch of the female during embryonic maturation until the neonates are fully developed and independent (Figure 8.1). They are released from the brood pouch after approximately 25 days at 17°C (personal observation). Neonates develop and grow through successive moulting, achieving adult maturity after approximately 5 moults (Gross *et al.* 2001). Further details can be found in Chapter 1, Section 1.1.

In summer (water temperatures between 10 and 15°C), *G. pulex* take 3-4 months to mature and in winter (water temperatures between 5 and 10°C), about 7 months (Hynes 1955). Welton and Clarke (1980) were able to clearly distinguish gender at a length of 4.5mm for females and set this size as the division point of juveniles and immature *G. pulex*. Mature females were distinguished by the presence of eggs. They were also able to establish onset of male maturity at 7.5mm and used this figure as a division point for immature and mature males as this was the minimum size male found in precopula. Welton and Clarke (1980) observed that ovigerous females were present all year round and had fully developed brood plates at 6.3 to7.2mm but most did not bear eggs until they reached between 7.3mm and 8.3mm in size, which is in agreement with Hynes (1955).

Figure 8.1: *G. pulex* in a precopulatory pairing with a *G. pulex* neonate
Photograph illustrating the comparative size of mature adults and newly released neonates. It was observed that as brood maturation time was similar to moult cycle length in females, *G. pulex* adults were often in a precopulatory position when neonates were released from the female brood pouch. Whether this is due to the lack of females with empty brood pouches or whether males are unable to detect females are gravid is not known. This behaviour has also been noted in a study by McCahon and Pascoe (1988).

8.1.1. Changes in gene expression over normal development

In the tobacco hornworm, *Manduca sexta*, there is developmental variation in the gene expression of the Halloween shade gene (Mss*hd*) which encodes ecdysone 20-monooxygenase, an enzyme that converts ecdysone into 20-hydroxyecdysone. Peak expression levels were observed in the fat body and midgut during the fifth instar and the beginning of pupal–adult development (Rewitz *et al.* 2006). The developmental gene expression of two isoforms of ecdysteroid receptor was recorded in the Mediterranean fruit fly, *Ceratitis capitata*. It was observed that both isoforms were present in embryos, apart from in the first few hours of embryogenesis. Two peaks were observed during metamorphosis, at larval–prepupal transition and during the second half of prepupal development, concurrently with 20-hydroxyecdysone peaks (Verras *et al.* 2002). In *Drosophila melanogaster*, no ecdysteroid receptors are present

in larval neurons, however, a high level of isoform EcR-B1 is present in larvae undergoing metamorphosis eventually switching to EcR-A isoform expression on transition form pupa to adult (Truman *et al.* 1994).

In the spruce budworm, *Choristoneura fumiferana*, developmental expression of glutathione-S-transferase (GST) was examined. Peak expression levels were observed in first instar larvae and diapausing second instar larvae, decreasing until the sixth instar but undetectable in pupae (Feng *et al.* 2001). Ultraspiracle gene (*usp*) expression in *C. fumiferana* and *D. melanogaster* is present in all embryonic, larval and pupal stages, peaking in sixth instar larvae for *C. fumiferana* and in the third late instar in *D. melanogaster* (Henrich *et al.* 1994; Perera *et al.* 1998). A developmentally regulated putative haemocyanin subunit has been discovered in the crab, *Cancer magister*. It is present in adults, but not in juveniles until the sixth instar (Durstewitz and Terwilliger 1997).

8.1.2. Effects of endocrine disrupting chemicals on normal development

The effects of endocrine disrupting chemicals (EDCs) on development in a wide range of species was extensively reviewed by Colborn (1993). Since that review, further research has been undertaken to determine the processes and pathways involved in order to understand the mechanisms of disruption after transgenerational and juvenile EDC exposure. Reproduction and development are the key processes known to be affected by EDC exposure, yet the mode of action and mechanisms of disruption remain unknown in many species. The importance of studying the effects of life cycle exposures due to the sensitivity of the developmental stages and limitations of invertebrate endocrinology, have been consistently reiterated (Segner *et al.* 2003; Kusk and Wollenberger 2007).

Exposure to 17α-ethinylestradiol, p-octylphenol, and tamoxifen was found to potently inhibit naupliar development in the copepod *Acartia tonsa* (Andersen *et al.* 2001). Additionally, benzo(a)pyrene (BaP), 4-nonylphenol (NP) and di(ethyl-hexyl)-phthalate (DEHP) were found to inhibit naupliar development in the copepod *Eurytemora affinis* (Forget-Leray *et al.* 2005). An ecotoxicological test procedure on the apple snail, *Marisa cornuarietis*, to examine the effects of EDCs on embryos has

been developed. Increased weight after hatching and a decreased heart rate were observed after exposure to bisphenol A and 17α-ethinylestradiol (Schirling *et al.* 2006). Degeneration of the epithelial cells, reduction in the number of veins, precocious cuticle formation and inhibition of growth of normal wing scales was evident in pupae of the spruce budworm, *Choristoneura fumiferana*, after exposure to tebufenozide. It was also found to induce the gene expression of hormone receptor 3 (CHR3) (Sundaram *et al.* 2002).

Studies on the effects of the fungicide, propiconazole on daphnid embryonic development showed interference in normal development through maternal exposure. Exposure elicited effects included underdeveloped second antennae, curved tail spine or a poorly formed carapace (Kast-Hutcheson *et al.* 2001). Similar observations occurred after exposure of *Daphnia magna* to testosterone; neonates produced by exposed females showed developmental abnormalities ranging from minor abnormalities of the carapace and antennae to severely developmentally arrested individuals (Mu and LeBlanc 2002).

Analysing changes between transcriptome profiles during developmental maturation in crustaceans will increase our understanding of the differential gene expression in response to life cycle changes.

8.2. SAMPLE COLLECTION

Amphipods were collected from a site near Cynrig hatchery (Environment Agency site; Chapter 7, Figure 7.3) near Llanfrynach, Powys by kick sampling as described in Chapter 2, Section 2.9.1. Further details of amphipod handling are described in Chapter 7, Section 7.2.

8.2.1. Developmental stage determination

For collection of neonates, paired animals were placed in containers with 80ml of deionised and dechlorinated tap water. An abundance of food (conditioned alder leaves) was provided for the mating pair; if there was not sufficient food available it was a common occurrence for the male to consume the female in her vulnerable post-moult state as also observed by Sexton (1928). Containers were checked daily; after separation of the male from the female, the female was placed in a separate container and regularly monitored for the presence of neonates (usually evident after ~25 days). Additionally, visibly gravid females (showing orange/pink discolouration in the area around the oostegites) from the mixed population were also placed in containers and monitored in the same way. Neonates were sacrificed within 7 days of birth and were between 1mm and 2mm in length.

Juveniles were hand selected from the mixed population by size; small juveniles were chosen at lengths between 3mm and 4mm, medium juveniles were between 5mm and 6mm. A random sample of amphipods were preserved in ethanol and measured to ensure length boundaries were being adhered to (Figure 8.2).

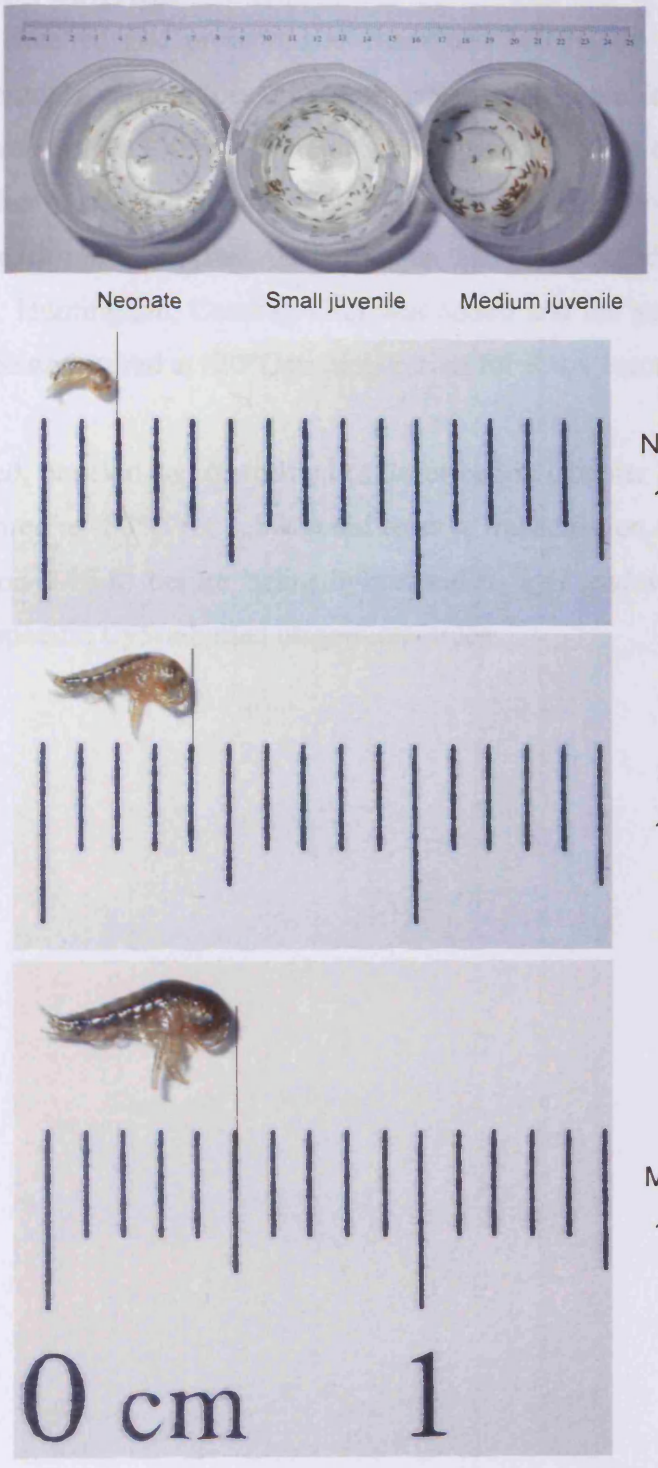


Figure 8.2: Different size *G. pulex* used in developmental gene expression profiling

8.3. SAMPLE PREPARATION

Amphipods were washed and preserved as described in Chapter 7, Section 7.4.1, except multiple animals were processed simultaneously and split into batches based on approximate wet weight (50mg). Neonates and juveniles were dried briefly after washing using filter paper, lifted off using a fine spatula and dropped into liquid nitrogen in a small mortar. After the nitrogen had evaporated, RNA*later*®ICE (~0.5ml; Ambion, Huntingdon, Cambs., UK) was added and the samples transferred to a microfuge tube and stored at -20°C in preparation for RNA extraction.

RNA was extracted, purified and quantified as described in Chapter 2, Sections 2.13.6 and 2.13.7 and stored at -80°C for subsequent reverse transcription and Cy3 labelling (Chapter 2, Section 2.13.8) before being hybridised to a *G. pulex* microarray slide alongside vector-specific Cy5-labelled oligonucleotides.

8.4. MICROARRAY ANALYSIS

Sample collection and storage is described in Sections 8.2 and 8.3. Before labelling, RNA was extracted as described in Chapter 2, Section 2.13.6 without the optional LiCl precipitation. Samples passing quality control (Chapter 2, Section 2.13.7) were labelled (Chapter 2, Section 2.13.8) using a 4:1:5 U:T:ACG ratio of dNTPs and hybridised (Chapter 2, Section 2.13.9) onto *G. pulex* microarray slides (Chapter 2, Section 2.13.3). Slides were scanned (Chapter 2, Section 2.13.10) and data stored as 2 (Cy3 and Cy5) TIFF (Tagged Image File Format) images.

8.4.1. Bioinformatics and Microarray interpretation

Images were analysed using SPOT (Version 2.0, CSIRO Mathematical and Information Sciences, Australia (Yang *et al.* 2001; Yang *et al.* 2002)) and data imported into GeneSpring ((Version GX 7.3; Agilent Technologies UK Ltd., Cheshire, UK) not including the control and blank data) as described in Chapter 2, Sections 2.13.11 and 2.13.12. Poor slides (determined from box plots) were removed from the dataset and the filtered data normalised as described in Chapter 2, Section 2.13.12.

8.4.1.1. Microarray quality validation

An MA plot was generated to demonstrate the raw and normalised filtered data of the experimental samples. An MA plot is defined as the log ratio vs. log median intensity where:

$$M = \log_2\left(\frac{Cy5}{Cy3}\right)$$

$$A = \frac{(\log_2 Cy5 + \log_2 Cy3)}{2}$$

The MA of raw data (Figure 8.3A and Figure 8.3C) shows the intensity dependent ratio and demonstrates spot artefacts and systematic variation. The ideal MA plot should have M values that are evenly distributed at 0 across the range of intensities (Dudoit *et al.* 2002). Raw data was normalised to remove systematic bias from

experimental artefacts as described in Chapter 2, Section 2.13.12. The MA plot demonstrates how the filtering and normalisation process created a data set distributed around M=0 (Figure 8.3B and Figure 8.3D).

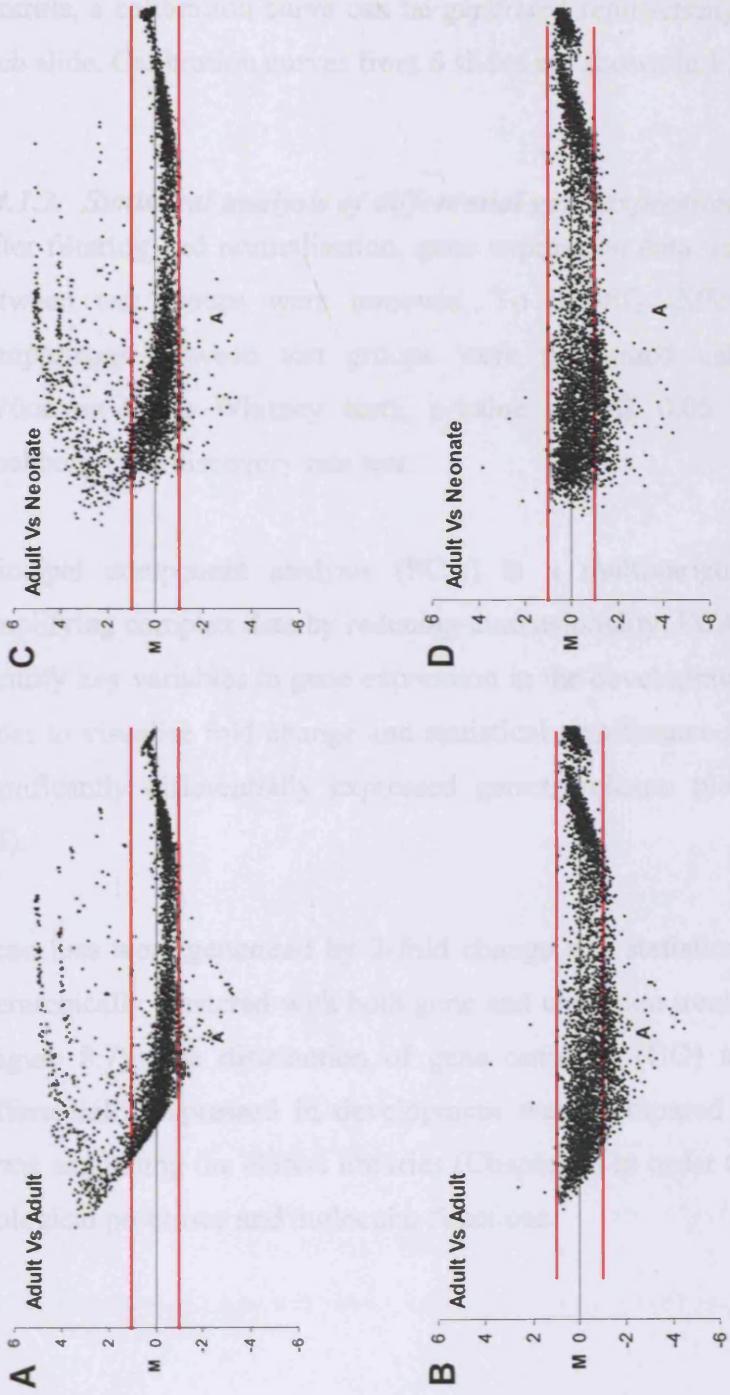


Figure 8.3: MA plots of developmental *G. pulex* microarray data before and after normalisation
A: Adult vs. Adult before normalisation; B: Adult vs. Adult after normalisation; C: Adult vs. Neonate before normalisation; D: Adult vs. Neonate after normalisation.

Internal synthetic control mRNA (Universal ScoreCard, GE Healthcare, Little Chalfont, Bucks., UK) was added to each RNA sample and reverse transcribed and indirectly incorporated with Cy3 simultaneously with the RNA. These control probes bind to complementary target sequences printed across the length and width of the microarray slides alongside the *G. pulex* DNA targets. Using the ratio ScoreCard controls, a calibration curve can be generated representing the range of detection for each slide. Calibration curves from 6 slides are shown in Figure 8.4.

8.4.1.2. Statistical analysis of differential gene expression

After filtering and normalisation, gene expression data that changed less than 2-fold between test groups were removed. To identify differentially expressed genes, comparisons between test groups were performed using a non-parametric test (Wilcoxon–Mann–Whitney test), p-value cut-off 0.05 using the Benjamini and Hochberg false discovery rate test.

Principal component analysis (PCA) is a multivariate statistical technique for simplifying complex data by reducing dimensionality. PCA analysis was performed to identify key variables in gene expression in the developmental dataset (Figure 8.5). In order to visualise fold change and statistical significance simultaneously and identify significantly differentially expressed genes, volcano plots were generated (Figure 8.6).

Gene lists were generated by 2-fold change and statistical analysis (see above) and hierarchically clustered with both gene and condition trees using a distance algorithm (Figure 8.7). The distribution of gene ontology (GO) terms in the lists of genes differentially expressed in development were compared to the distribution of GO terms annotating the cDNA libraries (Chapter 3) in order to identify over-represented biological processes and molecular functions.

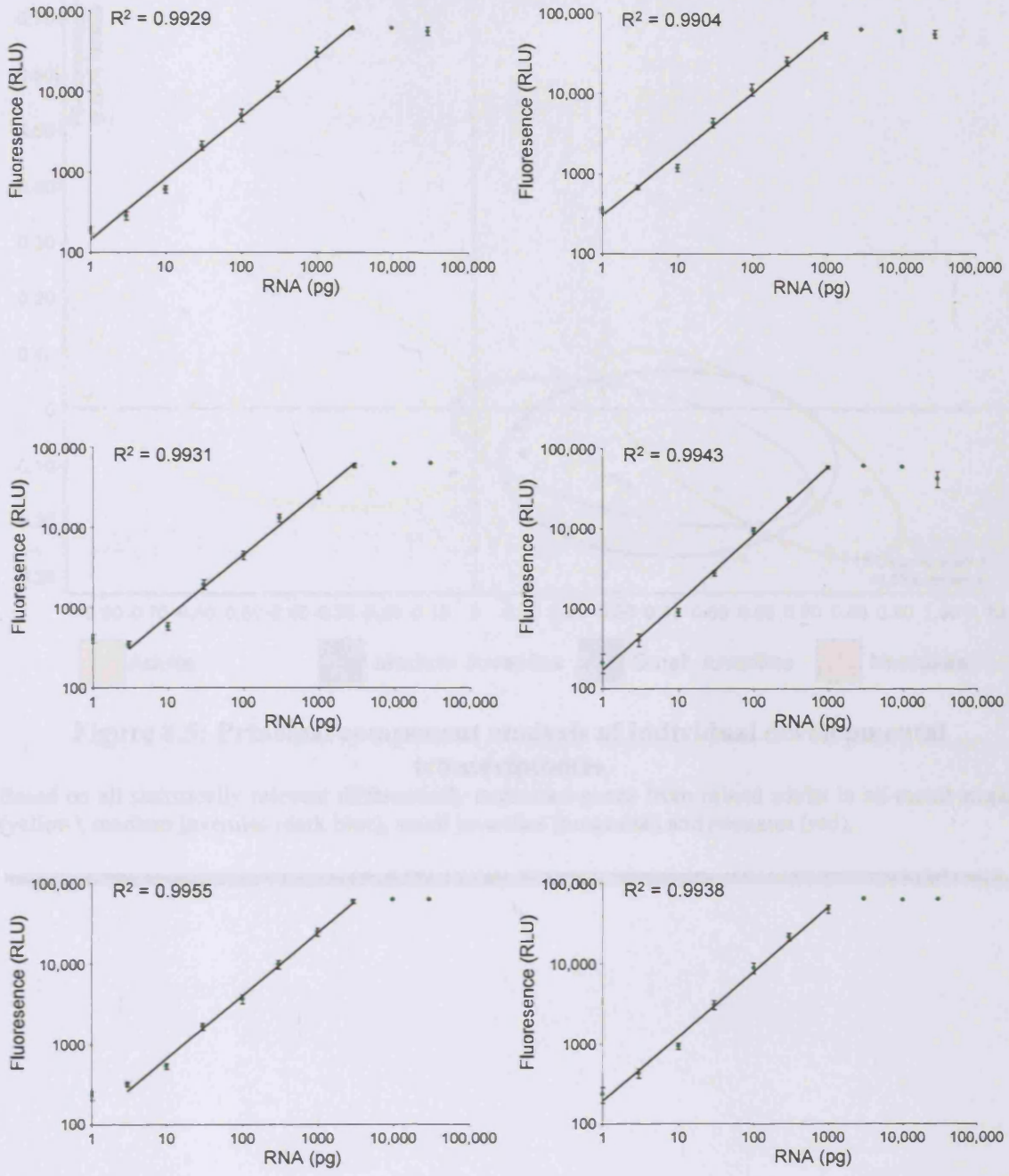


Figure 8.4: Universal ScoreCard calibration of 6 slides used to generate the Developmental dataset

Each ratio control was printed 10 times per slide; the mean signal intensity was used to generate calibration curves which are displayed with standard error bars.

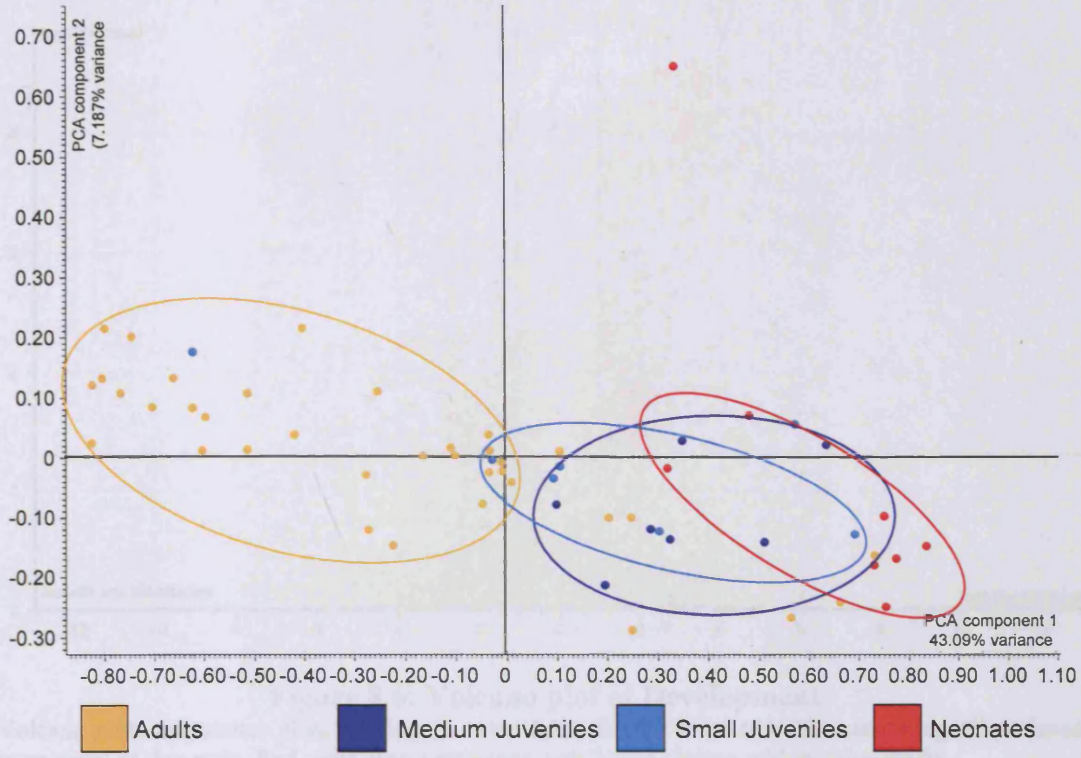


Figure 8.5: Principal component analysis of individual developmental transcriptomes

Based on all statistically relevant differentially expressed genes from mixed adults in all moult stages (yellow), medium juveniles (dark blue), small juveniles (turquoise) and neonates (red).

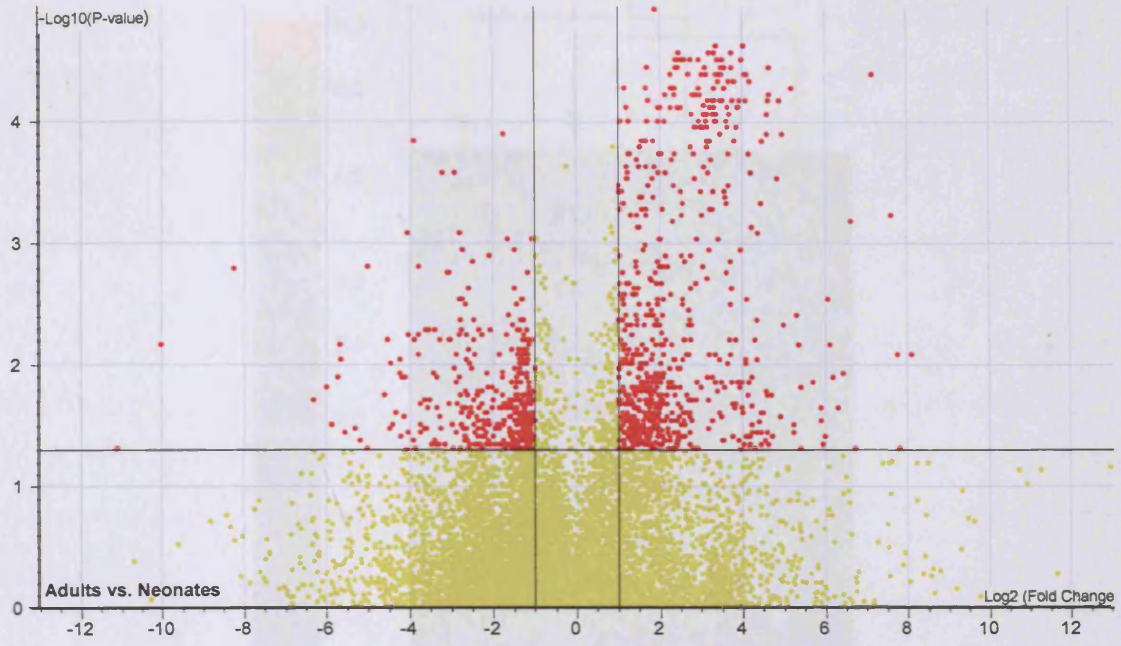


Figure 8.6: Volcano plot of Development

Volcano plots are scatter plots of –log (p-value 0.05) from 1-way ANOVA, assuming all variances were equal vs. log ratio. Red spots represent genes with 2-fold change with p-value <0.05.

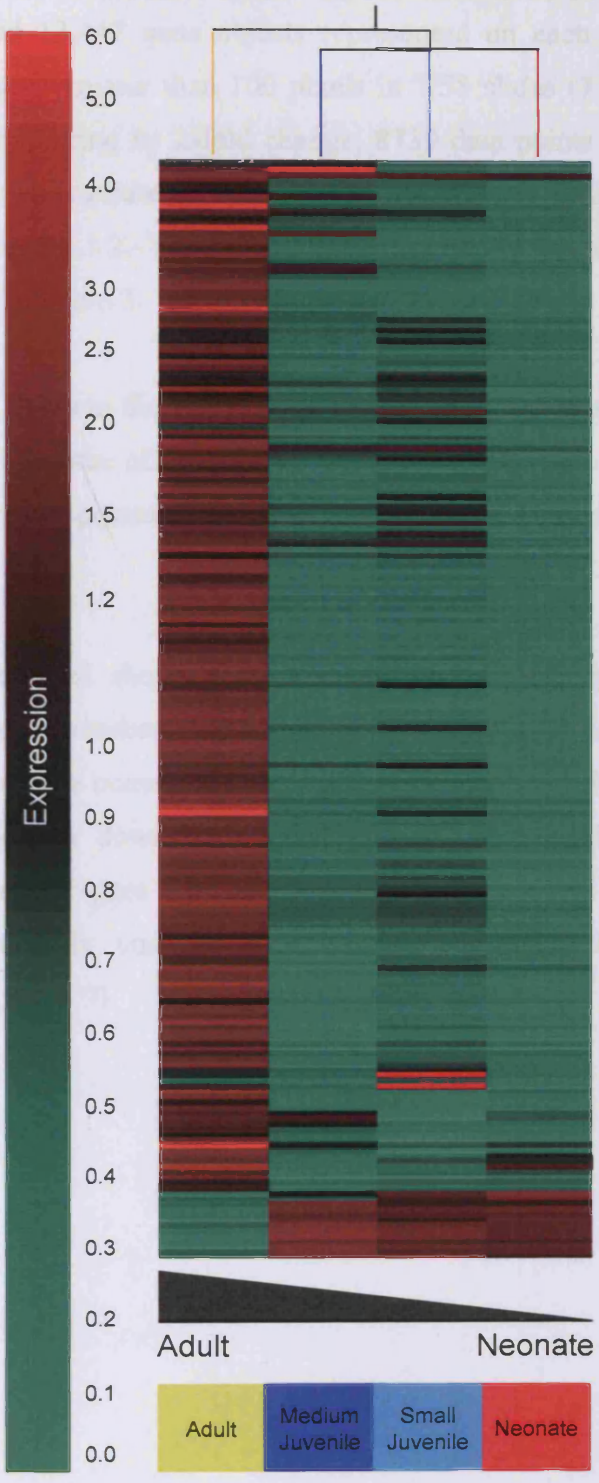


Figure 8.7: Global gene expression profiles for Developmental stage in *G. pulex*
Hierarchical tree of genes and heat map showing differential expression between the 4 life stages. The heat map columns left to right are: Mixed adults, Medium juveniles, Small juveniles and Neonates. Clustering was calculated using 'Spearman correlation' algorithm. Expression is average log signal intensity.

8.5. POTENTIAL BIOMARKERS OF DEVELOPMENTAL STAGE

Raw data contained 13,447 gene objects represented on each microarray slide of which all had a signal greater than 100 pixels in 7/58 slides (7 being the minimum sample size). After filtering by 2-fold change, 8739 data points remained (3090 up-regulated; 7301 down-regulated). This filtered dataset was statistically analysed as described in Section 8.4.1.2. The full list of differentially expressed genes (205) is given in Appendix B, Table 3.

Distribution of GO terms in the list of genes differentially expressed in developmental stages was compared to the cDNA library GO term distribution. There was no over-representation; therefore potential biomarkers were selected based on expression trend consistency.

13 genes were identified showing clear trends that potentially could be used as developmental stage biomarkers and were significantly altered during developmental stage although there were numerous others which, although not significantly different, showed clear trends. A downwards trend through from adults to neonates was observed in 19 genes (Figure 8.8) and 9 genes showing down-regulation in adults (Figure 8.9). Additionally, changes in vitellogenin expression through development were identified (Figure 8.9).

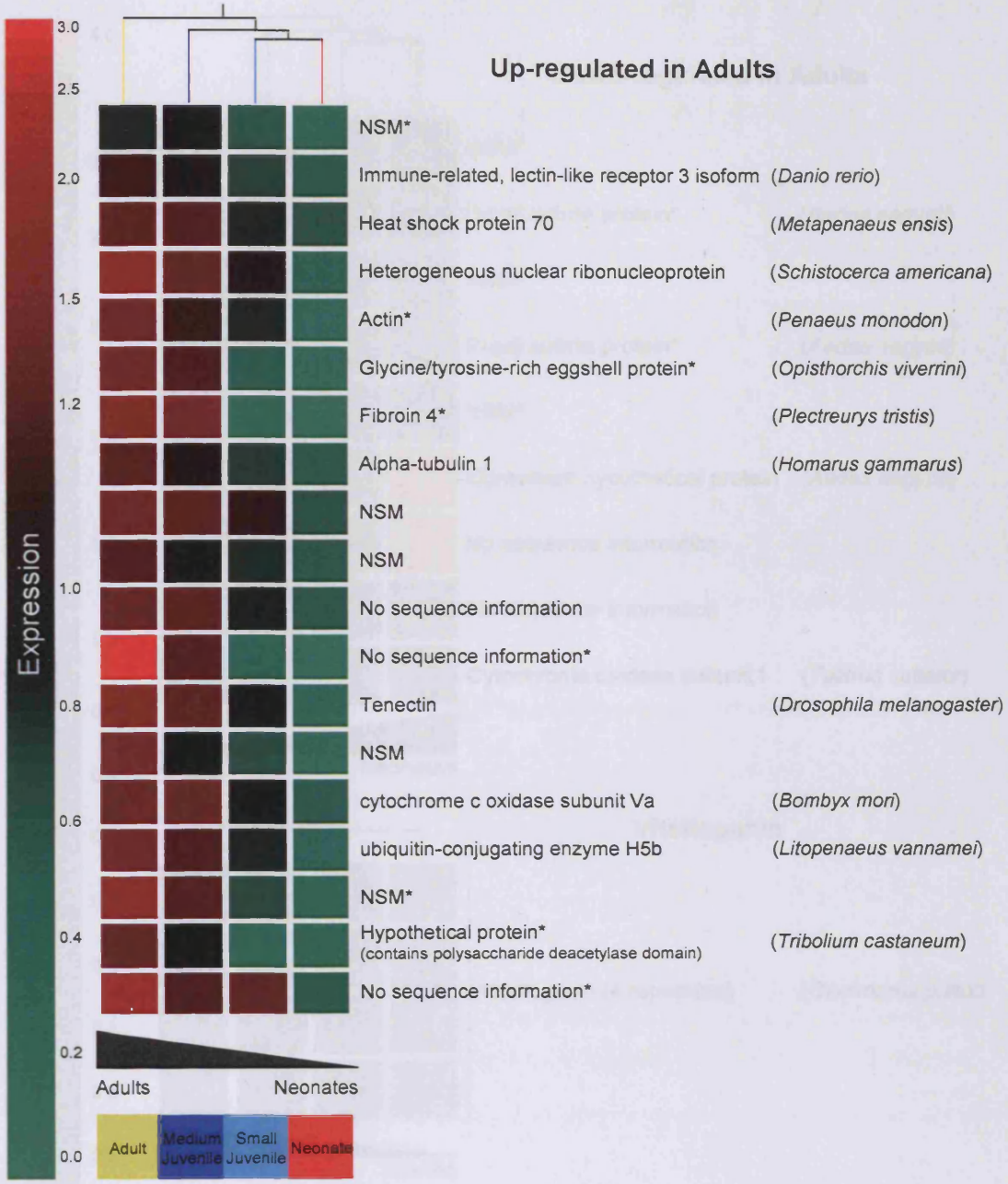


Figure 8.8: Genes potentially up-regulated in adults, increasing in expression through maturity

*Significantly different gene expression. NSM = No Significant Match against BLAST database. No sequence information describes PCR products that failed a sequence quality test (phred), this does not necessarily infer an inferior DNA target, but suggests a failure in the sequencing process due to the nature of high-throughput techniques.

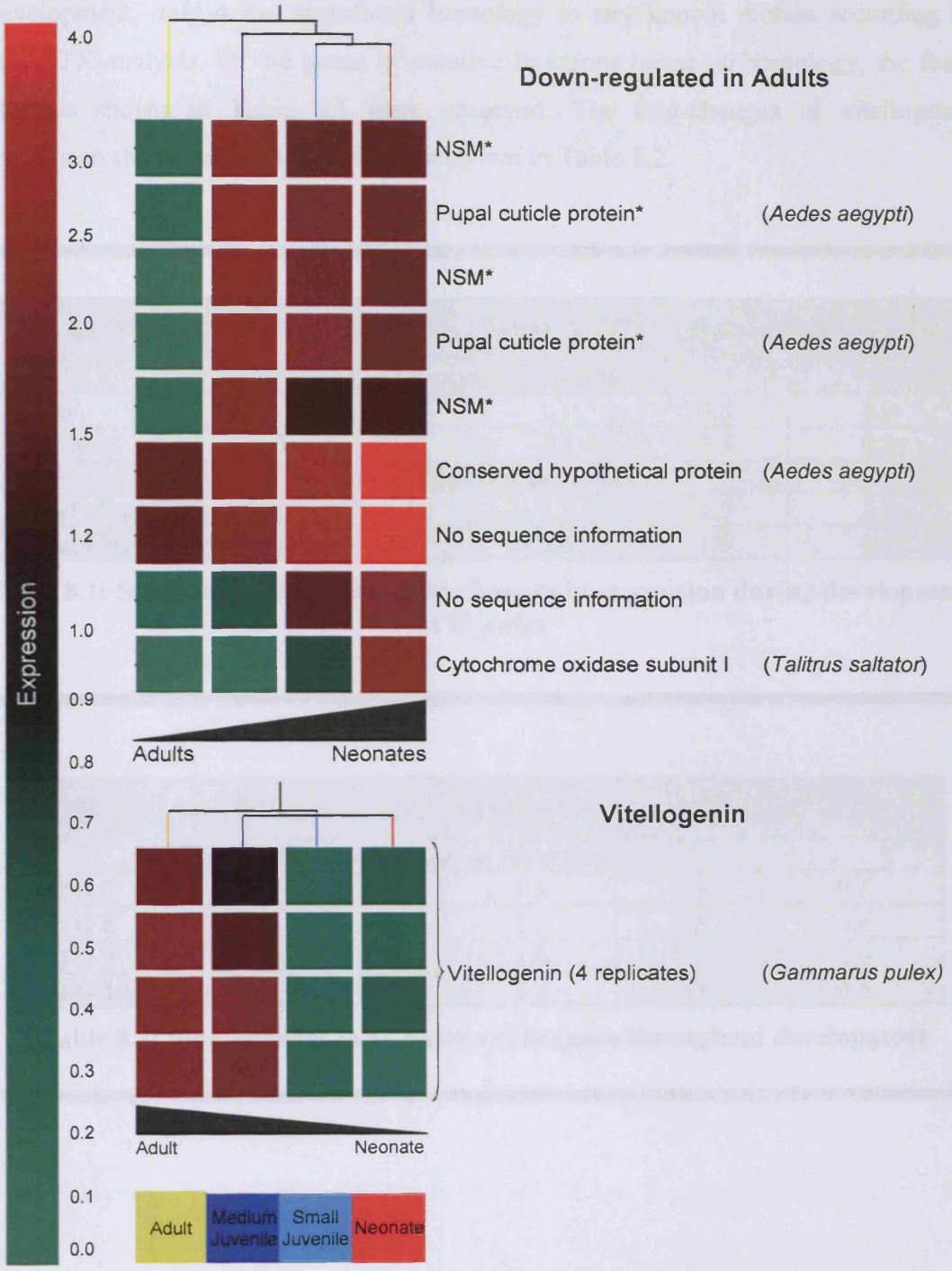


Figure 8.9: Genes potentially down-regulated in adults and changes in vitellogenin gene expression through development

*Significantly different gene expression. NSM = No Significant Match against BLAST database. No sequence information describes PCR products that failed a sequence quality test (phred), this does not necessarily infer an inferior DNA target, but suggests a failure in the sequencing process due to the nature of high-throughput techniques.

Of the 13 selected genes that were significantly altered in expression throughout development, only 4 had significant homology to any known protein according to BLASTX analysis. Of the genes of putative functions based on homology, the fold-changes shown in Table 8.1 were observed. The fold-changes of vitellogenin expression throughout development are shown in Table 8.2.

PUTATIVE FUNCTION	ADULTS	MED. JUV.	SM. JUV.	NEONATES
Up-regulated in Adults				
Actin	1.5	1.1	0.8	0.9
Fibroin 4	1.5	1.2	0.5	0.3
Down-regulated in Adults				
Pupal Cuticle Protein (1)	0.6	2.1	1.7	1.5
Pupal Cuticle Protein (2)	0.3	2.4	1.9	1.6

Table 8.1: Statistically significant fold-changes in expression during development of *G. pulex*

G. PULEX GENE ID	ADULTS	MED. JUV.	SM. JUV.	NEONATES
Vitellogenin (G. pulex)				
VTG 7	2.0	1.8	0.5	0.6
VTG 8	2.0	1.7	0.5	0.4
VTG 9	1.6	1.0	0.5	0.6
VTG 10	1.9	1.2	0.5	0.5

Table 8.2: Fold-changes in *G. pulex* vitellogenin throughout development

8.6. DISCUSSION

Microarrays allow changes in expression of thousands of genes to be analysed simultaneously, providing an excellent tool for global gene expression profiling. MA plots and calibration curves confirmed the success of the profiling process. A volcano plot has shown that differentially expressed genes were identified between adults and neonates, the most extreme datasets. The vast quantity of data produced by gene expression profiling presents a complex task to identify individual potential biomarkers, so PCA was performed to reduce dimensionality and statistical analysis was used to isolate the best potential biomarkers of development.

These *G. pulex* studies revealed 8739 potential biomarkers of developmental stage that were up- or down-regulated, however statistical analysis reduced this number to 205 genes. Of these genes a selection of 13 were hand picked which showed the most consistent expression patterns. One of the most interesting genes identified as being up-regulated during maturity and peaking in adults, was a putative glycine/tyrosine-rich eggshell protein, homologous to the functional protein of the trematode *Opisthorchis viverrini* (BLAST N; e $=8e^{-5}$). The gene is expressed in the vitelline follicles of *O. viverrini* and is already present in 2-week old juveniles; however adult maturity is reached at 4 weeks when egg-laying begins (Ruangsittichai *et al.* 2006). These trematodes are hermaphrodites and egg-layers and therefore their reproductive development is different to that of *G. pulex*. This putative protein may be involved in brood pouch development or oocyte development in *G. pulex*, which may contain proteins similar to trematode egg-shell, a protective cover surrounding the vitellocyte/oocyte cluster in *O. viverrini*.

Vitellogenin was identified as a possible biomarker of development due to up-regulation in medium juveniles and adults, therefore, as well as a confirmed female specific marker it can be used to identify the onset of maturity in female *G. pulex*.

Another gene of interest up-regulated in adults is fibroin 4, homologous to that of the primitive hunting spider, *Plectreurys tristis* (BLASTN; e = $5e^{-4}$) (Gatesy *et al.* 2001). Fibroin is the silk that spiders spin in order to make webs and produce egg cases and therefore of unknown putative function in *G. pulex*.

Of the genes down-regulated in adult *G. pulex* were 2 gene fragments homologous to the pupal cuticle protein of *A. aegypti* ($e = 2e^{-11}$). Although *G. pulex* lack a pupal phase, there is a possibility of a juvenile/neonatal isoform that is not expressed in adults. This putative 'pupal' cuticle protein isoform is different to the isoform up-regulated in Stage 3 of adult moult stage (Chapter 7, Figure 7.18). The 2 juveniles/neonatal isoform gene fragments are both part of cluster GAC02542, yet the moult stage 3 'pupal' cuticle gene fragment belongs to cluster GAC0513 (for clusters see http://ecoworm.bios.cf.ac.uk – '*Gammarus pulex* (2)').

The interpretation of gene profile data was inhibited by insufficient sequence information for amphipods and the number of uncharacterised proteins, leading to poor GO annotation. This discussion is a brief overview of some genes of interest; however there are a large number of genes identified that may be potential biomarkers of developmental stage. Further analysis of the remaining genes of interest shown in Appendix B, Table 3 may lead to further understanding the molecular mechanisms involved in *G. pulex* development.

CHAPTER 9

DETERMINING DIFFERENTIAL GENE EXPRESSION IN INTERSEX AND NORMAL *ECHINOGAMMARUS MARINUS*

9.1. INTRODUCTION

The marine amphipod *Echinogammarus marinus* is of the same genus as *Gammarus pulex* and is sometimes referred to as *Gammarus marinus*. They differ from *G. pulex* in habitat, diet, size and fecundity (Table 9.1) and are abundant along the coastline of the UK. They are an inter-tidal species most commonly found under stones and seaweed.

	GAMMARUS PULEX	*ECHINOGAMMARUS MARINUS*
Habitat	Freshwater	Marine
Diet	Leaf detritus	Decomposing seaweed
Size of adults (mm)	12-16[2]	20[4]
Fecundity (eggs/brood)	16[3]	22[1]

Table 9.1: Comparison between amphipods *G. pulex* and *E.marinus*

[1] Cheng, C. (1942). "On the fecundity of some gammarids." Journal of the Marine Biological Association of the United Kingdom 25: 467-475.
[2] Welton, J. S. and R. T. Clarke (1980). "Laboratory Studies on the Reproduction and Growth of the Amphipod, *Gammarus pulex* (L.)." Journal of Animal Ecology **49**(2): 581-592.
[3] Hynes, H. B. N. (1955). "The reproductive cycle of some British freshwater Gammaridae." Journal of Animal Ecology **24**: 352-387.
[4] Vlasblom, A. G. (1969). "A study of a population of *Marinogammarus marinus* (leach) in the oosterschelde." Netherlands Journal of Sea Research **4**(3): 317-338.

Although it has previously been demonstrated that size and fecundity are not consistently correlated in gammarids (Cheng 1942), in the same study is was observed that *E. marinus* produces more eggs than *G. pulex* with a mean of ~22 eggs per brood as observed between January and May. Studies in Scotland, UK have revealed intersex male and female *E. marinus* amongst normal populations at numerous sites along the coast of East Scotland. It was observed that the intersex animals, on average, had smaller brood sizes and reduced embryonic survival (Ford *et al.* 2003). Intersexuality is described in more detail in Chapter 1, Section 1.3.1.

Intersex *E. marinus* have been observed at reference (~5-8%) and more abundantly at polluted sites (~14-15%) (Ford *et al.* 2004). Further studies on the 'intersex state' of *E. marinus* revealed the underlying morphology. There were found to be two intersex phenotypes; intersex male and intersex female. The intersex males contained normal testes, seminal vesicles and *vas deferens* in addition to an oviduct; intersex females contained ovaries reduced in size due the presence of *vas deferens*. In the same study, intersex males were also found to be lacking the presence of an A-chain peptide of androgenic gland hormone, usually found in normal males; it was also undetectable in both intersex and normal females (Ford *et al.* 2005). Whether or not intersexuality in *E. marinus* is caused by pollution, presence of microsporidian feminising parasites or a combination is not known (Ford *et al.* 2006).

The mechanisms underlying intersexuality in *E. marinus* can be partially determined by comparing transcriptomic profiles for the four sexual phenotypes using microarray technology. It is possible that microarrays fabricated from one species can be used to assess the genetic profile of another species. This is achieved by cross-species hybridisation of the DNA probe from one species onto target DNA from another species. Due to DNA from *G. pulex* being spotted onto the microarray slides, genetic profiles would have to be created by cross-species hybridisation of *E. marinus* probe onto *G. pulex* targets.

Often referred to as heterologous hybridisation, this approach has successfully been used to identify evolutionarily conserved molecular mechanisms in distinct species by comparing bovine, mouse and frog ovarian gene expression on a multi-species cDNA microarray (Vallee *et al.* 2006). This technique has also been used to compare gene expression of human and porcine tissue using the Affymetrix Human Genome U133A GeneChip® (Affymetrix, UK Ltd., High Wycombe, Bucks., UK) yielding a low level of false positives (Shah *et al.* 2004) and in another study, distantly related mammalian species; cattle, pig and dog have been heterologously hybridised to the same Affymetrix GeneChip® platform (Ji *et al.* 2004). Gene profiles of *Xenopus laevis* have been successfully compared to *Xenopus tropicalis* using a *X. tropicalis* microarray system (Sartor *et al.* 2006).

9.2. SPECIMEN COLLECTION AND IDENTIFICATION

All *E. marinus* specimens were collected, visibly determined to be the correct species and sexual phenotype determined by Dr. Alex Ford (Environmental Research Institute, North Highland College, UHI Millennium Institute, Scotland UK). Normal and intersex male and female *E. marinus* were sampled from underneath stones and seaweed from Inverkeithing Bay, Scotland or Langland Bay, Gower Peninsula, South Wales from August 2006 to November 2006 (Figure 9.1).

Specimens were brought back to the laboratory, where sexual phenotype was ascertained after anaesthetising specimens using carbonated water (Ford *et al.* 2003). Intersex specimens were categorised as either intersex male or intersex female based upon specific characteristics as described by Ford *et al.* (2003) and stored in RNA*later*® or RNA*later*®-ICE (Ambion, Huntingdon, Cambs., UK) as described by the manufacturer.

9.2.1. Cytochrome oxidase I (COI) analysis

The genetic similarity was crudely tested by comparing cytochrome oxidase I (COI) DNA and amino acid sequences. PCR was performed as described in Chapter 2, Section 2.11.9.3 using primers COIa-H and COI-Gf (Appendix A, Table 6) and sequenced by Cardiff University Molecular Biology Support Unit as described in Chapter 2, Section 2.11.12. The DNA sequences were aligned with other amphipod DNA sequences using the same region of COI and a phylogenetic tree generated (Figure 9.2).

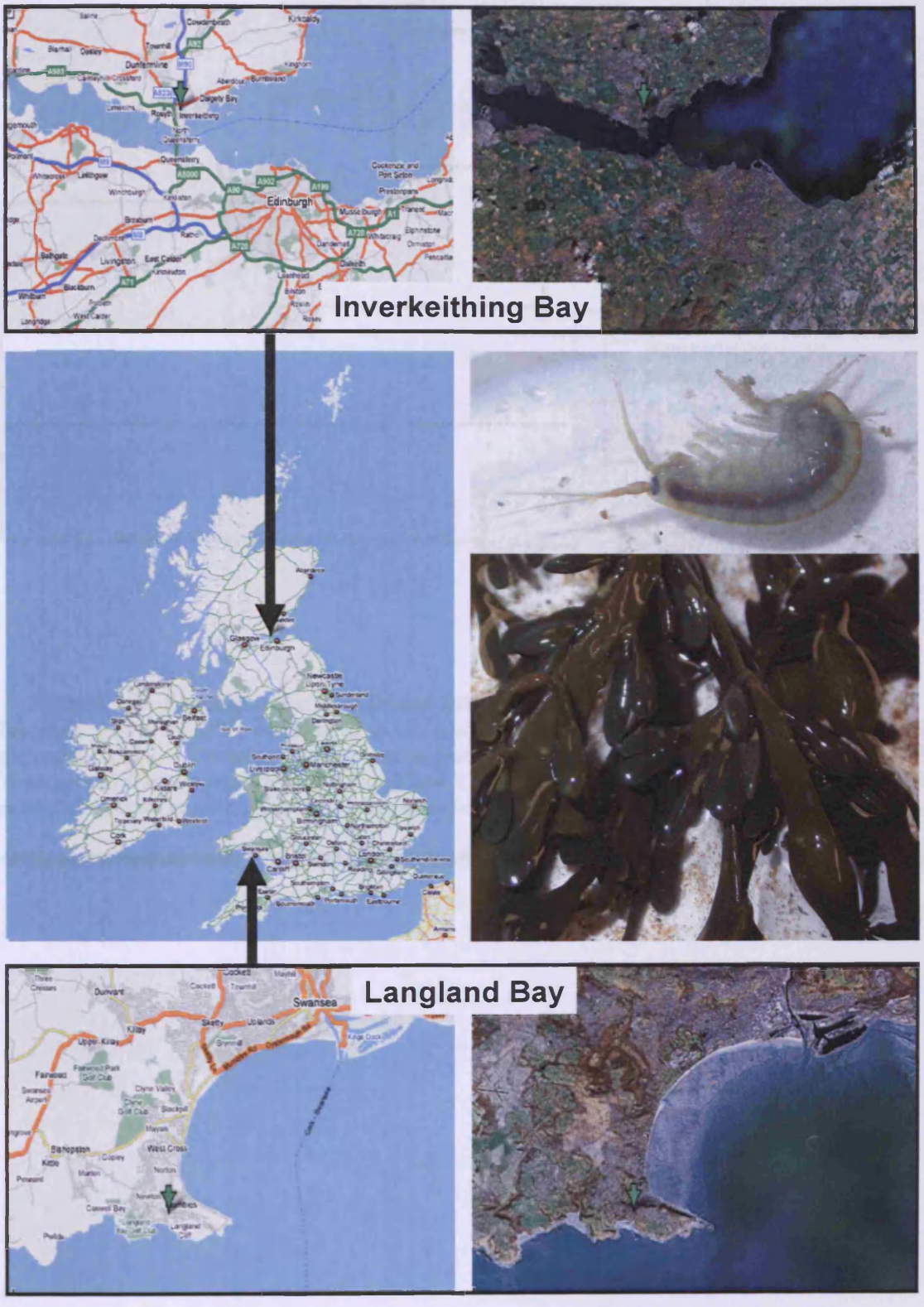


Figure 9.1: *E. marinus*, their habitat and sampling sites

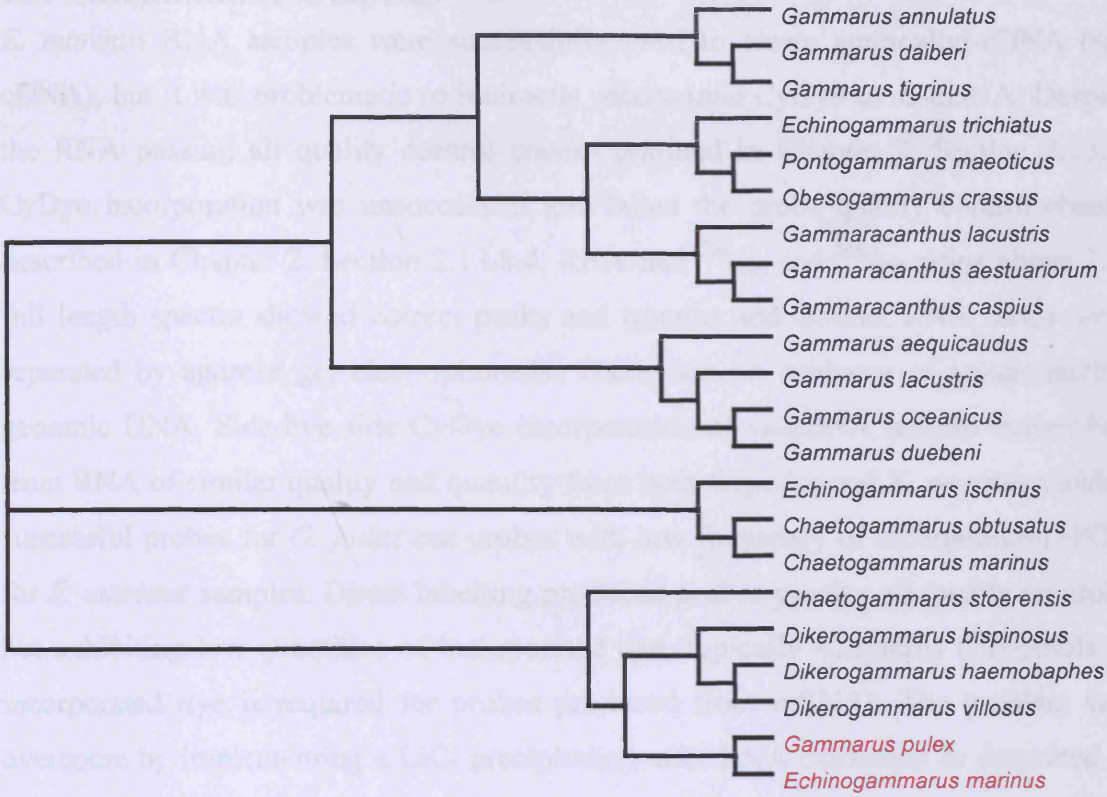


Figure 9.2: Phylogenetic tree based on cytochrome oxidase I DNA sequence
The same section of COI was aligned in all species described in the diagram using CLUSTALW alignment (EMBL - European Bioinformatics Institute, Hinxton, Cambs., UK). The exported DND (dendrogram) file was imported into Tree View, (Version 1.6, (Page 1996)) and a phylogenetic tree generated. Species in red highlight the amphipods used in this study.

9.3. MICROARRAY ANALYSIS

E. marinus RNA samples were successfully used to create aminoallyl-cDNA (aa-cDNA), but it was problematic to indirectly incorporate CyDye to aa-cDNA. Despite the RNA passing all quality control checks outlined in Chapter 2, Section 2.13.7, CyDye incorporation was unsuccessful and failed the probe quality control checks described in Chapter 2, Section 2.13.8.4. RNA had $^{260}/_{230}$ and $^{260}/_{280}$ ratios above 2.0, full length spectra showed correct peaks and troughs and distinct RNA bands were separated by agarose gel electrophoresis. There was no evidence of contaminating genomic DNA. Side-bye side CyDye incorporation of aa-cDNA reverse transcribed from RNA of similar quality and quantity from both *G. pulex* and *E. marinus* yielded successful probes for *G. pulex* but probes with low frequency of incorporation (FOI) for *E. marinus* samples. Direct labelling produced probes passing all quality controls, but exhibiting low quantities of incorporated dye, typically <20pmols (>20pmols of incorporated dye is required for probes produced from mRNA). The problem was overcome by implementing a LiCl precipitation after RNA extraction as described in Chapter 2, Section 2.13.6.1.

Sample collection and storage is described in Section 9.2. Before labelling, RNA was extracted as described in Chapter 2, Section 2.13.6 using optional LiCl precipitation. Samples passing quality control (Chapter 2, Section 2.13.7) were labelled (Chapter 2, Section 2.13.8) using a 4:1:5 U:T:ACG ratio of dNTPs and hybridised (Chapter 2, Section 2.13.9) onto *G. pulex* microarray slides (Chapter 2, Section 2.13.3). Slides were scanned (Chapter 2, Section 2.13.10) and data stored as 2 (Cy3 and Cy5) TIFF (Tagged Image File Format) images.

9.3.1. Bioinformatics and Microarray interpretation

Images were analysed using SPOT (Version 2.0, CSIRO Mathematical and Information Sciences, Australia (Yang *et al.* 2001; Yang *et al.* 2002)) and data imported into GeneSpring ((Version GX 7.3; Agilent Technologies UK Ltd., Cheshire, UK) not including the control and blank data) as described in Chapter 2, Sections 2.13.11 and 2.13.12. Poor slides (determined from box plots) were removed from the dataset and the filtered data normalised as described in Chapter 2, Section 2.13.12.

9.3.1.1. Microarray quality validation

An MA plot was generated to demonstrate the raw and normalised filtered data of the experimental samples. An MA plot is defined as the log ratio vs. log median intensity where:

$$M = \log_2\left(\frac{Cy5}{Cy3}\right)$$

$$A = \frac{(\log_2 Cy5 + \log_2 Cy3)}{2}$$

The MA of raw data (Figure 9.3A and Figure 9.3C) shows the intensity dependent ratio and demonstrates spot artefacts and systematic variation. The ideal MA plot should have M values that are evenly distributed at 0 across the range of intensities (Dudoit *et al.* 2002). Raw data was normalised to remove systematic bias from experimental artefacts as described in Chapter 2, Section 2.13.12. The MA plot demonstrates how the filtering and normalisation process created a data set distributed around M=0 (Figure 9.3B and Figure 9.3D).

Internal synthetic control mRNA (Universal ScoreCard, GE Healthcare, Little Chalfont, Bucks., UK) was added to each RNA sample and reverse transcribed and indirectly incorporated with Cy3 simultaneously with the RNA. These control probes bind to complementary target sequences printed across the length and width of the microarray slides alongside the *G. pulex* DNA targets. Using the ratio ScoreCard controls, a calibration curve can be generated representing the range of detection for each slide. Calibration curves from 6 slides are shown in Figure 9.4.

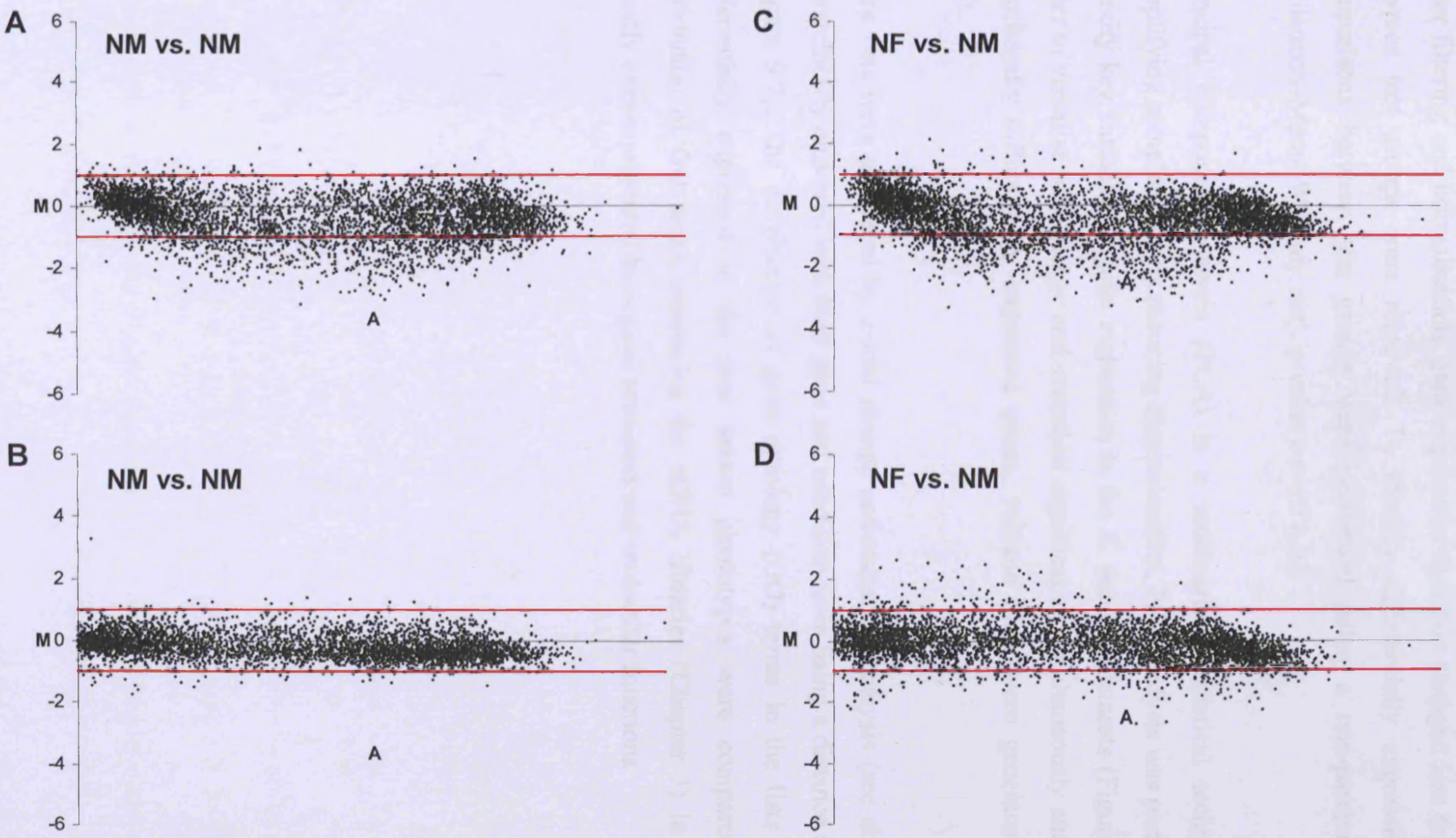


Figure 9.3: MA plots of male and female *G. pulex* microarray data before and after normalisation

A: Normal Male vs. Normal Male before normalisation; B: Normal Male vs. Normal Male after normalisation; C: Normal Female vs. Normal Male before normalisation; D: Normal Female vs. Normal Male after normalisation.

9.3.1.2. Statistical analysis of differential gene expression

After filtering and normalisation, gene expression data that changed less than 2-fold between test groups were removed. To identify differentially expressed genes, comparisons between test groups were performed using a non-parametric test (Wilcoxon–Mann–Whitney test), p-value cut-off 0.05.

Principal component analysis (PCA) is a multivariate statistical technique for simplifying complex data by reducing dimensionality. PCA analysis was performed to identify key variables in gene expression in the *E. marinus* datasets (Figure 9.5). In order to visualise fold change and statistical significance simultaneously and identify significantly differentially expressed genes, volcano plots were generated (Figure 9.6).

Gene lists were generated by 2-fold change and statistical analysis (see above) and hierarchically clustered with both gene and condition trees using a distance algorithm (Figure 9.7). The distribution of gene ontology (GO) terms in the lists of genes differentially expressed in the four sexual phenotypes were compared to the distribution of GO terms annotating the cDNA libraries (Chapter 3) in order to identify over-represented biological processes and molecular functions.

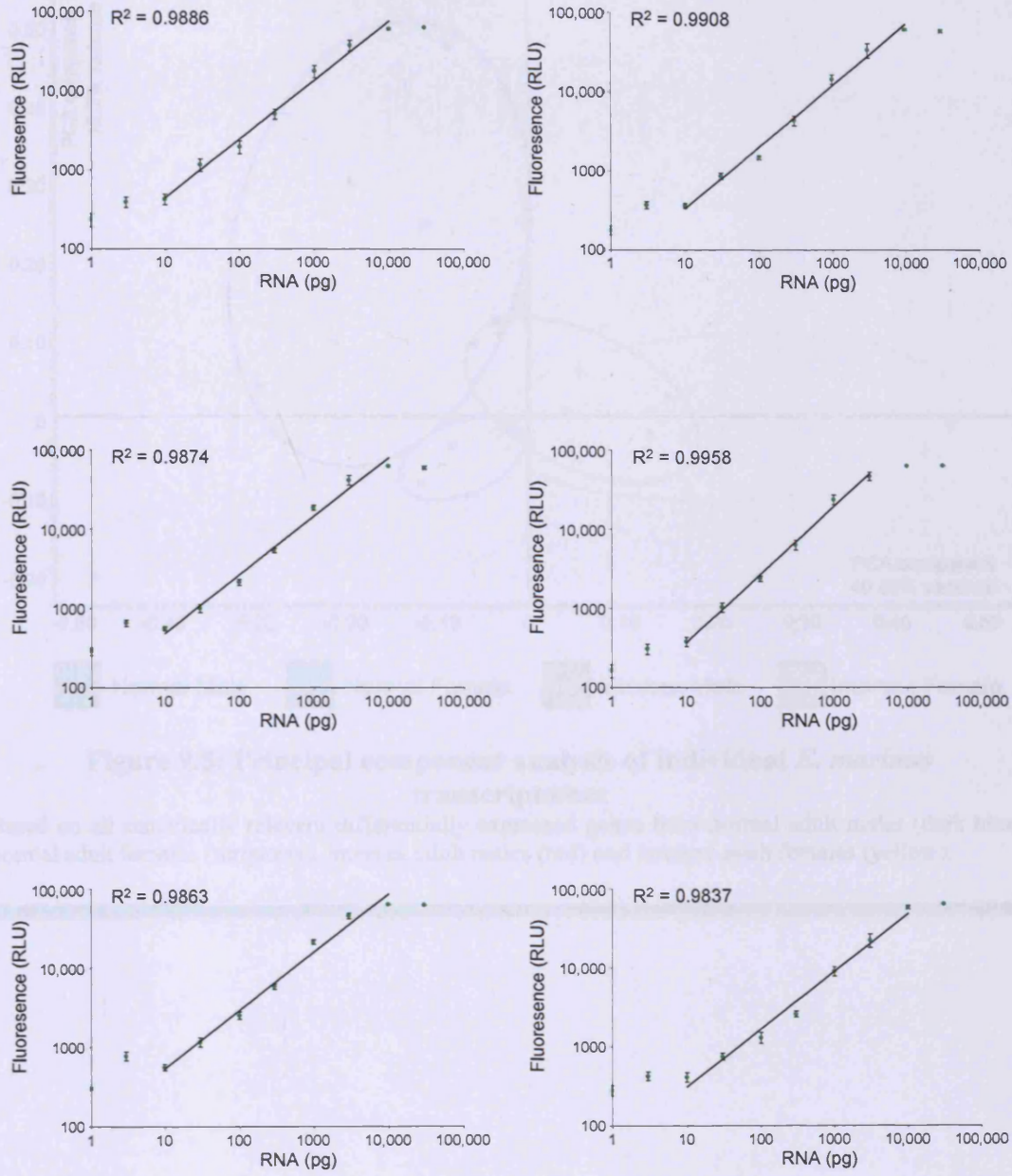


Figure 9.4: Universal ScoreCard calibration of 6 slides used to generate the *E. marinus* datasets

Each ratio control was printed 10 times per slide; the mean signal intensity was used to generate calibration curves which are displayed with standard error bars.

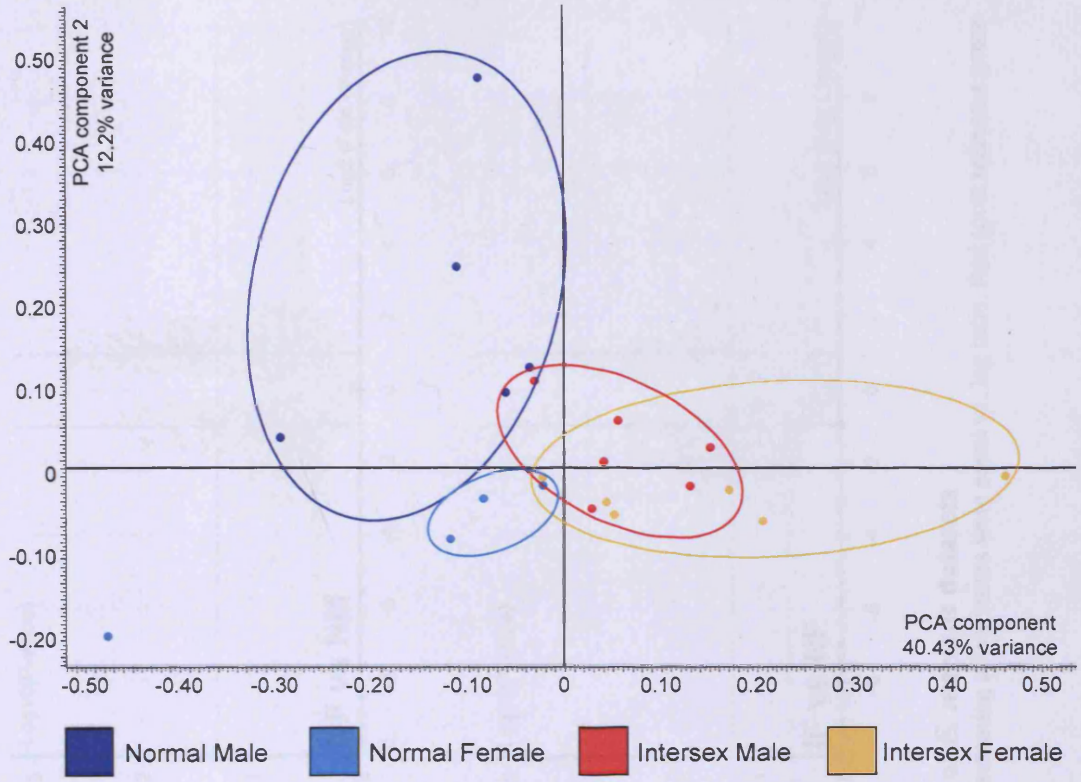


Figure 9.5: Principal component analysis of individual *E. marinus* transcriptomes

Based on all statistically relevant differentially expressed genes from normal adult males (dark blue), normal adult females (turquoise), intersex adult males (red) and intersex adult females (yellow).

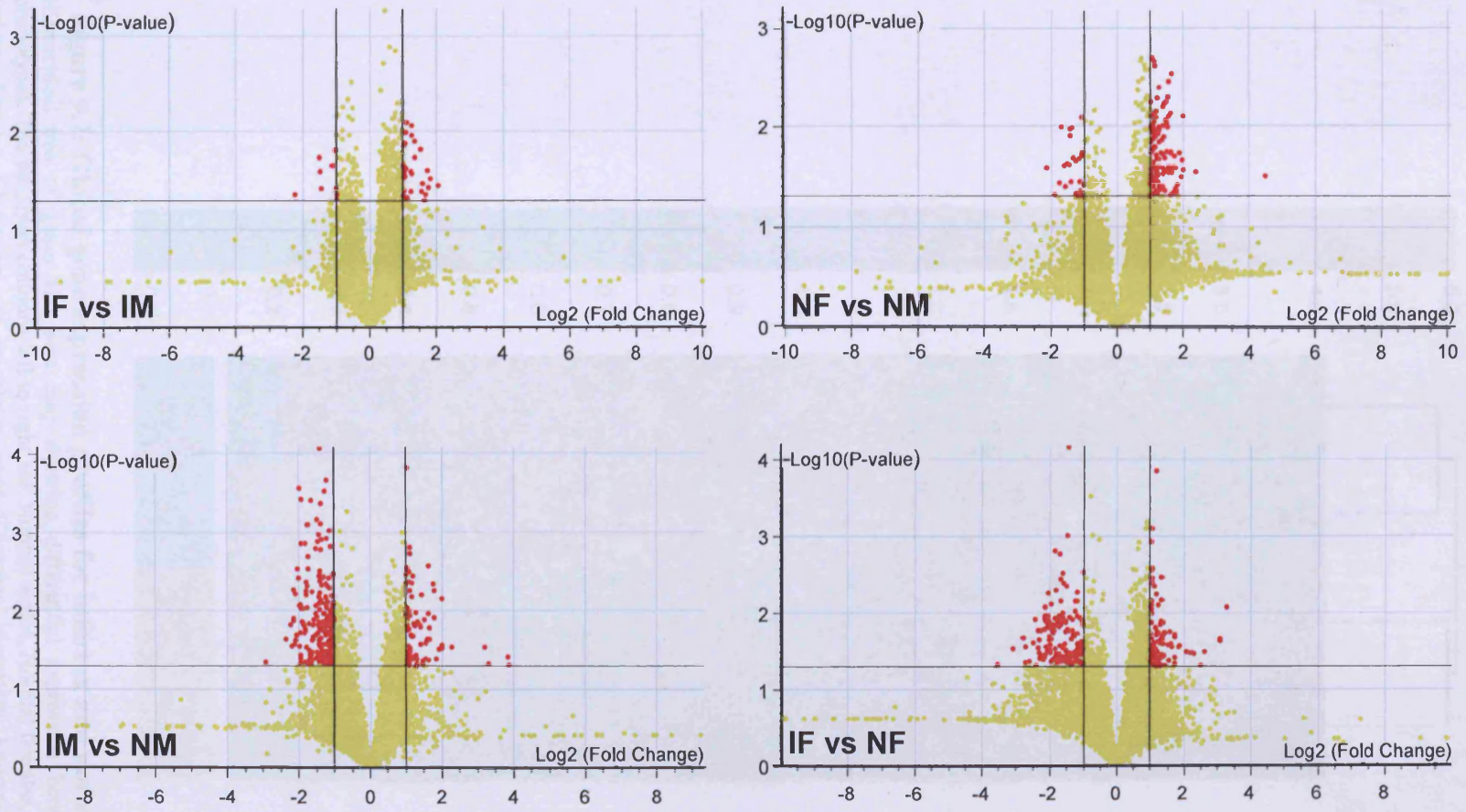


Figure 9.6: Volcano plot of *E. marinus* datasets

Volcano plots are scatter plots of –log (p-value 0.05) from 1-way ANOVA, assuming all variances were equal vs. log ratio. Red spots represent genes with 2-fold change with p-value <0.05.

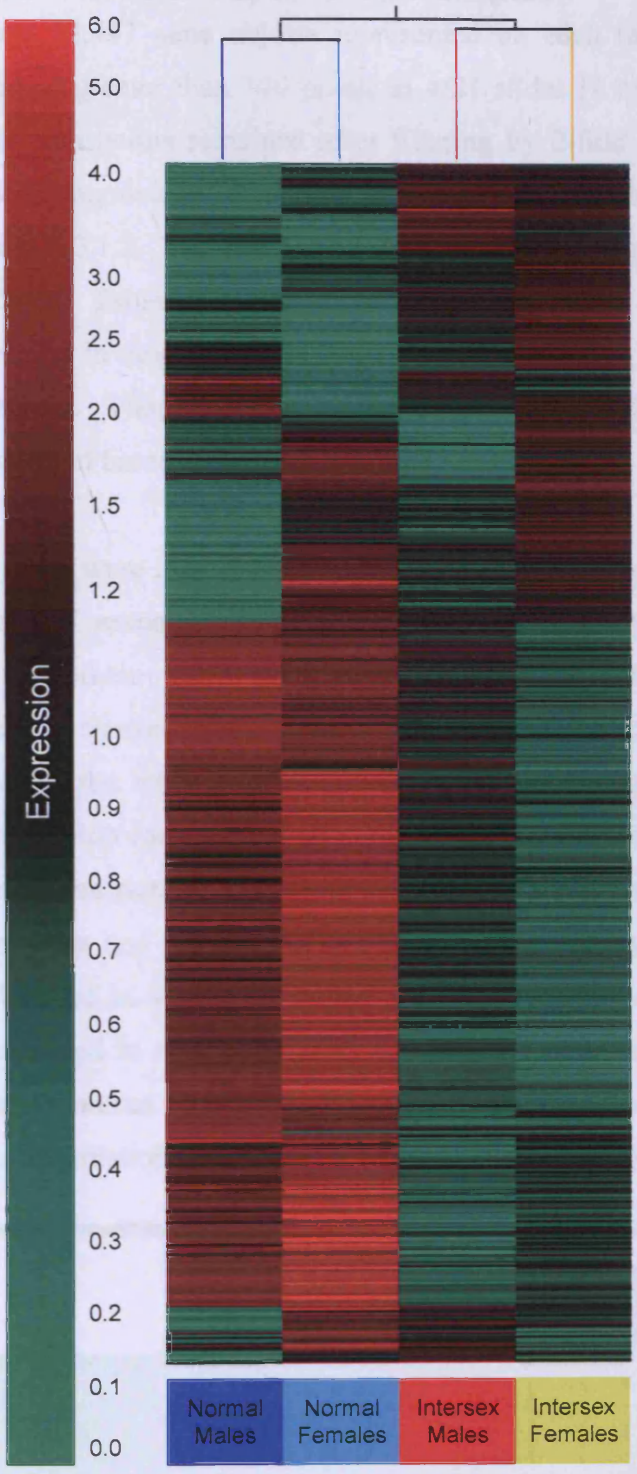


Figure 9.7: Global gene expression profiles for intersex and normal *E. marinus*
Hierarchical tree of genes and heat map showing differential expression between the 4 sexual phenotypes. The heat map columns left to right are: Normal males, Normal females, Intersex males and Intersex females. Clustering was calculated using 'Distance' algorithm. Expression is average log signal intensity.

9.4. POTENTIAL BIOMARKERS OF INTERSEX ANIMALS

Raw data contained 13,447 gene objects represented on each microarray slide of which all had a signal greater than 100 pixels in 4/21 slides (4 being the minimum sample size). 5307 data points remained after filtering by 2-fold change (1577 up-regulated; 4180 down-regulated). This filtered dataset was statistically analysed as described in Section 9.3.1.2. The full list of differentially expressed genes (555) is given in Appendix B, Table 4). Distribution of GO terms in the list of genes differentially expressed in developmental stages was compared to the cDNA library GO term distribution. There was no over-representation; therefore potential biomarkers were selected based on expression trend consistency.

A selection of 12 genes were identified showing clear trends that potentially could be used as biomarkers of sexual phenotype and were significantly altered between phenotypes, although there were numerous others which, although were not significantly different, showed clear trends. In intersex males, 5 genes were up-regulated compared to the other 3 phenotypes, separately, 3 genes showed normal male specific up-regulation compared to all other sexual phenotypes (Figure 9.8). Up-regulation in intersex and normal males was shown in 4 genes and 5 genes showed up-regulation in intersex and normal females (Figure 9.9). Up-regulation in normal *E. marinus* was observed in 4 genes, 4 genes were up-regulated in intersex females and 4 genes up-regulated in normal females and intersex males (Figure 9.10). Up-regulation in normal females compared to normal males was identified in 4 genes (Figure 9.11). This information is tabulated for easier comprehension in Table 9.1.

		UP-REGULATED			
		NM	NF	IM	IF
UP-REGULATED	NM	3	4	4	0
	NF	4	4	4	5
	IM	4	4	5	4
	IF	0	5	4	4

Table 9.2: Distribution of up-regulated potential biomarkers of sexual phenotype in *E. marinus*

NM = normal male; NF = normal female; IM = intersex male; IF = intersex female

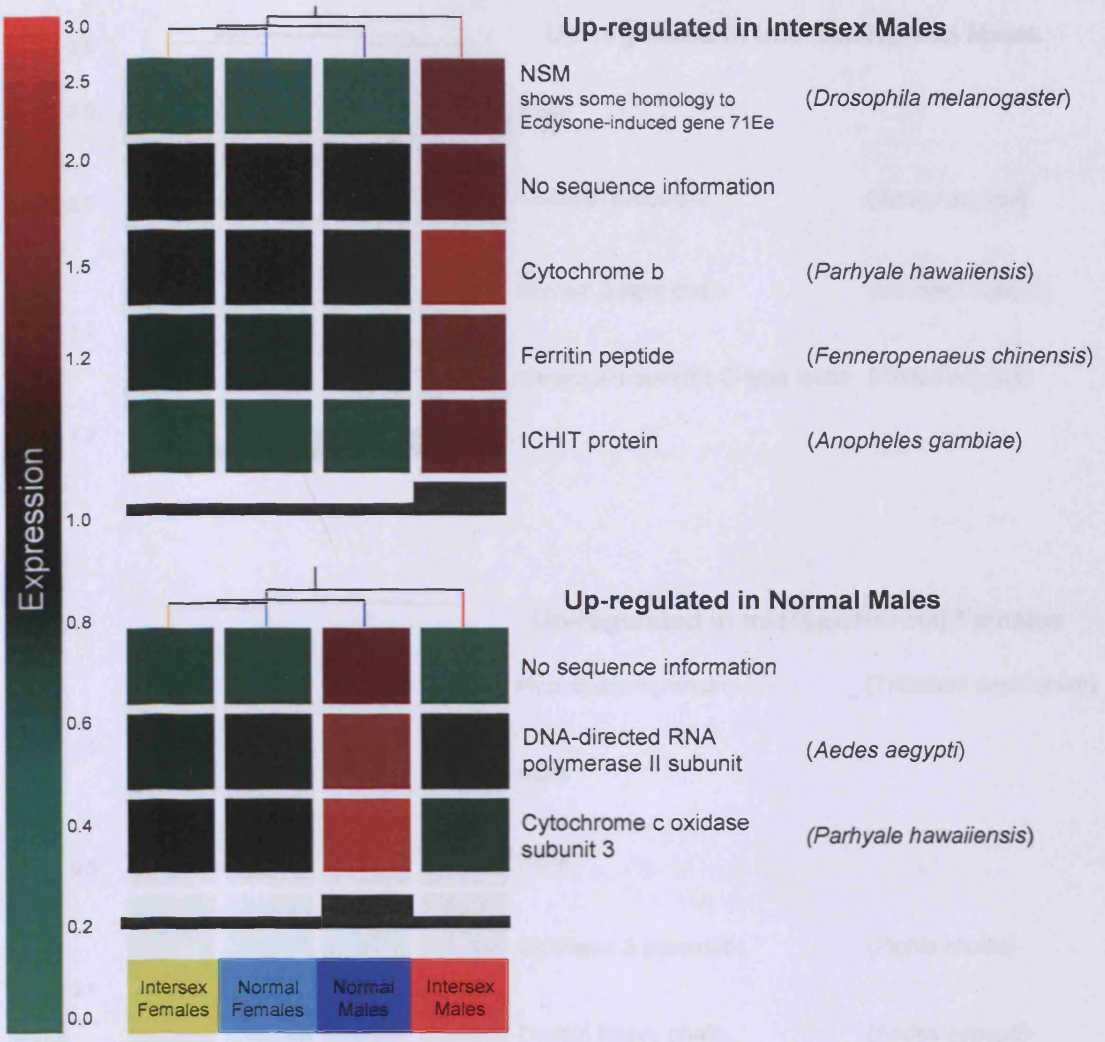


Figure 9.8: Genes potentially up-regulated in Intersex or Normal adult male *E. marinus*

NSM = no significant BLAST database match. No sequence information describes PCR products that failed a sequence quality test (phred), this does not necessarily infer an inferior DNA target, but suggests a failure in the sequencing process due to the nature of high-throughput techniques.

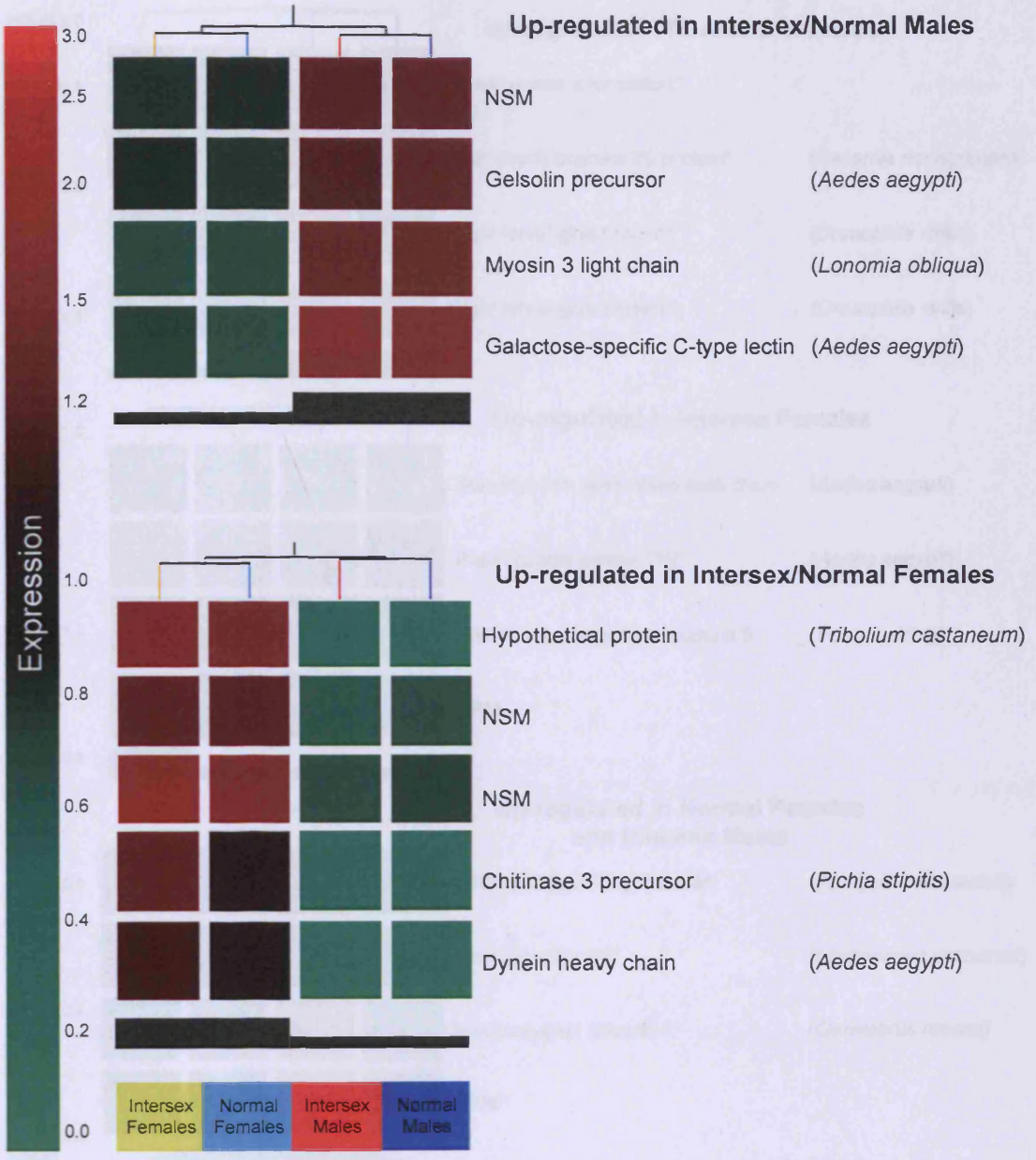


Figure 9.9: Genes potentially up-regulated in Intersex/Normal adult male and Intersex/Normal adult female *E. marinus*

NSM = no significant BLAST database match.

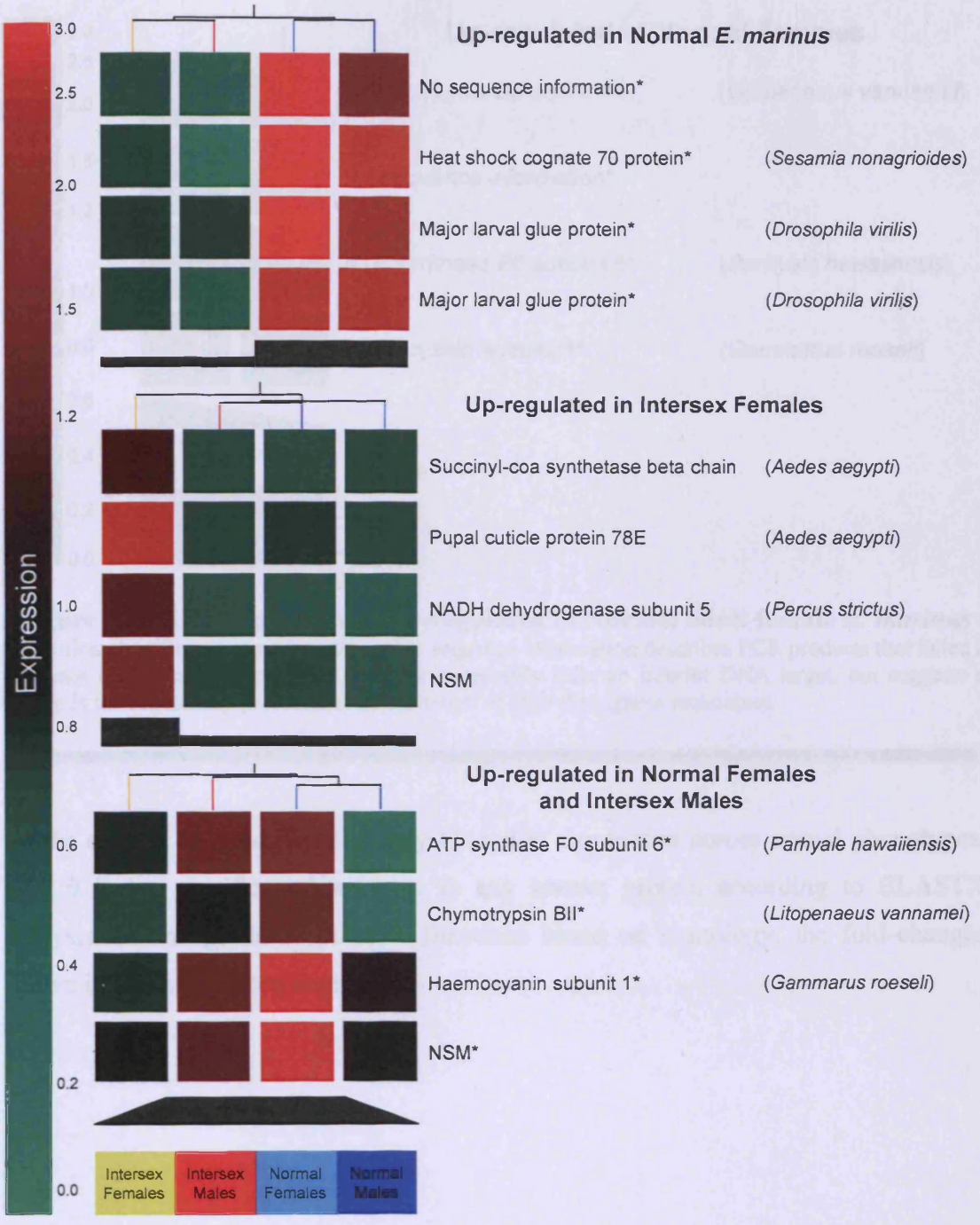


Figure 9.10: Genes potentially up-regulated in Normal, Intersex adult female or Normal Female/Intersex male *E. marinus*

*Significantly different gene expression; NSM = no significant BLAST database match. No sequence information describes PCR products that failed a sequence quality test (phred), this does not necessarily infer an inferior DNA target, but suggests a failure in the sequencing process due to the nature of high-throughput techniques.

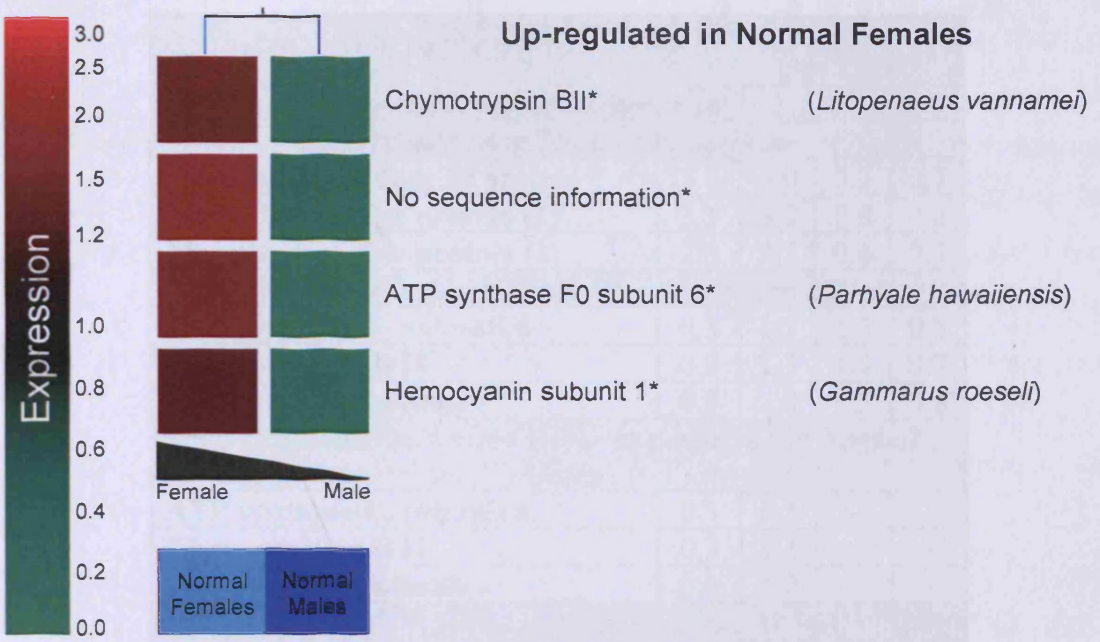


Figure 9.11: Genes potentially up-regulated in Normal adult female *E. marinus*
*Significantly different gene expression. No sequence information describes PCR products that failed a sequence quality test (phred), this does not necessarily infer an inferior DNA target, but suggests a failure in the sequencing process due to the nature of high-throughput techniques.

Of the chosen 12 genes significantly altered in expression across sexual phenotypes, only 9 had a significant homology to any known protein according to BLASTX analysis. Of the genes of putative functions based on homology, the fold-changes shown in Table 9.3 were observed.

PUTATIVE FUNCTION	NM	NF	IM	IF
Statistically significant				
Up-regulated in Normal E. marinus				
Heat shock cognate 70 protein	2.7	6.0	0.3	0.7
Major larval glue protein (1)	2.1	3.0	0.8	0.8
Major larval glue protein (2)	2.6	2.7	0.6	0.7
Up-regulated in Normal Females and Intersex Males				
ATP synthase F_0 subunit 6	0.3	1.5	1.3	0.8
Chymotrypsin B II	0.3	1.3	1.0	0.7
Haemocyanin subunit 1	0.9	2.8	1.3	0.8
Up-regulated in Normal Females compared to Normal Males				
ATP synthase F_0 subunit 6	0.3	1.5	-	-
Chymotrypsin B II	0.3	1.3	-	-
Haemocyanin subunit 1	0.4	1.2	-	-
Not statistically significant				
Up-regulated in Intersex Males				
ICHIT	0.7	0.7	1.2	0.7
NSM: ecdysone induced gene 71Ee	0.7	0.6	1.2	0.6
Up-regulated in Intersex Females				
Pupal cuticle protein 78E	0.8	0.7	0.7	2.7
Up-regulated in Intersex and Normal Males				
Gelsolin	1.3	0.8	1.2	0.8
Up-regulated in Intersex and Normal Females				
Chitinase 3 precursor	0.5	1.0	0.5	1.3

Table 9.3: Fold-changes in expression of potential biomarkers of sexual phenotype in *E. marinus*

9.5. DISCUSSION

Cross-species DNA microarray techniques allow thousands of simultaneous changes in gene expression to be identified, creating a global gene expression profile for phylogenetically distant species. In this study, the species used, *E. marinus*, was not as phylogenetically distant from *G. pulex* as species compared in previous studies by other research groups (reviewed in Buckley 2007). Therefore sequence similarity was likely to be high enough to allow heterologous hybridisation producing viable data. Although COI gene alignment doesn't provide a precise genome-wide comparison and is based entirely on one, highly conserved gene, it is often used to identify species genetically (Meyran *et al.* 1997). MA plots and calibration curves have confirmed the success of the profiling process showing a range of intensities signifying successful sequence recognition and binding over a wide range of expression levels. Volcano plots have shown that differentially expressed genes were identified between the 4 sexual phenotypes. The vast quantity of data produced by gene expression profiling presents a complex task to identify individual potential biomarkers, so PCA was performed to reduce dimensionality and statistical analysis was used to isolate the best potential biomarkers of phenotype.

These *E. marinus* studies revealed 5307 potential biomarkers of sexual phenotype that were up- or down-regulated, however statistical analysis reduced this number to 555 genes. Of these, selections of 12 genes were hand picked which showed the most consistent expression patterns. A further 21 genes showing a 2-fold up-regulation, but were not statistically relevant were chosen due to their consistent expression pattern and putative function.

Of particular interest was the heat shock cognate protein 70, up-regulated in normal *E. marinus* and homologous to the functional protein of the stem borer, *Sesamia nonagrioides* ($e = 7e^{-72}$). Heat shock cognate protein 70 (HSC70) is a member of the 70kDa heat shock protein family and is also known as heat shock protein 73 (HSP73). This family are involved in folding and trafficking newly synthesised proteins. HSC70 is expressed constitutively, regardless of physiological demand, in contrast to HSP70 induced by heat shock or stress; otherwise their function is believed similar (Goldfarb *et al.* 2006). This gene fragment also showed high homology to HSP70

from *Locusta migratoria* ($e = 9e^{-72}$) so it is highly likely that this up-regulation is of a heat shock protein family member, although identifying which particular protein will require further sequence analysis. Heat shock proteins also bind and stabilise steroid hormone receptors and act as a repressor of receptor activation and hold the receptor in a suitable conformation for ligand binding as stated in Chapter 1, Section 1.2.3.3. A disruption to this process may allow non-endogenous steroids to bind to hormone receptors causing subsequent pathways to be activated perhaps allowing a normally repressed action, such as the development of secondary sexual characteristics to develop.

The major larval glue protein up-regulated in normal male and female *E. marinus* is the same protein as up-regulated in Stage 1 of *G. pulex*, as described in Chapter 7, Section 7.8.4. It is used in *Drosophila* sp. to fix the pupal case to a substrate (Fraenkel and Brookes 1953) and it was suggested in the *G. pulex* study described in Chapter 7 that this gene product may be involved in creating and stabilising the new cuticle (Chapter 7, Section 7.8.4). Up-regulation in normal *E. marinus* suggests that the activity of this protein is reduced in intersex animals. Hypothetically, this could affect postmoult cuticle formation or stabilisation.

Another gene of interest was homologous to ATP synthase F_0 subunit 6 of the amphipod *Parhyale hawaiiensis* ($e = 5e^{-52}$). It was up-regulated in normal females and intersex males. The conclusions from a study in rats may explain a possible reason for these findings through its possible role in an oestrogen-related process. Other than the normal binding of oestrogen to its receptor (ER), alternative binding sites and interactions have been identified in rat brain mitochondrial fractions. A 23kDa membrane protein with the ability to bind 17β-oestradiol has been isolated and characterised as oligomycin-sensitivity conferring protein (OSCP), a subunit of F-type mitochondrial ATP synthase. This particular subunit is essential for proton gradient coupling across the F_0 to F_1 enzyme parts. This suggests an alternative mechanism for oestradiol and similar compounds by binding to ATP synthase affecting energy metabolism in rats (Zheng and Ramirez 1999). Stimulation of ATP synthase activity has been observed after the addition of oestradiol to rat liver preparations and inhibited by incubating diethylstilbestrol with rat liver, brain and heart preparations (Kipp and Ramirez 2001). Although the presence of oestrogen in gammarids has not

been confirmed, the presence of an oestrogen receptor orthologue and oestrogen-related receptors in similar organisms suggests a possible role for the steroid hormone or similar compounds (Maglich *et al.* 2001; Thornton *et al.* 2003).

A gene fragment that, when translated, shows similarity to ICHIT protein from the mosquito, *Anopheles gambiae*, was up-regulated in intersex males of *E. marinus* (e = $8e^{-5}$). In A. gambiae, ICHIT is a gene that encodes 2 putative chitin binding domains which are separated by a mucin region rich in poly-threonine. It is expressed in the midgut of pupae and adult mosquitoes and expression is induced by parasitic and bacterial infections. ICHIT shows a late immune response and is possibly involved in neutrophil/leucocyte trafficking in vertebrates and defence mechanisms against bacterial and parasitic attack in invertebrates (Dimopoulos *et al.* 1998). This evidence supports the observation by Ford *et al.* (2006) that microsporidian parasites may be one of the causative agents of the intersex phenotype in *E. marinus*. Up-regulation in intersex males suggests that the altered expression is due to a feminisation or demasculinisation process and a different mechanism is involved in the masculinisation of female *E. marinus*.

A multifaceted protein, gelsolin (e = 7.6) was identified in normal and intersex male *E. marinus*. Despite the low expect value, the majority of the translated protein showed good homology (65%) to gelsolin precursor from the yellow fever mosquito, *Aedes aegypti*. The reason behind the low 'Expect value' was the short length of the *G. pulex* target sequence, originating from the male subtractive library (Chapter 3, Section 3.3). Gelsolin binds to actin and is involved in cytoskeletal actin dynamics in vertebrates; it has also been identified in *Drosophila melanogaster*. Gelsolin has also been indicated in nuclear-receptor mediated intracellular signalling pathways (Archer *et al.* 2004). In human cells, gelsolin acts as an androgen receptor co-regulator through its binding to the DNA-binding domain and ligand-binding domain of the androgen receptor and enhances androgen receptor activity in the presence of androgen (Nishimura *et al.* 2003). The identification of a FXXFF motif in human gelsolin reiterated its role as an androgen receptor cofactor. The presence of a corresponding FXXLF docking site in the hydrophobic binding groove of the androgen receptor ligand-binding domain showed the ability of specific binding to the androgen receptor (van de Wijngaart *et al.* 2006). In spermatogenesis studies in

androgen receptor knockout mice, quantitative PCR using Sertoli cells (associated with developing spermatocytes/spermatids) revealed a reduction in gelsolin gene expression in postnatal mice (Wang *et al.* 2006). Another role of gelsolin is as a component of anchoring junctions in Sertoli cells (Guttman *et al.* 2002), therefore reduced gelsolin gene expression combined with reductions in claudin-II and occludin expression (major components of tight junctions), contributed to a decrease in the integrity of the blood-testis barrier. As a result of these experiments, it was suggested by Wang *et al.* (2006) that a functional androgen receptor and the presence of androgen are required to maintain the normal structure and function of tight and anchoring junctions. This would create an intact blood-testis barrier which is essential for successful spermatogenesis and fertility (Wang *et al.* 2006). The successful production of this versatile protein may be essential for effective spermatogenesis and release from the testis explaining the up-regulation of the putative gelsolin gene in normal and intersex male *E. marinus*. It has been observed that both intersex and normal male *E. marinus* are capable of producing sperm, however, the intersex males produce sperm at lower quantities; whether the intersex males are fertile is unknown (Ford *et al.* 2005).

A gene fragment putatively expressing a chitinase 3 precursor was up-regulated in intersex and normal female *E. marinus*, with amino acid sequence homology to a chitinase 3 precursor from the yeast, *Pichia stipitis* (e = 0.034). Chitinase was also found to be up-regulated in female *G. pulex* (Chapter 7) suggesting it may have a gender-specific role in female gammarids. The chitinase protein family in humans includes an oviduct-specific glycoprotein (OVP) which binds to the zona pellucida and plasma membrane of oocytes and developing embryos. Gene expression is thought to be driven by the presence of oestrogen due to the identification of oestrogen-responsive elements (EREs) in the promoter region (Agarwal *et al.* 2002). In sheep OVP showed no chitinase activity despite being similar in amino acid sequence to the chitinase protein family members (DeSouza 1995; DeSouza and Murray 1995). Despite the presence of *vas deferens* in intersex female *E. marinus*, the production of oocytes is still evident, though quantity and possibly quality is reduced (Ford *et al.* 2005). The presence of oestrogen has not been detected in gammarids; however this putative chitinase may play a similar role in oocyte and embryo

development explaining the up-regulation of gene expression in intersex and normal female *E. marinus* and female *G. pulex*.

In conclusion, an efficacious gammarid cross-species microarray platform has been developed and utilised successfully using *E. marinus* probes heterologously hybridised to a *G. pulex* target genes to identify possible biomarkers of sexual phenotype.

CHAPTER 10

GENERAL DISCUSSION AND FUTURE WORK

The aims of this project were to gain a better understanding of the endocrine system of *Gammarus pulex* at the molecular level and establish the genetic effects of endocrine disrupting chemicals and intersex state through the study of *Echinogammarus marinus*. The identification of potential molecular biomarkers was achieved through 2 main approaches; individual candidate gene identification and large scale genomics.

Understanding the processes involved in development, vitellogenesis, moulting and reproduction would increase the molecular knowledge of the normal reproductive and developmental biology of crustaceans, a requirement when studying endocrine disruption. Shedding of exuviate during ecdysis, at the end of the moult cycle, is concurrent with ovarian maturation, emphasising the significant relationship between moulting and reproduction. The moult cycle is linked to somatic and gonadal growth reiterating that the hormonal control of moulting and vitellogenesis is essential for both growth and reproduction.

By binding to native steroid hormone receptors, endocrine disrupting chemicals (EDCs) mimic hormones and act in direct competition with endogenous ligands in an agonistic or antagonistic manner, disrupting biological processes including behaviour, reproduction, development and homeostasis.

Gene expression profiling using microarray technology allows the molecular mechanisms underlying normal and disrupted gene responses and metabolic pathways in amphipods to be elucidated.

10.1 DEVELOPMENT OF A *GAMMARUS PULEX* LIBRARY

Two cDNA libraries of *G. pulex* genes have been developed to allow the identification of novel genes and to enhance the knowledge of *G. pulex* biology at the molecular level. In total 13,440 gene objects were isolated and expressed sequence tags generated, enriching the NCBI Entrez nucleotide records from 2097 (2038 of which were cytochrome oxidase or rRNA sequences used for phylogeny studies) to 14,443 through NCBI dBEST sequence submission. Therefore, 12,345 new expressed sequence tags (ESTs) and 1 CoreNucleotide partial GAPDH sequence were added to the global database.

The libraries consisted of 9600 full length clones and 3840 from male enriched suppressive subtractive hybridisation (SSH). The full length clones were isolated from a mixed cDNA library which represented cDNA populations from different stages of the moult cycle, development, and gender. Theoretically each clone represents an entire gene transcript in a directional orientation. The clonal vector insert is then sequenced from the 5′ end which, due to the library's directional nature always represents the N-terminal of the translated protein. The reasoning behind 5′ EST sequencing is the existence of the poly A tail at the 3′ end of the transcript. Sequencing from the 3′ end often causes a base shift effect giving false or uninterpretable sequence data. The male subtractive library had previously been generated using a suppressive subtraction hybridisation method. This approach utilises a male driver cDNA population which is enriched by the removal of cDNA fragments of equal abundance present in a female cDNA population. The remaining gene fragments should represent sequences that are higher in abundance in males than females.

With respect to number of clones picked compared to the number of positive inserts, there was an excellent success rate for the mixed library (92%) and the male subtractive library (82%). This was in spite of a bacterial contamination in the mixed library that threatened to prevent it from being used in this study. Using analytical microbiological techniques, the type of contamination was identified and a solid selective media used in order to preferentially allow the library clones to grow, whilst inhibiting the contaminating bacteria.

The ESTs were annotated with a gene ontology description. Gene ontology annotation provides descriptive terms to gene objects describing cellular component, molecular function and biological process. There was only one descriptive term that showed substantial differences in annotation frequency between libraries. This is surprising as the male subtractive library should represent only gene objects found in greater abundance than that of females. This is possibly due to there being no significant changes in male cDNA abundance when compared to females. The SSH procedure doesn't eliminate all non-differentially expressed genes (those genes found in both driver and tester populations) from the produced library; it merely subtracts one (driver; female) from the other (tester; male). If most genes were equal in abundance, the SSH library may not truly represent male expression. This phenomenon has previously been described using the same subtraction kit (Ji *et al.* 2002). In this study, only tester cDNA that was at least 5-fold higher in abundance than the driver cDNA was preferentially amplified. Ji (2002) recommended that the technique should only be used to detect 'dramatic alterations' in gene expression between cDNA populations. When there was less than a 5-fold increase in any cDNA abundance in the tester cDNA population, the resultant library consisted of randomly amplified cDNA, suggesting a 'random male' cDNA library had been produced in this *G. pulex* study. This indicated that males may express genes in similar abundance to females; however, it is likely that females also have an additional group of genes involved in ovarian maturation, brood nurture and vitellogenesis. This could not be determined due to deterioration of the female subtractive library preventing its use in this study.

It was expected that gene objects associated with the androgenic gland or secretion of the androgenic gland hormone may have been evident; however, the expression patterns of these components are not known and may only be transiently expressed during development in order to differentiate juveniles into males. It is well documented that the lack of an androgenic gland leads to female development and its presence leads to male development.

Gene objects involved in spermatogenesis were also expected to be abundant in males; however, they may not have been comparable with any known sequence available on the GenBank database, not found at greater than 5-fold abundance, or expressed at extremely low levels. Male specific genes were unlikely to be identified

on the basis of sequence when, to date, there are only a few proteins and genes identified as potentially being expressed only in males, including androgenic gland hormone and an insulin-like gene.

There were 12,187 successful sequences from newly picked libraries mxAA, mxAB, maSA, maSB and maSC. Of these sequences, 14% had annotations to the term cellular component, 29% had annotations to molecular function and 27% had biological process annotations, overall 35% had at least 1 GO annotation. Therefore, 65% had no annotation. After cluster analysis identified 3917 unique sequences, 18% were assigned an annotation to the term cellular component, 32% sequences were annotated to molecular function, 22% sequences were annotated to the term biological process and overall 37% were described with at least 1 gene ontology annotation.

The only identifiable effect of suppressive subtractive hybridisation was evident from the reduction of gene objects annotated to 'structural constituent of cuticle' from 13.5% in the mixed library to 0.91% in the male subtractive library. This is expected due to the presence of cuticle related genes in both males and females, therefore no abundance should have been evident in males. However, this appears to be the only category to which a substantial change in annotation distribution was observed and therefore it is possible this occurred by chance.

Annotation efficiency increased only slightly after GOs were assigned to clustered 'unique' sequences. However, it was likely that the annotations based on these unique sequences following CLOBB clustering, based on blast database matches (**C**luster **O**n the **B**asis of **B**LAST similarity; (Parkinson *et al.* 2002)), were more accurate due to an increase in the sequence length being annotated. GO annotation was enhanced in all cases by the use of an InterPro scan, meaning that GO annotations were based on BLAST homology as well as domain and motif recognition. Clustering was performed giving 2642 singletons in a total of 3917 clusters. It can be suggested that clusters approximately represent the abundance of each unique sequence in the transcriptome, therefore indicating that singletons are of relatively low abundance. Clusters containing the highest number of sequences therefore represent the most abundant transcripts.

There are 12,346 genes (all *G. pulex* nucleotide sequences from NCBI generated from this study) contained on the *G. pulex* PartiGene database: (http://wallace.cap.ed.ac.uk/NeglectedGenomes/ARTHROPODA/Crustacea.html/) representing 3917 clusters.

10.2 IDENTIFICATION OF CANDIDATE GENES

In order to assess the impact that endocrine disrupting chemicals (EDCs) have on the freshwater shrimp *G. pulex*, an alternative approach to identify potential biomarkers was employed. In addition to the large scale approach of cDNA libraries, individual candidate genes were sought that would give a genetic 'snapshot' of an individual animal. Through Rapid Amplification of cDNA Ends (RACE) PCR amplifications an 1874bp fragment of the putative HR3 receptor was isolated. RACE PCR allows the 5' or 3' end of a gene to be isolated when only a small section of gene sequence is known, as was the case for HR3, by the utilisation of PCR primers complementary to inserted adapter sequences attached to the cDNA ends.

This is the first sequence to be identified in *Gammarus* sp. as a possible nuclear receptor and it's expression levels may give information on the reproductive status of individuals due to the role HR3 plays in both moulting and vitellogenesis. A lack of the 5' DNA binding region, suggests that either there are significant problems with incomplete reverse transcription or that a novel ecdysteroid response gene is present in these crustaceans. It suggests that this protein may function in a novel way to HR3 proteins identified in other species. Quantitative polymerase chain reaction (QPCR) analysis could be used to identify the normal physiological expression patterns of this gene in order to determine its function and possible suitability as a biomarker of endocrine disruption. In addition, the isolation of this transcription factor gene could be the basis of future work in binding and protein expression experiments, determining the ligand (potentially a steroid hormone) and the responsive element to which it binds. This could then lead to the identification of the hormone-responsive target genes. An alternative approach to identifying this gene is to probe the cDNA microarray with appropriately labelled sequence; however, this will only be successful if that particular gene is represented on the array.

So far, attempts to isolate ecdysteroid receptor (EcR), ultraspiracle protein (USP) and oestrogen receptor (ER) through degenerate PCR have not been successful. Further optimisation and use of alternative cDNA populations as DNA templates may allow the amplification of the correct sequence. A balance between specificity of primers to the gene sequence and flexibility due to codon differences would need to be made in order to successfully isolate these genes. The presence of both EcR and USP are very likely in *G. pulex* due to their presence in similarly related species. However the presence of an oestrogen receptor gene is less certain.

10.3 ANALYSIS OF MOLECULAR BIOMARKERS

After the identification of 2 potential biomarkers of gender and moult stage and determination of the gene sequence, further analysis was performed in order to assess the viability of these genes as molecular biomarkers of endocrine disruption. The genes identified were further characterised using QPCR on individuals to give a genetic 'snapshot' of physiological status.

Two candidate genes have previously been identified in *G. pulex*, the exoskeletal arthrodial cuticle protein and the female yolk protein, vitellogenin. From QPCR analysis, vitellogenin was identified as a sex determination marker. Vitellogenin was 870x more highly expressed in morphologically identified females than in males identified using the same morphological identification method. It can therefore be used as a female sex specific marker in *G. pulex*.

Most of the potential vitellogenin and cuticle protein expression trends or dose responses to 17β-oestradiol or testosterone were masked by high error levels that are possibly due to differences in age/size, pairing status, exposure time and moult stage of individuals. The variation in cuticle protein expression over the moult cycle observed in Chapter 7 suggests that this may be a reason for the variable results gathered from the exposures. In future exposures these variables would have to be removed by a longer exposure period (i.e. 100 days) in order to increase the likelihood that the animal is exposed during the critical window for disruption. The characterisation of novel putative male, female, developmental stage and moult stage markers identified in Chapter 7 and Chapter 8 should be undertaken so that a multiple

biomarker approach can be used to take into account biological variation, ensuring that the chemical dose responses to EDCs are not masked by normal physiological alterations in gene expression levels. Moult stage variability could be overcome by using same stage (for example, intermoult) *G. pulex* of approximately the same weight (although this does not represent the same age between males and females) exposed to chemicals for a longer period.

Preliminary attempts to produce moult stage synchronous populations for exposure in this study have not been successful to date. Animals of similar size were transferred to 4°C with minimal light a few days after ecdysis to try and replicate the winter acquiescent period in which animals remain in intermoult. A small number of animals (5) had been separately adjusted to this temperature (on different days), they were returned to their original temperature of 17°C and monitored for evidence of ecdysis (visible exuviate), however 3 amphipods died before completing a moult cycle, probably due to the rapidity of the change in temperature which was based on the latency of water. The remaining animals did not moult at the same time as hypothesised (±9 days). This is probably due to the wild population used, individual variations and because *G. pulex* are non-clonal organisms. Due to the inherent problems with the small sample size of this preliminary experiment I would propose that a larger study should be performed, acclimatising the animals at a slower rate in order to determine whether a synchronous moult stage population is achievable.

During this exposure study, cytochrome oxidase sequence analysis was validated as a successful method for determining species and genotype. Due to the similarity in appearance of some species of gammarids, this is a desirable approach when studying *Gammarus* sp. at the molecular level to confirm species and genotype; any small changes in these parameters could significantly alter probe binding in specificity to target DNA potentially changing the gene expression response.

10.4 OPTIMISATION OF *GAMMARUS PULEX* MICROARRAYS

During the optimisation process, a successful method of extracting amphipod RNA by using RNA*later* ICE was identified. The main problem with extracting good quality RNA from amphipods is the hard exoskeleton. Not only does it prevent the successful penetration of preservation liquids such as RNA*later*, but certain components of the cuticle can have an inhibitory effect on downstream applications.

The first obstacle, involving non-penetration of storage solution, was overcome by cutting the sacrificed animal into 2 or 3 parts depending on amphipod size. Although this improved the stability of the stored material and reduced degradation by RNases, RNA yield was still not as high as expected when sample size was considered. A relatively new storage solution became available, RNA*later* ICE, which performed the same primary capacity as the standard RNA*later* solution, inhibiting the actions of RNases and keeping the RNA intact until the samples were required.

The main difference between the 2 storage solutions is the method of tissue storage. One of the key drawbacks of using standard RNA*later* is the samples have to be added at room temperature. For amphipods, with their protective exoskeleton, submersion in RNA*later* leads to death after a significant amount of time (>15 minutes). This is long enough to induce changes in gene expression, due to induction of a stress response, possibly interfering with the basal gene expression one is attempting to observe.

It is appropriate to suggest that animals in premoult are likely to take longer to expire than those animals in the sensitive postmoult state; additionally large amphipods will take longer to perish than smaller animals. This variation in stress response can be diminished by reducing the time taken for sacrifice, the ultimate approach being sacrifice into liquid nitrogen which immediately penetrates the exoskeleton causing instantaneous death. However, use of liquid nitrogen to rapidly sacrifice amphipods, means standard RNA*later* solution cannot be used, as RNA*later* is used at room temperature. By the time amphipods have defrosted to room temperature after nitrogen sacrifice, RNases begin to partially degrade the RNA.

By placing liquid nitrogen sacrificed animals in RNA*later* ICE, animals are able to defrost whilst at -20°C (RNases are relatively inactive at this temperature) where the storage solution remains liquid as it has an alcohol base and penetrates the defrosted tissue. An additional property of RNA*later* ICE is the softening of the exoskeleton which improves the effects of homogenisation, resulting in finer ground material and consequently a higher level of cellular lysis. This produces a significant increase in RNA yield when compared to the standard RNA*later* procedure (8x higher).

The *G. pulex* libraries described in Chapter 3 were used to create microarray slides spotted with target amplicons from which differential gene expression was determined between gender, moult stage, developmental stage (small juvenile, medium juvenile and adult) and cross-species hybridised *E. marinus*. Microarray technology allows the determination of differential gene expression of, in this case, 13,440 gene objects, to give a large quantity of data on both the normal and disrupted molecular responses of *G. pulex* or *E. marinus* through the creation of gene expression profiles. Other methods of analysing differential gene expression such as differential display are prone to false positives. AFLP (Amplified Fragment Length Polymorphisms) takes considerable time to optimise, requires the use of radioactive material and the expense of a fluorescent kit cannot necessarily be justified when compared to the cost of transcriptome profiling considering the vast amount of data collected from gene expression profiling.

The microarray optimisation procedure is a complex and detailed course of action, due to the sensitivity and variability of gene expression profiling, the sources of technical error and variation need to be systematically reduced. Each sample is required to be treated in exactly the same way when a large number of samples are used; randomisation of sample processing is the best approach for this. In addition, slides need to be as replicable as possible with low levels of between slide variations. Although there are normalisation methods available for post hybridisation data, there is a limit to variation and noise that data normalisation can correct without affecting biological variation. Therefore the robust method for printing and target preparation utilised in this study allowed for consistent and successful production of valid microarray data.

10.5 CHANGES IN GENE EXPRESSION IN ADULT *GAMMARUS PULEX*

In order to examine moult stage specific transcriptomic responses *G. pulex* were collected at set points throughout their moult cycle and moult stage specific cDNA produced. Amphipods within a size range of 7mm-12mm were incubated at 17°C with a 16 hour light, 8 hour dark cycle per day, the moult cycle length was ascertained for each individual and specimens collected at approximate stages C_2, C_4 / early D_0 and mid D_1, 2 samples were also collected at D_2/D_3 as close to predicted ecdysis as possible. The moult cycle for all animals was ~32±1 days, 28±1 days for females and 37±2 days for males at 17°C (mean±standard error). These samples were used to probe the *G. pulex* microarray slides to show differential gene expression within a gene expression profile based on the different moult cycle stages.

The second aim of these microarray experiments was to identify gender specific differences in gene expression. A volcano plot indicated that, surprisingly, only a small number of genes were differentially expressed between males and females compared to the substantial quantity between 'Stage 1' and 'Stage 3' amphipods. This may be due to the lack of female and male representation on the microarray through the absence of a female subtractive library and due to the inadequate SSH of the male subtractive library. Hierarchical clustering, represented by a dendrogram, did reveal distinct differences between the genders which were reflected by Principal Component Analysis (PCA). A large number of the female up-regulated 'clusters' were vitellogenin and haemocyanin, and the few up-regulated male gene objects represented by a small cluster were stated previously in Chapter 7. The data suggests that very few genes are involved in genetic differentiation in adult *G. pulex* and the majority of these are female-specific.

Through this moult stage and gender study, 3 ways of confirming or identifying gender have been developed or tested: QPCR, PCA and surface area/moult cycle length analysis. As first described in Chapter 5, vitellogenin gene expression levels in females were successfully used to determine gender amongst a mixture of male, female and undescribed amphipods. The method was able to identify the biological females (determined by precopulatory pairing position) and the biological males

(determined in the same way). Only 2 of the 36 animals fell outside the 95% confidence intervals for the male and female groups.

After microarray image analysis, PCA plot creation confirmed the QPCR results with male samples grouping to the left and females partitioned on the right. When the microarray data from unknown animals were added to this PCA plot, they fell neatly into the gender group that had been identified by QPCR. The 2 samples that were not identifiable by QPCR could be clearly identified by which side of the PCA plot they resided.

When a selection of animals from the same group were plotted onto a graph of surface area vs. moult cycle length, another gender specific pattern was observed. This was unexpected as animals were selected as pairs and assigned as male or female, or for the unknown samples, random animals of between 7mm and 12mm were selected from the mixed population. This was presumed to represent males and females of all sizes, however there is a clear divide between genders, evident from the graph. It was unexpected that all the larger animals were males and all the smaller animals were female. Although on average males are bigger than females, one would expect some of the animals from the smaller end of the size range to be young male adults and for there to be an overlap at the lower sizes consisting of females and young male adults. Why the distinctive divide was observed is not known. The possibility of animals of each gender being close to their maximum size due to the time of year collected is a possibility, but in order to remove any possibility of seasonal variation these samples were collected at different times of the year.

Another possibility is that males spend a relatively short time at smaller sizes and rapidly grow to a substantial size in order to be successful in precopulatory pairing and mating; males will only pair with females smaller then themselves. An alternative approach is to consider that males generally have a longer moult cycle length than females, which remains true even for males and females whose surface area was equal, indicating that this longer moult cycle length is not entirely size related. Further studies using area and moult cycle length could be used to determine if the rate of area increase compared to increase in moult cycle length is different between genders which could account for the gender separation.

It is possible that the hypothesis that males spend a shorter amount of time as juveniles and the observed increased moult cycle length in larger mature adults are related. As a male, more energy could be invested into growth rather than reproduction (there is a higher reproductive cost being a female due to oocyte production and brood nurture); it would in theory be better for a male to grow quickly to reach maturity faster. Hence, they may have shorter moults as juveniles than females, subsequently, when they reach maturity, male adults of reproductive size increase the time taken to go through one moult cycle. This hypothesis could be tested through monitoring moult cycle lengths and changes in surface area (i.e. size) through development from neonate to mature adults. This study provided 3 alternative techniques showing significant trends in relation to gender identification and determination, although the method of area vs. moult cycle length should be approached with caution until the reasons for such substantial differentiation can be ascertained.

Additionally, a more cost effective method for the determination of vitellogenin gene expression level has been optimised in *G. pulex*. Semi-quantitative PCR is a method used to visually assess the level of gene expression when compared to expression of a housekeeping gene such as glyceraldehyde-3-phosphate dehydrogenase (GAPDH). Both GAPDH and vitellogenin primer pairs are used within the same reaction creating a duplex PCR system for gender determination. Amplicon sizes were designed to be different so that they could be run within the same reaction vessel. By comparing amplification using 2 cDNA samples side-by-side, a female sample showed a clear band representing vitellogenin as early as cycle 15, 20 cycles earlier than the appearance of GAPDH. Higher cycle numbers represent a lower quantity of corresponding transcript, and consequently expression level. In the male sample, no vitellogenin gene fragment was observable; however GAPDH was visible earlier than that of the female sample, at cycle 25. This suggests that despite a larger quantity of cDNA being used as a DNA template in the male PCR, no vitellogenin amplicon was evident. It can be concluded that there was no vitellogenin expression in the males, in contrast with an observable gene expression level in the female, supporting the more costly QPCR results. Due to 5 samples being generated per specimen, in order to assess at which cycle amplicon bands are visible, large sample numbers would make this process very laborious. The use of this semi-quantitative method in pooled

samples could be very useful to assess average levels of vitellogenin in a mixed population, giving a rapid indication of gender bias, or more accurately, changes in vitellogenin expression trends in large random samples representing wild populations. Once a standard curve was produced, it could be used to measure vitellogenin expression responses across wild populations exposed to different EDCs. This approach remains to be validated.

The one statistically relevant gene of interest up-regulated in male *G. pulex*, when compared to females, was homologous to non-SMC element 1 homologue for *Xenopus tropicalis* which contains a domain found in subunits of the non-structural maintenance of chromosomes (SMC) complex. This is involved in genomic stability though DNA repair and DNA metabolism (Fujioka *et al.* 2002). It has been discovered in the grasshoppers *Eyprepocnemis plorans* and *Locusta migratoria*, that 3 cohesin subunits are evident in prophase I spermatocytes, 2 of these cohesins were 'non-structural maintenance of chromosome subunits' including RAD21, contributing to sister chromatid cohesion (Valdeolmillos *et al.* 2007). The non-SMC subunit RAD21 has also been observed to be mainly expressed in mitotically dividing cells in the testis (somatic cells and spermatogonia) in the medaka fish (Iwai *et al.* 2004). Although gene expression of RAD21 is not male-specific, its role in spermatogenesis in grasshoppers may suggests a reason for up-regulation in males.

Other markers of gender identified included chaoptin, up-regulated in females, and further confirmation that vitellogenin is also up-regulated in females. Short fragments of confirmed *G. pulex* vitellogenin gene (containing the same section amplified by QPCR) showed similar patterns of expression with female expression up-regulated by ~2.5x and down-regulated by ~0.3x in males. As expected, haemocyanin also appeared as up-regulated in females, although to a lesser extent than vitellogenin. This is most likely due to the sequence similarity between the von Willebrand domains contained in both genes. This highlights the problems of false positives using a platform such as DNA microarrays and highlights the importance of further sequence analysis and subsequent QPCR validation.

Moult stage was assessed using both cuticle protein gene expression analyses by QPCR and gene expression profiling using microarray technology. QPCR confirmed

cuticle protein as a moult cycle stage marker, being up-regulated in Stage 1 and decreasing to Stage 3. Other novel *G. pulex* genes that may be potential moult stage markers identified though gene expression profiling included chitinase, which was up-regulated in Stage 1 by a 2.2x fold change, decreasing to 1.5x in Stage 2 and 0.8x in Stage 3. Alternatively, the larval glue protein was up-regulated in stage 1 animals and α-amylase was up-regulated in Stage 2 animals. Additionally, the sensory appendage protein and piopio were up-regulated in Stage 1 of *G. pulex*.

This study has highlighted alterations in carbohydrate metabolic process, reproduction and development and changes in cuticle or chorion formation. PCA plots for moult staged animals revealed a clear circular trend representing the moult cycle stages. There were some outliers, but this would be expected considering the crude method used for moult stage identification. No morphological assessment of moult stage was undertaken, so if an individual animal (the data was generated on individual not pooled samples) delayed or accelerated their moult cycle, the amphipods may have been collected at the wrong stage. Delay of moulting has been observed in females, in the absence of males (Sutcliffe 1993). This may explain why the outliers always reside in the previous stage, for example, 2 'Stage 2' animals are found in 'Stage 1' and 2 'Stage 3' animals are found in 'Stage 2' indicating some individuals have slowed their moult cycle. In fact, 4 out of the 5 outliers found in the 'wrong' group are females, reinforcing this hypothesis. Volcano plots of moult stage, comparing 'Stage 1' to 'Stage 3' highlighted a suitable quantity of differentially expressed genes. Hierarchical clustering, represented by a dendrogram, showed clear moult stage trends indicating a large number of gene objects represented on the microarray slides are involved in the moult cycle highlighting the importance of cyclic ecdysis in this species.

The interpretation of gene profile data was inhibited by insufficient sequence information for amphipods and the number of uncharacterised proteins, leading to poor GO annotation. This study highlighted some genes of interest; however there are a large number of genes identified that may be potential biomarkers of gender and moult stage. Further analysis of the remaining genes of interest may lead to further understanding of the molecular mechanisms involved in sexual differentiation and moulting.

10.6 GENE EXPRESSION CHANGES IN DIFFERENT DEVELOPMENTAL STAGES OF *GAMMARUS PULEX*

Analysing changes between transcriptome profiles during developmental maturation in crustaceans has increased our understanding of the differential gene expression in response to developmental stage. PCA plot reveals adults grouped together on the left with juveniles and neonates overlapping from left to right. A volcano plot of the most extreme datasets, adults and neonates shows a large number of differentially regulated gene objects. Hierarchical clustering in the form of a dendrogram also highlights the large differences evident in adults compared to non-mature animals. This is not likely to be an issue of an increase in animal size and therefore an increase in gene expression, as all samples were normalised by putting an equal volume of RNA into the reverse transcription, and consequently an equal number of incorporated dye onto the slide (70pmols). The small number of up-regulated genes in the juveniles and neonates could be due to the distribution of RNA populations used to create the libraries. A significantly larger amount of adult RNA was used compared to juvenile and neonate.

A putative glycine/tyrosine-rich eggshell protein was up-regulated during maturity and peaked in adults, there was also an up-regulation in fibroin 4 in adults. Fibroin is silk that spiders spin in order to make webs and produce egg cases and therefore of unknown putative function in *G. pulex*, this homologue could be related to the 'amphipod silk' that some crustaceans create for their tubular nest or burrow, although its production in gammarids has not been recorded to date. A down-regulation of 2 gene fragments homologous to the pupal cuticle protein was evident in adult *G. pulex*. Vitellogenin up-regulation in animals above 5mm was evident; this correlates with onset of adult maturation.

The interpretation of gene profile data was inhibited by insufficient sequence information for amphipods and the number of uncharacterised proteins, leading to poor GO annotation. This discussion is a brief overview of some genes of interest; however there are a large number of genes identified that may be potential biomarkers of developmental stage. Further analysis of the remaining genes of interest shown in Appendix B, Table 3 may lead to further understanding the molecular mechanisms involved in *G. pulex* development.

10.7 DETERMINING DIFFERENTIAL GENE EXPRESSION IN INTERSEX AND NORMAL *ECHINOGAMMARUS MARINUS*

In addition to different sex, moult stages, EDC exposures and life stages a cross-species hybridisation reaction was performed with cDNA from *Echinogammarus marinus*, a similar species to *G. pulex*. This species has been used extensively in studying the development and occurrence of secondary sexual characteristics in amphipods (Ford *et al.* 2003; Ford *et al.* 2004a; Ford *et al.* 2004b; Ford and Fernandes 2005; Ford *et al.* 2005; Ford *et al.* 2006; Ford *et al.* 2007a; Ford *et al.* 2007b; Ford *et al.* 2008). Phylogenetic analysis using cytochrome oxidase I displayed a close similarity between *G. pulex* and *E. marinus* at the genetic level, suggesting they were suitable species for heterologous hybridization. Samples that had been identified as intersex were compared with cDNA from normal organisms. This work was possible due to collaboration with Dr Alex Ford (Environmental Research Institute, North Highland College, UHI Millennium Institute, Scotland UK). The results from this experiment gave an insight into the molecular mechanisms involved in the intersex state found in many EDC exposed organisms and highlighted novel biomarkers of endocrine disruption.

PCA analysis of intersex samples revealed two distinct clusters for normal males and normal females with both intersex phenotypes overlapping; this is contrary to morphological observations where intersex males and intersex females show distinct phenotypes (Ford *et al.* 2008). This indicates that a relatively small number of genes may be responsible for the formation of secondary sexual characteristics in normal males and females. Despite the results from PCA analysis, potential individual biomarkers of each phenotype, or a combination of phenotypes were identified. Volcano plots reflect the PCA plot with the fewest genes varying between the 2 intersex phenotypes. Differential expression between normal phenotypes was slightly higher, followed by differences between normal females and intersex females and normal males and intersex males. The volcano plots therefore suggest that there are greater differences between normal and intersex phenotypes for each males or females than between the 2 normal or 2 intersex phenotypes.

Hierarchical clustering, represented by a dendrogram, illustrated the similar and divisive gene objects. This also highlighted the similarity in gene expression profiles between the intersex males and females compared to the normal males and females. The main problem with the *E. marinus* gene expression profiling was the 2 different locations used to collect the specimens. The majority of intersex specimens were collected from Inverkeithing Bay, Scotland, which is in close proximity to a ship breaker's yard and a paper mill releasing effluent. High levels of PCBs and heavy metals have been recorded here (Ford *et al.* 2004b). The majority of normal specimens were collected from Langland Bay, Gower Peninsula, South Wales. In spite of this, they were collected at the same time of year to eliminate seasonal variation. The differing locations may have caused the results to appear contrary to morphological variations. For this reason any comparisons between groups must be further validated by a sequence specific technique such as QPCR.

The collection of animals from different sites arose after substantial problems with probe preparation. As previously explained, RNA*later* is used to store specimens until RNA extraction is required, RNA*later* ICE was not used in this case, and so animals for the entire experiment were collected and sacrificed at room temperature. This may have led to the problems described below. Initially all samples were collected from Inverkeithing Bay, Scotland. Significant degradation was identified in the normal phenotype samples, possibly due to the larger number of animals per vessel compared to the intersex samples. Due to the larger size of *E. marinus* compared to *G. pulex* the RNAlater solution insufficiently penetrated throughout the tissue to inactivate the RNase enzymes.

Due to time constraints, a second batch of animals were collected and identified from a local site (Langland Bay, Gower Peninsula, South Wales) after which RNA extraction was successful. This unexpected change in experimental design means data will have to be carefully analysed in order to prevent the identification of false positives arising from location variation. All normal vs. intersex putative biomarkers would have to be thoroughly investigated to ascertain whether it was a true biological response to the intersex phenotype.

After successful RNA extraction, a further problem was encountered when probes were prepared for microarray hybridisation. RNA passing all quality control checks was reverse transcribed using aminoallyl-dUTP yielding an acceptable quantity of aa-cDNA. However when coupling was attempted between aa-cDNA and Cy3 mono-reactive NHS ester, no incorporation of fluorescent dye was observed. After a number of alternative RNA extractions and use of alternative reverse transcriptase enzymes to check the previous steps, the key problem was identified as the final coupling step. *G. pulex* aa-cDNA was coupled alongside *E. marinus* aa-cDNA using identical protocols and reagents. Successful *G. pulex* probes were prepared, but yet again the *E. marinus* samples failed.

It was concluded that there must be an inhibitor of dye-coupling partitioning with the RNA, which cannot be removed by subsequent column purification. The contaminant was not evident from a spectral scan from 190nm to 650nm indicating it was not protein, phenol- or alcohol-based. The inhibitor could either be binding to the aminoallyl dUTP through the amino group, or binding to the CyDye via its ester group. There is a possibility that chitin was binding to the ester group of the CyDye via its amine group, preventing the dye from binding to the aminoallyl group incorporated into the cDNA. Chitin is a likely candidate due to its abundance in arthropodal exoskeletons. The larger size of *E. marinus* may have meant that an upper threshold of chitin was passed in these RNA samples. The problem was overcome by the addition of a lithium chloride precipitation step following the standard Tri Reagent RNA extraction method. LiCl is often used to selectively precipitate RNA away from DNA, carbohydrates and proteins.

In the cross-species hybridised microarrays, ATP synthase F_0 subunit 6 was seen to be up-regulated in normal females and intersex males which would possibly be useful as a marker of feminisation. ICHIT protein was up-regulated in intersex males of *E. marinus* with no up-regulation in any other group, suggesting it as an indicator of de-masculinisation. Gelsolin was up-regulated in intersex and normal males, therefore representing a possible male induced biomarker that still retains its male expression pattern despite the formation of secondary sexual characteristics. Similarly, a chitinase 3 precursor was up-regulated in intersex and normal female *E. marinus*, and therefore is a putative female induced biomarker that still retains its female expression

pattern despite the formation of secondary sexual characteristics. An efficacious gammarid cross-species microarray platform has been developed and utilised successfully using *E. marinus* probes heterologously hybridised to *G. pulex* target genes to identify possible biomarkers of sexual phenotype. Heat shock cognate protein 70 and the major larval glue protein were both observed to be consistently up-regulated in normal *E. marinus*, as previously explained. Further studies of these genes will have to be undertaken to assess whether the up-regulation is purely a site-specific up-regulation in specimens from Langland Bay.

10.8 CONCLUDING REMARKS AND FURTHER WORK

In addition to previous suggestions on possible supplementary work and the value of the collated results obtained during this study, there are a certain areas which would benefit from further work.

The potential biomarkers of physiological and disrupted status should be characterised further, by full length gene sequencing of each clone using PCR and RACE PCR. With the full length gene isolated, the translated amino acid sequence could be used to identify the gene more accurately through sequence homology. QPCR could then be used to assess the expression profiles of different individuals by using a gene-specific probe. The gene expression profiles of individuals with varying moult stage, developmental stage, gender and subsequently EDC exposed individuals may help to gain an understanding of the biological processes the biomarker is involved with and it's molecular function.

In addition to further gene characterisation, the development of a rapid assay to identify physiological status has the potential to produce a novel and rapid assay for biological assessment of individuals. Using a combination of multiple moult stage, gender and developmental genetic markers, representing different mechanisms, in one test an instant 'snapshot' of an individual's age, gender and moult stage would be provided. For example, ideally 3 biomarkers of each moult stage, gender and developmental stage (neonates, juveniles and adults) could create a specific 'pattern' of up- and down-regulated genes (i.e. visually red or green) providing ~25 'traffic-lights'. Once this test had been calibrated to normal individuals to identify normal

'patterns' of expression, exposed specimens could be analysed; any alteration in a single biomarker creating an unrecognised 'pattern' would indicate a disruption to the related mechanism or pathway.

Further libraries could be created to give a larger number of genes relating to a particular tissue, moult stage or developmental stage. Moult staged libraries of C_2 (Stage 1), late C_4/early D_0 (Stage 2), mid D_1 (Stage 3) and D_2/D_3 (Stage 4) and developmental libraries of neonates and juveniles would provide additional genetic information on these major physiological events throughout the life cycle. Intersex and normal SSH libraries of *E. marinus* would also be desirable to reflect pollutant disrupted individuals and those with secondary sexual characteristics. Using these libraries to generate new microarrays would potentially give a more efficient and relevant system for determining further biological pathways involved in endocrine disruption in amphipods.

This study has resulted in the submission of 12,345 *G. pulex* EST sequences to the NCBI GenBank database and the generation of the first *G. pulex* online genetic database:
http://wallace.cap.ed.ac.uk/NeglectedGenomes/ARTHROPODA/Crustacea.html.
Future work will build on these results to create an effective and novel tool for identifying molecular responses in *Gammarus pulex*.

The first nuclear receptor sequence to be identified in *Gammarus* sp. has been partially isolated, with the potential to give information on the reproductive status of individuals due to HR3s involvement in both moulting and vitellogenesis. Cuticle protein has been identified as a moult stage marker and vitellogenin as a sex determination marker. Attempts to isolate the EcR and USP should be continued through degenerate PCR in order to assess and potentially use these as biomarkers of normal and disrupted reproduction. The *G. pulex* microarray platform was successfully used in cross-species hybridisation with *E. marinus* to identify novel gender and intersex gene expression differences in this related gammarid species.

The experimental processes and results presented in this thesis have significantly enhanced not only the quantity of molecular knowledge of the freshwater amphipod

G. pulex, but also offers great insight into the genetic profiles of different genders, developmental and moult cycle stages in *G. pulex*. It also provides important information on endocrine disruption and intersex gene expression profiles in a related species *E. marinus*. A major potential output from this study is the production of a suite of novel and established molecular markers to potentially detect the early stages of endocrine disruption in the ubiquitous benthic invertebrate, *G. pulex* and related amphipods.

REFERENCES

Abramowitz, A. A., F. L. Hisaw and D. N. Papandrea (1944). "The occurence of a diabetogenic factor in the eyestalks of crustaceans." Biological Bulletin **86**(1): 1-5.

Agarwal, A., W. S. B. Yeung and K. F. Lee (2002). "Cloning and characterization of the human oviduct-specific glycoprotein (HuOGP) gene promoter." Molecular Human Reproduction **8**(2): 167-75.

Allen, Y., P. Matthiessen, A. P. Scott, S. Haworth, S. Feist and J. E. Thain (1999a). "The extent of oestrogenic contamination in the UK estuarine and marine environments--further surveys of flounder." Science of the Total Environment **233**(1-3): 5-20.

Allen, Y., A. P. Scott, P. Matthiessen, S. Haworth, J. E. Thain and S. Feist (1999b). "Survey of estrogenic activity in United Kingdom estuarine and coastal waters and its effects on gonadal development of the flounder *Platichthys flesus*." Environmental Toxicology and Chemistry **18**(8): 1791-800.

Altschul, S. F., Thomas L. Madden, Alejandro A. Schäffer, Jinghui Zhang, Zheng Zhang, Webb Miller, and David J. Lipman (1997). "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs." Nucleic Acids Research **25**: 3389-402.

Alzieu, C. (2000). "Impact of Tributyltin on Marine Invertebrates." Ecotoxicology **9**(1): 71-76.

Andersen, H. R., L. Wollenberger, B. Halling-Sorensen and K. O. Kusk (2001). "Development of copepod nauplii to copepodites--a parameter for chronic toxicity including endocrine disruption." Environmental Toxicology and Chemistry **20**(12): 2821-9.

Andersen, S. O. (1999). "Exoskeletal proteins from the crab, *Cancer pagurus*." Comparative Biochemistry and Physiology Part A: Molecular and Integrative Physiology **123**(2): 203-11.

Archer, S. K., C. A. Behm, C. Claudianos and H. D. Campbell (2004). "The Flightless I protein and the gelsolin family in nuclear hormone receptor-mediated signalling." Biochemical Society Transactions **32**: 940-42.

Ashburner, M., C. A. Ball, J. A. Blake, D. Botstein, H. Butler, J. M. Cherry, A. P. Davis, K. Dolinski, S. S. Dwight, J. T. Eppig, M. A. Harris, D. P. Hill, L. Issel-Tarver, A. Kasarskis, S. Lewis, J. C. Matese, J. E. Richardson, M. Ringwald, G. M. Rubin and G. Sherlock (2000). "Gene ontology: tool for the unification of biology. The Gene Ontology Consortium." Nature Genetics **25**(1): 25-9.

Azumi, K., M. Fujie, T. Usami, Y. Miki and N. Satoh (2004). "A cDNA microarray technique applied for analysis of global gene expression profiles in tributyltin-exposed ascidians." Marine Environmental Research **58**(2-5): 543-6.

Badiee, A., H. G. Eiken, V. M. Steen and R. Lovlie (2003). "Evaluation of five different cDNA labeling methods for microarrays using spike controls." BMC Biotechnology **3**: 23.

Baker, M. E. (1988). "Invertebrate vitellogenin is homologous to human von Willebrand factor." Biochemical Journal **256**(3): 1059-61.

Barki, A., I. Karplus, R. Manor and A. Sagi (2006). "Intersexuality and behavior in crayfish: the de-masculinization effects of androgenic gland ablation." Hormones and Behavior **50**(2): 322-31.

Barnard, J. L. and G. S. Karaman (1991). "The Families and Genera of Marine Gammaridean Amphipoda:(except Marine Gammaroids)." Records of the Australian Museum Supplement **13**(1/2): 1-866.

Bede, J. C., P. E. A. Teal, W. G. Goodman and S. S. Tobe (2001). "Biosynthetic Pathway of Insect Juvenile Hormone III in Cell Suspension Cultures of the Sedge *Cyperus iria*." Plant Physiology **127**(2): 584-93.

Billinghurst, Z., A. S. Clare, K. Matsumura and M. H. Depledge (2000). "Induction of cypris major protein in barnacle larvae by exposure to 4-n-nonylphenol and 17β-oestradiol." Aquatic Toxicology **47**(3-4): 203-12.

Block, D. S., A. C. Bejarano and G. Thomas Chandler (2003). "Ecdysteroid concentrations through various life-stages of the meiobenthic harpacticoid copepod, *Amphiascus tenuiremis* and the benthic estuarine amphipod, *Leptocheirus plumulosus*." General and Comparative Endocrinology **132**(1): 151-60.

Blockwell, S. J., D. Pascoe and E. J. Taylor (1996). "Effects of lindane on the growth of the freshwater amphipod *Gammarus pulex* (L.)." Chemosphere **32**(9): 1795-803.

Blockwell, S. J., E. J. Taylor, I. Jones and D. Pascoe (1998). "The influence of fresh water pollutants and interaction with *Asellus aquaticus* (L.) on the feeding activity of *Gammarus pulex* (L.)." Archives of Environmental Contamination and Toxicology **34**(1): 41-7.

Bollache, L. and F. Cezilly (2004). "State-dependent pairing behaviour in male *Gammarus pulex* (L.) (Crustacea, Amphipoda): effects of time left to moult and prior pairing status." Behavioural Processes **66**(2): 131-37.

Bollenbacher, W. E., S. L. Smith, J. J. Wielgus and L. I. Gilbert (1977). "Evidence for an α-ecdysone cytochrome P-450 mixed function oxidase in insect fat body mitochondria." Nature **268**(5621): 660-63.

Bowman, C. J., K. J. Kroll, M. J. Hemmer, L. C. Folmar and N. D. Denslow (2000). "Estrogen-induced vitellogenin mRNA and protein in sheepshead minnow (*Cyprinodon variegatus*)." General and Comparative Endocrinology **120**(3): 300-13.

Breitholtz, M. and B.-E. Bengtsson (2001). "Oestrogens have no Hormonal Effect on the Development and Reproduction of the Harpacticoid Copepod *Nitocra spinipes*." Marine Pollution Bulletin **42**(10): 879-86.

Brian, J. V., C. A. Harris, M. Scholze, T. Backhaus, P. Booy, M. Lamoree, G. Pojana, N. Jonkers, T. Runnalls, A. Bonfa, A. Marcomini and J. P. Sumpter (2005). "Accurate prediction of the response of freshwater fish to a mixture of estrogenic chemicals." Environmental Health Perspectives **113**(6): 721-8.

Brodie, D. A. and K. Halcrow (1977). "The ultrastructure of the sinus gland of *Gammarus oceanicus* (Crustacea: Amphipoda)." Cell and Tissue Research **182**(4): 557-64.

Brown, R. J., M. Conradi and M. H. Depledge (1999). "Long-term exposure to 4-nonylphenol affects sexual differentiation and growth of the amphipod *Corophium volutator* (Pallas, 1766)." Science of the Total Environment **233**(1): 77-88.

Buckley, B. A. (2007). "Comparative environmental genomics in non-model species: using heterologous hybridization to DNA-based microarrays." Journal of Experimental Biology **210**(Pt 9): 1602-6.

Bulau, P., I. Meisen, B. Reichwein-Roderburg, J. Peter-Katalinic and R. Keller (2003). "Two genetic variants of the crustacean hyperglycemic hormone (CHH) from the Australian crayfish, *Cherax destructor*: detection of chiral isoforms due to posttranslational modification." Peptides **24**(12): 1871-9.

Bultelle, F., M. Panchout, F. Leboulenger and J. M. Danger (2002). "Identification of differentially expressed genes in *Dreissena polymorpha* exposed to contaminants." Marine Environmental Research **54**(3-5): 385-9.

Bustin, S. A. (2000). "Absolute quantification of mRNA using real-time reverse transcription polymerase chain reaction assays." Journal of Molecular Endocrinology **25**(2): 169-93.

Chan, S.-M. (1998). "Cloning of a shrimp (*Metapenaeus ensis*) cDNA encoding a nuclear receptor superfamily member: an insect homologue of E75 gene." FEBS Letters **436**(3): 395-400.

Chan, S., S. M. Rankin and L. L. Keeley (1988). "Characterization of the Molt Stages in *Penaeus vannamei*: Setogenesis and Hemolymph Levels of Total Protein, Ecdysteroids, and Glucose." Biological Bulletin **175**(2): 185.

Chang, E. S. (1993). "Comparative endocrinology of molting and reproduction: insects and crustaceans." Annual Review of Entomology **38**: 161-80.

Chang, E. S. (1995). "Physiological and biochemical changes during the molt cycle in decapod crustaceans: an overview." Journal of Experimental Marine Biology and Ecology **193**(1-2): 1-14.

Chang, E. S. (2001). "Crustacean Hyperglycemic Hormone Family: Old Paradigms and New Perspectives." Integrative and Comparative Biology **41**(3): 380-88.

Charniaux-Cotton, H. (1985). "Vitellogenesis and Its Control in Malacostracan Crustacea." Integrative and Comparative Biology **25**(1): 197-206.

Cheek, A. O., T. H. Brouwer, S. Carroll, S. Manning, J. A. McLachlan and M. Brouwer (2001). "Experimental evaluation of vitellogenin as a predictive biomarker for reproductive disruption." Environmental Health Perspectives **109**(7): 681-90.

Chen, H. Y., R. D. Watson, J. C. Chen, H. F. Liu and C. Y. Lee (2007). "Molecular characterization and gene expression pattern of two putative molt-inhibiting hormones from *Litopenaeus vannamei*." General and Comparative Endocrinology **151**(1): 72-81.

Chen, Y. N., D. Y. Tseng, P. Y. Ho and C. M. Kuo (1999). "Site of vitellogenin synthesis determined from a cDNA encoding a vitellogenin fragment in the freshwater giant prawn, *Macrobrachium rosenbergii*." Molecular Reproduction and Development **54**(3): 215-22.

Cheng, C. (1942). "On the fecundity of some gammarids." Journal of the Marine Biological Association of the United Kingdom **25**: 467-75.

Cho, W.-L., M. Z. Kapitskaya and A. S. Raikhel (1995). "Mosquito ecdysteroid receptor: Analysis of the cDNA and expression during vitellogenesis." Insect Biochemistry and Molecular Biology **25**(1): 19-27.

Chung, A. C. K., D. S. Durica, S. W. Clifton, B. A. Roe and P. M. Hopkins (1998a). "Cloning of crustacean ecdysteroid receptor and retinoid-X receptor gene homologs and elevation of retinoid-X receptor mRNA by retinoic acid." Molecular and Cellular Endocrinology **139**(1-2): 209-27.

Chung, A. C. K., D. S. Durica and P. M. Hopkins (1998b). "Tissue-Specific Patterns and Steady-State Concentrations of Ecdysteroid Receptor and Retinoid-X-Receptor mRNA during the Molt Cycle of the Fiddler Crab, *Uca pugilator*." General and Comparative Endocrinology **109**(3): 375-89.

Chung, J. S., H. Dircksen and S. G. Webster (1999). "A remarkable, precisely timed release of hyperglycemic hormone from endocrine cells in the gut is associated with ecdysis in the crab *Carcinus maenas*." Proceedings of the National Academy of Sciences of the United States of America **96**(23): 13103-07.

Chung, J. S. and S. G. Webster (2003). "Moult cycle-related changes in biological activity of moult-inhibiting hormone (MIH) and crustacean hyperglycaemic hormone (CHH) in the crab, *Carcinus maenas*. From target to transcript." European Journal of Biochemistry **270**(15): 3280-8.

Colborn, T., F. S. vom Saal and A. M. Soto (1993). "Developmental effects of endocrine-disrupting chemicals in wildlife and humans." Environmental Health Perspectives **101**(5): 378-84.

Cold, A. and V. E. Forbes (2004). "Consequences of a short pulse of pesticide exposure for survival and reproduction of *Gammarus pulex*." Aquatic Toxicology **67**(3): 287-99.

Conesa, A., S. Gotz, J. M. Garcia-Gomez, J. Terol, M. Talon and M. Robles (2005). "Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research." Bioinformatics **21**(18): 3674-76.

Copeland, P. A., J. P. Sumpter, T. K. Walker and M. Croft (1986). "Vitellogenin levels in male and female rainbow trout (*Salmo gairdneri* Richardson) at various stages of the reproductive cycle." Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology **83**(2): 487-93.

Corpet, F. (1988). "Multiple sequence alignment with hierarchical clustering." Nucleic Acids Research **16**(22): 10881-90.

Cruz, J., D. Martin and X. Belles (2007). "Redundant ecdysis regulatory functions of three nuclear receptor HR3 isoforms in the direct-developing insect *Blattella germanica*." Mechanisms of Development **124**(3): 180-9.

D'Avino, P. P., S. Crispi, L. Cherbas, P. Cherbas and M. Furia (1995). "The moulting hormone ecdysone is able to recognize target elements composed of direct repeats." Molecular and Cellular Endocrinology **113**(1): 1-9.

Dam, E., B. Styrishave, K. F. Rewitz and O. Andersen (2006). "Intermoult duration affects the susceptibility of shore crabs *Carcinus maenas* (L.) to pyrene and their ability to metabolise it." Aquatic Toxicology **80**(3): 290-7.

de Kleijn, D. P., K. P. Janssen, S. L. Waddy, R. Hegeman, W. Y. Lai, G. J. Martens and F. Van Herp (1998). "Expression of the crustacean hyperglycaemic hormones and the gonad-inhibiting hormone during the reproductive cycle of the female American lobster *Homarus americanus*." Journal of Endocrinology **156**(2): 291-8.

Denslow, N. D., C. J. Bowman, R. J. Ferguson, H. S. Lee, M. J. Hemmer and L. C. Folmar (2001a). "Induction of gene expression in sheepshead minnows (*Cyprinodon variegatus*) treated with 17β-estradiol, diethylstilbestrol, or ethinylestradiol: the use of mRNA fingerprints as an indicator of gene regulation." General and Comparative Endocrinology **121**(3): 250-60.

Denslow, N. D., H. S. Lee, C. J. Bowman, M. J. Hemmer and L. C. Folmar (2001b). "Multiple responses in gene expression in fish treated with estrogen." Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology **129**(2-3): 277-82.

Desbrow, C., E. J. Routledge, G. C. Brighty, J. P. Sumpter and M. Waldock (1998). "Identification of estrogenic chemicals in STW effluent. 1. Chemical fractionation and in vitro biological screening." Environmental Science & Technology **32**(11): 1549-58.

DeSouza, M. M. (1995). "An estrogen-dependent secretory protein, which shares identity with chitinases, is expressed in a temporally and regionally specific manner in the sheep oviduct at the time of fertilization and embryo development." Endocrinology **136**(6): 2485-96.

DeSouza, M. M. and M. K. Murray (1995). "An estrogen-dependent sheep oviductal glycoprotein has glycan linkages typical of sialomucins and does not contain chitinase activity." Biology of Reproduction **53**(6): 1517-26.

Devaraj, H. and A. Natarajan (2006). "Molecular mechanisms regulating molting in a crustacean." FEBS Journal **273**(4): 839-46.

Dimopoulos, G., D. Seeley, A. Wolf and F. C. Kafatos (1998). "Malaria infection of the mosquito *Anopheles gambiae* activates immune-responsive genes during critical transition stages of the parasite life cycle." EMBO (European Molecular Biology Organization) Journal **17**: 6115-23.

Dougherty, J. J., R. K. Puri and D. O. Toft (1984). "Polypeptide components of two 8 S forms of chicken oviduct progesterone receptor." Journal of Biological Chemistry **259**(12): 8004-9.

Drinking Water Inspectorate (2002). Annual Report for Wales.

Dudley, A. M., J. Aach, M. A. Steffen and G. M. Church (2002). "Measuring absolute expression with microarrays with a calibrated reference sample and an extended signal intensity range." Proceedings of the National Academy of Sciences of the United States of America **99**(11): 7554-9.

Dudoit, S., Y. H. Yang, M. J. Callow and T. P. Speed (2002). "Statistical methods for identifying differentially expressed genes in replicated cDNA microarray experiments." Statistica Sinica **12**: 111-39.

Durica, D. S. and P. M. Hopkins (1996). "Expression of the genes encoding the ecdysteroid and retinoid receptors in regenerating limb tissues from the fiddler crab, *Uca pugilator*." Gene **171**(2): 237-41.

Durica, D. S., X. Wu, G. Anilkumar, P. M. Hopkins and A. C. K. Chung (2002). "Characterization of crab EcR and RXR homologs and expression during limb regeneration and oocyte maturation." Molecular and Cellular Endocrinology **189**(1-2): 59-76.

Durstewitz, G. and N. B. Terwilliger (1997). "cDNA cloning of a developmentally regulated hemocyanin subunit in the crustacean *Cancer magister* and phylogenetic analysis of the hemocyanin gene family." Molecular Biology and Evolution **14**(3): 266-76.

EC (1997). European Workshop on the Impact of Endocrine Disrupters on Human Health and Wildlife. Report of the Proceedings 2-4 December 1996, Weybridge, UK (EUR 17549), Brussels, Belgium., European Commission.

Eidem, J. K., H. Kleivdal, K. Kroll, N. Denslow, R. van Aerle, C. Tyler, G. Panter, T. Hutchinson and A. Goksoyr (2006). "Development and validation of a direct homologous quantitative sandwich ELISA for fathead minnow (*Pimephales promelas*) vitellogenin." Aquatic Toxicology **78**(2): 202-06.

El Haj, A. J., S. L. Tamone, M. Peake, P. Sreenivasula Reddy and E. S. Chang (1997). "An ecdysteroid-responsive gene in a lobster - a potential crustacean member of the steroid hormone receptor superfamily." Gene **201**(1-2): 127-35.

Elke, C., P. Rauch, M. Spindler-Barth and K. D. Spindler (1999). "DNA-binding properties of the ecdysteroid receptor-complex (EcR/USP) of the epithelial cell line from *Chironomus tentans*." Archives of Insect Biochemistry and Physiology **41**(3): 124-33.

Eystathioy, T., L. Swevers and K. Iatrou (2001). "The orphan nuclear receptor BmHR3A of *Bombyx mori*: hormonal control, ovarian expression and functional properties." Mechanisms of Development **103**(1-2): 107-15.

Fairs, N. J., R. P. Evershed, P. T. Quinlan and L. J. Goad (1989). "Detection of unconjugated and conjugated steroids in the ovary, eggs, and haemolymph of the decapod crustacean *Nephrops norvegicus*." General and Comparative Endocrinology **74**(2): 199-208.

Fang, Y., A. E. Fliss, D. M. Robins and A. J. Caplan (1996). "Hsp90 regulates androgen receptor hormone binding affinity *in vivo*." Journal of Biological Chemistry **271**(45): 28697-702.

Feng, Q., K. G. Davey, S. D. P. A, T. R. Ladd, A. Retnakaran, B. L. Tomkins, S. Zheng and S. R. Palli (2001). "Developmental expression and stress induction of glutathione S-transferase in the spruce budworm, *Choristoneura fumiferana*." Journal of Insect Physiology **47**(1): 1-10.

Fingerman, M. (1987). "The endocrine mechanisms of crustaceans." Journal of Crustacean Biology **7**(1): 1-24.

Fliss, A. E., S. Benzeno, J. Rao and A. J. Caplan (2000). "Control of estrogen receptor ligand binding by Hsp90." Journal of Steroid Biochemistry and Molecular Biology **72**(5): 223-30.

Ford, A. T., T. F. Fernandes, S. A. Rider, P. A. Read, C. D. Robinson and I. M. Davies (2003). "Reproduction in the amphipod, *Echinogammarus marinus*: a comparison between normal and intersex specimens." Journal of the Marine Biological Association of the United Kingdom **83**(05): 937-40.

Ford, A. T., T. F. Fernandes, P. A. Read, C. D. Robinson and I. M. Davies (2004a). "The costs of intersexuality: a crustacean perspective." Marine Biology **145**(5): 951-57.

Ford, A. T., T. F. Fernandes, S. A. Rider, P. A. Read, C. D. Robinson and I. M. Davies (2004b). "Endocrine disruption in a marine amphipod? Field observations of intersexuality and de-masculinisation." Marine Environmental Research **58**(2-5): 169-73.

Ford, A. T., T. P. Rodgers-Gray, I. M. Davies, A. M. Dunn, P. A. Read, C. D. Robinson, J. E. Smith and T. F. Fernandes (2005). "Abnormal gonadal morphology in intersex, *Echinogammarus marinus* (Amphipoda): a possible cause of reduced fecundity?" Marine Biology **147**(4): 913-18.

Ford, A. T. and T. F. Fernandes (2005). "Notes on the Occurrence of Intersex in Amphipods." Hydrobiologia **548**(1): 313-18.

Ford, A. T., T. F. Fernandes, C. D. Robinson, I. M. Davies and P. A. Read (2006). "Can industrial pollution cause intersexuality in the amphipod, *Echinogammarus marinus*?" Marine Pollution Bulletin **53**(1-4): 100-06.

Ford, A. T., P. A. Read, T. L. Jones, F. Michino, Y. Pang and T. F. Fernandes (2007a). "An investigation into intersex amphipods and a possible association with aquaculture." Marine Environmental Research.

Ford, A. T., I. Martins and T. F. Fernandes (2007b). "Population level effects of intersexuality in the marine environment." Science of the Total Environment **374**(1): 102-11.

Ford, A. T., C. Sambles and P. Kille (2008). "Intersexuality in crustaceans: genetic, individual and population level effects." Marine Environmental Research **In Press.**

Forget-Leray, J., I. Landriau, C. Minier and F. Leboulenger (2005). "Impact of endocrine toxicants on survival, development, and reproduction of the estuarine copepod *Eurytemora affinis* (Poppe)." Ecotoxicology and Environmental Safety **60**(3): 288-94.

Fossi, M. C., S. Casini, L. Marsili, G. Neri, G. Mori, S. Ancora, A. Moscatelli, A. Ausili and G. Notarbartolo-di-Sciara (2002). "Biomarkers for endocrine disruptors in three species of Mediterranean large pelagic fish." Marine Environmental Research **54**(3-5): 667-71.

Fraenkel, G. and V. J. Brookes (1953). "The Process by Which the Puparia of Many Species of Flies Become Fixed to a Substrate." Biological Bulletin **105**(3): 442-49.

Francois, E., D.-Y. Wang, R. Fulthorpe, S. N. Liss and E. A. Edwards (2003). "DNA microarrays for detecting endocrine-disrupting compounds." Biotechnology Advances **22**(1-2): 17-26.

Fujioka, Y., Y. Kimata, K. Nomaguchi, K. Watanabe and K. Kohno (2002). "Identification of a Novel Non-structural Maintenance of Chromosomes (SMC) Component of the SMC5-SMC6 Complex Involved in DNA Repair." Journal of Biological Chemistry **277**(24): 21585-91.

Gagne, F., C. Blaise and J. Pellerin (2005). "Altered exoskeleton composition and vitellogenesis in the crustacean *Gammarus* sp. collected at polluted sites in the Saguenay Fjord, Quebec, Canada." Environmental Research **98**(1): 89-99.

Garcia-Reyero, N., D. Raldua, L. Quiros, G. Llaveria, J. Cerda, D. Barcelo, J. O. Grimalt and B. Pina (2004). "Use of vitellogenin mRNA as a biomarker for endocrine disruption in feral and cultured fish." Analytical and Bioanalytical Chemistry **378**(3): 670-5.

Gatesy, J., C. Hayashi, D. Motriuk, J. Woods and R. Lewis (2001). "Extreme Diversity, Conservation, and Convergence of Spider Silk Fibroin Sequences." Science **291**(5513): 2603.

Gaxiola, G., G. Cuzon, T. Garcia, G. Taboada, R. Brito, M. E. Chimal, A. Paredes, L. Soto, C. Rosas and A. van Wormhoudt (2005). "Factorial effects of salinity, dietary carbohydrate and moult cycle on digestive carbohydrases and hexokinases in *Litopenaeus vannamei* (Boone, 1931)." Comparative Biochemistry and Physiology Part A: Molecular and Integrative Physiology **140**(1): 29-39.

Georget, V., B. Terouanne, J. C. Nicolas and C. Sultan (2002). "Mechanism of antiandrogen action: key role of hsp90 in conformational change and transcriptional activity of the androgen receptor." Biochemistry **41**(39): 11824-31.

Ghekiere, A., T. Verslycke and C. Janssen (2006). "Effects of methoprene, nonylphenol, and estrone on the vitellogenesis of the mysid *Neomysis integer*." General and Comparative Endocrinology **147**(2): 190-5.

Gibbs, P. E., G. W. Bryan and P. L. Pascoe (1991). "TBT-induced imposex in the dogwhelk, *Nucella lapillus*: Geographical uniformity of the response and effects." Marine Environmental Research **32**(1-4): 79-87.

Giesy, J. P., S. L. Pierens, E. M. Snyder, S. Miles-Richardson, V. J. Kramer, S. A. Snyder, K. M. Nichols and D. A. Villeneuve (2000). "Effects of 4-nonylphenol on fecundity and biomarkers of estrogenicity in fathead minnows (*Pimephales promelas*)." Environmental Toxicology and Chemistry **19**(5): 1368-77.

Goldfarb, S. B., O. B. Kashlan, J. N. Watkins, L. Suaud, W. Yan, T. R. Kleyman and R. C. Rubenstein (2006). "Differential effects of Hsc70 and Hsp70 on the intracellular trafficking and functional expression of epithelial sodium channels." Proceedings of the National Academy of Sciences of the United States of America **103**(15): 5817-22.

Gooding, M. P. and G. A. LeBlanc (2001). "Biotransformation and disposition of testosterone in the eastern mud snail *Ilyanassa obsoleta*." General and Comparative Endocrinology **122**(2): 172-80.

Graham, L. D., P. A. Pilling, R. E. Eaton, J. J. Gorman, C. Braybrook, G. N. Hannan, A. Pawlak-Skrzecz, L. Noyce, G. O. Lovrecz, L. Lu and R. J. Hill (2007). "Purification and characterization of recombinant ligand-binding domains from the ecdysone receptors of four pest insects." Protein Expression and Purification **53**(2): 309-24.

Greve, P., O. Sorokine, T. Berges, C. Lacombe, A. Van Dorsselaer and G. Martin (1999). "Isolation and amino acid sequence of a peptide with vitellogenesis inhibiting activity from the terrestrial isopod *Armadillidium vulgare* (Crustacea)." General and Comparative Endocrinology **115**(3): 406-14.

Gross-Sorokin, M. Y., E. P. Grist, M. Cooke and M. Crane (2003). "Uptake and depuration of 4-nonylphenol by the benthic invertebrate *Gammarus pulex*: how important is feeding rate?" Environmental Science & Technology **37**(10): 2236-41.

Gross-Sorokin, M. Y., S. D. Roast and G. C. Brighty (2006). "Assessment of feminization of male fish in English rivers by the Environment Agency of England and Wales." Environmental Health Perspectives **114 Suppl 1**: 147-51.

Gross, M. Y., D. S. Maycock, M. C. Thorndyke, D. Morritt and M. Crane (2001). "Abnormalities in sexual development of the amphipod *Gammarus pulex* (L.) found below sewage treatment works." Environmental Toxicology and Chemistry **20**(8): 1792-7.

Groth, D., H. Lehrach and S. Hennig (2004). "GOblet: a platform for Gene Ontology annotation of anonymous sequence data." Nucleic Acids Research **32**(1): W313-W17.

Gu, P. L., S. S. Tobe, B. K. C. Chow, K. H. Chu, J. G. He and S. M. Chan (2002). "Characterization of an additional molt inhibiting hormone-like neuropeptide from the shrimp *Metapenaeus ensis*." Peptides **23**(11): 1875-83.

Guttman, J. A., P. Janmey and A. W. Vogl (2002). "Gelsolin-evidence for a role in turnover of junction-related actin filaments in Sertoli cells." Journal of Cell Science **115**(3): 499-505.

Hagger, J. A., M. H. Depledge, J. Oehlmann, S. Jobling and T. S. Galloway (2006). "Is there a causal association between genotoxicity and the imposex effect?" Environmental Health Perspectives **114 Suppl 1**: 20-6.

Harries, J. E., D. A. Sheahan, S. Jobling, P. Matthiessen, P. Neall, E. J. Routledge, R. Rycroft, J. P. Sumpter and T. Tylor (1996). "A survey of estrogenic activity in United Kingdom inland waters." Environmental Toxicology and Chemistry **15**(11): 1993-2002.

Hasegawa, Y., E. Hirose and Y. Katakura (1993). "Hormonal Control of Sexual Differentiation and Reproduction in Crustacea." American Zoologist **33**(3): 403-11.

Haverinen, M., S. Passinen, H. Syvala, S. Pasanen, T. Manninen, P. Tuohimaa and T. Ylikomi (2001). "Heat shock protein 90 and the nuclear transport of progesterone receptor." Cell Stress & Chaperones **6**(3): 256-62.

Heckmann, L. H., N. Friberg and H. W. Ravn (2005). "Relationship between biochemical biomarkers and pre-copulatory behaviour and mortality in *Gammarus pulex* following pulse-exposure to λ-cyhalothrin." Pest Management Science **61**(7): 627-35.

Heckmann, L. H., R. Connon, T. H. Hutchinson, S. J. Maund, R. M. Sibly and A. Callaghan (2006). "Expression of target and reference genes in *Daphnia magna* exposed to ibuprofen." BMC Genomics **7**: 175.

Hegde, P., R. Qi, K. Abernathy, C. Gay, S. Dharap, R. Gaspard, J. Earle-Hughes, E. Snesrud, N. Lee and J. Quackenbush (2000). "A concise guide to cDNA microarray analysis-II." BioTechniques **29**(3): 548-62.

Henrich, V. C., A. A. Szekely, S. J. Kim, N. E. Brown, C. Antoniewski, M. A. Hayden, J. A. Lepesant and L. I. Gilbert (1994). "Expression and function of the ultraspiracle (*usp*) gene during development of *Drosophila melanogaster*." Developmental Biology **165**(1): 38-52.

Henrich, V. C., M. E. Vogtli, C. Antoniewski, M. Spindler-Barth, S. Przibilla, M. Noureddine and M. Lezzi (2000). "Developmental effects of a chimeric ultraspiracle gene derived from *Drosophila* and *Chironomus*." Genesis **28**(3-4): 125-33.

Homola, E. and E. S. Chang (1997). "Methyl Farnesoate: Crustacean Juvenile Hormone in Search of Functions." Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology **117**(3): 347-56.

Hong, E. J., S. H. Park, K. C. Choi, P. C. Leung and E. B. Jeung (2006). "Identification of estrogen-regulated genes by microarray analysis of the uterus of immature rats exposed to endocrine disrupting chemicals." Reproductive Biology and Endocrinology **4**: 49.

Huberman, A. (2000). "Shrimp endocrinology. A review." Aquaculture **191**(1-3 SU -): 191-208.

Huggett, J., K. Dheda, S. Bustin and A. Zumla (2005). "Real-time RT-PCR normalisation; strategies and considerations." Genes and Immunity **6**(4): 279-84.

Hynes, H. B. N. (1955). "The reproductive cycle of some British freshwater Gammaridae." Journal of Animal Ecology **24**: 352-87.

Iguchi, T., H. Watanabe and Y. Katsu (2006). "Application of ecotoxicogenomics for studying endocrine disruption in vertebrates and invertebrates." Environmental Health Perspectives **114 Suppl 1**: 101-5.

Iguchi, T., H. Watanabe and Y. Katsu (2007). "Toxicogenomics and ecotoxicogenomics for studying endocrine disruption and basic biology." General and Comparative Endocrinology.

Imhof, M. O., S. Rusconi and M. Lezzi (1993). "Cloning of a *Chironomus tentans* cDNA encoding a protein (cEcRH) homologous to the *Drosophila melanogaster* ecdysteroid receptor (dEcR)." Insect Biochemistry and Molecular Biology **23**(1): 115-24.

Iwai, T., J. Lee, A. Yoshii, T. Yokota, K. Mita and M. Yamashita (2004). "Changes in the expression and localization of cohesin subunits during meiosis in a non-mammalian vertebrate, the medaka fish." Gene Expression Patterns **4**(5): 495-504.

Janer, G., J. Bachmann, J. Oehlmann, U. Schulte-Oehlmann and C. Porte (2006a). "The effect of organotin compounds on gender specific androstenedione metabolism in the freshwater ramshorn snail *Marisa cornuarietis*." Journal of Steroid Biochemistry and Molecular Biology **99**(2-3): 147-56.

Janer, G., A. Lyssimachou, J. Bachmann, J. Oehlmann, U. Schulte-Oehlmann and C. Porte (2006b). "Sexual dimorphism in esterified steroid levels in the gastropod *Marisa cornuarietis*: The effect of xenoandrogenic compounds." Steroids **71**(6): 435-44.

Ji, W., M. B. Wright, L. Cai, A. Flament and K. Lindpaintner (2002). "Efficacy of SSH PCR in isolating differentially expressed genes." BMC Genomics **3**(1): 12.

Ji, W., W. Zhou, K. Gregg, N. Yu, S. Davis and S. Davis (2004). "A method for cross-species gene expression analysis with high-density oligonucleotide arrays." Nucleic Acids Research **32**(11): e93.

Jobling, S., D. Casey, T. Rogers-Gray, J. Oehlmann, U. Schulte-Oehlmann, S. Pawlowski, T. Baunbeck, A. P. Turner and C. R. Tyler (2004). "Comparative responses of molluscs and fish to environmental estrogens and an estrogenic effluent." Aquatic Toxicology **66**(2): 207-22.

Jobling, S., R. Williams, A. Johnson, A. Taylor, M. Gross-Sorokin, M. Nolan, C. R. Tyler, R. van Aerle, E. Santos and G. Brighty (2006). "Predicted exposures to steroid estrogens in U.K. rivers correlate with widespread sexual disruption in wild fish populations." Environmental Health Perspectives **114 Suppl 1**: 32-9.

Johnson, I., J. M. Weeks and P. Kille (2005). Endocrine disruption in aquatic and terrestrial invertebrates (UC 4906/6). DEFRA Final Report (UC 4906/6).

Jones, G. and P. A. Sharp (1997). "Ultraspiracle: an invertebrate nuclear receptor for juvenile hormones." Proceedings of the National Academy of Sciences of the United States of America **94**(25): 13499-503.

Ju, Z., M. C. Wells and R. B. Walter (2007). "DNA microarray technology in toxicogenomics of aquatic models: methods and applications." Comparative Biochemistry and Physiology Part C: Pharmacology, Toxicology and Endocrinology **145**(1): 5-14.

Junera, H., C. Zerbib, M. Martin and J.-J. Meusy (1977). "Evidence for control of vitellogenin synthesis by an ovarian hormone in *Orchestia gammarella* (Pallas), crustacea; Amphipoda." General and Comparative Endocrinology **31**(4): 457-62.

Jungmann, D., V. Ladewig, K. U. Ludwichowski, P. Petzsch and R. Nagel (2004). "Intersexuality in *Gammarus fossarum* KOCH-a common inducible phenomenon?" Archiv für Hydrobiologie **159**(4): 511-29.

Kallivretaki, E., R. Eggen, S. Neuhauss, M. Alberti, U. Kausch and H. Segner (2006). "Aromatase in zebrafish: a potential target for endocrine disrupting chemicals." Marine Environmental Research **62 Suppl**: S187-90.

Kapitskaya, M., S. Wang, D. E. Cress, T. S. Dhadialla, A. S. Raikhel, W. L. Cho and M. Z. Kapitskaya (1996). "The mosquito ultraspiracle homologue, a partner of ecdysteroid receptor heterodimer: cloning and characterization of isoforms expressed during vitellogenesis." Molecular and Cellular Endocrinology **121**(2): 119-32.

Kapitskaya, M. Z., C. Li, K. Miura, W. Segraves and A. S. Raikhel (2000). "Expression of the early-late gene encoding the nuclear receptor HR3 suggests its involvement in regulating the vitellogenic response to ecdysone in the adult mosquito." Molecular and Cellular Endocrinology **160**(1-2): 25-37.

Kast-Hutcheson, K., C. V. Rider and G. A. LeBlanc (2001). "The fungicide propiconazole interferes with embryonic development of the crustacean *Daphnia magna*." Environmental Toxicology and Chemistry **20**(3): 502-9.

Kato, Y., K. Kobayashi, S. Oda, N. Tatarazako, H. Watanabe and T. Iguchi (2007). "Cloning and characterization of the ecdysone receptor and ultraspiracle protein from the water flea *Daphnia magna*." Journal of Endocrinology **193**(1): 183-94.

Kelly, A., M. J. Hatcher and A. M. Dunn (2004). "Intersexuality in the amphipod *Gammarus duebeni* results from incomplete feminisation by the vertically transmitted parasitic sex ratio distorter *Nosema granulosis*." Evolutionary Ecology **18**(2): 121-32.

Khalaila, I., S. Weil and A. Sagi (1999). "Endocrine balance between male and female components of the reproductive system in intersex *Cherax quadricarinatus* (Decapoda: Parastacidae)." The Journal of Experimental Zoology **283**(3): 286-94.

Khamnamtong, B., S. Thumrungtanakit, S. Klinbunga, T. Aoki, I. Hirono and P. Menasveta (2006). "Identification of sex-specific expression markers in the giant tiger shrimp (*Penaeus monodon*)." Journal of Biochemistry and Molecular Biology **39**(1): 37-45.

Khan, R. L., G. E. Gonye, G. Gao and J. S. Schwaber (2006). "A universal reference sample derived from clone vector for improved detection of differential gene expression." BMC Genomics **7**: 109.

King, D. S. and J. B. Siddall (1969). "Conversion of α-ecdysone to p-ecdysone by crustaceans and insects." Nature **221**: 955-56.

Kipp, J. L. and V. D. Ramirez (2001). "Effect of estradiol. diethylstilbestrol, and resveratrol on F0F1-ATPase activity from mitochondrial preparations of rat heart, liver, and brain." Endocrine **15**(2): 165-75.

Kirby, M. F., Y. T. Allen, R. A. Dyer, S. W. Feist, I. Katsiadaki, P. Matthiessen, A. P. Scott, A. Smith, G. D. Stentiford, J. E. Thain, K. V. Thomas, L. Tolhurst and M. J. Waldock (2004). "Surveys of plasma vitellogenin and intersex in male flounder (*Platichthys flesus*) as measures of endocrine disruption by estrogenic contamination in United Kingdom estuaries: temporal trends, 1996 to 2001." Environmental Toxicology and Chemistry **23**(3): 748-58.

Klein, B., G. Le Moullac, D. Sellos and A. Van Wormhoudt (1996). "Molecular cloning and sequencing of trypsin cDNAs from *Penaeus vannamei* (Crustacea, Decapoda): Use in assessing gene expression during the moult cycle." The International Journal of Biochemistry & Cell Biology **28**(5): 551-63.

Koelle, M. R., W. S. Talbot, W. A. Segraves, M. T. Bender, P. Cherbas and D. S. Hogness (1991). "The *Drosophila* EcR gene encodes an ecdysone receptor, a new member of the steroid receptor superfamily." Cell **67**(1): 59-77.

Kostrouchova, M., M. Krause, Z. Kostrouch and J. E. Rall (1998). "CHR3: a *Caenorhabditis elegans* orphan nuclear hormone receptor required for proper epidermal development and molting." Development **125**(9): 1617-26.

Kostrouchova, M., M. Krause, Z. Kostrouch and J. E. Rall (2001). "Nuclear hormone receptor CHR3 is a critical regulator of all four larval molts of the nematode *Caenorhabditis elegans*." Proceedings of the National Academy of Sciences of the United States of America **98**(13): 7360-5.

Kreutzweiser, D. P., S. S. Capell, K. L. Wainio-Keizer and D. C. Eichenberg (1994). "Toxicity of a new molt-inducing insecticide (RH-5992) to aquatic macroinvertebrates." Ecotoxicology and Environmental Safety **28**(1): 14-24.

Kuhn, E. (2001). "From Library Screening to Microarray Technology: Strategies to Determine Gene Expression Profiles and to Identify Differentially Regulated Genes in Plants." Annals of Botany **87**(2): 139-55.

Kusk, K. O. and L. Wollenberger (2007). "Towards an internationally harmonized test method for reproductive and developmental effects of endocrine disrupters in marine copepods." Ecotoxicology **16**(1): 183-95.

Lachaise, F., A. L. Roux, M. Hubert and R. Lafont (1993). "The Molting Gland of Crustaceans: Localization, Activity, and Endocrine Control (A Review)." Journal of Crustacean Biology **13**(2): 198-234.

Lacombe, C., P. Greve and G. Martin (1999). "Overview on the sub-grouping of the crustacean hyperglycemic hormone family." Neuropeptides **33**(1): 71-80.

LaFont, R. (2000). "The Endocrinology of Invertebrates." Ecotoxicology **9**(1): 41-57.

Langelan, R. E., J. E. Fisher, K. Hiruma, S. R. Palli and L. M. Riddiford (2000). "Patterns of MHR3 Expression in the Epidermis during a Larval Molt of the Tobacco Hornworm *Manduca sexta*." Developmental Biology **227**(2): 481-94.

Larkin, P., L. C. Folmar, M. J. Hemmer, A. J. Poston, H. S. Lee and N. D. Denslow (2002a). "Array technology as a tool to monitor exposure of fish to xenoestrogens." Marine Environmental Research **54**(3-5): 395-9.

Larkin, P., T. Sabo-Attwood, J. Kelso and N. D. Denslow (2002b). "Gene expression analysis of largemouth bass exposed to estradiol, nonylphenol, and p,p'-DDE." Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology **133**(4): 543-57.

Larkin, P., L. C. Folmar, M. J. Hemmer, A. J. Poston and N. D. Denslow (2003). "Expression profiling of estrogenic compounds using a sheepshead minnow cDNA macroarray." EHP Toxicogenomics **111**(6): 839-46.

Laufer, H. and A. Sagi (1991). "Juvenile hormone-like compounds and reproduction in male and female crustaceans: with implications for aquaculture." Bulletin of the Institute of Zoology, Academia Sinica **16**(Monograph): 541-51.

Laufer, H., J. S. B. Ahl and A. Sagi (1993). "The Role of Juvenile Hormones in Crustacean Reproduction." Integrative and Comparative Biology **33**(3): 365-74.

Lee, C., J. G. Na, K. C. Lee and K. Park (2002). "Choriogenin mRNA induction in male medaka, *Oryzias latipes* as a biomarker of endocrine disruption." Aquatic Toxicology **61**(3-4): 233-41.

Lee, K. J., R. D. Watson and R. D. Roer (1998). "Molt-inhibiting hormone mRNA levels and ecdysteroid titer during a molt cycle of the blue crab, *Callinectes sapidus*." Biochemical and Biophysical Research Communications **249**(3): 624-7.

Leu, J. H., C. C. Chang, J. L. Wu, C. W. Hsu, I. Hirono, T. Aoki, H. F. Juan, C. F. Lo, G. H. Kou and H. C. Huang (2007). "Comparative analysis of differentially expressed genes in normal and white spot syndrome virus infected *Penaeus monodon*." BMC Genomics **8**(1): 120.

Liu, L., H. Laufer, Y. Wang and T. Hayes (1997). "A neurohormone regulating both methyl farnesoate synthesis and glucose metabolism in a crustacean." Biochemical and Biophysical Research Communications **237**(3): 694-701.

Livak, K. J. and T. D. Schmittgen (2001). "Analysis of relative gene expression data using real-time quantitative PCR and the 2-ΔΔCt method." Methods **25**(4): 402-08.

Lockhart, D. J., H. Dong, M. C. Byrne, M. T. Follettie, M. V. Gallo, M. S. Chee, M. Mittmann, C. Wang, M. Kobayashi and H. Horton (1996). "Expression monitoring by hybridization to high-density oligonucleotide arrays." Nature Biotechnology **14**(13): 1675-80.

Maglich, J. M., A. Sluder, X. Guan, Y. Shi, D. D. McKee, K. Carrick, K. Kamdar, T. M. Willson and J. T. Moore (2001). "Comparison of complete nuclear receptor sets from the human, *Caenorhabditis elegans* and *Drosophila* genomes." Genome Biology **2**(8): 1–0029.

Maitland, P. S. (1966). "Notes on the biology of *Gammarus pulex* in the River Endrick." Hydrobiologia **28**(1): 142-52.

Maki, A., S. Sawatsubashi, S. Ito, Y. Shirode, E. Suzuki, Y. Zhao, K. Yamagata, A. Kouzmenko, K.-i. Takeyama and S. Kato (2004). "Juvenile hormones antagonize ecdysone actions through co-repressor recruitment to EcR/USP heterodimers." Biochemical and Biophysical Research Communications **320**(1): 262-67.

Malbouisson, J. F. C., T. W. K. Young and A. W. Bark (1994). "Disruption of precopula in *Gammarus pulex* as a result of brief exposure to γ-hexachlorocyclohexane (Lindane)." Chemosphere **28**(11): 2011-20.

Manor, R., S. Weil, S. Oren, L. Glazer, E. D. Aflalo, T. Ventura, V. Chalifa-Caspi, M. Lapidot and A. Sagi (2007). "Insulin and gender: an insulin-like gene expressed exclusively in the androgenic gland of the male crayfish." General and Comparative Endocrinology **150**(2): 326-36.

Matthiessen, P. and P. E. Gibbs (1998). "Critical appraisal of the evidence for tributyl tin-mediated endocrine disruption in mollusks." Environmental Toxicology and Chemistry **17**(1): 37-43.

Matthiessen, P., Y. Allen, S. Bamber, J. Craft, M. Hurst, T. Hutchinson, S. Feist, I. Katsiadaki, M. Kirby, C. Robinson, S. Scott, J. Thain and K. Thomas (2002). "The impact of oestrogenic and androgenic contamination on marine organisms in the United Kingdom--summary of the EDMAR programme. Endocrine Disruption in the Marine Environment." Marine Environmental Research **54**(3-5): 645-9.

Mattson, M. P. and E. Spaziani (1985). "Characterization of molt-inhibiting hormone (MIH) action on crustacean Y-organ segments and dispersed cells in culture and a bioassay for MIH activity." The Journal of Experimental Zoology **236**(1): 93-101.

McCahon, C. P. and D. Pascoe (1988a). "Cadmium toxicity to the freshwater amphipod *Gammarus pulex* (L.) during the moult cycle." Freshwater Biology **19**: 197-203.

McCahon, C. P. and D. Pascoe (1988b). "Culture techniques for three freshwater macroinvertebrate species and their use in toxicity tests." Chemosphere **17**(12): 2471-80.

McCahon, C. P. and D. Pascoe (1988c). "Use of *Gammarus pulex* (L.) in safety evaluation tests: Culture and selection of a sensitive life stage." Ecotoxicology and Environmental Safety **15**(3): 245-52.

McCahon, P. and D. Pascoe (1988d). "Increased Sensitivity to Cadmium of the Freshwater Amphipod *Gammarus pulex*(L.) During the Reproductive Period." Aquatic Toxicology **13**(3).

McClellan-Green, P., J. Romano and E. Oberdorster (2007). "Does gender really matter in contaminant exposure? A case study using invertebrate models." Environmental Research **104**(1): 183-91.

Meier, R. (1988). "A universal and efficient protocol for the isolation of RNA from tissues and cultured cells." Nucleic Acids Research **16**(5): 2340.

Meyran, J. C., M. Monnerot and P. Taberlet (1997). "Taxonomic status and phylogenetic relationships of some species of the genus *Gammarus* (Crustacea, Amphipoda) deduced from mitochondrial DNA sequences." Molecular Phylogenetics and Evolution **8**(1): 1-10.

Minakuchi, C., Y. Nakagawa, M. Kiuchi, A. Seino, S. Tomita and M. Kamimura (2003). "Molecular cloning and expression analysis of ultraspiracle (USP) from the rice stem borer *Chilo suppressalis*." Insect Biochemistry and Molecular Biology **33**(1): 41-49.

Morcillo, Y. and C. Porte (1999). "Evidence of endocrine disruption in the imposex-affected gastropod *Bolinus brandaris*." Environmental Research **81**(4): 349-54.

Morgan, T. H. (1920). "Variations in the Secondary Sexual Characters of the Fiddler Crab." The American Naturalist **54**(632): 220-46.

Mu, X. and G. A. LeBlanc (2002). "Developmental toxicity of testosterone in the crustacean *Daphnia magna* involves anti-ecdysteroidal activity." General and Comparative Endocrinology **129**(2): 127-33.

Muncke, J. and R. I. L. Eggen (2006). "Vitellogenin 1 mRNA as an early molecular biomarker for endocrine disruption in developing zebrafish (*Danio rerio*)." Environmental Toxicology and Chemistry **25**(10): 2734-41.

Murado, M. A. and J. A. Vazquez (2006). "The notion of hormesis and the dose-response theory: A unified approach." Journal of Theoretical Biology.

Naderi, A., A. A. Ahmed, N. L. Barbosa-Morais, S. Aparicio, J. D. Brenton and C. Caldas (2004). "Expression microarray reproducibility is improved by optimising purification steps in RNA amplification and labelling." BMC Genomics **5**: 9.

Nishimura, K., H. J. Ting, Y. Harada, T. Tokizane, N. Nonomura, H. Y. Kang, H. C. Chang, S. Yeh, H. Miyamoto and M. Shin (2003). "Modulation of Androgen Receptor Transactivation by Gelsolin A Newly Identified Androgen Receptor Coregulator 1." Cancer Research **63**(16): 4888-94.

Nuwaysir, E. F., M. Bittner, J. Trent, J. C. Barrett and C. A. Afshari (1999). "Microarrays and toxicology: The advent of toxicogenomics." Molecular Carcinogenesis **24**(3): 153-59.

Oberdorster, E., M. Brouwer and J. A. McLachlan (2000). "PAH effects on grass shrimp, *Palaeomonetes pugio*: interactions with the ecdysteroid system." Marine Environmental Research **50**(1-5): 194-95.

Oberdorster, E. and A. O. Cheek (2001). "Gender benders at the beach: endocrine disruption in marine and estuarine organisms." Environmental Toxicology and Chemistry **20**(1): 23-36.

Oberdorster, E. and P. McClellan-Green (2002). "Mechanisms of imposex induction in the mud snail, *Ilyanassa obsoleta*: TBT as a neurotoxin and aromatase inhibitor." Marine Environmental Research **54**(3-5): 715-18.

Oehlmann, J., P. Fioroni, E. Stroben and B. Markert (1996). "Tributyltin (TBT) effects on *Ocinebrina aciculata* (Gastropoda: Muricidae): imposex development, sterilization, sex change and population decline." Science of the Total Environment **188**(2-3): 205-23.

Oehlmann, J., P. Di Benedetto, M. Tillmann, M. Duft, M. Oetken and U. Schulte-Oehlmann (2007). "Endocrine disruption in prosobranch molluscs: evidence and ecological relevance." Ecotoxicology **16**(1): 29-43.

Ogura, T., C. Minakuchi, Y. Nakagawa, G. Smagghe and H. Miyagawa (2005). "Molecular cloning, expression analysis and functional confirmation of ecdysone receptor and ultraspiracle from the Colorado potato beetle *Leptinotarsa decemlineata*." FEBS Journal **272**(16): 4114-28.

Okumura, T., K. Yamano and K. Sakiyama (2007). "Vitellogenin gene expression and hemolymph vitellogenin during vitellogenesis, final maturation, and oviposition in female kuruma prawn, *Marsupenaeus japonicus*." Comparative Biochemistry and Physiology Part A: Molecular and Integrative Physiology.

Ollivaux, C., J. Vinh, D. Soyez and J.-Y. Toullec (2006). "Crustacean hyperglycemic and vitellogenesis-inhibiting hormones in the lobster *Homarus gammarus*. Implications for structural and functional evolution of a neuropeptide family." FEBS Journal **273**(10): 2151-60.

Page, R. D. M. (1996). "Tree View: An application to display phylogenetic trees on personal computers." Bioinformatics **12**(4): 357.

Parkinson, J., D. Guiliano and M. Blaxter (2002). "Making sense of EST sequences by CLOBBing them." BMC Bioinformatics **3**(1): 31.

Parkinson, J., A. Anthony, J. Wasmuth, R. Schmid, A. Hedley and M. Blaxter (2004). "PartiGene--constructing partial genomes." Bioinformatics **20**(9): 1398-404.

Parthasarathy, R. and S. R. Palli (2007). "Stage- and cell-specific expression of ecdysone receptors and ecdysone-induced transcription factors during midgut remodeling in the yellow fever mosquito, *Aedes aegypti*." Journal of Insect Physiology **53**(3): 216-29.

Pascoe, D., T. J. Kedwards, S. J. Maund, E. Muthi and E. J. Taylor (1994). "Laboratory and field evaluation of a behavioural bioassay: the *Gammarus pulex*(L.) precopula separation (GaPPS) test." Water Research **28**(2): 369-72.

Perera, S. C., S. R. Palli, T. R. Ladd, P. J. Krell and A. Retnakaran (1998). "The ultraspiracle gene of the spruce budworm, *Choristoneura fumiferana*: cloning of cDNA and developmental expression of mRNA." Developmental Genetics **22**(2): 169-79.

Pickford, K. A., R. E. Thomas-Jones, B. Wheals, C. R. Tyler and J. P. Sumpter (2003). "Route of exposure affects the oestrogenic response of fish to 4-tert-nonylphenol." Aquatic Toxicology **65**(3): 267-79.

Poulton, M. and D. Pascoe (1990). "Disruption of precopula in *Gammarus pulex* (L.) -- Development of a behavioural bioassay for evaluating pollutant and parasite induced stress." Chemosphere **20**(3-4): 403-15.

Poynton, H. C., J. R. Varshavsky, B. Chang, G. Cavigiolio, S. Chan, P. S. Holman, A. V. Loguinov, D. J. Bauer, K. Komachi, E. C. Theil, E. J. Perkins, O. Hughes and C. D. Vulpe (2007). "*Daphnia magna* ecotoxicogenomics provides mechanistic insights into metal toxicity." Environmental Science & Technology **41**(3): 1044-50.

Puinean, A. M., P. Labadie, E. M. Hill, M. Osada, M. Kishida, R. Nakao, A. Novillo, I. P. Callard and J. M. Rotchell (2006). "Laboratory exposure to 17β-estradiol fails to induce vitellogenin and estrogen receptor gene expression in the marine invertebrate *Mytilus edulis*." Aquatic Toxicology **79**(4): 376-83.

Puinean, A. M. and J. M. Rotchell (2006). "Vitellogenin gene expression as a biomarker of endocrine disruption in the invertebrate, *Mytilus edulis*." Marine Environmental Research.

Purdom, C. E., P. A. Hardiman, V. V. J. Bye, N. C. Eno, C. R. Tyler and J. P. Sumpter (1994). "Estrogenic Effects of Effluents from Sewage Treatment Works." Chemistry and Ecology **8**(4): 275-85.

Quilang, J., S. Wang, P. Li, J. Abernathy, E. Peatman, Y. Wang, L. Wang, Y. Shi, R. Wallace and X. Guo (2007). "Generation and analysis of ESTs from the eastern oyster, *Crassostrea virginica* Gmelin and identification of microsatellite and SNP markers." BMC Genomics **8**(1): 157.

Quinitio, E. T., A. Hara, K. Yamauchi and S. Nakao (1994). "Changes in the steroid hormone and vitellogenin levels during the gametogenic cycle of the giant tiger shrimp *Penaeus monodon*. ." Comparative Biochemistry and Physiology: 21–26.

Quinn, B., F. Gagne, M. Costello, C. McKenzie, J. Wilson and C. Mothersill (2004). "The endocrine disrupting effect of municipal effluent on the zebra mussel (*Dreissena polymorpha*)." Aquatic Toxicology **66**(3): 279-92.

Raviv, S., S. Parnes, C. Segall, C. Davis and A. Sagi (2006). "Complete sequence of *Litopenaeus vannamei* (Crustacea: Decapoda) vitellogenin cDNA and its expression in endocrinologically induced sub-adult females." General and Comparative Endocrinology **145**(1): 39-50.

Reaka, M. L. (1975). "Molting in stomatopod crustaceans. I. Stages of the molt cycle, setagenesis, and morphology." Journal of Morphology **146**(1): 55-80.

Reichert, K. and R. Menzel (2005). "Expression profiling of five different xenobiotics using a *Caenorhabditis elegans* whole genome microarray." Chemosphere **61**(2): 229-37.

Reinke, R., D. E. Krantz, D. Yen and S. L. Zipursky (1988). "Chaoptin, a cell surface glycoprotein required for *Drosophila* photoreceptor cell morphogenesis, contains a repeat motif found in yeast and human." Cell **52**(2): 291-301.

Retnakaran, A., I. Gelbic, M. Sundaram, W. Tomkins, T. Ladd, M. Primavera, Q. Feng, B. Arif, R. Palli and P. Krell (2001). "Mode of action of the ecdysone agonist tebufenozide (RH-5992), and an exclusion mechanism to explain resistance to it." Pest Management Science **57**(10): 951-7.

Rewitz, K. F., R. Rybczynski, J. T. Warren and L. I. Gilbert (2006). "Developmental expression of *Manduca* shade, the P450 mediating the final step in molting hormone synthesis." Molecular and Cellular Endocrinology **247**(1-2): 166-74.

Ribeiro, R. C., P. J. Kushner and J. D. Baxter (1995). "The nuclear hormone receptor gene superfamily." Annual Review of Medicine **46**: 443-53.

Rickman, D. S., C. J. Herbert and L. P. Aggerbeck (2003). "Optimizing spotting solutions for increased reproducibility of cDNA microarrays." Nucleic Acids Research **31**(18): e109.

Robertson, H. M., R. Martos, C. R. Sears, E. Z. Todres, K. K. O. Walden and J. B. Nardi (1999). "Diversity of odourant binding proteins revealed by an expressed sequence tag project on male *Manduca sexta* moth antennae." Insect Molecular Biology **8**(4): 501-18.

Rodgers-Gray, T. P., J. E. Smith, A. E. Ashcroft, R. E. Isaac and A. M. Dunn (2004). "Mechanisms of parasite-induced sex reversal in *Gammarus duebeni*." International Journal for Parasitology **34**(6): 747-53.

Royer, V., S. Fraichard and H. Bouhin (2002). "A novel putative insect chitinase with multiple catalytic domains: hormonal regulation during metamorphosis." Biochemical Journal **366**(Pt 3): 921-28.

Ruangsittichai, J., V. Viyanant, S. Vichasri-Grams, P. Sobhon, S. Tesana, E. S. Upatham, A. Hofmann, G. Korge and R. Grams (2006). "*Opisthorchis viverrini*: Identification of a glycine-tyrosine rich eggshell protein and its potential as a diagnostic tool for human opisthorchiasis." International Journal for Parasitology **36**(13): 1329-39.

Sambrook, J., Fritsch, E.F. & Maniatis, T (1989). Molecular cloning: a laboratory manual, Cold Spring Harbor Press.

Santos, E. A. and R. Keller (1993). "Crustacean hyperglycemic hormone (CHH) and the regulation of carbohydrate metabolism: Current perspectives." Comparative Biochemistry and Physiology Part A: Molecular and Integrative Physiology **106**(3): 405-11.

Santos, E. M., G. C. Paull, K. J. Van Look, V. L. Workman, W. V. Holt, R. van Aerle, P. Kille and C. R. Tyler (2007a). "Gonadal transcriptome responses and physiological consequences of exposure to oestrogen in breeding zebrafish (*Danio rerio*)." Aquatic Toxicology.

Santos, E. M., V. L. Workman, G. C. Paull, A. L. Filby, K. J. Van Look, P. Kille and C. R. Tyler (2007b). "Molecular basis of gender and reproductive status in breeding zebrafish." Physiological Genomics.

Sartor, M. A., A. M. Zorn, J. A. Schwanekamp, D. Halbleib, S. Karyala, M. L. Howell, G. E. Dean, M. Medvedovic, C. R. Tomlinson and O. Journals (2006). "A new method to remove hybridization bias for interspecies comparison of global gene expression profiles uncovers an association between mRNA sequence divergence and differential gene expression in *Xenopus*." Nucleic Acids Research **34**(1): 185-200.

Schirling, M., A. Bohlen, R. Triebskorn and H. R. Kohler (2006). "An invertebrate embryo test with the apple snail *Marisa cornuarietis* to assess effects of potential developmental and endocrine disruptors." Chemosphere **64**(10): 1730-8.

Segner, H., K. Caroll, M. Fenske, C. R. Janssen, G. Maack, D. Pascoe, C. Schafers, G. F. Vandenbergh, M. Watts and A. Wenzel (2003). "Identification of endocrine-disrupting effects in aquatic vertebrates and invertebrates: report from the European IDEA project." Ecotoxicology and Environmental Safety **54**(3): 302-14.

Seol, W., H. S. Choi and D. D. Moore (1996). "An orphan nuclear hormone receptor that lacks a DNA binding domain and heterodimerizes with other receptors." Science **272**(5266): 1336-9.

Seol, W., B. Hanstein, M. Brown and D. D. Moore (1998). "Inhibition of Estrogen Receptor Action by the Orphan Receptor SHP (Short Heterodimer Partner)." Molecular Endocrinology **12**(10): 1551-57.

Sexton, E. W. (1928). "On the rearing and breeding of *Gammarus* in laboratory conditions." Journal of the Marine Biological Association of the United Kingdom **15**: 33–35.

Shah, G., M. Azizian, D. Bruch, R. Mehta and D. Kittur (2004). "Cross-species comparison of gene expression between human and porcine tissue, using single microarray platform--preliminary results." Clinical Transplantation **18 Suppl 12**: 76-80.

Skinner, D. M. (1962). "The structure and metabolism of a crustacean integumentary tissue during a molt cycle." Biological Bulletin **123**(3): 635.

Snyder, M. J. (1998). "Identification of a New Cytochrome P450 Family, CYP45, from the Lobster, *Homarus americanus*, and Expression Following Hormone and Xenobiotic Exposures." Archives of Biochemistry and Biophysics **358**(2): 271-76.

Snyder, M. J. and E. P. Mulder (2001). "Environmental endocrine disruption in decapod crustacean larvae: hormone titers, cytochrome P450, and stress protein responses to heptachlor exposure." Aquatic Toxicology **55**(3-4): 177-90.

Soetaert, A., L. N. Moens, K. Van der Ven, K. Van Leemput, B. Naudts, R. Blust and W. M. De Coen (2005). "Molecular impact of propiconazole on *Daphnia magna* using a reproduction-related cDNA array." Comparative Biochemistry and Physiology Part C: Pharmacology, Toxicology and Endocrinology.

Soetaert, A., L. N. Moens, K. Van der Ven, K. Van Leemput, B. Naudts, R. Blust and W. M. De Coen (2006). "Molecular impact of propiconazole on *Daphnia magna* using a reproduction-related cDNA array." Comparative Biochemistry and Physiology Part C: Pharmacology, Toxicology and Endocrinology.

Soetaert, A., K. van der Ven, L. N. Moens, T. Vandenbrouck, P. van Remortel and W. M. De Coen (2007). "*Daphnia magna* and ecotoxicogenomics: Gene expression profiles of the anti-ecdysteroidal fungicide fenarimol using energy-, molting-and life stage-related cDNA libraries." Chemosphere.

Sole, M., C. Porte and D. Barcelo (2000). "Vitellogenin induction and other biochemical responses in carp, *Cyprinus carpio*, after experimental injection with 17α-ethynylestradiol." Archives of Environmental Contamination and Toxicology **38**(4): 494-500.

Sole, M., D. Raldua, D. Barcelo and C. Porte (2003). "Long-term exposure effects in vitellogenin, sex hormones, and biotransformation enzymes in female carp in relation to a sewage treatment works." Ecotoxicology and Environmental Safety **56**(3): 373-80.

Spooner, N., P. E. Gibbs, G. W. Bryan and L. J. Goad (1991). "The effect of tributyltin upon steroid titres in the female dogwhelk, *Nucella lapillus*, and the development of imposex." Marine Environmental Research **32**(1): 37-49.

Steel, C. G. H. (1980). "Mechanisms of Coordination between Moulting and Reproduction in Terrestrial Isopod Crustacea." Biological Bulletin **159**(1): 206-18.

Sterrenburg, E., R. Turk, J. M. Boer, G. B. van Ommen and J. T. den Dunnen (2002). "A common reference for cDNA microarray hybridizations." Nucleic Acids Research **30**(21): e116.

Subramoniam, T. (2000). "Crustacean ecdysteriods in reproduction and embryogenesis." Comparative Biochemistry and Physiology Part C: Pharmacology, Toxicology and Endocrinology **125**(2): 135-56.

Summavielle, T., P. R. R. Monteiro, M. A. Reis-Henriques and J. Coimbra (2003). "In vitro metabolism of steroid hormones by ovary and hepatopancreas of the crustacean Penaeid shrimp *Marsupenaeus japonicus*." Scientia Marina **67**(3): 299-306.

Sumpter, J. P. (1995). "Feminized responses in fish to environmental estrogens." Toxicology Letters **82-83**: 737-42.

Sumpter, J. P. and S. Jobling (1995). "Vitellogenesis as a biomarker for estrogenic contamination of the aquatic environment." Environmental Health Perspectives **103 Suppl 7**: 173-8.

Sumpter, J. P. (1998). "Xenoendocrine disrupters -- environmental impacts." Toxicology Letters **102-103**: 337-42.

Sundaram, M., S. R. Palli, P. J. Krell, S. S. Sohi, T. S. Dhadialla and A. Retnakaran (1998). "Basis for selective action of a synthetic molting hormone agonist, RH-5992 on lepidopteran insects." Insect Biochemistry and Molecular Biology **28**(9): 693-704.

Sundaram, M., S. R. Palli, G. Smagghe, I. Ishaaya, Q.-L. Feng, M. Primavera, W. L. Tomkins, P. J. Krell and A. Retnakaran (2002). "Effect of RH-5992 on adult development in the spruce budworm, *Choristoneura fumiferana*." Insect Biochemistry and Molecular Biology **32**(2): 225-31.

Sutcliffe, D. W. (1993). "Reproduction in *Gammarus* (Crustacea, Amphipoda): Female Strategies." Freshwater Forum **3**: 26-64.

Suzuki, Y., T. Matsuoka, Y. Iimura and H. Fujiwara (2002). "Ecdysteroid-dependent expression of a novel cuticle protein gene BMCPG1 in the silkworm, *Bombyx mori*." Insect Biochemistry and Molecular Biology **32**(6): 599-607.

Swida, U. (1990). "Glue protein genes in *Drosophila virilis*: their organization, developmental control of transcription and specific mRNA degradation." Development **108**(2): 269-80.

Tamone, S. L. and E. S. Chang (1993). "Methyl farnesoate stimulates ecdysteroid secretion from crab Y-organs *in vitro*." General and Comparative Endocrinology **89**(3): 425-32.

Taylor, M. R., P. Holmes, R. Duarte-Davidson, C. D. N. Humfrey and P. T. C. Harrison (1999). "A research strategy for investigating the ecological significance of endocrine disruption: report of a UK workshop." Science of the Total Environment **233**(1-3): 181-91.

Terwilliger, N. B., L. Dangott and M. Ryan (1999). "Cryptocyanin, a crustacean molting protein: evolutionary link with arthropod hemocyanins and insect hexamerins." Proceedings of the National Academy of Sciences of the United States of America **96**(5): 2013-8.

Terwilliger, N. B., M. Ryan and M. R. Phillips (2006). "Crustacean hemocyanin gene family and microarray studies of expression change during eco-physiological stress." Integrative and Comparative Biology **46**(6): 991.

Thornton, J. W., E. Need and D. Crews (2003). "Resurrecting the Ancestral Steroid Receptor: Ancient Origin of Estrogen Signaling." Science **301**(5640): 1714-17.

Timmermans, M. J., M. E. de Boer, B. Nota, T. E. de Boer, J. Marien, R. M. Klein-Lankhorst, N. M. van Straalen and D. Roelofs (2007). "Collembase: a repository for springtail genomics and soil quality assessment." BMC Genomics **8**(1): 341.

Truman, J. W., W. S. Talbot, S. E. Fahrbach and D. S. Hogness (1994). "Ecdysone receptor expression in the CNS correlates with stage-specific responses to ecdysteroids during *Drosophila* and *Manduca* development." Development **120**(1): 219-34.

Tsang, W. S., L. S. Quackenbush, B. K. Chow, S. H. Tiu, J. G. He and S. M. Chan (2003). "Organization of the shrimp vitellogenin gene: evidence of multiple genes and tissue specific expression by the ovary and hepatopancreas." Gene **303**: 99-109.

Tyler, C. R., R. van Aerle, T. H. Hutchinson, S. Maddix and H. Trip (1999). "An in vivo testing system for endocrine disruptors in fish early life stages using induction of vitellogenin." Environmental Toxicology and Chemistry **18**(2): 337-47.

Valdeolmillos, A. M., A. Viera, J. Page, I. Prieto, J. L. Santos, M. T. Parra, M. M. Heck, A. C. Martinez, J. L. Barbero, J. A. Suja and J. S. Rufas (2007). "Sequential loading of cohesin subunits during the first meiotic prophase of grasshoppers." PLoS Genetics **3**(2): e28.

Vallack, H. W., D. J. Bakker, I. Brandt, E. Brostrom-Lunden, A. Brouwer, K. R. Bull, C. Gough, R. Guardans, I. Holoubek and B. Jansson (1998). "Controlling persistent organic pollutants-what next?" Environmental Toxicology and Pharmacology **6**(3): 143-75.

Vallee, M., C. Robert, S. Methot, M. F. Palin and M. A. Sirard (2006). "Cross-species hybridizations on a multi-species cDNA microarray to identify evolutionarily conserved genes expressed in oocytes." BMC Genomics **7**(1): 113.

van de Wijngaart, D. J., M. E. van Royen, R. Hersmus, A. C. W. Pike, A. B. Houtsmuller, G. Jenster, J. Trapman and H. J. Dubbink (2006). "Novel FXXFF and FXXMF Motifs in Androgen Receptor Cofactors Mediate High Affinity and Specific Interactions with the Ligand-binding Domain." Journal of Biological Chemistry **281**(28): 19407.

Verras, M., P. Gourzi, A. Zacharopoulou and A. C. Mintzas (2002). "Developmental profiles and ecdysone regulation of the mRNAs for two ecdysone receptor isoforms in the Mediterranean fruit fly *Ceratitis capitata*." Insect Molecular Biology **11**(6): 553-65.

Verslycke, T., K. De Wasch, H. F. De Brabander and C. R. Janssen (2002). "Testosterone Metabolism in the Estuarine Mysid *Neomysis integer* (Crustacea; Mysidacea): Identification of Testosterone Metabolites and Endogenous Vertebrate-Type Steroids." General and Comparative Endocrinology **126**(2): 190-99.

Vogtli, M., M. O. Imhof, N. E. Brown, P. Rauch, M. Spindler-Barth, M. Lezzi and V. C. Henrich (1999). "Functional characterization of two Ultraspiracle forms (CtUSP-1 and CtUSP-2) from *Chironomus tentans*." Insect Biochemistry and Molecular Biology **29**(10): 931-42.

Wainwright, G., S. G. Webster, M. C. Wilkinson, J. S. Chung and H. H. Rees (1996). "Structure and Significance of Mandibular Organ-inhibiting Hormone in the Crab, *Cancer pagurus*. ." Journal of Biological Chemistry **271**(22): 12749.

Wang, R. S., S. Yeh, L. M. Chen, H. Y. Lin, C. Zhang, J. Ni, C. C. Wu, P. A. di Sant'Agnese, K. L. deMesy-Bentley and C. R. Tzeng (2006). "Androgen Receptor in Sertoli Cell Is Essential for Germ Cell Nursery and Junctional Complex Formation in Mouse Testes." Endocrinology **147**(12): 5624.

Warren, J. T., A. Petryk, G. Marques, M. Jarcho, J. P. Parvy, C. Dauphin-Villemant, M. B. O'Connor and L. I. Gilbert (2002). "Molecular and biochemical characterization of two P450 enzymes in the ecdysteroidogenic pathway of *Drosophila melanogaster*." Proceedings of the National Academy of Sciences of the United States of America **99**(17): 11043-48.

Watson, R. D. and E. Spaziani (1985). "Biosynthesis of ecdysteroids from cholesterol by crab Y-organs, and eyestalk suppression of cholesterol uptake and secretory activity, *in vitro*." General and Comparative Endocrinology **59**(1): 140-48.

Watson, R. D., K. J. Lee, Q. Shihong, L. Ming, U. Heidi R, R. Robert D and S. Eugene (2001). "Molecular Cloning, Expression, and Tissue Distribution of Crustacean Molt-Inhibiting Hormone." Integrative and Comparative Biology **41**(3): 407-17.

Watts, M. M., D. Pascoe and K. Carroll (2001). "Survival and precopulatory behaviour of *Gammarus pulex* (L.) exposed to two xenoestrogens." Water Research **35**(10): 2347-52.

Watts, M. M., D. Pascoe and K. Carroll (2002). "Population responses of the freshwater amphipod *Gammarus pulex* (L.) to an environmental estrogen, 17α-ethinylestradiol." Environmental Toxicology and Chemistry **21**(2): 445-50.

Watts, M. M., D. Pascoe and K. Carroll (2003). "Exposure to 17 α-ethinylestradiol and bisphenol A--effects on larval moulting and mouthpart structure of *Chironomus riparius*." Ecotoxicology and Environmental Safety **54**(2): 207-15.

Weller, J., G. C. Sun, B. Zhou, Q. Lan, K. Hiruma, L. M. Riddiford, P. Cherbas and J. W. Truman (2001). "Isolation and developmental expression of two nuclear receptors, MHR4 and betaFTZ-F1, in the tobacco hornworm, *Manduca sexta*." Insect Biochemistry and Molecular Biology **31**(8): 827-37.

Welton, J. S. (1979). "Life-history and production of the amphipod *Gammarus pulex* in a Dorset chalk stream." Freshwater Biology **9**: 263-75.

Welton, J. S. and R. T. Clarke (1980). "Laboratory Studies on the Reproduction and Growth of the Amphipod, *Gammarus pulex* (L.)." Journal of Animal Ecology **49**(2): 581-92.

West, J. M. (1997). "Ultrastructural and Contractile Activation Properties of Crustacean Muscle Fibres Over the Moult Cycle." Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology **117**(3): 333-45.

White, K. P., P. Hurban, T. Watanabe and D. S. Hogness (1997). "Coordination of *Drosophila* metamorphosis by two ecdysone-induced nuclear receptors." Science **276**(5309): 114-7.

Willoughby, L. G. and D. W. Sutcliffe (1976). "Experiments on feeding and growth of the amphipod *Gammarus pulex* (L.) related to its distribution in the River Duddon." Freshwater Biology **6**: 577-86.

Wren, J. (2006). Life-cycle toxicity testing in *Caenorhabditis elegans* - Comparative effects on traits and their mechanistic basis. Biosciences Department. Cardiff, Wales, UK, Cardiff University. **PhD Thesis**.

Wright, D. A. (1980). "Calcium balance in premoult and post-moult *Gammarus pulex* (Amphipoda)." Freshwater Biology **10**(6): 571-79.

Wrobel, G., J. Schlingemann, L. Hummerich, H. Kramer, P. Lichter and M. Hahn (2003). "Optimization of high-density cDNA-microarray protocols by 'design of experiments'." Nucleic Acids Research **31**(12): e67.

Wu, X., P. M. Hopkins, S. R. Palli and D. S. Durica (2004). "Crustacean retinoid-X receptor isoforms: distinctive DNA binding and receptor-receptor interaction with a cognate ecdysteroid receptor." Molecular and Cellular Endocrinology **218**(1-2): 21-38.

Yang, G. P., D. T. Ross, W. W. Kuang, P. O. Brown and R. J. Weigel (1999). "Combining SSH and cDNA microarrays for rapid identification of differentially expressed genes." Nucleic Acids Research **27**(6): 1517-23.

Yang, Y. H., M. J. Buckley and T. P. Speed (2001). "Analysis of cDNA microarray images." Briefings in Bioinformatics **2**(4): 341.

Yang, Y. H., M. J. Buckley, S. Dudoit and T. P. Speed (2002). "Comparison of methods for image analysis on cDNA microarray data." Journal of Computational and Graphical Statistics **11**(1): 108-36.

Yeh, M. S., C. J. Huang, J. H. Leu, Y. C. Lee and I. H. Tsai (1999). "Molecular cloning and characterization of a hemolymph clottable protein from tiger shrimp (*Penaeus monodon*)." European Journal of Biochemistry **266**(2): 624-33.

Zapata-Perez, O., M. Del-Rio, J. Dominguez, R. Chan, V. Ceja and G. Gold-Bouchot (2005). "Preliminary studies of biochemical changes (ethoxycoumarin O-deethylase activities and vitellogenin induction) in two species of shrimp (*Farfantepenaeus duorarum* and *Litopenaeus setiferus*) from the Gulf of Mexico." Ecotoxicology and Environmental Safety **61**(1): 98-104.

Zeleny, C. (1905). "Compensatory regulation." The Journal of Experimental Zoology **2**: 1-102.

Zerulla, M., R. Lange, T. Steger-Hartmann, G. Panter, T. Hutchinson and D. R. Dietrich (2002). "Morphological sex reversal upon short-term exposure to endocrine modulators in juvenile fathead minnow (*Pimephales promelas*)." Toxicology Letters **131**(1-2): 51-63.

Zhao, X. F., J. X. Wang, X. L. Xu, Z. M. Li and C. J. Kang (2004). "Molecular cloning and expression patterns of the molt-regulating transcription factor HHR3 from *Helicoverpa armigera*." Insect Molecular Biology **13**(4): 407-12.

Zheng, J. and V. D. Ramirez (1999). "Purification and identification of an estrogen binding protein from rat brain: oligomycin sensitivity-conferring protein (OSCP), a subunit of mitochondrial F0F1-ATP synthase/ATPase." Journal of Steroid Biochemistry and Molecular Biology **68**(1-2): 65-75.

Zhu, J., K. Miura, L. Chen and A. S. Raikhel (2003). "Cyclicity of mosquito vitellogenic ecdysteroid-mediated signaling is modulated by alternative dimerization of the RXR homologue Ultraspiracle." Proceedings of the National Academy of Sciences of the United States of America **100**(2): 544-49.

Zou, E. and M. Fingerman (1997a). "Effects of Estrogenic Xenobiotics on Molting of the Water Flea, *Daphnia magna*." Ecotoxicology and Environmental Safety **38**(3): 281-85.

Zou, E. and M. Fingerman (1997b). "Synthetic estrogenic agents do not interfere with sex differentiation but do inhibit molting of the cladoceran *Daphnia magna*." Bulletin of Environmental Contamination and Toxicology **58**(4): 596-602.

Zou, E. and R. Bonvillain (2004). "Chitinase activity in the epidermis of the fiddler crab, *Uca pugilator*, as an in vivo screen for molt-interfering xenobiotics." Comparative Biochemistry and Physiology Part C: Pharmacology, Toxicology and Endocrinology **139**(4): 225-30.

Zou, E. (2005). "Impacts of Xenobiotics on Crustacean Molting: The Invisible Endocrine Disruption." Integrative and Comparative Biology **45**(1): 33.

APPENDIX A

MATERIALS

Table 1: Biological Reagents and Kits

REAGENT	SUPPLIER
THERMO-FAST® 96-well plates (Skirted, Semi-skirted, Detection), adhesive PCR foil seals, gas permeable adhesive seals, 1ml, 200µl filter pipette tips	*AbGene, Surrey, UK.*
50ml Falcon tubes, 1ml pipette tips	*Alpha Laboratories, Eastleigh, Hants, UK.*
RNA*later*®-ICE, 5-(3-aminoallyl)-dUTP	*Ambion, Huntingdon, Cambs., UK*
Cy5-dUTP, mRNA Purification Kit, SizeSep™ 400 Spun Columns, CyScribe™ GFX™ Purification Kit, Universal ScoreCard, Universal ScoreCard DNA, CyDye Post-Labeling Reactive Dye Packs	*Amersham, Bucks., UK*
ABI Prism™ Dye Terminator Cycle Sequencing Reaction Ready Kit	*Applied Biosystems, Foster City, CA, USA*
Triton X-100	*BDH, VWR International Ltd, Poole, Dorset.*
Agarose powder	*Bioline Ltd., London, UK.*
TEMED	*BIO-RAD, Herts, UK*
Tris, DTT	*Boehringer-Mannheim Ltd., Lewes, East Sussex*
SeaPrep® agarose	*Cambrex Bio Science, Nottingham, UK.*
Advantage®2 PCR Kit, dNTPs (Ultra Pure)	*Clontech UK, Basingstoke, UK*
UltraGAPS™ coated slides	*Corning B.B., The Netherlands*
UVettes	*Eppendorf UK Ltd., Cambridge, UK.*
Ethanol, Ispropanol, Uvettes, Methanol	*Fisher*
Acetic acid, Chloroform, EDTA, Ethidium bromide, Glycerol, Hydrochloric acid, Isopropanol, SDS, Sodium chloride, Sodium hydroxide	*Fisons Scientific Equipment UK Ltd., Loughborough, UK.*
Disposable Sterile Universal Tubes, 1.5ml microcentrifuge tubes, 10µl filter pipette tips	*Greiner, Stonehouse, UK*
Ethanol	*Hayman Ltd., Witham, Essex, UK.*
DH5α™ Competent Cells, IPTG, 10x TAE, SuperScript™ II Reverse Transcriptase, First-strand reverse transcriptase buffer (5X), 0.1M DTT, E-Gel® Electrophoresis reagents, E-Gel® Low Range Molecular Marker, SYBR Green, PicoGreen	*Invitrogen Ltd., Paisley, UK.*

REAGENT	SUPPLIER
0.22µM Nucleopore™ filters, Montage® Multiscreen $PCR_{\mu 96}$ cleanup plates, Montage® SEQ_{96} Sequencing Reaction Cleanup Kit	*Millipore U.K. Ltd., Watford, Hertfordshire, U.K.*
M13 sequencing oligonucleotides, Custom Synthesised oligonucleotides, Anchored oligo dT, Random Hexamers	*MWG Biotech UK., Milton Keynes, UK.*
Pellet Paint® NF Co-Precipitant	*Novagen, Merck Chemicals Limited, Nottingham,*
Protogel®	*UK)*
100bp, 1kb DNA ladder, Restriction enzymes and their buffers	*New England Biolabs (UK) Ltd. Herts., UK.*
MacConkey agar No. 3	*Oxoid Ltd., Basingstoke, Hants., UK.*
dNTPs, loading dye, $MgCl_2$ (25mM), 10X Mg-free buffer, MMLV reverse transcriptase, RNAsin, 5X RT buffer, molecular mass markers (ΦX174 DNA/Hae III and Lambda DNA/EcoR I + Hind III) pGEM®-T Vector System, Restriction enzymes, Wizard plus SV Minipreps kit, dNTPs (dATP, dCTP, dGTP, dTTP), T4 DNA ligase, PCR purification kit, Gel purification kit.	*Promega Ltd., Southampton, UK.*
QIAquick gel purification kit, RNeasy Mini Kit	*Qiagen Ltd., Crawley, West Sussex, UK.*
Nexterion® Slide E, Nexterion® Block E, Nexterion® Spot Modified.	*SCHOTT Jena^{er} Glas GmbH, Jena, Germany.*
Ammonium persulphate (APS), ampicillin, BSA, Chloroform, Igepal-630, isoamyl alcohol, CHROMASOLV® Plus water for HPLC, Kanamycin, LB agar, LB broth, Mineral oil, Phenol, polyethylenimine (PEI), SOB media, 20X SSC, Tri® Reagent, sodium acetate, β-mercaptoethanol, Ethidium bromide, Tween, Coomassie Blue, Glycine, SDS, DMSO	*Sigma-Aldrich, Gillingham, Dorset, UK.*
10µl, 200µl pipette tips, 1ml, 200µl filter pipette tips	*Starlab UK Ltd., Milton Keynes, UK.*
pBluescript® II XR cDNA Library Construction Kit, XL10-Gold® ultracompetent cells	*Stratagene Europe, Amsterdam, The Netherlands*

Table 2: List of routinely employed solutions, reagents and media

PROCEDURE	SOLUTION	COMPONENTS
Agarose gel electrophoresis (Section 2.11.11)	Electrophoresis loading dye	*50% (v/v) Glycerol, 1% (w/v) Bromophenol blue, 1X TAE (sterilised)*
	10x TAE	*40mM Tris-acetate, 1mM Na_2-EDTA (pH 7.6)*
Reverse transcriptase - Polymerase chain reaction (RT-PCR) (Section 2.11.6)	5x MMLV-RT buffer (Promega, Southampton, UK)	*50mM Tris-HCl (pH 8.3), 75mM KCl, 3mM $MgCl_2$, 10mM DTT*
Polymerase chain reaction (PCR) (Section 2.11.9)	10X *Taq* polymerase buffer	*500mM KCl, 100mM Tris-HCl (pH 9.0 at 25 °C), 1% (v/v) Triton®X-100*
Wizard® Plus SV miniprep kit (Promega, Southampton, UK) (Section 2.11.2)	Cell resuspension solution	*50mM Tris (pH 7.5), 10mM EDTA, 100µg/ml RNase A*
	Cell lysis solution	*0.2M NaOH, 1% (w/v) SDS*
	Neutralisation solution	*4.09M Guanidine Hydrochloride, 0.759M Potassium Acetate, 2.12M Glacial Acetic Acid, (adjusted to pH 4.2)*
	Column wash	*60mM Potassium Acetate, 8.3mM Tris-HCl (pH 7.5), 0.04mM EDTA, 60% (v/v) Ethanol*
Quantitative PCR (Section 2.11.13)	10X QPCR buffer	*12.5µM ROX, 125mM Tris-HCl (pH8.3), 50mM $MgCl_2$, 625mM KCl.*

PROCEDURE	SOLUTION	COMPONENTS
cDNA Library (Section 2.12)	10x Triton® free PCR buffer	*500mM KCl,* *100mM Tris-HCl (pH 9.0 at 25 °C)*
	Library loading dye	*10mM Tris-HCl, pH 7.5* *1mM EDTA* *0.005% (w/v) bromophenol blue* *10% (v/v) glycerol*
cDNA Array (Section 2.13)	Aminoallyl dUTP / dNTP mix	*Make Fresh* *Ratio: 4:1:5, U:T:ACG* *40µl aa-dUTP (50mM)* *25µl dATP, dCTP, dGTP* *5µl dTTP (all 100mM)* *80µl sterile filtered water*
	Wash Buffer A	*0.1% (v/v) Triton®X-100*
	Wash Buffer B	*1mM HCl*
	Wash Buffer C	*10mM KCl*
	Blocking Solution B	*1% (w/v) BSA* *0.1% (w/v) SDS* *5X SSC Buffer*
	2X Hybridisation Buffer	*50% (v/v) formamide* *0.2% (w/v) SDS* *10X SSC Buffer*

Table 3: DNA marker fragment sizes (bp)

DNA MARKER	FRAGMENT SIZES (BP)	SUPPLIER
λDNA/*Hin*dIII	23130, 9416, 6557, 4361, 2322, 2027, 564, 125	*Promega, Southampton, UK*
ϕX174/*Hae*III	1353, 1078, 872, 603, 310, 281, 271, 234, 194, 148, 72	*Promega, Southampton, UK*
100bp	1517 (45ng), 1200 (35ng), 1000 (95ng), 900 (27ng), 800 (24ng), 700 (21ng), 600 (18ng), 517 (97ng), 500 (97ng), 400, (38ng), 300, (29ng), 200 (25ng), 100 (48ng)	*New England BioLabs (UK) Ltd. Herts., UK*
1kb	10002 (42ng), 8001 (42ng), 6001 (50ng), 5001 (42ng), 4001 (33ng), 3001 (125ng), 2000 (48ng), 1500 (36ng), 1000 (42ng), 517 (42ng), 500 (42ng)	*New England BioLabs (UK) Ltd. Herts., UK*

Table 4: Bacterial cloning vectors

VECTOR	SELECTIVE MARKER	CONTEXT OF USE	SOURCE
pGEM-T	Amp	Cloning of PCR products (**Error! Reference source not found.**)	*Promega, Southampton, UK*
pBluescript® SK II+	Amp	cDNA library cloning (**Error! Reference source not found.**)	*Stratagene Europe, Amsterdam, The Netherlands*

Table 5: Bacterial genotypes

E. COLI STRAIN	GENOTYPE
DH5α™ (Invitrogen Ltd., Paisley, UK)	F⁻ ϕ80*lac*ZΔM15 Δ(*lac*ZYA-*arg*F)U169 *deo*R *rec*A1 *end*A1 *hsd*R17(rk⁻, mk+) *pho*A *sup*E44 *thi*-1 *gyr*A96 *rel*A1 λ⁻
XL10-Gold® (Stratagene Europe, Amsterdam, The Netherlands)	Tet^{r}Δ(*mcrA*)*183*Δ(*mcrCB-hsdSMR-mrr*)*173 endA1 supE44 thi-1 recA1 gyrA96 relA1 lac* Hte [F' *proAB lacI*^{q}Z*ΔM15* Tn*10* (Tet^r) Amy Cam^r]

Table 6: Primers used in this study.
All primers were purchased from MWG Biotech UK, Milton Keynes, UK.

Primer	DNA Sequence (5′ → 3′)	Melting temperature	Annealing temperature
Random hexamer	NNNNNN	N/A	N/A
M13 forward	CAGGAAACAGCTATGAC	59.4°C	56.0°C
M13 reverse	GTAAAACGACGGCCAGT	57.9°C	56.0°C
COIa-H	AGTATAAGCGTCTGGGTAGTC	57.9°C	54.0°C
COI-Gf	ACAGTTGGTATAGACATTGATAC	55.3°C	54.0°C
Libn_for	CCCCCCCTCGAGTTTTTTT	56.7°C	56.0°C
Libn_rev	GCAGGAATTCGGCACGAG	58.2°C	56.0°C
T7	TAATACGACTCACTATAGGG	53.2°C	50.0°C
SP6	CATTTAGGTGACACTATA	50.2°C	50.0°C
Gp_Vg_F	TCAAGTTCAGCACCAACAGG	57.3°C	57.0°C
Gp_Vg_R	TCACGCTACCGATCACTTTG	57.3°C	57.0°C
Gp_GAPDH_F	AGTCCACTGGGGTTTTCACA	57.3°C	57.0°C
Gp_GAPDH_R	TCGGTGTATCCAAGGTAGCC	59.4°C	57.0°C
GpHR3-5_1	TGCTTGGTCATGCCCTCGACGGAC	67.8°C	variable
GpHR3-5_2	CTCGACGGACTCCTGCATGGCGC	69.6°C	variable
GpHR3-5_3	CCAGCAGCTTGGCCAGCACCACAC	69.6°C	variable
UPM	CTAATACGACTCACTATAGGGCAAGCAGTGGTATCAACGCAGAGT AND CTAATACGACTCACTATAGGGC	N/A	N/A
NUP	AAGCAGTGGTATCAACGCAGAGT	N/A	N/A

Table 7: Degenerate code

Degenerate letter	Nucleic acids
I	deoxyInosine A+C+G+T
N	A+C+G+T
H	A+C+T
B	C+G+T
V	A+C+G
D	T+G+A
R	A+G
M	A+C
S	C+G
Y	C+T
K	G+T
W	A+T

Table 8: Degenerate primers used in this study

[1] MWG Biotech UK, Milton Keynes, UK.
[2] Operon Biotechnologies GmbH, Cologne, Germany.

Candidate Gene	Primer	DNA Sequence (5′ → 3′)	Supplier
EcR	UpEcR_F	TGYGGYGAYMGNGCYTCYGGYTAYC	*MWG Biotech*[1]
	UpEcR_R	CATRCCSACRGCSAGRCAYTTYTT	*MWG Biotech*[1]
	GpEcR-1F	CACCTGTGAAGGNTGYAAAGG	*Operon*[2]
	GpEcR-1R	CATVACCTCISATGAGCAGGC	*Operon*[2]
	GpEcR-2F	ATGTAYATGMGRCGIAAGTGTCAGG	*Operon*[2]
	GpEcR-2R	GCCTTTAGYARIGTTATCTGGTCC	*Operon*[2]
	EcR_CEMDY_F1	GGTTTCTTCMGGCGGAG	*MWG Biotech*[1]
	EcR_SGYHY_F2	GARATGGACATGTAYATGCG	*MWG Biotech*[1]
	EcR_TILTV_R2	GCTGIACTGTKAGGATCGT	*MWG Biotech*[1]
	EcR_DQITL_R1	TCGCARCATCATSACCTC	*MWG Biotech*[1]
USP	UpUSP_F	TGYGGYGAYMGNGCYTCYGGYAARC	*MWG Biotech*[1]
	UpUSP_R	TTYTCYTGRACSGCYTCNCKYTTCAT	*MWG Biotech*[1]
	GpUSP-1F	CAAACATTAYGGHGTITACAGC	*Operon*[2]
	GpUSP-1R	CACCAGCTCMGABAGIACGC	*Operon*[2]
	GpUSP-2F	TTCTTCAAGCGSACAGTGC	*Operon*[2]
	GpUSP-2R	GCAATAAGCAIYTCGTTCC	*Operon*[2]
ER	GpER-1F	GCTGTCTGCAGTGAYTATGC	*Operon*[2]
	GpER-1R	AAGCCAGGAAYYTTCTTGG	*Operon*[2]
	GpER-2Fa	GTCCTGTGAAGGHTGCAAAGC	*Operon*[2]
	GpER2Fg	GTCCTGTGAAGGHTGCAAGGC	*Operon*[2]
	GpER-2R	ATCATGTACCAGYTCCTTGTCGG	*Operon*[2]
	Gp2ER-1Ftag	GGTGTKGCGAGYTGTGAGG	*MWG Biotech*[1]
	Gp2ER-1Fctc	GCGTKGCGTCYTGTGAGG	*MWG Biotech*[1]
	Gp2ER-2F	CGAGTACACSTGTCCCGC	*MWG Biotech*[1]
	Gp2ER-1R	GACRATXTCRTCCGCCTGCC	*MWG Biotech*[1]

Table 9: RACE primers used in this study

*UPM (universal primer mix) was 0.4mM UPM_Long and 2mM UPM_Short.
All oligonucleotides were obtained from MWG Biotech UK, Milton Keynes, UK.

Primer	DNA Sequence (5′ → 3′)	Supplier
GpHR3-5_1	TGCTTGGTCATGCCCTCGACGGAC	*MWG Biotech*
GpHR3-5_2	CTCGACGGACTCCTGCATGGCGC	*MWG Biotech*
GpHR3-5_3	CCAGCAGCTTGGCCAGCACCACAC	*MWG Biotech*
HHR3-RACE5	CCTCGTGCCGAATTCCTGCAGCC	*MWG Biotech*
Gp_HR3_nest_1	TCGGCGCTGGGGATATCGAAATTGG	*MWG Biotech*
Gp_HR3_nest_2	CGGCGGGGGTGGGACTGAAGCTGTA	*MWG Biotech*
Gp_HR3_nest_3	TAGGCGTCAGGAGAGGTGGGCGTGGA	*MWG Biotech*
Gp_HR3_nest_4	CGAGGCCGCCCTCATCTGGCTCT	*MWG Biotech*
UPM_Long*	CTAATACGACTCACTATAGGGCAAGCAGTGGTATCAACGCAGAGT	*MWG Biotech*
UPM_Short*	CTAATACGACTCACTATAGGGC	*MWG Biotech*
NUP	AAGCAGTGGTATCAACGCAGAGT	*MWG Biotech*

APPENDIX B

GENE LISTS

Table 1: Gene list of up and down-regulated objects due to moult cycle

Systematic	Fold Change			Cluster	Sequence Description	max eValue
	1	2	3			
Gp_maSA_29G08	39.06	-8.79	-0.73	GAC00001	---NA---	-
Gp_maSA_52A05	-0.53	-1.69	0.01	GAC00001	---NA---	-
Gp_maSA_42E05	2.01	0.95	-11.21	GAC00001	---NA---	-
Gp_mxAA_34D05	-6.52	0.74	-10.89	GAC00001	---NA---	-
Gp_maSA_42B10	2.31	1.08	-0.13	GAC00001	---NA---	-
Gp_maSA_33E06	-6.19	-8.49	-10.77	GAC00001	---NA---	-
Gp_maSA_33G02	0.07	0.67	-11.07	GAC00001	---NA---	-
Gp_maSA_24B07	-0.20	0.45	-10.84	GAC00001	---NA---	-
Gp_maSC_02A01	-9.15	0.67	0.62	GAC00001	---NA---	-
Gp_mxAA_61C01	-4.36	-9.72	-0.18	GAC00017	actin	1.00E-136
Gp_maSA_01C03	2.52	1.15	-11.34	GAC00022	cg31626-isoform a	1.00E-06
Gp_mxAA_95B11	2.40	0.82	-11.40	GAC00034	---NA---	1.00E-04
Gp_maSA_25C11	-1.99	0.83	-0.40	GAC00035	hemocyanin subunit 1	1.00E-137
Gp_maSA_31C11	53.53	0.31	-0.35	GAC00056	---NA---	-
Gp_mxAA_36F03	4.51	2.31	0.57	GAC00057	hemocyanin subunit 1	1.00E-154
Gp_maSA_35F04	1.82	0.54	0.24	GAC00057	hemocyanin subunit 1	1.00E-154
Gp_mxAA_51C06	-1.08	-0.83	-0.03	GAC00057	hemocyanin subunit 1	1.00E-154
Gp_mxAA_60G05	-0.24	1.63	1.18	GAC00057	hemocyanin subunit 1	1.00E-154
Gp_maSC_04E01	-1.03	0.26	0.72	GAC00062	zinc metalloproteinase	1.00E-42
Gp_mxAA_26B02	-0.12	-1.05	-0.30	GAC00062	zinc metalloproteinase	1.00E-42
Gp_mxAA_53E10	1.68	0.29	-0.04	GAC00065	cuticle protein	1.00E-12
Gp_mxAB_05D02	1.00	0.82	-0.15	GAC00074	alpha 2a pancreatic	1.00E-118
Gp_maSA_45B01	-6.16	2.66	-0.34	GAC00082	---NA---	-
Gp_maSA_02D10	0.66	0.67	-4.57	GAC00092	---NA---	-
Gp_maSB_03G08	1.09	0.19	0.06	GAC00093	---NA---	-
Gp_maSA_35E03	-2.26	1.02	-11.15	GAC00101	---NA---	-
Gp_mxAA_05E03	1.00	0.19	-0.09	GAC00114	---NA---	1.00E+00
Gp_maSB_03E03	1.07	-0.04	-0.34	GAC00114	---NA---	1.00E+00
Gp_mxAA_75B01	1.17	0.57	-11.25	GAC00114	---NA---	1.00E+00
Gp_mxAA_101E04	1.86	0.43	-0.18	GAC00114	---NA---	1.00E+00
Gp_mxAA_93H10	2.30	0.79	-0.02	GAC00114	---NA---	1.00E+00
Gp_mxAB_05B11	1.22	0.16	0.04	GAC00141	---NA---	-
Gp_mxAA_56H10	1.43	0.36	-0.10	GAC00141	---NA---	-
Gp_mxAA_61G05	2.22	0.25	-0.19	GAC00141	---NA---	-
Gp_maSA_49D04	2.21	1.30	-0.06	GAC00141	---NA---	-
Gp_maSA_50C02	1.20	0.48	-0.26	GAC00141	---NA---	-
Gp_mxAA_81H05	1.96	0.79	-0.12	GAC00141	---NA---	-
Gp_mxAA_71F10	18.74	0.14	10.91	GAC00141	---NA---	-
Gp_maSA_28D05	2.62	1.53	0.13	GAC00141	---NA---	-
Gp_mxAA_05A07	1.07	0.32	0.01	GAC00141	---NA---	-
Gp_mxAA_58F04	2.65	0.73	-0.16	GAC00141	---NA---	-
Gp_mxAA_89C08	1.92	0.62	-11.16	GAC00141	---NA---	-
Gp_mxAA_86F01	1.78	0.97	-0.32	GAC00162	atph+mitochondrial f1beta polypeptide	1.00E-78
Gp_maSA_41B11	0.79	1.51	0.05	GAC00225	rh28_orysjdead-box atp-dependent rna helicase 28	1.00E+00
Gp_maSA_23F02	0.60	1.68	-0.06	GAC00232	trypsin	1.00E-89
Gp_maSA_26C03	1.17	1.56	-0.19	GAC00232	trypsin	1.00E-89
Gp_mxAA_57B02	1.75	1.84	-11.53	GAC00232	trypsin	1.00E-89
Gp_mxAA_05F05	-0.19	-1.65	-0.68	GAC00243	cytoskeletal actinb	1.00E-145
Gp_mxAA_93E06	1.05	-0.54	-2.26	GAC00243	cytoskeletal actinb	1.00E-145
Gp_mxAA_10G01	-0.15	-2.28	-0.69	GAC00268	actin	1.00E-127
Gp_maSB_03G01	1.16	0.30	0.29	GAC00296	---NA---	-
Gp_maSA_21H02	0.71	0.89	-11.06	GAC00363	iron-sulfur cluster scaffold homolog	1.00E-14
Gp_maSA_22C10	0.46	0.02	1.87	GAC00380	---NA---	-

Systematic	Fold Change			Cluster	Sequence Description	max eValue
	1	2	3			
Gp_mxAA_101H05	1.75	0.65	-0.02	GAC00388	alpha-amylase	1.00E-95
Gp_mxAA_19H01	1.23	1.15	-0.01	GAC00388	alpha-amylase	1.00E-95
Gp_mxAA_90B01	6.34	0.78	-0.06	GAC00388	alpha-amylase	1.00E-95
Gp_mxAA_60H02	-7.53	-0.56	3.11	GAC00390	---NA---	-
Gp_maSA_22H07	-8.28	0.75	0.84	GAC00414	---NA---	-
Gp_mxAA_77F02	-0.49	-2.60	0.30	GAC00436	glycosyl hydrolase family7	1.00E-93
Gp_mxAA_47E10	1.14	-0.56	-0.30	GAC00438	cuticle proprotein	1.00E-51
Gp_maSA_23D11	1.46	-0.03	0.06	GAC00448	---NA---	-
Gp_maSA_23F07	-7.46	0.16	0.12	GAC00457	---NA---	-
Gp_maSA_23G05	-7.37	9.47	0.47	GAC00464	actin	-
Gp_maSA_23G08	1.28	-1.91	1.05	GAC00466	---NA---	-
Gp_maSA_45D01	1.79	1.54	-0.05	GAC00466	---NA---	-
Gp_mxAA_45H07	-0.97	-1.02	-0.02	GAC00467	serine protease	1.00E-49
Gp_maSA_46B11	0.14	-1.23	0.84	GAC00481	fatty acid binding protein 10	1.00E-13
Gp_maSA_24B03	-1.86	-1.04	1.41	GAC00487	---NA---	-
Gp_maSA_24C09	-0.24	1.58	-10.48	GAC00501	---NA---	-
Gp_mxAA_13D09	1.03	0.51	0.17	GAC00772	actin	1.00E-122
Gp_maSA_29C10	-6.61	1.07	0.04	GAC00772	actin	1.00E-122
Gp_mxAA_59H10	-0.64	0.95	1.51	GAC00772	actin	1.00E-122
Gp_maSA_33C05	-0.73	2.41	0.23	GAC00796	---NA---	-
Gp_maSA_36F12	-0.66	-0.12	1.38	GAC00807	reverse transcriptase	1.00E-20
Gp_maSA_32A10	1.33	0.72	0.06	GAC00949	---NA---	-
Gp_maSA_33A06	1.30	1.09	-1.31	GAC01002	---NA---	-
Gp_maSA_34H02	-1.73	0.00	23.31	GAC01113	---NA---	-
Gp_maSA_35G05	-0.40	39.99	-0.63	GAC01173	---NA---	-
Gp_maSA_37F04	-0.64	-0.18	16.53	GAC01290	---NA---	-
Gp_maSA_38F03	13.81	0.15	-0.32	GAC01352	---NA---	-
Gp_mxAA_41G01	-1.12	-1.32	-0.22	GAC01379	---NA---	-
Gp_mxAA_22A06	-2.00	-0.43	0.29	GAC01379	---NA---	-
Gp_mxAA_13A12	-0.98	-1.16	-0.29	GAC01379	---NA---	-
Gp_mxAA_75B09	-1.18	-1.66	-0.18	GAC01379	---NA---	-
Gp_mxAA_16C09	0.10	-0.12	1.22	GAC01379	---NA---	-
Gp_mxAA_22D07	-1.47	-1.11	0.01	GAC01379	---NA---	-
Gp_maSA_39H04	14.57	1.02	-0.45	GAC01392	---NA---	-
Gp_maSA_41C01	1.80	0.80	-11.25	GAC01487	---NA---	-
Gp_maSA_41G07	-0.01	-1.28	0.64	GAC01523	protein unq6350 pro21055 homolog precursor	1.00E+00
Gp_maSA_44D10	3.64	4.57	7.67	GAC01682	---NA---	-
Gp_mxAA_50A05	1.67	0.25	0.01	GAC01715	mgc79767 protein	1.00E-59
Gp_mxAA_92A08	0.16	-1.21	-0.32	GAC01746	actin 1	1.00E-11
Gp_maSA_45H01	1.50	0.83	-0.08	GAC01767	clot_penmohemolymph clottable protein precursor	1.00E-08
Gp_maSA_46D07	-0.62	-1.15	0.21	GAC01793	---NA---	-
Gp_maSA_50H07	8.34	0.72	-10.89	GAC02014	---NA---	-
Gp_maSB_01D01	1.23	-9.09	-1.67	GAC02204	---NA---	-
Gp_mxAA_95A12	1.64	-1.60	-0.38	GAC02329	heterogeneous nuclear ribonucleoprotein	1.00E-12
Gp_mxAA_90C02	-1.04	-0.45	0.04	GAC02376	ferritin	1.00E-69
Gp_maSB_07H04	11.72	-0.72	0.33	GAC02431	---NA---	-
Gp_mxAA_74A03	7.87	-0.46	0.12	GAC02437	---NA---	-
Gp_mxAA_94D12	-4.39	1.53	-1.84	GAC02437	---NA---	-
Gp_mxAA_12A03	4.33	0.72	0.22	GAC02437	---NA---	-
Gp_mxAA_50D08	-7.88	-0.51	0.25	GAC02437	---NA---	-
Gp_mxAA_78H09	-7.15	0.66	0.94	GAC02437	---NA---	-
Gp_mxAA_04B05	1.32	0.34	-0.07	GAC02437	---NA---	-
Gp_mxAA_33A11	1.29	-9.88	-2.60	GAC02437	---NA---	-
Gp_mxAA_76C04	1.18	-0.70	-0.10	GAC02437	---NA---	-
Gp_mxAA_04H02	-0.70	-0.69	1.48	GAC02438	ribonuclease e	1.00E-05
Gp_mxAA_77A06	4.34	1.00	0.09	GAC02447	---NA---	-
Gp_mxAA_20G09	7.29	1.14	-0.13	GAC02465	pupal cuticle protein	1.00E-18
Gp_mxAA_73H03	-6.70	0.78	-0.62	GAC02465	pupal cuticle protein	1.00E-18
Gp_mxAA_17D01	0.16	-1.31	-0.32	GAC02474	---NA---	-
Gp_mxAA_04E08	4.55	-0.60	-0.19	GAC02478	---NA---	1.00E-27
Gp_mxAA_04E09	1.30	-1.06	-0.80	GAC02479	---NA---	-
Gp_mxAA_27E04	7.61	0.72	-0.31	GAC02483	cg3355-isoform a	1.00E-43
Gp_mxAA_84H08	0.37	0.44	-11.19	GAC02485	---NA---	-
Gp_mxAA_69C10	0.33	-3.05	-0.16	GAC02489	c-type lectin	1.00E-17
Gp_mxAB_06H07	-2.03	-0.03	-0.37	GAC02492	cytochrome oxidase subunit 3	1.00E-55
Gp_mxAA_36C11	16.48	0.33	-0.29	GAC02496	---NA---	1.00E+00
Gp_mxAA_14E06	0.07	0.51	-22.22	GAC02496	---NA---	1.00E+00
Gp_mxAA_42H07	1.57	0.41	0.08	GAC02497	cytochrome b	1.00E-88

Systematic	Fold Change			Cluster	Sequence Description	max eValue
	1	2	3			
Gp_mxAA_35G12	0.46	-3.85	0.52	GAC02506	type iii restrictionres subunit	1.00E-11
Gp_mxAA_20G04	-1.10	-0.62	-0.20	GAC02514	---NA---	-
Gp_mxAA_55E08	-20.87	1.04	28.13	GAC02539	serine collagenase 1 precursor	1.00E-57
Gp_mxAA_80H11	0.44	-9.02	-11.14	GAC02547	cg30045-pa	1.00E-25
Gp_mxAA_95D10	0.41	-1.10	-0.01	GAC02547	cg30045-pa	1.00E-25
Gp_mxAA_05H01	-1.24	-1.35	-0.24	GAC02551	---NA---	-
Gp_mxAA_53B02	0.18	-0.30	1.20	GAC02624	glutamine synthetase	1.00E-105
Gp_mxAA_32C10	0.36	-1.84	3.04	GAC02644	ribosomal protein sa	1.00E-89
Gp_mxAA_38A11	-7.12	-8.86	2.05	GAC02668	tenascin xb	1.00E+00
Gp_mxAA_93F12	0.52	-0.69	-11.97	GAC02680	---NA---	-
Gp_mxAA_50E11	-0.53	-8.92	0.78	GAC02680	---NA---	-
Gp_mxAA_63F12	0.73	0.31	-10.97	GAC02735	---NA---	-
Gp_mxAA_21C01	1.54	0.10	-0.27	GAC02751	ubiquitin-conjugating enzyme e2d 3 (ubc4 5yeast)isoform cra_a	1.00E-79
Gp_mxAA_99A04	-7.58	0.76	0.89	GAC02871	gdp dissociation inhibitor 2	1.00E-70
Gp_mxAA_70F01	0.38	0.33	-11.03	GAC02950	---NA---	-
Gp_mxAA_90H10	-4.91	-0.08	-11.10	GAC02980	fatty acid binding protein 4	1.00E-14
Gp_mxAA_94G01	-0.83	-1.20	-0.24	GAC03008	---NA---	-
Gp_mxAA_27A05	-0.27	5.19	0.99	GAC03014	non-smc element 1 homolog	1.00E-20
Gp_mxAA_70G12	-0.21	1.93	1.38	GAC03014	non-smc element 1 homolog	1.00E-20
Gp_mxAA_68D01	-0.59	1.28	1.70	GAC03014	non-smc element 1 homolog	1.00E-20
Gp_mxAA_75D04	-0.68	1.81	1.62	GAC03014	non-smc element 1 homolog	1.00E-20
Gp_mxAA_11D05	0.34	1.11	-0.12	GAC03111	---NA---	-
Gp_mxAA_22D09	-0.53	-9.38	1.04	GAC03140	---NA---	-
Gp_mxAA_19F03	4.56	0.61	-0.06	GAC03414	ribonuclease e	1.00E-07
Gp_mxAA_20H10	0.51	1.22	0.07	GAC03473	muscle-specific protein 300 cg33715-isoform d	1.00E-07
Gp_mxAA_50F09	-8.03	-0.54	0.21	GAC03505	---NA---	-
Gp_mxAA_70C07	-0.64	4.07	0.94	GAC03505	---NA---	-
Gp_mxAA_46H11	0.51	0.25	-1.08	GAC03528	---NA---	-
Gp_mxAA_49D02	1.61	-9.49	1.21	GAC03687	---NA---	-
Gp_mxAA_54E11	2.80	0.85	-0.22	GAC03726	nadh dehydrogenase subunit 1	1.00E-66
Gp_mxAA_28H09	-7.59	-0.30	0.84	GAC03764	heat shock protein hsp40	1.00E-13
Gp_mxAA_31B11	-10.48	-0.43	0.23	GAC03836	tail-specific thyroid hormone up-regulated (gene 5)	1.00E+00
Gp_mxAA_32D07	-0.38	-9.12	0.53	GAC03876	ubiquitin family protein	1.00E-97
Gp_mxAA_32F05	-7.32	0.01	-1.29	GAC03884	elongation factor 1-alpha	1.00E-93
Gp_mxAA_35D05	-0.68	-1.01	-0.24	GAC03966	differentiation regulator	1.00E+00
Gp_mxAA_37G07	-0.93	-1.03	0.02	GAC04054	orf2-encoded protein	1.00E-17
Gp_mxAA_90E09	0.12	0.23	1.22	GAC04065	---NA---	-
Gp_mxAA_96G05	3.07	2.23	-10.87	GAC04206	---NA---	-
Gp_mxAA_48B02	-7.01	-0.01	-0.86	GAC04432	1-acylglycerol-3-phosphate o-acyltransferase 4 (lysophosphatidic aciddelta)	1.00E-14
Gp_mxAA_49A10	1.35	-0.44	-0.49	GAC04474	cuticle proprotein	1.00E-53
Gp_mxAA_49C07	3.94	-0.19	-1.54	GAC04485	ribonuclease e	1.00E-07
Gp_mxAA_73D01	0.90	-1.03	-0.82	GAC04650	dead (asp-glu-ala-asp) box polypeptide 55	1.00E-51
Gp_mxAA_91C03	0.23	0.81	1.64	GAC04741	---NA---	-
Gp_mxAA_60H11	2.53	1.46	0.21	GAC05075	---NA---	-
Gp_mxAA_93G12	-0.11	-18.16	1.00	GAC05534	---NA---	-
Gp_mxAA_73C05	5.29	0.03	0.11	GAC05629	---NA---	-
Gp_mxAA_78H03	-0.75	-1.18	0.01	GAC05886	cortical granule protein with ldl-receptor-like repeats	1.00E-05
Gp_mxAA_80G01	5.71	0.70	-0.07	GAC05975	---NA---	-
Gp_mxAA_84A08	-14.03	0.65	1.25	GAC06127	---NA---	-
Gp_mxAA_90C04	0.22	0.21	6.75	GAC06428	---NA---	-
Gp_mxAA_96F03	-7.39	0.21	0.52	GAC06719	---NA---	1.00E-05
Gp_mxAA_97E02	-6.98	0.53	1.27	GAC06769	---NA---	-
Gp_mxAA_99E11	-2.07	-0.43	13.04	GAC06859	---NA---	-
Gp_mxAA_44B11	-0.25	2.21	2.03	GAC07218	non-smc element 1 homolog	1.00E-16
Gp_mxAB_04A02	-21.68	0.02	-0.03	GAC07239	---NA---	-
Gp_mxAB_08B08	3.06	2.99	0.21	GAC07331	fatty acid binding protein 4	1.00E-14
Gp_mxAB_08H03	0.08	-0.42	1.88	GAC07391	rna polymerase ii large subunit	1.00E-04
Gp_mxAB_10C05	0.16	-2.22	-0.27	GAC07506	-	1.00E-170
Gp_mxAB_11A04	0.07	-0.46	3.04	GAC07569	---NA---	-

Table 2: Gene list of up and down-regulated objects due to gender

Systematic	Fold Change		Cluster	Sequence Description	max eValue
	Female	Male			
Gp_maSA_25B02	-8.7	0.3	GAC00001	—NA—	-
Gp_maSA_38B10	0.1	-15.7	GAC00001	—NA—	-
Gp_maSA_42B06	56.8	-0.2	GAC00001	—NA—	-
Gp_mxAA_16A12	2.4	-0.9	GAC00001	—NA—	-
Gp_maSA_47B06	-1.2	0	GAC00003	cytochrome c oxidase subunit iii	1.00E-91
Gp_maSA_01B05	17.4	-0.4	GAC00013	—NA—	-
Gp_maSA_02B10	-0.1	-1.4	GAC00041	—NA—	-
Gp_maSA_52E07	4.4	-0.1	GAC00056	—NA—	-
Gp_maSA_48D08	-2.9	0.2	GAC00056	—NA—	-
Gp_mxAA_50A11	0.9	-9.2	GAC00057	hemocyanin subunit 1	1.00E-154
Gp_maSB_07H06	1.3	-0.4	GAC00067	hemocyanin subunit 1	1.00E-21
Gp_maSA_44C11	2.3	3.4	GAC00070	chitinase 3 precursor	1.00E+00
Gp_maSA_38G07	-1.2	-0.2	GAC00075	clot_penmohemolymph clottable protein precursor	1.00E-08
Gp_maSA_33G07	0.2	-7.9	GAC00081	chaoptic cg1744-pa	1.00E-05
Gp_maSA_36E11	2.3	-1.6	GAC00081	chaoptic cg1744-pa	1.00E-05
Gp_maSA_28A06	3.6	-0.5	GAC00103	protein	1.00E-09
Gp_mxAA_10H10	2.6	-8.1	GAC00114	—NA—	1.00E+00
Gp_mxAA_46F06	-2.5	-7.6	GAC00114	—NA—	1.00E+00
Gp_maSA_02G06	1.5	-0.1	GAC00114	—NA—	1.00E+00
Gp_maSB_03E03	0.2	-1.2	GAC00114	—NA—	1.00E+00
Gp_mxAA_81E01	1.1	-0.7	GAC00141	—NA—	-
Gp_mxAA_10F09	2.1	-0.4	GAC00141	—NA—	-
Gp_mxAA_56H10	1.1	-0.2	GAC00141	—NA—	-
Gp_maSA_28D05	1.8	-0.3	GAC00141	—NA—	-
Gp_mxAA_91B12	2.7	0.2	GAC00162	atph+mitochondrial f1beta polypeptide	1.00E-78
Gp_maSA_43E12	14.2	0.6	GAC00198	—NA—	-
Gp_maSA_49A03	1	0.1	GAC00225	rh28_orysjdead-box atp-dependent rna helicase 28	1.00E+00
Gp_mxAA_29H06	-1.2	-0.6	GAC00232	trypsin	1.00E-89
Gp_mxAA_82G02	-5.3	0.1	GAC00243	cytoskeletal actinb	1.00E-145
Gp_mxAA_93E04	-3.9	1.3	GAC00243	cytoskeletal actinb	1.00E-145
Gp_maSA_49B01	5.4	-0.1	GAC00287	—NA—	-
Gp_maSA_37H11	0.7	-15.1	GAC00449	—NA—	1.00E+00
Gp_maSC_02A02	2.1	0.3	GAC00466	—NA—	-
Gp_maSA_47B12	4	-0.4	GAC00477	—NA—	-
Gp_maSA_24E11	2.8	0.8	GAC00514	—NA—	-
Gp_maSA_25B12	-1.6	-0.9	GAC00558	—NA—	1.00E+00
Gp_mxAA_92G01	3.1	-1.9	GAC00590	alpha 4	1.00E-28
Gp_maSA_51B12	-10.2	0.3	GAC00601	—NA—	-
Gp_maSA_30C08	-4.1	-0.3	GAC00605	igsf22 protein	1.00E+00
Gp_mxAA_75G10	-6.5	0.1	GAC00611	secreted protein	1.00E+00
Gp_maSA_27F10	-4.8	29.9	GAC00662	—NA—	-
Gp_maSA_27H12	-20.2	0	GAC00673	—NA—	-
Gp_maSA_28B06	-1.9	-0.3	GAC00686	alcohol dehydrogenase	1.00E-09
Gp_maSA_28C05	-0.1	-1.2	GAC00695	-	1.00E-170
Gp_maSA_29C08	80.4	-1.2	GAC00770	—NA—	-
Gp_maSA_52B04	0.8	-7.8	GAC00810	nadh dehydrogenase subunit 2	1.00E-28
Gp_mxAA_47F05	-9.9	0.7	GAC00898	troponin c 25d	1.00E-27
Gp_maSA_31C07	0.7	-7.5	GAC00903	muscle lim protein	1.00E-23
Gp_maSA_34A08	1.4	-0.3	GAC01035	—NA—	-
Gp_maSA_34C07	-1.8	0.3	GAC01079	c-type lectin	1.00E-10
Gp_maSA_34G06	-13.2	0.1	GAC01109	—NA—	-
Gp_maSB_07F04	1.2	0.1	GAC01184	glutathione peroxidase 1	1.00E-05
Gp_maSA_36G11	0	-9	GAC01239	chitin binding protein	1.00E-24
Gp_maSA_37E07	-10.4	0.2	GAC01283	—NA—	-
Gp_mxAA_80A02	37.6	-2.6	GAC01379	—NA—	-
Gp_maSA_43F11	-0.3	1.1	GAC01637	brix domain containing 1	1.00E+00
Gp_mxAA_50A05	1.3	-0.3	GAC01715	mgc79767 protein	1.00E-59
Gp_maSA_48A10	0.3	-1.3	GAC01870	—NA—	-
Gp_maSA_48H11	0.8	-7.7	GAC01914	ferric-chelate reductase 1	1.00E-16
Gp_maSA_50G04	10.7	0.1	GAC02006	—NA—	-
Gp_maSA_51C12	-1.9	-0.1	GAC02040	—NA—	-
Gp_maSA_51D12	-11.1	-0.5	GAC02048	—NA—	-
Gp_maSA_51G04	1.6	-0.4	GAC02072	maltase-like protein agm1	1.00E-26
Gp_mxAA_34C04	7	-0.1	GAC02087	actin	1.00E-27
Gp_maSC_05E03	0.3	-7.7	GAC02099	—NA—	-
Gp_maSA_53E05	1	-7.5	GAC02167	—NA—	-
Gp_maSB_02D09	-2.4	0.2	GAC02243	—NA—	-
Gp_maSB_03G10	1.3	-0.7	GAC02306	c-type lectin	1.00E-17
Gp_mxAA_94D12	1.7	-7.6	GAC02437	—NA—	-
Gp_mxAA_29C11	-2.1	-0.8	GAC02437	—NA—	-
Gp_mxAA_04B05	1.5	-0.1	GAC02437	—NA—	-

Systematic	Fold Change		Cluster	Sequence Description	max eValue
	Female	Male			
Gp_mxAA_91G06	2	-1.2	GAC02437	---NA---	-
Gp_mxAA_76A07	1.7	-0.3	GAC02450	cd209-like protein	1.00E-04
Gp_mxAA_41B04	-10.7	0.4	GAC02459	---NA---	-
Gp_mxAA_04D04	-3	-7.3	GAC02465	pupal cuticle protein	1.00E-18
Gp_mxAA_20G09	6.5	-0.4	GAC02465	pupal cuticle protein	1.00E-18
Gp_mxAA_46C01	0.1	-3.2	GAC02470	pupal cuticle protein	1.00E-19
Gp_mxAA_32E10	1.6	-0.8	GAC02470	pupal cuticle protein	1.00E-19
Gp_mxAA_66F06	-1.2	0.1	GAC02472	---NA---	-
Gp_mxAA_52A11	3.6	0.1	GAC02492	cytochrome oxidase subunit 3	1.00E-55
Gp_mxAA_12D02	1.9	-0.4	GAC02492	cytochrome oxidase subunit 3	1.00E-55
Gp_mxAA_39C01	-0.5	-1.2	GAC02495	glycine rich protein	1.00E-06
Gp_mxAA_28E07	-1.9	-0.6	GAC02495	glycine rich protein	1.00E-06
Gp_mxAA_95F08	-10.7	-0.2	GAC02497	cytochrome b	1.00E-88
Gp_mxAA_42H07	1.1	-0.6	GAC02497	cytochrome b	1.00E-88
Gp_mxAA_23F07	0.6	1.8	GAC02526	---NA---	-
Gp_mxAA_82B02	4.5	-0.9	GAC02527	cg31973-isoform a	1.00E-44
Gp_mxAA_39G04	-2.8	-8.4	GAC02542	pupal cuticle protein	1.00E-19
Gp_mxAA_80H11	0.3	-15.2	GAC02547	cg30045-pa	1.00E-25
Gp_mxAA_75G05	-1.1	0.1	GAC02548	---NA---	-
Gp_mxAA_20A06	8.3	-0.5	GAC02561	pupal cuticle protein	1.00E-19
Gp_mxAA_59E10	7.6	-0.6	GAC02628	---NA---	-
Gp_mxAA_07F12	56.9	-0.7	GAC02644	ribosomal protein sa	1.00E-89
Gp_mxAA_97C08	-1.1	0.5	GAC02684	nadh dehydrogenase subunit 5	1.00E-27
Gp_mxAA_87E07	-0.1	-2.2	GAC02728	apolipoprotein d	1.00E-17
Gp_mxAA_85C02	-1	-0.2	GAC02739	---NA---	-
Gp_mxAA_21C01	1.8	-0.1	GAC02751	ubiquitin-conjugating enzyme e2d 3 (ubc4 5yeast)isoform cra_a	1.00E-79
Gp_mxAA_100B08	0.8	-15.4	GAC02763	cuticular protein	1.00E-18
Gp_maSC_03C07	0.4	-8.1	GAC02863	actin 1	1.00E-10
Gp_maSC_03H11	0	-1.1	GAC02894	---NA---	-
Gp_mxAA_55D03	0.7	-8.6	GAC02969	---NA---	-
Gp_mxAA_100E06	0.8	-7.6	GAC02977	tubulin alpha-3 chain (alpha-tubulin 3)	1.00E-115
Gp_mxAA_13A02	1.2	-8.2	GAC02980	fatty acid binding protein 4	1.00E-14
Gp_mxAA_35A03	2	-9	GAC03010	acid alpha glucosidase	1.00E-61
Gp_mxAA_75D04	-0.9	1	GAC03014	non-smc element 1 homolog	1.00E-20
Gp_mxAA_92F05	-2.2	0.3	GAC03014	non-smc element 1 homolog	1.00E-20
Gp_mxAA_66D03	-1.6	-0.5	GAC03014	non-smc element 1 homolog	1.00E-20
Gp_mxAA_27A05	-1.3	1.8	GAC03014	non-smc element 1 homolog	1.00E-20
Gp_mxAA_23E08	1.1	-5.1	GAC03108	hemocyanin subunit 1	1.00E-113
Gp_mxAA_12H08	1.2	-0.3	GAC03173	---NA---	-
Gp_mxAA_14B06	-1.3	-0.2	GAC03223	atph+mitochondrial f1o subunit (oligomycin sensitivity conferring protein)	1.00E-49
Gp_mxAA_26B12	0.8	-15.1	GAC03236	---NA---	-
Gp_mxAA_15D09	-1.7	-0.7	GAC03273	---NA---	-
Gp_mxAA_71F08	2	-5.1	GAC03285	---NA---	1.00E+00
Gp_mxAA_93E12	1.2	-8	GAC03285	---NA---	1.00E+00
Gp_mxAA_94E02	3.4	-8.2	GAC03285	---NA---	1.00E+00
Gp_mxAB_06G01	2.4	-0.9	GAC03285	---NA---	1.00E+00
Gp_mxAB_04C12	8.9	-1.1	GAC03285	---NA---	1.00E+00
Gp_mxAA_84H06	6.9	-9.5	GAC03285	---NA---	1.00E+00
Gp_mxAA_29F11	-1.4	-0.2	GAC03321	---NA---	-
Gp_mxAA_21G12	-1.1	-0.3	GAC03503	---NA---	-
Gp_mxAA_28F05	-10.9	-0.4	GAC03752	---NA---	-
Gp_mxAA_29H11	1.6	0	GAC03793	---NA---	-
Gp_mxAA_61F05	1.8	0	GAC03822	fcgbp protein	1.00E-08
Gp_mxAA_32G03	3.5	-2.4	GAC03822	fcgbp protein	1.00E-08
Gp_mxAA_92A11	7.4	-16.1	GAC03822	fcgbp protein	1.00E-08
Gp_mxAA_30H09	-3	-1	GAC03827	si:ch211-	1.00E-20
Gp_mxAA_34D06	-1.1	-0.4	GAC03938	ift172 intraflagellar transport particle protein 172	1.00E-55
Gp_mxAA_36E05	-0.3	-2.3	GAC04001	protein	1.00E-10
Gp_mxAA_36H04	0	-4.5	GAC04017	---NA---	-
Gp_mxAA_71A10	-12.3	1	GAC04065	---NA---	-
Gp_mxAA_38F10	-0.7	-1.3	GAC04087	---NA---	-
Gp_mxAB_05G06	3.2	-0.4	GAC04308	vitellogenin fused with superoxide dismutase	1.00E-04
Gp_mxAA_51H05	-12.4	-0.7	GAC04312	---NA---	-
Gp_mxAA_94C08	2.4	-7.7	GAC04336	mucin-like peritrophin	1.00E-08
Gp_mxAA_53G08	1.1	-0.9	GAC04346	clot_penmohemolymph clottable protein precursor	1.00E+00
Gp_mxAA_78A04	-11.1	2	GAC04386	---NA---	1.00E-12
Gp_mxAA_46H12	1.1	-0.4	GAC04386	---NA---	1.00E-12
Gp_mxAA_48H04	-1.1	-0.4	GAC04467	set and mynd domain containing 3	1.00E-44

Systematic	Fold Change		Cluster	Sequence Description	max eValue
	Female	Male			
Gp_mxAA_52D11	-9.9	0.4	GAC04649	minichromosome maintenance complex component 6	1.00E-126
Gp_mxAA_58C06	-1.3	-7.7	GAC04951	—NA—	-
Gp_mxAA_72E07	-0.8	1.2	GAC05076	—NA—	-
Gp_mxAA_91B08	11.6	-0.3	GAC05108	fip1 like 1	1.00E+00
Gp_mxAA_63H10	-1.5	-0.5	GAC05213	—NA—	-
Gp_mxAA_72A11	-0.4	-8.5	GAC05573	—NA—	-
Gp_mxAA_74D08	2	-0.8	GAC05685	bubblegum related protein	1.00E-46
Gp_mxAA_75G07	-1.2	-0.1	GAC05740	gaba-gated ion channel	1.00E+00
Gp_mxAA_77D12	-1.3	4.3	GAC05814	#NAME?	1.00E-04
Gp_mxAA_77H11	-2.6	0.8	GAC05837	cg31626-isoform a	1.00E+00
Gp_mxAA_84C06	-10.7	-0.3	GAC06139	chitinase 10	1.00E-15
Gp_mxAA_92C04	1.6	-0.2	GAC06506	—NA—	-
Gp_mxAA_93C10	0.3	-7.7	GAC06559	serine protease	1.00E-66
Gp_mxAA_95A03	-10.4	0.3	GAC06629	—NA—	1.00E+00
Gp_mxAA_95D02	-1.9	-0.2	GAC06646	ahe structure oft ph	1.00E-46
Gp_mxAA_96H11	-1.1	-0.2	GAC06738	—NA—	-
Gp_mxAB_02E04	1.7	-0.1	GAC06973	—NA—	-
Gp_mxAB_05B10	-1.3	0.1	GAC07091	microsomal glutathione s-transferase 3	1.00E-30
Gp_mxAB_06C04	-0.2	-1.1	GAC07158	—NA—	-
Gp_mxAB_07A06	2	-0.5	GAC07203	—NA—	-
Gp_mxAB_07C02	-0.4	-1.3	GAC07249	—NA—	-
Gp_mxAB_09G04	0.5	-8.1	GAC07465	—NA—	-
Gp_mxAB_02E03	-2	0.7	N/A	N/A	N/A
Gp_mxAB_04A01	-24.7	5.1	N/A	N/A	N/A
Gp_mxAB_04G08	-0.2	-19.1	N/A	N/A	N/A
Gp_maSA_47C06	0.3	-1.5	N/A	N/A	N/A
Gp_maSA_46A10	3.6	-0.2	N/A	N/A	N/A
Gp_mxAA_69F07	-10.1	-0.9	N/A	N/A	N/A
Gp_mxAA_62E05	-10.9	0	N/A	N/A	N/A
Gp_maSA_24G04	-10.1	-0.3	N/A	N/A	N/A
Gp_mxAA_77H04	-20.4	0.5	N/A	N/A	N/A
Gp_mxAA_43B09	2.3	-0.2	N/A	N/A	N/A
Gp_mxAB_03C09	-10.8	0.7	N/A	N/A	N/A
Gp_maSB_06C02	10.1	5.2	N/A	N/A	N/A
Gp_mxAA_77C01	-1.2	0.5	N/A	N/A	N/A
Gp_maSA_48C01	0.4	-7.9	N/A	N/A	N/A
Gp_mxAA_53A06	2.8	-0.5	N/A	N/A	N/A
Gp_mxAA_33C02	3.2	-0.5	N/A	N/A	N/A
Gp_mxAA_73E11	-21.2	-0.6	N/A	N/A	N/A
Vg10	1.5	-2.4	N/A	N/A	N/A
Vg7	1.5	-2.3	N/A	N/A	N/A
Gp_mxAA_98B05	0.9	-7.6	N/A	N/A	N/A
Gp_mxAA_08E04	-10	0.4	N/A	N/A	N/A
Vg9	1.3	-2.7	N/A	N/A	N/A
Vg6	1.6	-1.6	N/A	N/A	N/A
Vg8	1.5	-2.3	N/A	N/A	N/A
Gp_mxAA_28C05	-1.4	0	N/A	N/A	N/A
Gp_maSB_05F10	0.4	-8.5	N/A	N/A	N/A
Gp_mxAA_26B04	-1.8	-0.8	N/A	N/A	N/A

Table 3: Gene list of up and down-regulated objects due to development

Systematic	Fold Change				Cluster	Sequence description	max eValue
	Adult	JM	JS	Neo			
Gp_maSA_42H05	21.7	-0.1	-0.6	-0.3	GAC00001	---NA---	-
Gp_mxAA_82C01	1.1	-3.7	-3	-3.7	GAC00001	---NA---	-
Gp_maSA_30A06	0.3	-1.2	14.8	-2.8	GAC00001	---NA---	-
Gp_maSA_34G05	1.1	-2.7	-0.3	-1.7	GAC00001	---NA---	-
Gp_mxAA_81F09	-0.9	1	0.1	0.2	GAC00004	---NA---	-
Gp_mxAA_64B07	0	-1.1	-15.2	-2.1	GAC00057	hemocyanin subunit 1	1.00E-154
Gp_mxAA_47A05	0.7	-0.5	-0.1	-1.1	GAC00057	hemocyanin subunit 1	1.00E-154
Gp_maSA_42E06	0.2	-14.9	-2	-1.5	GAC00074	alpha 2a pancreatic	1.00E-118
Gp_mxAA_33E11	0.5	-1.2	-0.8	-2.7	GAC00114	---NA---	1.00E+00
Gp_maSA_46E11	0.9	0.2	-0.4	-1.3	GAC00228	---NA---	-
Gp_mxAA_12C07	0.7	-1.2	-0.7	-2.3	GAC00232	trypsin	1.00E-89
Gp_mxAA_73E05	1.8	-4.9	-1.7	-6.9	GAC00243	cytoskeletal actinb	1.00E-145
Gp_mxAA_59E11	1	-0.7	-0.3	-2.3	GAC00243	cytoskeletal actinb	1.00E-145
Gp_maSB_02B10	0.9	-0.6	-0.7	-1.6	GAC00297	---NA---	-
Gp_maSB_07B12	0.5	-0.6	-0.9	-1.5	GAC00336	---NA---	-
Gp_maSA_26B03	0.6	-1.7	-0.4	-0.2	GAC00336	---NA---	-
Gp_mxAA_31E06	1.1	-1.6	-1	-3.6	GAC00338	phosphoenolpyruvate carboxykinase	1.00E-119
Gp_maSB_07E08	0.5	-1.7	-1.7	-1.8	GAC00338	phosphoenolpyruvate carboxykinase	1.00E-119
Gp_maSA_28H01	0.5	-0.9	-1.2	-1.6	GAC00338	phosphoenolpyruvate carboxykinase	1.00E-119
Gp_mxAA_45H04	0.9	-3.2	-1.2	-4.6	GAC00338	phosphoenolpyruvate carboxykinase	1.00E-119
Gp_maSA_35H09	0.7	-2.7	-1.4	-3.4	GAC00338	phosphoenolpyruvate carboxykinase	1.00E-119
Gp_mxAA_101A07	0.6	-22.1	-2.6	-6.4	GAC00338	phosphoenolpyruvate carboxykinase	1.00E-119
Gp_maSA_37A01	0.6	-1.2	-1.1	-2.3	GAC00338	phosphoenolpyruvate carboxykinase	1.00E-119
Gp_maSA_21D04	0.7	-3.3	-5.1	-2	GAC00338	phosphoenolpyruvate carboxykinase	1.00E-119
Gp_mxAA_19F05	1.7	-1.2	-0.8	-2	GAC00338	phosphoenolpyruvate carboxykinase	1.00E-119
Gp_maSA_22F10	0.7	-1.2	-1.8	-3	GAC00338	phosphoenolpyruvate carboxykinase	1.00E-119
Gp_mxAA_93E09	1.5	-1.8	-1.2	-3.1	GAC00338	phosphoenolpyruvate carboxykinase	1.00E-119
Gp_mxAA_10C11	0.8	-2.4	-1.8	-5.3	GAC00338	phosphoenolpyruvate carboxykinase	1.00E-119
Gp_mxAA_54G02	0.7	-3.5	-2.3	-4.1	GAC00338	phosphoenolpyruvate carboxykinase	1.00E-119
Gp_maSA_34H09	0.6	-1.6	-1.5	-2.1	GAC00371	---NA---	-
Gp_mxAA_60E04	1.1	-5.6	-2.6	-5.1	GAC00411	---NA---	-
Gp_mxAA_101F09	1.5	-7	0	-13.4	GAC00467	serine protease	1.00E-49
Gp_maSB_03D09	0.7	-0.3	-0.2	-12.9	GAC00530	immune-lectin-like receptor 3 isoform b	1.00E-04
Gp_mxAA_07H10	0.6	-3.8	-2.7	-9.4	GAC00530	immune-lectin-like receptor 3 isoform b	1.00E-04
Gp_maSA_25F06	0.3	-1.9	-0.4	-1.6	GAC00573	---NA---	-
Gp_mxAA_15G03	1.1	-1	-0.9	-3.2	GAC00574	---NA---	-
Gp_mxAA_22F10	-0.1	-1.2	-1.2	-2.1	GAC00574	---NA---	-
Gp_mxAA_69G01	1.1	-3.5	-1.3	-2.9	GAC00595	elongation factor 1 alpha	1.00E-112
Gp_mxAA_100H06	1.5	-2.6	-2	-4.7	GAC00611	secreted protein	1.00E+00
Gp_maSA_47E10	0.1	-1.2	-0.7	-0.6	GAC00621	---NA---	-
Gp_mxAA_72D07	0.2	-1.2	-0.9	-1.9	GAC00627	---NA---	-
Gp_mxAA_72D04	0	-1	-0.8	-1.6	GAC00627	---NA---	-
Gp_mxAA_53F08	0.2	-0.6	-2.5	-1.7	GAC00627	---NA---	-
Gp_mxAA_07H06	0	-0.9	-0.1	-1.4	GAC00634	lipopolysaccharide and beta--glucan binding protein	1.00E-88
Gp_mxAA_06C05	23.7	-3.1	-3.3	-4.7	GAC00681	phosphoenolpyruvate carboxykinase	1.00E-125
Gp_mxAA_39G03	1	-5.5	-1.9	-4.2	GAC00681	phosphoenolpyruvate carboxykinase	1.00E-125
Gp_maSA_31D07	1	-1.2	-2.2	-3.2	GAC00681	phosphoenolpyruvate carboxykinase	1.00E-125
Gp_mxAA_09G05	0.4	-2	-1.7	-4.7	GAC00681	phosphoenolpyruvate carboxykinase	1.00E-125
Gp_maSA_36E04	2.5	-1.8	-2.3	-0.9	GAC00681	phosphoenolpyruvate carboxykinase	1.00E-125
Gp_mxAA_32E05	2.3	-2.8	-1.1	-3.6	GAC00681	phosphoenolpyruvate carboxykinase	1.00E-125
Gp_mxAA_28E04	1.2	-3.8	-2	-6.1	GAC00681	phosphoenolpyruvate carboxykinase	1.00E-125
Gp_mxAA_61D09	0.9	-7.1	-1.3	-4.2	GAC00681	phosphoenolpyruvate carboxykinase	1.00E-125
Gp_mxAA_22D01	0.1	-4	-4.2	-3.1	GAC00681	phosphoenolpyruvate carboxykinase	1.00E-125
Gp_mxAA_06C08	0.3	-3.5	-5.2	-5.9	GAC00681	phosphoenolpyruvate carboxykinase	1.00E-125
Gp_mxAA_83D12	0.8	-2.1	-2.7	-4.7	GAC00681	phosphoenolpyruvate carboxykinase	1.00E-125
Gp_mxAA_75D03	1.6	-4.9	-2.8	-5.9	GAC00681	phosphoenolpyruvate carboxykinase	1.00E-125
Gp_mxAA_09D06	1.3	-2	-3	-4.7	GAC00681	phosphoenolpyruvate carboxykinase	1.00E-125
Gp_mxAB_05G04	1.3	-1.9	-1	-2.9	GAC00681	phosphoenolpyruvate carboxykinase	1.00E-125
Gp_mxAA_93B12	1.4	-2.2	-1.5	-3.4	GAC00681	phosphoenolpyruvate carboxykinase	1.00E-125
Gp_mxAA_09B09	0.7	-3.6	-2.8	-5.5	GAC00681	phosphoenolpyruvate carboxykinase	1.00E-125
Gp_mxAA_75C05	1	-3.1	-1.9	-4.1	GAC00681	phosphoenolpyruvate carboxykinase	1.00E-125
Gp_mxAA_80C09	1.5	-2.9	-2.4	-4.1	GAC00681	phosphoenolpyruvate carboxykinase	1.00E-125
Gp_mxAA_26C09	0.5	-4.3	-3.2	-5.1	GAC00681	phosphoenolpyruvate carboxykinase	1.00E-125
Gp_maSA_28A10	0.8	-2.2	-1.7	-5.6	GAC00681	phosphoenolpyruvate carboxykinase	1.00E-125
Gp_mxAA_05A12	0.7	-3.2	-2.1	-4.8	GAC00681	phosphoenolpyruvate carboxykinase	1.00E-125
Gp_maSB_07B05	0.9	-2	-2	-2	GAC00681	phosphoenolpyruvate carboxykinase	1.00E-125
Gp_mxAA_06G09	0.4	-3.4	-3.2	-7.4	GAC00681	phosphoenolpyruvate carboxykinase	1.00E-125
Gp_mxAA_22E04	1.1	-3.3	-2.8	-4.2	GAC00681	phosphoenolpyruvate carboxykinase	1.00E-125

Systematic	Fold Change				Cluster	Sequence description	max eValue
	Adult	JM	JS	Neo			
Gp_mxAA_84G11	1.2	-2.4	-3.1	-4.9	GAC00681	phosphoenolpyruvate carboxykinase	1.00E-125
Gp_maSC_02E03	0.6	-20.2	-0.7	-1.8	GAC00711	phosphoenolpyruvate carboxykinase	1.00E-23
Gp_maSA_41C10	1.5	-14.4	-0.2	-0.8	GAC00711	phosphoenolpyruvate carboxykinase	1.00E-23
Gp_maSA_29B01	1.3	-0.8	-15.1	-2.4	GAC00757	---NA---	-
Gp_mxAA_54H09	0.5	0.1	-0.2	-1.6	GAC00772	actin	1.00E-122
Gp_mxAA_33E05	-3.2	0.6	0.4	0.4	GAC00786	---NA---	-
Gp_maSA_30D01	1	-2.3	-0.8	-3.4	GAC00845	---NA---	-
Gp_mxAA_51A12	0	-0.8	-0.2	-2.9	GAC01065	---NA---	-
Gp_mxAA_95E09	0.2	-0.9	-0.3	-1.3	GAC01144	carboxypeptidase b	1.00E-80
Gp_mxAA_69H12	0.9	-4.1	-1.5	-3.6	GAC01166	hemocyanin subunit 1	1.00E-145
Gp_mxAA_99C07	0.5	-3.4	-2.1	-3.9	GAC01166	hemocyanin subunit 1	1.00E-145
Gp_maSA_36B09	0.8	-1.1	-0.4	-2.2	GAC01201	---NA---	-
Gp_maSA_47A08	0.7	-3.1	-2.2	-3.3	GAC01239	chitin binding protein	1.00E-24
Gp_maSA_38F03	4.6	-0.7	-1.2	-1.5	GAC01352	---NA---	-
Gp_maSA_39A02	3.6	-2.1	-1.1	-4.4	GAC01374	serine protease	1.00E-15
Gp_maSA_42B05	0.7	-0.9	-0.6	-6.5	GAC01546	---NA---	-
Gp_maSA_42C10	1.8	-2	-1	-2.7	GAC01561	protease m1 zinc metalloprotease	1.00E-10
Gp_maSA_42H06	3.2	-2.3	-1.1	0.3	GAC01595	---NA---	-
Gp_mxAA_70H04	1.1	-2.3	-1.4	-4.4	GAC01748	glutathione s-transferase	1.00E-52
Gp_maSA_50G08	1.1	-1.2	-1.1	-3.8	GAC02009	---NA---	-
Gp_maSA_52D10	2.8	-3.5	-0.9	-1.7	GAC02107	---NA---	-
Gp_mxAA_55F04	0.9	-2.5	-1.2	-5.9	GAC02329	heterogeneous nuclear ribonucleoprotein	1.00E-12
Gp_mxAA_10G11	0.5	-0.4	0.1	-2.7	GAC02376	ferritin	1.00E-69
Gp_mxAA_44D12	0	-0.6	0	-1.3	GAC02376	ferritin	1.00E-69
Gp_mxAA_62G07	0.3	-0.4	0.2	-1.5	GAC02376	ferritin	1.00E-69
Gp_mxAA_90C02	0.1	-0.3	-0.2	-1.2	GAC02376	ferritin	1.00E-69
Gp_mxAA_53H02	0.1	-0.9	-0.1	-1.1	GAC02376	ferritin	1.00E-69
Gp_mxAA_101C12	0	-0.4	-0.3	-1.2	GAC02376	ferritin	1.00E-69
Gp_mxAA_34E07	-0.3	-1.6	-1.3	-2.7	GAC02437	---NA---	-
Gp_mxAA_34E05	0.1	-0.5	-0.7	-1.8	GAC02437	---NA---	-
Gp_mxAA_51C02	-0.1	-0.9	-0.5	-1.1	GAC02437	---NA---	-
Gp_mxAA_42D05	1.2	-2.4	-2.1	-6.9	GAC02437	---NA---	-
Gp_mxAA_100B09	1	-0.6	-1.8	-2	GAC02437	---NA---	-
Gp_mxAA_11D11	-0.8	-13.6	0.2	0.9	GAC02437	---NA---	-
Gp_mxAA_25G08	0.8	-7.8	-1.1	-2.3	GAC02489	c-type lectin	1.00E-17
Gp_mxAA_74F06	1.4	-2	-0.5	-2.9	GAC02495	glycine rich protein	1.00E-06
Gp_mxAA_47G03	-6.9	1.2	0.7	0.6	GAC02526	---NA---	-
Gp_mxAA_42G06	1.4	0.5	0.4	0.2	GAC02526	---NA---	-
Gp_mxAA_17E12	0.3	-0.3	-0.3	-12.9	GAC02527	cg31973-isoform a	1.00E-44
Gp_mxAA_27F02	0.4	-0.1	-1.4	-1.1	GAC02527	cg31973-isoform a	1.00E-44
Gp_mxAA_30H03	0.6	-4.8	-3.1	-6.2	GAC02532	oxysterol binding	1.00E-39
Gp_mxAA_11E01	-2.6	1.4	0.9	0.6	GAC02542	pupal cuticle protein	1.00E-19
Gp_mxAA_84E03	-0.7	1.2	0.7	0.5	GAC02542	pupal cuticle protein	1.00E-19
Gp_mxAA_44C08	-3.8	0.6	0.4	0.3	GAC02547	cg30045-pa	1.00E-25
Gp_mxAA_13B02	0.8	-1.2	-1.9	-4.6	GAC02551	---NA---	-
Gp_mxAA_08G05	2.4	-0.6	-0.4	-1.7	GAC02561	pupal cuticle protein	1.00E-19
Gp_mxAA_33H10	1.3	-2.7	-1.2	-6.8	GAC02585	---NA---	-
Gp_mxAA_06E06	0.6	-0.3	0	-3.3	GAC02590	mgc80281 protein	1.00E-13
Gp_mxAA_79B09	0.7	-0.6	-1.4	-4.5	GAC02602	complement componentr subcomponent	1.00E-05
Gp_mxAA_07A05	0.7	-3.7	-2.7	-5.6	GAC02615	protein	1.00E-14
Gp_mxAA_57C01	0.8	-0.6	-0.3	-1.8	GAC02628	---NA---	-
Gp_mxAA_89H01	0.1	-1.2	-0.7	-0.8	GAC02628	---NA---	-
Gp_mxAB_05G03	1.3	-0.6	-0.3	-1.1	GAC02628	---NA---	-
Gp_mxAA_58F01	0.8	-3.1	-0.7	-1.9	GAC02628	---NA---	-
Gp_mxAA_62F10	1	-0.9	-0.5	-1.4	GAC02628	---NA---	-
Gp_mxAA_71H06	0.8	-1	-0.5	-1.9	GAC02628	---NA---	-
Gp_mxAA_45F08	0.8	-1	-0.5	-1.9	GAC02628	---NA---	-
Gp_mxAA_23E07	1.2	-0.7	-0.2	-13.5	GAC02628	---NA---	-
Gp_mxAA_71E01	0.3	-1.1	-0.7	-1	GAC02628	---NA---	-
Gp_mxAA_58B08	4.9	-1	-0.2	-1.4	GAC02628	---NA---	-
Gp_mxAA_46B11	0.5	0.2	-0.9	-2.3	GAC02628	---NA---	-
Gp_mxAA_30B12	0.7	-0.8	-0.5	-13.6	GAC02628	---NA---	-
Gp_mxAA_34E01	0.2	-0.8	0.1	-1.5	GAC02632	---NA---	-
Gp_mxAA_77F06	-8.9	0.6	0.2	0.7	GAC02686	---NA---	-
Gp_mxAA_83G07	2.9	-11.4	-1.1	-5.1	GAC02702	alpha-2-macroglobulin precursor splice variant 1 tam	1.00E-11
Gp_mxAA_42B11	1	-5.3	-2.2	-4.8	GAC02706	gasp precursor	1.00E-63
Gp_mxAA_82A10	1.2	-2.4	-2.1	-6.3	GAC02706	gasp precursor	1.00E-63
Gp_mxAA_46E10	90.3	-0.6	-0.8	-1.1	GAC02706	gasp precursor	1.00E-63
Gp_mxAA_11D02	-3.1	0	0.5	1.3	GAC02727	c-type lectin	1.00E-32

Systematic	Fold Change				Cluster	Sequence description	max eValue
	Adult	JM	JS	Neo			
Gp_mxAA_100B12	0.5	-4.1	-1.6	-5.6	GAC02764	---NA---	-
Gp_mxAA_44D10	1.5	-0.3	0.7	-0.6	GAC02786	translation elongation factor 2	1.00E-94
Gp_maSC_03B04	0.6	-1.9	-1.8	-4.8	GAC02855	actin	1.00E-31
Gp_maSC_03F08	0	-2.3	-1.2	-2.1	GAC02884	---NA---	-
Gp_maSC_05A02	0.6	98.2	-1.5	-3.9	GAC02933	---NA---	-
Gp_mxAA_14B08	0.1	-0.6	-0.4	-2.5	GAC02980	fatty acid binding protein 4	1.00E-14
Gp_mxAA_74F07	0.9	-0.7	-0.4	-3.1	GAC03042	---NA---	-
Gp_mxAA_82B05	-0.2	-1.2	103.6	-1.7	GAC03121	---NA---	-
Gp_mxAA_19C07	0	-1	-0.3	-1.6	GAC03395	zeelin1 cg6803-isoform d	1.00E-06
Gp_mxAA_20D04	0.9	-3.9	-3.4	-7.6	GAC03445	pupal cuticle protein	1.00E-18
Gp_mxAA_74G04	0.7	-0.3	-0.3	-1.1	GAC03452	chemosensory protein csp1	1.00E+00
Gp_mxAA_78E07	1.2	-0.6	0	-4.6	GAC03467	---NA---	-
Gp_mxAA_29G05	0.2	-3.6	-1.7	-6.1	GAC03788	---NA---	-
Gp_mxAA_31B04	-1.4	-3.2	-0.5	-3.4	GAC03834	---NA---	-
Gp_mxAA_31E10	0.5	-4	-1.1	-0.5	GAC03848	---NA---	-
Gp_mxAA_69A05	0.1	-4.6	-2.1	-5.4	GAC04065	---NA---	-
Gp_mxAA_47B09	1.7	-3.8	-2.2	-4	GAC04107	---NA---	-
Gp_mxAA_42F08	-3	1.1	0.5	0.8	GAC04220	calponin transgelin	1.00E-44
Gp_mxAA_42F09	0.8	0.5	-2.9	-5.9	GAC04221	---NA---	-
Gp_mxAA_74H12	0.9	-3.5	-0.9	-3	GAC04363	---NA---	-
Gp_mxAA_46E06	1.4	-1.8	-1.8	-3.7	GAC04370	peripheral myelin protein 2	1.00E-22
Gp_mxAA_47D05	1.3	-3.1	-1.8	-5.3	GAC04403	-	-
Gp_mxAA_51C03	1	-0.8	-0.7	-14	GAC04583	novel member of the keratin associated protein 4family	1.00E-04
Gp_mxAA_51F05	1.7	-2.6	-0.5	-4.7	GAC04600	---NA---	-
Gp_mxAA_52C06	1.6	-1.7	-1.9	-4.8	GAC04638	---NA---	-
Gp_mxAA_73H06	0.6	0.3	-14	-0.5	GAC04649	minichromosome maintenance complex component 6	1.00E-126
Gp_mxAA_60D10	1	-5.4	-1.9	-5.2	GAC05049	---NA---	1.00E+00
Gp_mxAA_99G04	1.9	-16.2	-1.5	-3.3	GAC05203	---NA---	-
Gp_mxAA_77G07	1	-4.5	-1.5	-5.2	GAC05404	muscle actin	1.00E-126
Gp_mxAA_70G02	-6	1.8	0.9	0.6	GAC05514	---NA---	-
Gp_mxAA_77B06	0.3	-4.6	-4.1	-5.7	GAC05802	---NA---	-
Gp_mxAA_77D06	1.4	-3.8	-1.6	-5.9	GAC05813	endonuclease reverse transcriptase	1.00E+00
Gp_mxAA_77E10	0.6	-3.3	-2.3	-5.2	GAC05820	ribonuclease e	1.00E-09
Gp_mxAA_78A07	1	-2.1	-2.3	-3.3	GAC05839	---NA---	-
Gp_mxAA_78H02	1	-3.5	-2.5	-4.3	GAC05885	apolipoprotein d	1.00E-13
Gp_mxAA_79E01	0.4	-1.2	0	-1.8	GAC05915	eukaryotic translation initiation factorsubunit m	1.00E-67
Gp_mxAA_83B03	0.6	-6.5	-1.4	-3.1	GAC06081	---NA---	-
Gp_mxAA_88C01	1	-1.4	-1.1	-2.6	GAC06342	cellulosomal scaffoldin anchoring	1.00E-06
Gp_mxAA_94E09	1.1	-16	-15.2	-2.2	GAC06611	-	-
Gp_mxAA_95G10	1.6	-2.5	-1.1	-2.8	GAC06673	legumain	1.00E-10
Gp_mxAB_06F11	0	-0.4	-0.4	-1.3	GAC07184	merozoite surface protein 1	1.00E-07
Gp_mxAB_07B07	0.4	-0.6	-0.2	-1.1	GAC07214	---NA---	-
Gp_mxAB_07F05	0.3	-2.6	-1.1	-3.2	GAC07284	cg31973-isoform a	1.00E-34
Gp_mxAB_07G04	-3.3	1.1	1.2	0.2	GAC07294	muscle actin	1.00E-130
Gp_mxAB_07G12	-0.7	1.1	0.4	0.4	GAC07302	---NA---	1.00E+00
Gp_mxAB_08B06	0.8	-2	-1.3	-2.8	GAC07329	actin	1.00E-126
Gp_mxAB_08B10	1.2	-10.3	-1.6	-7.2	GAC07333	---NA---	-
Gp_mxAB_08E01	1	-2.7	-1.9	-3.2	GAC07357	udp glycosyltransferase 2polypeptide a1	1.00E-31
Gp_mxAB_08H03	0.7	-1.6	-0.8	-1.7	GAC07391	rna polymerase ii large subunit	1.00E-04
Gp_mxAB_08H08	0.7	-0.8	-0.3	-13.6	GAC07396	---NA---	-
Gp_mxAB_09D11	-0.2	-0.7	0.1	-1.6	GAC07439	muscle-specific protein 300 cg33715-isoform d	1.00E-05
Gp_mxAB_09F11	0.6	-0.2	0.2	-2	GAC07461	ferritin	1.00E-70
Gp_mxAB_11A07	-0.2	-0.1	-0.6	-1.4	GAC07572	---NA---	-
Gp_mxAB_11F11	1.2	-1.1	-2	-2.5	GAC07633	---NA---	-
Gp_mxAA_08G03	0.4	0	-13.8	-0.5	N/A	N/A	N/A
Gp_maSA_45G07	1.1	-1.4	-1.3	-0.1	N/A	N/A	N/A
Gp_mxAA_84H09	0.8	-4.3	-1.9	-6.3	N/A	N/A	N/A
Gp_mxAA_22D05	1	-3.7	-4.6	-2.2	N/A	N/A	N/A
Gp_mxAA_23F01	2	0.2	-1.2	-2.2	N/A	N/A	N/A
Gp_mxAA_93A04	1.2	0.2	0.3	-0.6	N/A	N/A	N/A
Gp_mxAA_43E05	0.4	-6.1	-0.7	-3.1	N/A	N/A	N/A
Gp_maSB_06B02	0.3	-2.3	-1	-2.1	N/A	N/A	N/A
Gp_mxAA_27A07	1.4	-1.7	-0.9	-2.4	N/A	N/A	N/A
Gp_mxAA_91C02	0.4	-3.5	-2	-5.4	N/A	N/A	N/A
Gp_maSB_01G02	0.8	-1.3	-14.9	-2.9	N/A	N/A	N/A
Gp_mxAA_48D10	0.4	-0.7	0	-1.7	N/A	N/A	N/A
Gp_mxAA_30F11	0.8	-2.5	-1.6	-3.9	N/A	N/A	N/A
Gp_mxAA_89A10	0.5	-1	-0.4	-1.3	N/A	N/A	N/A

Table 4: Gene list of up and down-regulated objects due to intersex/normal phenotypes

Systematic	Fold Change					Sequence Description	max eValue
	IF	IM	NF	NM	Cluster		
Gp_mxAA_09H03	0.26	-0.20	-1.39	2.30	GAC00001	---NA---	-
Gp_maSA_34E05	0.32	0.01	0.18	-11.91	GAC00001	---NA---	-
Gp_maSA_52C04	0.23	0.34	0.60	-19.86	GAC00001	---NA---	-
Gp_mxAA_95A09	0.03	-0.18	2.43	0.78	GAC00001	---NA---	-
Gp_mxAA_82A04	0.44	0.47	-3.31	-20.52	GAC00001	---NA---	-
Gp_maSA_51F07	-0.09	0.13	0.34	-1.06	GAC00001	---NA---	-
Gp_mxAA_52F03	-1.06	-0.26	4.46	1.07	GAC00001	---NA---	-
Gp_mxAA_45B12	-0.71	-0.55	1.54	0.97	GAC00001	---NA---	-
Gp_maSA_30A07	-1.07	0.26	0.12	0.56	GAC00001	---NA---	-
Gp_mxAA_16B10	-0.33	0.07	3.94	3.27	GAC00001	---NA---	-
Gp_maSA_23H07	0.04	-16.26	0.39	-19.95	GAC00001	---NA---	-
Gp_mxAA_81F11	-0.12	-0.31	3.09	0.87	GAC00001	---NA---	-
Gp_maSA_28G09	-0.14	0.16	-0.30	-1.02	GAC00001	---NA---	-
Gp_maSA_51F10	-0.06	-0.23	1.00	1.01	GAC00001	---NA---	-
Gp_maSA_29E12	0.13	0.08	0.96	1.82	GAC00001	---NA---	-
Gp_mxAA_65F03	-0.48	0.16	0.86	1.57	GAC00001	---NA---	-
Gp_maSA_43F02	0.12	-0.10	0.05	-1.81	GAC00003	cytochrome c oxidase subunit iii	1.00E-91
Gp_mxAA_96E07	0.23	-17.60	0.03	0.93	GAC00008	myosin regulatory light chain 2 smooth muscle	1.00E-36
Gp_mxAA_46H08	-0.22	-0.22	3.85	0.74	GAC00008	myosin regulatory light chain 2 smooth muscle	1.00E-36
Gp_maSA_37E03	-0.06	-0.52	1.47	-0.91	GAC00017	actin	1.00E-136
Gp_maSA_32D05	-0.33	0.30	1.75	-0.12	GAC00035	hemocyanin subunit 1	1.00E-137
Gp_maSA_25C11	-0.03	-0.19	0.31	-20.32	GAC00035	hemocyanin subunit 1	1.00E-137
Gp_mxAA_25E09	-16.62	-1.32	3.27	0.58	GAC00041	---NA---	-
Gp_maSA_48F11	0.06	-0.19	0.13	-20.37	GAC00052	a disintegrin and metalloproteinase with thrombospondin motifs like	1.00E-11
Gp_mxAA_10B03	-0.33	0.21	0.36	1.06	GAC00052	a disintegrin and metalloproteinase with thrombospondin motifs like	1.00E-11
Gp_mxAA_52H11	-3.03	-1.29	3.13	0.69	GAC00057	hemocyanin subunit 1	1.00E-154
Gp_mxAA_101D03	-4.50	-1.00	6.27	1.05	GAC00057	hemocyanin subunit 1	1.00E-154
Gp_mxAA_44E08	-0.29	-0.20	3.55	0.42	GAC00057	hemocyanin subunit 1	1.00E-154
Gp_maSA_35F04	-1.06	-0.40	1.26	0.65	GAC00057	hemocyanin subunit 1	1.00E-154
Gp_mxAA_57G06	-1.10	-0.21	2.61	0.67	GAC00057	hemocyanin subunit 1	1.00E-154
Gp_mxAA_53E03	-0.98	-0.35	2.16	0.23	GAC00057	hemocyanin subunit 1	1.00E-154
Gp_mxAA_54F03	-1.03	-0.17	1.69	-0.07	GAC00057	hemocyanin subunit 1	1.00E-154
Gp_mxAA_36F07	-1.69	-0.95	1.25	0.36	GAC00057	hemocyanin subunit 1	1.00E-154
Gp_mxAA_58E04	-0.42	-0.48	3.23	0.64	GAC00057	hemocyanin subunit 1	1.00E-154
Gp_mxAA_62G06	-0.69	-1.09	3.97	1.33	GAC00057	hemocyanin subunit 1	1.00E-154
Gp_mxAA_53B07	-0.45	-0.25	1.34	0.74	GAC00057	hemocyanin subunit 1	1.00E-154
Gp_mxAA_62H08	-1.38	-0.65	5.40	1.87	GAC00057	hemocyanin subunit 1	1.00E-154
Gp_mxAA_37D02	-0.73	-0.61	1.29	2.49	GAC00057	hemocyanin subunit 1	1.00E-154
Gp_mxAA_48H12	-0.84	0.11	1.62	1.32	GAC00057	hemocyanin subunit 1	1.00E-154
Gp_mxAA_15C09	-1.13	-0.59	0.61	1.54	GAC00057	hemocyanin subunit 1	1.00E-154
Gp_mxAA_58D08	-0.32	6.97	1.96	0.10	GAC00057	hemocyanin subunit 1	1.00E-154
Gp_maSA_39E12	0.08	-0.31	0.03	-20.18	GAC00070	chitinase 3 precursor	1.00E+00
Gp_maSA_34D02	0.08	0.49	-49.88	-0.97	GAC00070	chitinase 3 precursor	1.00E+00
Gp_maSA_44C11	-0.11	0.15	133.98	-0.76	GAC00070	chitinase 3 precursor	1.00E+00
Gp_maSA_20C05	0.04	0.25	0.05	-1.01	GAC00070	chitinase 3 precursor	1.00E+00
Gp_maSA_34D03	-0.54	-0.02	2.31	1.45	GAC00070	chitinase 3 precursor	1.00E+00
Gp_maSA_32F04	-0.58	-0.38	2.85	0.35	GAC00077	hemocyanin subunit 1	1.00E-107
Gp_mxAA_10B08	-0.61	-0.38	1.46	0.67	GAC00077	hemocyanin subunit 1	1.00E-107
Gp_maSA_27E08	0.18	-0.26	-0.33	-1.94	GAC00078	---NA---	-
Gp_maSC_05F12	-0.31	0.08	1.07	3.41	GAC00105	actin	1.00E-131
Gp_maSA_02F06	-0.32	0.06	2.88	2.49	GAC00105	actin	1.00E-131
Gp_maSA_26C06	-0.46	-0.41	2.19	0.73	GAC00105	actin	1.00E-131
Gp_maSA_47H12	-0.57	0.45	0.55	1.15	GAC00110	---NA---	-
Gp_maSA_02G05	0.01	0.17	-0.91	-19.76	GAC00113	erpl protein	1.00E-06
Gp_mxAA_41D07	0.17	-16.48	-0.35	-0.76	GAC00114	---NA---	1.00E+00
Gp_mxAA_29E09	-0.76	-1.28	0.77	0.73	GAC00114	---NA---	1.00E+00
Gp_mxAA_96G10	-0.18	-0.56	3.19	0.42	GAC00140	hemocyanin subunit 1	1.00E-111
Gp_mxAA_45H08	-0.47	-0.28	4.72	1.72	GAC00141	---NA---	-
Gp_maSA_34A05	0.24	-0.05	-0.51	-1.91	GAC00146	---NA---	-
Gp_maSA_27F09	0.09	0.05	-0.91	-1.28	GAC00171	---NA---	-
Gp_mxAA_29A03	-1.40	-0.60	1.56	0.92	GAC00205	cytochrome c oxidase subunit i	1.00E-102
Gp_mxAA_46E04	0.24	-5.35	0.44	0.70	GAC00220	atp synthase f0 subunit 6	1.00E-55
Gp_maSA_42H03	0.24	0.60	-0.30	-20.04	GAC00222	---NA---	-
Gp_mxAA_61G04	-0.56	-0.21	0.40	1.63	GAC00228	---NA---	-
Gp_maSA_53H01	-0.79	0.63	1.66	1.52	GAC00228	---NA---	-
Gp_mxAA_16G09	-0.91	-0.74	2.04	0.98	GAC00228	---NA---	-
Gp_mxAA_96A01	0.86	-1.41	0.97	-0.59	GAC00232	trypsin	1.00E-89
Gp_mxAA_10H06	-0.78	-1.27	1.56	0.84	GAC00243	cytoskeletal actinb	1.00E-145
Gp_maSB_06G12	-0.58	-0.53	1.84	0.85	GAC00243	cytoskeletal actinb	1.00E-145
Gp_mxAA_94C11	-0.55	0.27	1.41	2.05	GAC00243	cytoskeletal actinb	1.00E-145
Gp_maSA_19F06	2.48	-0.03	-0.04	-1.01	GAC00243	cytoskeletal actinb	1.00E-145
Gp_maSA_19G05	-0.29	-0.28	2.56	1.15	GAC00254	reverse transcriptase	1.00E-13
Gp_maSA_20B01	0.04	-1.11	-0.15	-2.05	GAC00266	---NA---	-
Gp_maSB_03F12	0.05	0.37	-1.39	0.17	GAC00283	arginine kinase	1.00E-12

Systematic	Fold Change				Cluster	Sequence Description	max eValue
	IF	IM	NF	NM			
Gp_maSA_20G07	0.23	-0.01	0.14	-1.62	GAC00303	endonuclease reverse transcriptase	1.00E-07
Gp_maSA_20G08	0.41	-17.01	-0.20	-20.39	GAC00304	trypsin	1.00E-30
Gp_mxAA_101A07	-0.28	-1.56	1.88	0.43	GAC00338	phosphoenolpyruvate carboxykinase	1.00E-119
Gp_mxAA_81G12	-0.50	0.30	0.67	1.61	GAC00374	proline-rich extensin-like family protein	1.00E-07
Gp_maSA_38H08	-0.34	0.13	-0.46	42.36	GAC00376	---NA---	-
Gp_maSA_22D02	-0.03	0.07	-0.67	-2.13	GAC00383	---NA---	-
Gp_mxAA_58A02	-0.11	-0.64	1.07	0.49	GAC00388	alpha-amylase	1.00E-95
Gp_maSA_22D12	-1.96	0.01	0.31	-1.09	GAC00389	---NA---	-
Gp_maSA_22G03	0.35	-16.88	-6.72	-0.24	GAC00403	---NA---	-
Gp_mxAA_68C07	-1.73	-1.57	0.77	1.04	GAC00435	ribosomal protein s27a	1.00E-63
Gp_mxAA_39H08	-0.59	-0.10	2.39	1.51	GAC00436	glycosyl hydrolase family7	1.00E-93
Gp_mxAA_06C06	-0.49	-0.46	3.00	0.69	GAC00436	glycosyl hydrolase family7	1.00E-93
Gp_mxAA_39H11	-0.38	0.15	1.21	1.86	GAC00436	glycosyl hydrolase family7	1.00E-93
Gp_mxAA_11C03	-0.68	-0.40	1.84	0.48	GAC00436	glycosyl hydrolase family7	1.00E-93
Gp_mxAA_10C10	-0.86	-0.43	3.09	1.08	GAC00436	glycosyl hydrolase family7	1.00E-93
Gp_mxAA_43H07	-0.77	-0.41	2.96	0.18	GAC00436	glycosyl hydrolase family7	1.00E-93
Gp_mxAA_06B06	-0.99	-0.39	2.35	0.97	GAC00436	glycosyl hydrolase family7	1.00E-93
Gp_mxAA_85F08	-0.85	0.35	2.75	1.21	GAC00436	glycosyl hydrolase family7	1.00E-93
Gp_mxAA_17A12	-0.30	-0.19	4.34	1.60	GAC00436	glycosyl hydrolase family7	1.00E-93
Gp_mxAA_31G11	0.37	-16.62	8.18	1.00	GAC00436	glycosyl hydrolase family7	1.00E-93
Gp_mxAA_13G09	-0.29	-0.22	5.01	1.27	GAC00436	glycosyl hydrolase family7	1.00E-93
Gp_mxAA_14G05	-0.34	0.01	4.88	1.30	GAC00436	glycosyl hydrolase family7	1.00E-93
Gp_mxAA_77F02	0.38	0.60	-0.17	-20.83	GAC00436	glycosyl hydrolase family7	1.00E-93
Gp_mxAA_80D08	-16.34	-0.18	2.58	0.54	GAC00436	glycosyl hydrolase family7	1.00E-93
Gp_mxAA_11A03	-0.43	-0.14	2.04	0.37	GAC00436	glycosyl hydrolase family7	1.00E-93
Gp_mxAA_14A05	-0.72	-0.19	3.38	0.83	GAC00436	glycosyl hydrolase family7	1.00E-93
Gp_mxAA_40E03	-0.67	-0.40	1.85	1.65	GAC00436	glycosyl hydrolase family7	1.00E-93
Gp_mxAA_40F04	-0.98	-0.99	1.17	0.12	GAC00436	glycosyl hydrolase family7	1.00E-93
Gp_mxAA_19D04	-1.14	-0.51	1.12	0.60	GAC00436	glycosyl hydrolase family7	1.00E-93
Gp_mxAA_09H12	-0.88	-0.69	1.84	0.35	GAC00436	glycosyl hydrolase family7	1.00E-93
Gp_mxAA_06A04	-0.29	0.06	1.20	0.78	GAC00436	glycosyl hydrolase family7	1.00E-93
Gp_mxAB_06E02	-0.07	-0.49	2.45	1.19	GAC00436	glycosyl hydrolase family7	1.00E-93
Gp_mxAA_15E04	-0.33	0.04	2.33	0.37	GAC00436	glycosyl hydrolase family7	1.00E-93
Gp_mxAA_31H02	-0.75	-0.12	2.24	0.37	GAC00436	glycosyl hydrolase family7	1.00E-93
Gp_mxAA_36E02	-0.30	-0.30	2.05	1.11	GAC00436	glycosyl hydrolase family7	1.00E-93
Gp_mxAA_85A08	-1.25	-0.29	2.83	0.83	GAC00436	glycosyl hydrolase family7	1.00E-93
Gp_mxAA_40F02	-0.08	-0.31	3.14	4.74	GAC00436	glycosyl hydrolase family7	1.00E-93
Gp_mxAA_30B11	-0.50	-0.17	4.31	1.46	GAC00436	glycosyl hydrolase family7	1.00E-93
Gp_mxAA_22C05	-0.33	-0.10	1.04	0.36	GAC00436	glycosyl hydrolase family7	1.00E-93
Gp_mxAA_23C03	-0.24	-0.31	3.38	0.91	GAC00436	glycosyl hydrolase family7	1.00E-93
Gp_mxAA_43A04	-0.02	-0.24	3.69	1.36	GAC00436	glycosyl hydrolase family7	1.00E-93
Gp_mxAA_49G05	-0.97	-0.41	4.55	2.07	GAC00436	glycosyl hydrolase family7	1.00E-93
Gp_mxAA_05G04	-0.33	-0.17	1.00	1.06	GAC00436	glycosyl hydrolase family7	1.00E-93
Gp_mxAA_22D10	-0.42	-0.92	2.27	0.88	GAC00436	glycosyl hydrolase family7	1.00E-93
Gp_mxAA_09B06	-0.40	-0.68	1.66	1.57	GAC00436	glycosyl hydrolase family7	1.00E-93
Gp_mxAA_101D08	-0.48	-0.24	3.92	0.54	GAC00436	glycosyl hydrolase family7	1.00E-93
Gp_mxAA_54D06	-0.74	-0.38	6.77	1.27	GAC00436	glycosyl hydrolase family7	1.00E-93
Gp_mxAA_39C09	-0.40	-0.24	2.32	1.28	GAC00436	glycosyl hydrolase family7	1.00E-93
Gp_mxAA_84H02	-1.14	-0.49	2.12	0.75	GAC00436	glycosyl hydrolase family7	1.00E-93
Gp_mxAA_84F11	-0.27	-1.01	4.94	2.49	GAC00436	glycosyl hydrolase family7	1.00E-93
Gp_mxAA_52D07	-0.42	-0.14	2.27	0.98	GAC00436	glycosyl hydrolase family7	1.00E-93
Gp_mxAA_98A07	-0.20	-0.10	1.11	10.54	GAC00436	glycosyl hydrolase family7	1.00E-93
Gp_mxAA_10A12	-1.08	-0.45	0.83	0.90	GAC00436	glycosyl hydrolase family7	1.00E-93
Gp_mxAA_85D09	-0.25	-0.22	2.16	0.29	GAC00436	glycosyl hydrolase family7	1.00E-93
Gp_mxAA_94F02	-0.49	-0.23	2.22	0.32	GAC00436	glycosyl hydrolase family7	1.00E-93
Gp_mxAA_22H02	0.32	-4.13	0.25	-1.26	GAC00437	-	1.00E-170
Gp_maSA_35D08	-0.10	-1.84	4.69	0.02	GAC00466	---NA---	-
Gp_mxAA_83E02	-0.53	-0.14	2.27	0.55	GAC00467	serine protease	1.00E-49
Gp_maSA_25G06	-1.05	-0.39	1.47	0.70	GAC00579	---NA---	-
Gp_maSA_25G11	-16.31	-0.17	-24.70	0.71	GAC00582	deuterolysin metalloproteasefamily protein	1.00E-12
Gp_maSA_26C10	0.06	-16.63	0.02	-0.92	GAC00603	---NA---	-
Gp_mxAA_11A10	0.14	0.24	-1.29	-0.17	GAC00640	tropomyosin	1.00E-14
Gp_mxAA_82E03	-0.51	-0.22	2.25	0.38	GAC00640	tropomyosin	1.00E-14
Gp_mxAA_23A03	-0.22	-0.23	0.60	1.52	GAC00640	tropomyosin	1.00E-14
Gp_maSA_44H08	-0.94	0.38	0.83	1.32	GAC00640	tropomyosin	1.00E-14
Gp_mxAA_60A05	0.07	0.77	-24.33	-21.05	GAC00640	tropomyosin	1.00E-14
Gp_mxAA_37H07	-0.27	-0.26	1.90	0.12	GAC00640	tropomyosin	1.00E-14
Gp_maSA_27D03	-0.25	-0.67	0.73	-6.64	GAC00642	---NA---	-
Gp_maSA_27D08	0.18	-0.04	-0.28	-13.75	GAC00646	dead (asp-glu-ala-asp) box polypeptide 48	1.00E-29
Gp_maSA_27F03	-0.24	-32.96	2.33	-20.60	GAC00657	---NA---	-
Gp_maSA_27G01	-0.18	-0.04	0.10	-1.42	GAC00663	---NA---	-
Gp_mxAA_84G11	-0.42	-0.19	1.21	-0.66	GAC00681	phosphoenolpyruvate carboxykinase	1.00E-125
Gp_mxAA_76D05	-0.50	-0.47	2.65	0.53	GAC00681	phosphoenolpyruvate carboxykinase	1.00E-125
Gp_mxAA_100E05	0.07	-0.11	-1.01	0.54	GAC00681	phosphoenolpyruvate carboxykinase	1.00E-125

Systematic	Fold Change				Cluster	Sequence Description	max eValue
	IF	IM	NF	NM			
Gp_mxAA_18D04	-0.67	0.14	1.25	2.92	GAC00692	myosin heavynonmuscle or smooth muscle	1.00E-33
Gp_mxAA_80C10	-0.23	-1.45	2.95	1.48	GAC00692	myosin heavynonmuscle or smooth muscle	1.00E-33
Gp_mxAA_36B08	-0.51	0.29	1.08	1.26	GAC00692	myosin heavynonmuscle or smooth muscle	1.00E-33
Gp_mxAA_39E03	-0.92	-0.19	1.57	1.09	GAC00692	myosin heavynonmuscle or smooth muscle	1.00E-33
Gp_mxAA_10G07	-0.85	-0.33	0.84	1.41	GAC00692	myosin heavynonmuscle or smooth muscle	1.00E-33
Gp_mxAA_70B11	-0.27	-0.21	0.68	1.23	GAC00692	myosin heavynonmuscle or smooth muscle	1.00E-33
Gp_mxAA_18G01	-0.49	-0.25	0.70	1.39	GAC00692	myosin heavynonmuscle or smooth muscle	1.00E-33
Gp_mxAA_06E01	-0.98	0.15	0.33	1.03	GAC00692	myosin heavynonmuscle or smooth muscle	1.00E-33
Gp_mxAA_38F02	-0.70	-0.14	1.96	0.75	GAC00692	myosin heavynonmuscle or smooth muscle	1.00E-33
Gp_mxAA_72D02	-1.03	-0.47	0.24	0.90	GAC00692	myosin heavynonmuscle or smooth muscle	1.00E-33
Gp_mxAA_25C01	0.04	0.13	-25.45	-0.16	GAC00692	myosin heavynonmuscle or smooth muscle	1.00E-33
Gp_mxAA_05D01	-0.56	-0.29	1.59	0.46	GAC00692	myosin heavynonmuscle or smooth muscle	1.00E-33
Gp_mxAA_35E01	-0.83	-0.36	1.08	2.25	GAC00692	myosin heavynonmuscle or smooth muscle	1.00E-33
Gp_mxAA_30D08	-0.31	-0.07	1.87	1.93	GAC00692	myosin heavynonmuscle or smooth muscle	1.00E-33
Gp_mxAA_47E09	-0.16	0.08	2.57	4.55	GAC00692	myosin heavynonmuscle or smooth muscle	1.00E-33
Gp_mxAA_97H01	-0.78	0.16	1.13	1.34	GAC00692	myosin heavynonmuscle or smooth muscle	1.00E-33
Gp_mxAA_14D08	-0.71	-0.06	1.76	2.84	GAC00692	myosin heavynonmuscle or smooth muscle	1.00E-33
Gp_mxAA_18D02	-1.26	-0.14	1.08	1.69	GAC00692	myosin heavynonmuscle or smooth muscle	1.00E-33
Gp_mxAA_96C03	0.12	-0.04	0.36	-1.74	GAC00692	myosin heavynonmuscle or smooth muscle	1.00E-33
Gp_mxAA_71H09	-0.45	-0.14	1.82	2.44	GAC00692	myosin heavynonmuscle or smooth muscle	1.00E-33
Gp_mxAA_56H11	-0.72	-0.11	0.80	2.83	GAC00692	myosin heavynonmuscle or smooth muscle	1.00E-33
Gp_mxAA_35D03	-1.10	-0.16	0.76	1.48	GAC00692	myosin heavynonmuscle or smooth muscle	1.00E-33
Gp_mxAA_97G11	-0.71	-0.56	1.47	2.12	GAC00692	myosin heavynonmuscle or smooth muscle	1.00E-33
Gp_maSA_40D05	-0.46	0.46	0.65	1.36	GAC00748	myosin light chain	1.00E-38
Gp_mxAA_44G06	-0.99	-1.28	3.00	0.81	GAC00748	myosin light chain	1.00E-38
Gp_mxAA_66G08	-0.78	0.58	1.05	1.64	GAC00748	myosin light chain	1.00E-38
Gp_mxAA_13D09	-0.70	-0.57	1.05	0.91	GAC00772	actin	1.00E-122
Gp_mxAA_40H03	-0.37	-1.28	1.14	0.61	GAC00772	actin	1.00E-122
Gp_mxAA_57F12	-0.24	-0.40	1.59	1.61	GAC00772	actin	1.00E-122
Gp_mxAA_20E01	-0.44	-0.21	1.98	2.18	GAC00782	cytochrome c	1.00E-45
Gp_maSA_29D12	0.59	-0.11	-1.80	-0.54	GAC00782	cytochrome c	1.00E-45
Gp_mxAA_14A11	0.29	-0.04	-18.71	-0.04	GAC00782	cytochrome c	1.00E-45
Gp_mxAA_85D03	-0.01	0.00	-0.60	-19.99	GAC00782	cytochrome c	1.00E-45
Gp_mxAA_33H07	-2.80	-0.30	1.48	0.45	GAC00786	---NA---	-
Gp_mxAA_07G05	-0.81	-0.57	2.70	1.02	GAC00786	---NA---	-
Gp_mxAA_76E03	0.44	0.11	-1.19	0.06	GAC00786	---NA---	-
Gp_mxAA_84B07	-0.85	-0.61	1.50	0.41	GAC00838	hemocyanin subunit 1	1.00E-101
Gp_mxAA_53B01	-0.72	-0.30	1.14	0.54	GAC00838	hemocyanin subunit 1	1.00E-101
Gp_mxAA_23F03	-0.22	-0.02	1.61	1.65	GAC00903	muscle lim protein	1.00E-23
Gp_mxAA_18H12	-0.46	-0.66	1.57	0.45	GAC00903	muscle lim protein	1.00E-23
Gp_mxAA_77C05	-0.28	0.23	0.22	-20.33	GAC00903	muscle lim protein	1.00E-23
Gp_maSA_31F10	0.55	0.65	0.46	-20.75	GAC00929	---NA---	-
Gp_maSA_41E02	0.20	-16.79	0.04	-0.55	GAC00932	---NA---	-
Gp_maSA_42G08	0.11	-0.04	-2.16	0.07	GAC01035	---NA---	-
Gp_maSA_33H08	0.10	0.01	-1.16	-0.73	GAC01052	---NA---	-
Gp_mxAA_47F11	0.57	0.10	1.33	-39.70	GAC01053	---NA---	-
Gp_mxAA_24C10	0.55	0.08	-1.76	-0.56	GAC01065	---NA---	-
Gp_mxAA_77F11	-0.66	0.27	-1.27	0.45	GAC01136	arginine kinase	1.00E-63
Gp_mxAA_32F09	-0.28	-0.51	3.94	1.19	GAC01136	arginine kinase	1.00E-63
Gp_mxAB_04H06	-1.34	0.19	6.32	0.68	GAC01139	cytochrome b	1.00E-83
Gp_mxAA_70A03	-0.28	-0.04	0.24	-1.57	GAC01166	hemocyanin subunit 1	1.00E-145
Gp_mxAA_68A02	-0.53	-0.25	2.37	0.25	GAC01166	hemocyanin subunit 1	1.00E-145
Gp_mxAA_67B05	-0.33	-0.04	3.35	0.54	GAC01166	hemocyanin subunit 1	1.00E-145
Gp_mxAA_87H04	1.00	-0.17	0.27	-1.02	GAC01166	hemocyanin subunit 1	1.00E-145
Gp_mxAB_05H09	-0.77	-0.41	1.45	1.08	GAC01166	hemocyanin subunit 1	1.00E-145
Gp_mxAA_96F09	-0.02	-0.14	3.34	1.06	GAC01166	hemocyanin subunit 1	1.00E-145
Gp_mxAA_88H05	-0.12	0.33	4.06	0.94	GAC01166	hemocyanin subunit 1	1.00E-145
Gp_mxAA_88F04	-0.10	-0.63	1.66	0.61	GAC01166	hemocyanin subunit 1	1.00E-145
Gp_mxAA_93G10	-1.04	-0.05	3.93	1.14	GAC01166	hemocyanin subunit 1	1.00E-145

Systematic	Fold Change				Cluster	Sequence Description	max eValue
	IF	IM	NF	NM			
Gp_mxAA_66G03	-1.05	-0.49	3.79	0.89	GAC01166	hemocyanin subunit 1	1.00E-145
Gp_mxAA_84C02	-1.55	-0.80	2.96	1.05	GAC01166	hemocyanin subunit 1	1.00E-145
Gp_mxAA_96G11	-1.57	-1.37	2.16	0.83	GAC01166	hemocyanin subunit 1	1.00E-145
Gp_mxAA_67E08	-1.37	-1.14	2.10	0.85	GAC01166	hemocyanin subunit 1	1.00E-145
Gp_mxAA_66C07	-0.78	0.65	4.04	0.63	GAC01166	hemocyanin subunit 1	1.00E-145
Gp_mxAA_37H04	-0.87	-0.61	1.68	0.69	GAC01197	nadh dehydrogenase subunit 1	1.00E-58
Gp_maSA_36C01	0.18	-1.35	-0.01	-0.44	GAC01204	---NA---	-
Gp_mxAA_21G10	-0.03	-0.05	0.33	-20.30	GAC01235	hla-b associated transcript 1	1.00E-98
Gp_maSA_44F01	0.07	-1.36	0.63	-0.18	GAC01320	---NA---	-
Gp_mxAA_21G08	-0.31	-0.30	1.14	0.78	GAC01379	---NA---	-
Gp_maSA_39A09	-0.01	2.47	-25.03	-21.17	GAC01381	---NA---	-
Gp_maSA_39G09	-0.36	-16.82	0.74	0.24	GAC01423	map kinase-interacting serine threonine kinase 2	1.00E-44
Gp_mxAA_46A04	-0.36	0.06	7.93	1.36	GAC01443	zinc metalloproteinase	1.00E-49
Gp_mxAA_50A06	-0.08	-16.77	11.21	1.27	GAC01463	---NA---	-
Gp_maSB_06D11	0.18	-0.53	0.27	-1.82	GAC01463	---NA---	-
Gp_maSA_41F02	0.47	0.05	-8.64	-0.16	GAC01513	---NA---	-
Gp_maSA_43C08	-0.14	-0.04	1.75	-0.12	GAC01612	---NA---	-
Gp_maSA_49F12	0.00	0.14	-1.06	1.77	GAC01628	---NA---	-
Gp_maSA_51E12	-0.51	-3.28	1.27	-0.45	GAC01629	sarcoplasmic calcium-binding protein	-
Gp_maSA_43F11	0.06	0.02	-0.06	-6.24	GAC01637	brix domain containing 1	1.00E+00
Gp_mxAA_74C12	-0.31	-0.08	2.68	2.54	GAC01715	mgc79767 protein	1.00E-59
Gp_mxAA_25A06	-0.19	-1.26	5.52	0.51	GAC01715	mgc79767 protein	1.00E-59
Gp_mxAA_56F03	-0.09	-0.12	0.04	-2.19	GAC01715	mgc79767 protein	1.00E-59
Gp_mxAA_08H10	-0.45	-0.05	0.35	-2.16	GAC01715	mgc79767 protein	1.00E-59
Gp_mxAA_56B04	-0.08	-0.09	-0.30	-40.03	GAC01715	mgc79767 protein	1.00E-59
Gp_maSA_45F01	0.32	-0.08	-0.02	-1.32	GAC01751	---NA---	-
Gp_maSA_45G01	0.20	-0.02	-0.36	-4.18	GAC01760	---NA---	-
Gp_maSA_46G08	-0.19	0.16	-0.41	-1.00	GAC01814	---NA---	-
Gp_mxAA_98C05	-0.46	-0.10	2.01	0.36	GAC02009	---NA---	-
Gp_maSA_51G10	0.39	0.36	0.29	-2.88	GAC02074	---NA---	-
Gp_maSA_52C01	-1.88	0.06	0.33	-1.28	GAC02097	---NA---	-
Gp_mxAA_74E06	-0.59	0.75	0.54	1.26	GAC02168	myosin light chain 2	1.00E-16
Gp_mxAA_45E01	-0.98	-1.17	1.85	0.66	GAC02214	hemocyanin subunit 1	1.00E-120
Gp_mxAB_04G09	0.08	-0.08	3.31	0.60	GAC02255	hemocyanin subunit 1	1.00E-87
Gp_mxAA_37A03	-0.64	-1.30	0.12	-0.27	GAC02331	isoform cra_b	1.00E-09
Gp_mxAA_93F08	0.01	-0.03	0.39	-20.74	GAC02376	ferritin	1.00E-69
Gp_mxAA_47C02	0.06	0.20	-1.11	0.29	GAC02376	ferritin	1.00E-69
Gp_mxAA_13D11	-4.78	-16.78	-4.75	1.58	GAC02437	---NA---	-
Gp_mxAA_33B01	0.35	-16.57	-0.27	-0.32	GAC02437	---NA---	-
Gp_mxAA_52A04	-0.34	-0.38	2.11	0.10	GAC02437	---NA---	-
Gp_mxAA_25H06	-0.10	-1.26	0.28	-0.51	GAC02437	---NA---	-
Gp_mxAA_60C11	-0.45	0.53	2.81	0.52	GAC02437	---NA---	-
Gp_mxAA_69F09	-0.38	-16.58	1.09	0.69	GAC02437	---NA---	-
Gp_mxAA_41H04	-0.47	-0.35	1.67	1.15	GAC02437	---NA---	-
Gp_mxAA_95F09	-0.49	0.04	3.17	1.84	GAC02437	---NA---	-
Gp_mxAA_24A11	-0.23	-0.13	2.63	0.72	GAC02437	---NA---	-
Gp_mxAA_07A09	0.05	-1.75	-0.42	-0.13	GAC02437	---NA---	-
Gp_mxAA_20B03	-2.38	-0.40	2.77	0.96	GAC02437	---NA---	-
Gp_mxAA_34E05	0.25	-0.01	-24.88	-0.01	GAC02437	---NA---	-
Gp_mxAA_38C02	-0.12	0.28	1.43	-0.88	GAC02437	---NA---	-
Gp_mxAA_24G08	-0.34	-0.29	1.29	1.22	GAC02437	---NA---	-
Gp_mxAA_11D11	0.58	-0.26	-1.29	1.96	GAC02437	---NA---	-
Gp_mxAA_54G09	-1.12	-1.03	1.82	1.09	GAC02437	---NA---	-
Gp_mxAA_64G11	-1.10	-1.00	5.50	1.27	GAC02437	---NA---	-
Gp_mxAA_09G04	-0.57	0.09	1.72	1.25	GAC02437	---NA---	-
Gp_mxAA_84C01	0.12	-0.10	1.85	0.72	GAC02441	tyrosine 3-monooxygenase tryptophan 5-monooxygenase activationzeta polypeptide	1.00E-122
Gp_mxAA_38H06	0.29	0.15	-25.43	0.32	GAC02450	cd209-like protein	1.00E-04
Gp_mxAA_81B06	-0.70	-0.26	3.29	1.23	GAC02451	zinc finger protein	1.00E-10
Gp_mxAA_04C04	-0.42	-1.06	3.87	2.08	GAC02458	transglutaminase 1	1.00E-15
Gp_mxAA_09H04	-0.35	0.17	1.39	0.69	GAC02465	pupal cuticle protein	1.00E-18
Gp_mxAA_76A03	0.22	0.45	-2.97	-0.60	GAC02465	pupal cuticle protein	1.00E-18
Gp_mxAA_48F01	2.75	0.38	-26.45	-18.16	GAC02465	pupal cuticle protein	1.00E-18
Gp_mxAA_24D07	-0.18	-16.74	0.31	0.13	GAC02465	pupal cuticle protein	1.00E-18
Gp_mxAA_41G07	-0.60	-0.37	2.62	-0.04	GAC02470	pupal cuticle protein	1.00E-19
Gp_mxAA_42D07	0.11	-0.10	-0.13	-11.11	GAC02474	---NA---	-
Gp_mxAA_73H05	-0.34	-0.38	1.33	0.79	GAC02483	cg3355-isoform a	1.00E-43
Gp_mxAA_66H09	0.14	-1.37	0.54	1.39	GAC02484	cytochrome oxidase subunit i	1.00E-108
Gp_mxAA_47B04	-0.78	0.06	0.67	2.57	GAC02484	cytochrome oxidase subunit i	1.00E-108
Gp_mxAB_07B04	0.34	-0.49	-0.42	-1.35	GAC02484	cytochrome oxidase subunit i	1.00E-108
Gp_mxAA_79D09	0.03	-0.54	1.99	1.25	GAC02484	cytochrome oxidase subunit i	1.00E-108
Gp_mxAA_61C08	-16.65	-0.57	0.27	0.30	GAC02484	cytochrome oxidase subunit i	1.00E-108
Gp_mxAA_65H11	0.50	-33.82	0.95	-20.40	GAC02485	---NA---	-
Gp_mxAA_46D09	0.01	-16.91	1.27	-19.24	GAC02485	---NA---	-
Gp_mxAA_11C07	0.04	0.14	-1.05	-0.34	GAC02492	cytochrome oxidase subunit 3	1.00E-55
Gp_mxAA_32B09	0.07	-16.99	0.31	-15.27	GAC02492	cytochrome oxidase subunit 3	1.00E-55
Gp_mxAA_34G02	-0.94	-0.48	3.80	0.58	GAC02492	cytochrome oxidase subunit 3	1.00E-55
Gp_mxAA_36E12	0.40	0.09	0.83	-20.17	GAC02492	cytochrome oxidase subunit 3	1.00E-55
Gp_mxAA_76E05	-0.30	0.11	-1.05	-0.28	GAC02495	glycine rich protein	1.00E-06

Systematic	Fold Change				Cluster	Sequence Description	max eValue
	IF	IM	NF	NM			
Gp_mxAA_12D11	0.23	0.09	0.15	-1.54	GAC02496	---NA---	1.00E+00
Gp_mxAA_74C06	-0.18	-16.41	-2.29	0.69	GAC02497	cytochrome b	1.00E-88
Gp_mxAA_47H09	1.30	-0.46	1.60	-19.79	GAC02497	cytochrome b	1.00E-88
Gp_mxAA_25B05	0.20	0.08	0.68	-39.77	GAC02497	cytochrome b	1.00E-88
Gp_mxAA_74A01	0.12	-0.29	0.09	-14.22	GAC02497	cytochrome b	1.00E-88
Gp_mxAA_11H09	-0.79	-0.55	2.62	-0.06	GAC02497	cytochrome b	1.00E-88
Gp_mxAB_05H01	0.37	-0.16	1.26	-0.55	GAC02503	cytochrome b	1.00E-81
Gp_mxAA_05B06	0.20	-0.03	-0.20	-1.13	GAC02511	---NA---	-
Gp_mxAA_33A10	-1.67	-0.42	2.36	0.26	GAC02527	cg31973-isoform a	1.00E-44
Gp_mxAA_24C02	0.17	0.34	0.01	-20.02	GAC02527	cg31973-isoform a	1.00E-44
Gp_mxAA_05E01	-0.67	-0.14	2.24	0.86	GAC02530	beta 2c	1.00E-126
Gp_mxAA_12D12	-0.02	0.14	2.73	1.02	GAC02531	ribosomal protein l3	1.00E-103
Gp_mxAA_26C07	1.14	0.22	-27.61	0.33	GAC02542	pupal cuticle protein	1.00E-19
Gp_mxAA_34H05	-0.93	-0.33	1.23	0.79	GAC02542	pupal cuticle protein	1.00E-19
Gp_mxAA_59D08	-0.47	-0.19	3.32	0.04	GAC02547	cg30045-pa	1.00E-25
Gp_mxAA_50H05	0.25	-0.35	0.07	-19.98	GAC02547	cg30045-pa	1.00E-25
Gp_mxAA_30F05	-1.33	-0.27	1.41	0.85	GAC02547	cg30045-pa	1.00E-25
Gp_mxAA_09G03	0.12	0.17	-0.09	-1.41	GAC02547	cg30045-pa	1.00E-25
Gp_mxAA_08G06	-0.85	-0.15	1.53	1.49	GAC02547	cg30045-pa	1.00E-25
Gp_mxAA_20A06	0.10	-0.16	-1.37	-0.36	GAC02561	pupal cuticle protein	1.00E-19
Gp_mxAA_30C05	-0.81	-0.24	2.18	0.36	GAC02564	gastrokine 1 precursor	1.00E+00
Gp_mxAA_09H11	0.17	-0.23	-0.02	-1.15	GAC02572	succinate-adp-beta subunit	1.00E-91
Gp_mxAA_30G01	-0.63	-0.52	3.75	0.98	GAC02572	succinate-adp-beta subunit	1.00E-91
Gp_mxAA_08F02	0.15	0.03	0.18	-20.17	GAC02572	succinate-adp-beta subunit	1.00E-91
Gp_mxAB_04F10	0.24	-0.08	0.81	-1.46	GAC02579	---NA---	-
Gp_mxAA_84F02	0.17	-0.33	-1.15	-20.10	GAC02623	---NA---	-
Gp_mxAA_25F07	-0.28	-0.55	3.98	2.99	GAC02630	pupal cuticle protein	1.00E-16
Gp_mxAA_68B11	-0.18	-0.25	1.33	0.23	GAC02630	pupal cuticle protein	1.00E-16
Gp_mxAA_36C10	1.10	0.14	-25.25	0.67	GAC02630	pupal cuticle protein	1.00E-16
Gp_mxAA_30D03	-16.46	-0.24	-2.25	1.15	GAC02638	cuticle	1.00E-07
Gp_mxAA_21H04	-0.89	-0.10	1.42	0.82	GAC02644	ribosomal protein sa	1.00E-89
Gp_mxAA_43D04	-0.05	-0.40	4.04	0.69	GAC02657	loc780781 protein	1.00E-04
Gp_mxAA_59A12	-0.36	-0.14	3.37	0.25	GAC02662	---NA---	-
Gp_mxAA_74A06	-0.76	-0.49	1.19	0.29	GAC02681	c-type lectin	1.00E-74
Gp_mxAA_42D04	-0.55	-0.38	0.93	1.37	GAC02684	nadh dehydrogenase subunit 5	1.00E-27
Gp_mxAA_08D06	-0.32	-0.35	1.01	0.61	GAC02685	catalase	1.00E-107
Gp_mxAA_08E02	-0.22	0.23	-1.26	0.53	GAC02689	---NA---	-
Gp_mxAA_11B07	-0.74	-0.50	3.12	1.79	GAC02689	---NA---	-
Gp_mxAA_39F08	0.24	-0.23	0.28	-19.78	GAC02689	---NA---	-
Gp_mxAA_08E08	-1.06	-0.20	0.63	1.68	GAC02693	ribonuclease e	1.00E-08
Gp_mxAA_14E12	0.35	-0.60	-1.02	-0.09	GAC02702	alpha-2-macroglobulin precursor splice variant 1 tam	1.00E-11
Gp_mxAA_18G12	0.09	-0.10	-1.00	-20.59	GAC02702	alpha-2-macroglobulin precursor splice variant 1 tam	1.00E-11
Gp_mxAA_13D01	-0.12	0.19	0.86	1.43	GAC02702	alpha-2-macroglobulin precursor splice variant 1 tam	1.00E-11
Gp_mxAA_15H12	0.02	-0.76	1.17	1.46	GAC02705	---NA---	-
Gp_mxAA_08H11	0.24	-0.02	-0.20	-1.30	GAC02710	rna polymerase ii large subunit	1.00E-06
Gp_mxAA_97B01	-16.18	-0.21	1.06	-3.31	GAC02717	---NA---	-
Gp_mxAA_69B02	0.20	-0.15	-0.71	-20.96	GAC02729	troponin t-2	1.00E-11
Gp_mxAA_84F05	0.12	0.06	0.09	-2.58	GAC02730	mannose receptor	1.00E+00
Gp_mxAA_94F04	-15.58	0.26	-0.06	-21.68	GAC02730	mannose receptor	1.00E+00
Gp_mxAA_69E09	-0.38	-0.55	2.15	0.61	GAC02737	atp synthase f0 subunit 6	1.00E-51
Gp_mxAA_65B08	-0.20	0.34	0.47	-2.50	GAC02737	atp synthase f0 subunit 6	1.00E-51
Gp_mxAA_87A03	0.13	-0.08	0.12	-20.25	GAC02737	atp synthase f0 subunit 6	1.00E-51
Gp_mxAA_85C02	0.63	0.11	-24.90	-3.07	GAC02739	---NA---	-
Gp_maSC_04G01	0.19	-0.12	-0.49	-1.07	GAC02754	---NA---	-
Gp_maSC_04E12	0.03	-0.08	-1.60	-1.02	GAC02920	---NA---	-
Gp_maSC_05B06	0.09	0.39	-2.48	-0.75	GAC02943	---NA---	-
Gp_mxAA_59E03	-0.04	0.26	-1.29	-1.16	GAC02980	fatty acid binding protein 4	1.00E-14
Gp_mxAA_87D10	-0.09	-3.48	0.52	-0.13	GAC02987	---NA---	-
Gp_mxAA_100G08	0.33	40.00	0.00	-20.59	GAC02990	---NA---	-
Gp_mxAA_100H05	0.36	-0.13	-24.97	-0.55	GAC02996	---NA---	-
Gp_mxAA_35A03	0.69	-2.20	-0.35	-19.74	GAC03010	acid alpha glucosidase	1.00E-61
Gp_mxAA_45C12	-0.51	-0.09	0.07	1.45	GAC03014	non-smc element 1 homolog	1.00E-20
Gp_mxAA_62E10	0.14	-0.36	-1.16	-0.11	GAC03014	non-smc element 1 homolog	1.00E-20
Gp_mxAA_59G04	-1.59	-1.00	1.20	0.64	GAC03014	non-smc element 1 homolog	1.00E-20
Gp_mxAA_101E03	0.23	-0.22	-0.32	-21.36	GAC03025	von willebrand factor type a domain containing protein	1.00E-28
Gp_mxAA_25E10	-0.08	-0.38	1.36	0.04	GAC03028	glycine rich protein	1.00E-07
Gp_mxAA_86A02	-0.82	-0.80	1.30	0.65	GAC03028	glycine rich protein	1.00E-07
Gp_mxAA_68G07	-0.47	-0.57	2.48	1.69	GAC03039	---NA---	-
Gp_mxAA_51D07	0.17	-0.26	-0.26	-20.63	GAC03041	---NA---	-
Gp_mxAA_101G07	-0.09	-0.03	0.84	1.46	GAC03042	---NA---	-
Gp_mxAA_15C04	-0.10	0.00	-1.40	0.28	GAC03070	---NA---	-
Gp_mxAA_95F05	-0.52	0.06	4.27	0.79	GAC03082	---NA---	-
Gp_mxAA_26H11	-0.23	0.30	2.45	0.71	GAC03121	---NA---	-
Gp_mxAA_12B08	-0.39	-0.27	1.49	0.43	GAC03143	keratin associated protein 5-4	1.00E-11
Gp_mxAA_34B09	0.08	-0.32	-1.02	-0.25	GAC03144	cg4662-isoform b	1.00E-21
Gp_mxAA_31A07	0.14	-0.02	0.24	-20.41	GAC03186	---NA---	-
Gp_mxAA_29D03	-0.01	0.02	-1.36	0.20	GAC03201	rd rna-binding protein	1.00E-30
Gp_mxAA_15C01	0.24	-0.17	-24.80	-0.76	GAC03263	ribonuclease e	1.00E-09

Systematic	Fold Change					Sequence Description	max eValue
	IF	IM	NF	NM	Cluster		
Gp_mxAB_03C08	-0.32	-0.43	5.54	0.69	GAC03268	---NA---	-
Gp_mxAA_16A04	-0.38	0.46	2.02	2.02	GAC03296	---NA---	-
Gp_mxAA_16B12	-0.53	-1.05	8.92	1.45	GAC03306	pt repeat family protein	1.00E-05
Gp_mxAA_30H02	-0.17	-0.18	6.10	0.51	GAC03321	---NA---	-
Gp_mxAA_16H01	-1.31	-0.94	1.39	0.79	GAC03324	ae013749_4membrane protein	1.00E+00
Gp_mxAA_78H05	-0.33	-0.23	1.71	1.00	GAC03334	---NA---	-
Gp_mxAA_17B11	-1.53	-16.78	0.66	-40.04	GAC03340	---NA---	-
Gp_mxAB_05A07	0.43	0.24	-25.91	-0.62	GAC03367	large low complexity protein with proline alanine-rich repeat	1.00E-08
Gp_mxAA_18C07	0.12	-0.13	-25.06	-0.10	GAC03372	---NA---	-
Gp_mxAA_42C12	-1.45	-0.08	1.26	0.92	GAC03386	cuticle protein	1.00E-13
Gp_mxAA_60B04	-0.76	-0.94	0.86	1.73	GAC03397	dnak protein	1.00E-97
Gp_mxAA_19G03	0.62	-1.03	-0.18	-0.32	GAC03420	ankyrin repeat domain-containing protein 28	1.00E+00
Gp_mxAA_91E07	-1.18	-0.54	2.40	1.09	GAC03426	nadh dehydrogenase subunit 2	1.00E-46
Gp_mxAA_67F10	-0.81	0.52	1.35	1.19	GAC03426	nadh dehydrogenase subunit 2	1.00E-46
Gp_mxAA_20A08	-0.66	-0.18	1.86	0.26	GAC03432	---NA---	-
Gp_mxAA_20D04	-7.98	-0.05	0.42	0.90	GAC03445	pupal cuticle protein	1.00E-18
Gp_mxAA_20F02	-1.06	-0.60	2.01	0.94	GAC03460	---NA---	-
Gp_mxAA_20G07	-0.32	-0.20	1.15	0.05	GAC03467	---NA---	-
Gp_mxAA_21C03	-0.08	0.49	2.68	3.60	GAC03483	arthrodial cuticle protein	1.00E-27
Gp_mxAA_21G03	0.15	0.32	-24.96	-1.07	GAC03499	---NA---	-
Gp_mxAA_44H06	-0.36	-2.16	5.04	1.69	GAC03607	heat shock protein 70	1.00E-72
Gp_mxAA_48D04	-0.04	-16.63	0.66	0.05	GAC03611	90-kda heat shock protein	1.00E-100
Gp_mxAA_25H07	-0.19	0.19	0.87	1.18	GAC03646	mitochondrial ribosomal protein l18	1.00E-18
Gp_mxAA_84H04	-0.12	0.25	-25.85	-0.39	GAC03683	luminal binding protein	1.00E-93
Gp_mxAA_47E02	0.07	-0.26	-2.20	0.09	GAC03695	cellulose synthase operon c-like	1.00E-22
Gp_mxAA_30G12	0.09	-0.33	-0.08	-2.25	GAC03822	fcgbp protein	1.00E-08
Gp_mxAA_52B01	-0.53	-0.69	1.41	0.75	GAC03822	fcgbp protein	1.00E-08
Gp_mxAA_65F08	-1.51	0.02	1.62	1.00	GAC03848	---NA---	-
Gp_mxAA_32D07	-0.20	0.35	2.22	-39.33	GAC03876	ubiquitin family protein	1.00E-97
Gp_mxAA_32D10	-0.55	-0.33	1.31	1.69	GAC03877	mitochondrial ribosomal protein l38	1.00E-23
Gp_mxAA_34B03	-0.63	0.52	1.09	1.29	GAC03928	glycyl-trna synthetase	1.00E-17
Gp_mxAA_98C11	0.16	0.32	-0.11	-2.51	GAC03967	---NA---	-
Gp_mxAA_35F01	0.30	0.00	0.52	-39.71	GAC03973	atpase 6	1.00E-41
Gp_mxAA_97D02	0.30	-1.34	0.49	-0.60	GAC03988	istidine-rich glycoprotein precursor	1.00E+00
Gp_mxAA_36C01	-16.58	0.34	-0.81	-0.68	GAC03989	polycystic kidney disease 1-like 2	1.00E-27
Gp_mxAA_59F05	0.15	-0.21	-24.86	3.14	GAC04065	---NA---	-
Gp_mxAA_38B03	0.98	-0.46	1.77	-0.39	GAC04065	---NA---	-
Gp_mxAA_38E07	0.00	-0.22	0.55	-1.15	GAC04082	protein	1.00E-07
Gp_mxAA_71E02	-0.03	-0.15	0.09	-1.09	GAC04107	---NA---	-
Gp_mxAA_73H11	0.03	0.25	-25.03	-0.66	GAC04107	---NA---	-
Gp_mxAA_40A05	-0.71	-0.31	2.09	0.92	GAC04137	---NA---	-
Gp_mxAA_41A11	-1.15	-0.52	1.44	0.05	GAC04165	---NA---	-
Gp_mxAA_42H09	-0.45	0.29	0.19	1.39	GAC04230	---NA---	-
Gp_mxAA_43C10	-0.02	-16.78	1.37	0.60	GAC04250	keratin associated protein 5-4	1.00E-11
Gp_mxAA_44C10	0.07	0.43	-0.41	-22.19	GAC04281	---NA---	-
Gp_mxAA_96C05	-1.28	-0.74	1.80	1.26	GAC04310	ribosomal protein l23	1.00E-55
Gp_mxAA_51E07	-0.62	-0.68	0.96	1.64	GAC04332	methionine adenosyltransferasealpha	1.00E-47
Gp_mxAA_87D07	-1.46	0.00	3.91	1.16	GAC04336	mucin-like peritrophin	1.00E-08
Gp_mxAA_54G08	0.02	0.00	-3.38	-1.05	GAC04387	---NA---	-
Gp_mxAA_64B11	0.37	-0.73	1.01	0.34	GAC04390	---NA---	-
Gp_mxAA_47D03	-16.73	-0.27	0.59	-0.10	GAC04402	growth hormone inducible transmembrane protein	1.00E-11
Gp_mxAB_06H01	0.20	-1.55	-0.20	-0.19	GAC04446	---NA---	-
Gp_mxAA_49A10	-0.04	-1.15	1.99	1.75	GAC04474	cuticle proprotein	1.00E-53
Gp_mxAA_50D04	-0.63	-0.91	8.84	3.64	GAC04495	mgc137594 protein	1.00E-05
Gp_mxAA_57F02	0.12	-0.14	0.41	-13.02	GAC04538	keratin associated protein	1.00E+00
Gp_mxAA_50F05	-16.46	-0.32	0.95	-0.50	GAC04556	hydroxysteroid dehydrogenase like 1	1.00E-48
Gp_mxAA_73G10	-0.66	-0.73	3.33	1.85	GAC04561	-	1.00E-170
Gp_mxAA_59D04	-1.52	-0.39	3.83	0.89	GAC04608	cg3541-isoform b	1.00E-08
Gp_mxAA_51G07	0.25	-16.90	0.28	-0.09	GAC04608	cg3541-isoform b	1.00E-08
Gp_mxAA_52A06	-0.05	-0.21	-2.03	0.02	GAC04624	---NA---	-
Gp_mxAA_95F02	-1.06	-0.30	1.40	1.94	GAC04656	endonuclease-reverse transcriptase	1.00E-07
Gp_mxAA_87B07	0.40	0.06	-0.24	-1.37	GAC04684	endocuticle structural glycoproteinbd-1	1.00E-18
Gp_mxAA_55E09	-0.49	-0.68	2.94	1.22	GAC04823	polynucleotide kinase- 3-phosphatase	1.00E-05
Gp_mxAA_55H09	-0.50	-0.42	0.17	1.99	GAC04843	---NA---	-
Gp_mxAA_56D11	0.63	-0.34	0.35	-1.20	GAC04871	small subunit ribosomal protein 1	1.00E-55
Gp_mxAA_95B09	-0.07	-0.01	1.90	1.47	GAC04916	ahe structure oft ph	1.00E-24
Gp_mxAA_57H04	0.09	-0.55	-0.82	-20.33	GAC04935	---NA---	-
Gp_mxAA_57H06	0.03	-0.15	-1.20	-0.67	GAC04937	enoyl coenzyme ashortmitochondrial	1.00E-75
Gp_mxAA_58A01	0.06	0.39	-0.87	-20.22	GAC04940	protein	1.00E-18

Systematic	Fold Change					Sequence Description	max eValue
	IF	IM	NF	NM	Cluster		
Gp_mxAA_59G03	-0.90	-1.73	2.89	1.04	GAC05016	dna-damage-inducible transcript 4	1.00E-04
Gp_mxAA_61D04	0.31	-0.02	0.32	-1.21	GAC05090	fam108a1 protein	1.00E-57
Gp_mxAA_91H01	-0.17	-0.07	2.35	0.16	GAC05115	elongation factor 1-alpha	1.00E-113
Gp_mxAA_69C01	-1.98	-0.62	2.31	0.67	GAC05123	---NA---	-
Gp_mxAA_67A03	0.33	0.28	-25.01	-19.79	GAC05147	---NA---	-
Gp_mxAA_63B04	0.62	-0.59	-24.99	-0.29	GAC05173	---NA---	-
Gp_mxAA_64E09	1.03	-0.74	-25.70	-0.34	GAC05241	---NA---	-
Gp_mxAA_64G01	-1.90	-1.39	1.14	0.56	GAC05251	proteasomealpha type 1	1.00E-95
Gp_mxAA_64H09	0.69	-16.37	0.45	-1.24	GAC05263	---NA---	-
Gp_mxAA_65G02	-0.09	-0.27	1.12	0.17	GAC05270	---NA---	1.00E+00
Gp_mxAA_65B10	0.20	0.02	0.33	-1.35	GAC05279	---NA---	-
Gp_mxAA_91D04	-0.32	-0.76	3.21	1.84	GAC05383	---NA---	-
Gp_mxAA_67H02	0.78	-1.35	-24.50	-20.17	GAC05386	---NA---	-
Gp_mxAA_71G09	-1.40	-1.42	0.82	0.62	GAC05404	muscle actin	1.00E-126
Gp_mxAA_77G07	-0.65	-0.13	0.30	1.09	GAC05404	muscle actin	1.00E-126
Gp_mxAA_68C05	-0.26	0.19	-0.41	-1.24	GAC05405	---NA---	1.00E+00
Gp_mxAA_68H06	0.13	0.15	-1.98	-0.77	GAC05432	normocyte binding protein 2a	1.00E+00
Gp_mxAA_69A01	-1.17	-0.58	2.44	0.72	GAC05437	---NA---	1.00E+00
Gp_mxAA_73C02	0.06	0.07	-0.66	-1.71	GAC05450	---NA---	-
Gp_mxAA_70H12	0.35	0.13	-1.20	-1.47	GAC05522	---NA---	1.00E+00
Gp_mxAA_72F10	-0.25	0.14	0.47	1.60	GAC05601	gdp-mannose pyrophosphorylase a	1.00E-12
Gp_mxAA_75G04	0.37	-3.33	1.64	-19.39	GAC05739	clathrin heavy chain	1.00E-46
Gp_mxAA_76D02	-0.09	-0.21	3.27	0.76	GAC05766	ahe structure oft ph	1.00E-51
Gp_mxAA_76F01	-6.08	-0.06	2.71	2.56	GAC05779	---NA---	-
Gp_mxAA_77E07	-0.52	-0.24	3.45	0.56	GAC05818	beta-ureidopropionase isoform 1	1.00E-93
Gp_mxAA_78H03	0.90	0.28	0.60	-2.95	GAC05886	cortical granule protein with ldl-receptor-like repeats	1.00E-05
Gp_mxAA_79B06	-16.81	-16.55	4.59	1.10	GAC05898	lyr motif containing 4	1.00E-14
Gp_mxAA_79C07	0.29	0.04	-1.05	0.60	GAC05905	---NA---	-
Gp_mxAA_80H08	-0.01	0.26	-1.01	-1.92	GAC05971	vacuolar atp synthase subunit g	1.00E-20
Gp_mxAA_83C01	-0.57	0.09	-0.28	-19.67	GAC06088	breast cancer metastasis-suppressor 1-like	1.00E-06
Gp_mxAA_85E04	0.37	-0.29	-1.35	-0.20	GAC06202	---NA---	-
Gp_mxAA_86D06	-0.46	-0.27	4.43	0.19	GAC06246	superoxide dismutase	1.00E+00
Gp_mxAA_88E02	0.93	-2.98	2.39	-0.82	GAC06356	---NA---	-
Gp_mxAA_88F03	0.63	-16.46	0.97	-2.07	GAC06360	protein	1.00E-09
Gp_mxAA_89E04	0.17	-0.03	-0.56	-20.04	GAC06402	---NA---	-
Gp_mxAA_93C09	0.30	-0.40	-18.34	-0.59	GAC06558	ahe structure oft ph	1.00E-62
Gp_mxAA_99A05	0.00	-0.02	-0.16	-1.34	GAC06563	---NA---	-
Gp_mxAA_94H11	-0.74	0.17	0.89	1.37	GAC06627	---NA---	-
Gp_mxAA_96H11	0.65	-0.25	0.28	-20.86	GAC06738	---NA---	-
Gp_mxAA_97H04	0.31	-0.02	0.36	-20.37	GAC06784	---NA---	-
Gp_mxAA_98B07	0.34	0.36	-0.30	-20.19	GAC06800	alkaline phosphatase	1.00E-66
Gp_mxAB_05D03	-0.13	0.46	1.39	2.02	GAC07103	---NA---	-
Gp_mxAA_94B11	-0.73	-0.16	2.27	0.87	GAC07232	---NA---	-
Gp_mxAB_08B01	0.42	-0.37	0.32	-1.12	GAC07324	---NA---	-
Gp_mxAB_08B04	-1.65	-0.20	0.53	1.49	GAC07327	---NA---	1.00E+00
Gp_mxAB_08C10	0.07	0.09	-0.19	-20.21	GAC07344	loc556764 protein	1.00E-98
Gp_mxAB_08D02	0.28	-0.25	0.19	-1.04	GAC07347	tryptophan-rich antigen (pv-fam-a)	1.00E+00
Gp_mxAB_08E03	0.08	-0.42	0.55	-20.67	GAC07359	cytochrome oxidase subunit i	1.00E-68
Gp_mxAB_08E12	0.15	-0.09	1.21	-0.69	GAC07366	cg3419-isoform a	1.00E-06
Gp_mxAB_08G02	-1.46	-0.30	0.28	0.47	GAC07378	---NA---	-
Gp_mxAB_09B10	0.29	-0.45	0.44	-19.76	GAC07419	cleavage stimulation3 pre-ma subunittau	1.00E-10
Gp_mxAB_09C02	0.59	-0.28	0.31	-20.18	GAC07422	cell division protein	1.00E+00
Gp_mxAB_09C10	-1.45	0.02	1.08	0.58	GAC07429	chromosome 10 open reading frame 59	1.00E-13
Gp_mxAB_09D09	-0.25	1.24	1.72	1.60	GAC07437	---NA---	-
Gp_mxAB_09D12	0.09	-0.41	0.06	-2.35	GAC07440	loc407663 protein	1.00E-08
Gp_mxAB_09E01	-0.41	-0.52	2.27	1.06	GAC07441	hemocyanin subunit 1	1.00E-115
Gp_mxAB_09G07	-16.23	0.30	0.74	-0.29	GAC07468	---NA---	1.00E-04
Gp_mxAB_09H10	0.04	0.34	0.06	-1.70	GAC07481	cg3419-isoform a	1.00E-07
Gp_mxAB_10B08	-0.47	-0.04	3.85	0.38	GAC07499	glycosyl hydrolase family7	1.00E-70
Gp_mxAB_10D04	-1.11	-0.64	1.42	1.47	GAC07516	hemocyanin subunit 1	1.00E-139
Gp_mxAB_10E02	-1.40	-1.45	1.33	1.23	GAC07524	muscle actin	1.00E-127
Gp_mxAB_10F10	0.10	-0.58	-1.05	0.98	GAC07541	nadh dehydrogenase subunit 4	1.00E-59
Gp_mxAB_11B10	-0.85	-0.42	1.14	0.71	GAC07586	---NA---	-
Gp_mxAB_11C09	-0.42	-0.08	1.83	0.44	GAC07596	pupal cuticle protein	1.00E-19
Gp_mxAB_11E07	-0.34	0.10	2.65	0.69	GAC07617	glucocerebrosidase precursor	1.00E-37
Gp_mxAB_11H11	-0.28	0.01	3.18	-0.04	GAC07655	---NA---	-
Gp_mxAA_17B01	0.31	-1.20	-0.12	-0.76	N/A	N/A	N/A
Gp_maSA_32D02	0.13	0.95	-1.12	-0.82	N/A	N/A	N/A
Gp_mxAA_62G05	-0.62	-0.52	5.46	2.01	N/A	N/A	N/A
Gp_mxAA_89A03	-1.62	-1.90	0.23	0.67	N/A	N/A	N/A
Gp_mxAA_69C07	0.36	-0.14	3.01	0.28	N/A	N/A	N/A
Gp_mxAA_78B01	-0.42	0.25	0.87	2.22	N/A	N/A	N/A
Gp_mxAA_78G02	0.08	0.21	-1.24	-0.16	N/A	N/A	N/A
Gp_mxAB_01C02	0.46	0.25	1.03	-0.84	N/A	N/A	N/A
Gp_mxAA_92B02	0.28	0.18	-2.33	-0.06	N/A	N/A	N/A
Gp_maSB_05B06	0.10	0.14	-0.28	-1.30	N/A	N/A	N/A
Gp_mxAA_06E04	-1.57	-1.64	0.39	1.06	N/A	N/A	N/A

Systematic	Fold Change				Cluster	Sequence Description	max eValue
	IF	IM	NF	NM			
Gp_mxAA_91G07	0.11	-0.37	0.32	-20.01	N/A	N/A	N/A
Gp_mxAA_62F03	-0.06	0.30	3.80	0.80	N/A	N/A	N/A
Gp_mxAA_90A05	-1.49	-1.78	0.87	0.95	N/A	N/A	N/A
Gp_mxAA_80C05	-16.53	0.12	0.10	-0.92	N/A	N/A	N/A
Gp_mxAA_96D10	0.25	-1.41	-0.17	-1.41	N/A	N/A	N/A
Gp_mxAA_77C02	0.06	-0.07	3.45	0.42	N/A	N/A	N/A
Gp_mxAA_81E06	-0.82	-0.22	1.27	1.35	N/A	N/A	N/A
Gp_mxAB_04E01	-0.79	1.26	-18.24	-1.13	N/A	N/A	N/A
Gp_mxAA_24H03	-0.16	0.20	1.25	2.25	N/A	N/A	N/A
Gp_maSA_47D09	0.24	-0.03	0.05	-1.62	N/A	N/A	N/A
Gp_mxAA_65G10	-0.33	-0.14	0.20	1.13	N/A	N/A	N/A
Gp_mxAA_79E03	-1.25	-1.29	1.04	1.41	N/A	N/A	N/A
Gp_mxAA_90F05	-0.18	-0.40	2.34	-3.09	N/A	N/A	N/A
Gp_mxAA_27A02	-0.55	0.05	1.15	2.01	N/A	N/A	N/A
Gp_mxAB_03G02	0.60	-0.07	10.24	-20.22	N/A	N/A	N/A
Gp_mxAA_20C01	-0.41	-0.40	1.14	2.91	N/A	N/A	N/A
Gp_mxAA_69F07	0.01	-12.93	0.75	-1.34	N/A	N/A	N/A
Gp_mxAA_80F02	-0.68	-0.59	1.70	0.24	N/A	N/A	N/A
Gp_mxAA_20H12	-0.58	-0.57	1.15	2.05	N/A	N/A	N/A
Gp_mxAA_91F08	-0.89	-0.48	1.40	0.47	N/A	N/A	N/A
Gp_mxAA_23C02	-0.56	-0.20	1.20	-0.18	N/A	N/A	N/A
Gp_mxAA_24F03	0.00	-0.06	1.43	2.93	N/A	N/A	N/A
Gp_maSB_06C12	0.54	-0.23	-0.04	-20.20	N/A	N/A	N/A
Gp_mxAA_40H02	-0.15	0.74	-24.97	-0.77	N/A	N/A	N/A
Gp_maSA_46A09	0.10	-1.05	5.51	1.27	N/A	N/A	N/A
Gp_mxAA_44A03	0.40	-2.24	0.63	-0.40	N/A	N/A	N/A
Gp_mxAA_62G03	-0.03	-0.03	0.80	-1.61	N/A	N/A	N/A
Gp_mxAA_84C10	-0.43	0.04	3.52	2.03	N/A	N/A	N/A
Gp_mxAB_03B04	0.16	-16.53	0.91	-0.77	N/A	N/A	N/A
Gp_mxAA_98E11	0.15	-0.10	-1.33	-0.20	N/A	N/A	N/A
Gp_mxAB_10E05	-0.03	0.43	-0.57	-7.85	N/A	N/A	N/A
Gp_mxAA_89F05	-0.92	-0.32	2.08	0.67	N/A	N/A	N/A
Gp_mxAA_43B04	0.49	-1.12	-1.18	0.06	N/A	N/A	N/A
Gp_mxAA_65B11	-0.10	-17.00	4.09	1.35	N/A	N/A	N/A
Gp_mxAB_03E08	-0.16	-0.22	1.03	0.79	N/A	N/A	N/A
Gp_maSA_43B04	0.13	0.39	-0.37	-20.15	N/A	N/A	N/A
Gp_mxAA_40G03	-1.50	-0.79	1.73	0.62	N/A	N/A	N/A
Gp_mxAA_21D04	-0.54	0.24	1.32	0.96	N/A	N/A	N/A
Gp_mxAA_18E02	0.11	-0.27	1.98	0.46	N/A	N/A	N/A
Gp_mxAA_47B07	0.04	-0.93	-1.29	0.01	N/A	N/A	N/A
Gp_mxAA_91H03	-0.49	-0.74	3.56	1.12	N/A	N/A	N/A
Gp_mxAA_77H04	0.91	-16.84	0.94	-1.75	N/A	N/A	N/A
Gp_mxAA_28A07	-0.13	-0.52	3.25	1.05	N/A	N/A	N/A
Gp_mxAA_35G01	-0.15	-0.53	6.68	1.45	N/A	N/A	N/A
Gp_mxAA_93B11	-0.47	-0.88	2.28	0.51	N/A	N/A	N/A
Gp_mxAA_59A03	-0.06	-0.62	3.57	1.79	N/A	N/A	N/A