Gene duplications, divergence and recombination shape adaptive evolution of the fish ectoparasite, *Gyrodactylus bullatarudis*

Mateusz Konczal[1*], Karolina J. Przesmycka[1], Ryan S. Mohammed[2], Karl P. Phillips[3,4], Francisco Camara[5,6], Sebastian Chmielewski[1], Christoph Hahn[7], Roderic Guigo[5,6], Jo Cable[8], Jacek Radwan[1]

[1] Evolutionary Biology Group, Faculty of Biology, Adam Mickiewicz University, 61-614 Poznań, Poland

[2] The University of the West Indies Zoology Museum, Department of Life Sciences, Faculty of Science and Technology, UWI, St. Augustine, Trinidad and Tobago, WI.

[3] School of Biological, Earth & Environmental Sciences, University College Cork, Cork, Ireland

[4] Marine Institute, Furnace, Newport, Co. Mayo, Ireland

[5] Centre for Genomic Regulation (CRG), Dr. Aiguader 88, 08003 Barcelona, Spain.

[6] Universitat Pompeu Fabra (UPF), Barcelona, Spain

[7] Institute of Biology, University of Graz, Austria

[8] School of Biosciences, Cardiff University, Cardiff, CF10 3AX, UK

* Corresponding author

E-mail: mateusz.konczal@amu.edu.pl (MK)

Abstract

Determining the molecular basis of parasite adaptation to its host is an important component in understanding host-parasite coevolution and the epidemiology of parasitic infections. Here, we investigate short- and long-term adaptive evolution in the eukaryotic parasite, *Gyrodactylus bullatarudis*, infecting Caribbean guppies (*Poecilia reticulata*), by comparing the reference genome of Tobagonian *G. bullatarudis* with other Platyhelminthes, and by analyzing resequenced samples from local Trinidadian populations. At the macroevolutionary timescale, we observed duplication of G-protein and serine proteases genes, which are likely important in host-parasite arms races. Serine protease also showed strong evidence of ongoing, diversifying selection at the microevolutionary timescale. Furthermore, our analyses revealed that a hybridization event, involving two divergent genomes, followed by recombination has dramatically affected the genetic composition of Trinidadian populations. The recombinant genotypes invaded Trinidad and replaced local parasites in all populations. We localized more than 300 genes in regions fixed in local populations for variants of different origin, possibly due to diversifying selection pressure from local host populations. In addition, around 70 genes were localized in regions identified as heterozygous in some, but not all, individuals. This pattern is consistent with a very recent spread of recombinant parasites. Overall, our results are consistent with the notion that recombination between divergent genomes can result in particularly successful parasites.

Keywords

Recombination, hybrids, parasite, adaptive evolution, reference genome, admixture

Introduction

Parasites are ubiquitous in wild animals, and can profoundly alter the physiology, behaviour and reproductive success of hosts. Parasites represent around 40% of described species (Dobson, Lafferty, Kuris, Hechinger, & Jetz, 2008) and virtually every metazoan host supports at least one parasite species (Poulin & Morand, 2004). Consequently, parasites play key roles in ecosystem functioning (Frainer, McKie, Amundsen, Knudsen, & Lafferty, 2018; Hatcher, Dick, & Dunn, 2012), and understanding parasite evolution is therefore a major component of eco-evolutionary research.

In contrast to viruses and bacteria, relatively little is known about molecular evolutionary dynamics in eukaryotic parasites (Criscione, Poulin, & Blouin, 2005; Wit & Gilleard, 2017), despite parasitism being one of the most common lifestyles amongst eukaryotes (Windsor, 2002). Although genomic approaches are increasingly applied to parasitological research, such studies have been predominantly clinical, focusing on the evolution of drug resistance among human and livestock parasites (Choi et al., 2016; Coghlan et al., 2019; Cole & Viney, 2018; Hupalo et al., 2016; Small et al., 2016). Such studies provide little information about parasite evolution in natural populations, where anthropogenic influence is limited and where avoiding host immune defenses is the main selective pressure. This pressure has implications well beyond simple adaptation, affecting the evolution of sex (Hamilton, 2006; Morran, Schmidt, Gelarden, Parrish, & Lively, 2011), sexual selection (Hamilton & Zuk, 1982), speciation (Venditti, Meade, & Pagel, 2010) and the maintenance of genetic variation in populations (Woolhouse, Webster, Domingo, Charlesworth, & Levin, 2002).

Genomics has recently emerged as a key tool for systematic investigation of parasite evolution. Firstly, it allows us to study the evolutionary and demographic history, revealing pathways of infection. Secondly, it provides resources for studying the genetic structure of parasite populations, understanding of which is crucial for assessing local host adaptation or testing hypotheses on the role of parasites in ecological speciation (El Nagar & Maccoll, 2016). Thirdly, genomic scans can identify loci under selection. The molecular basis of host-pathogen coevolution is of particular interest to evolutionary biologists testing Red Queen scenarios of host-parasite coevolution (Papkou et al., 2019; Woolhouse et al., 2002). Fourthly, genomics can be used to detect phenotypically cryptic species, diverged lineages, or hybrids in sexually reproducing parasites. Hybridization, in particular, is increasingly being recognized as an important source of raw material for natural selection in parasite evolution (King, Stelkens, Webster, Smith, & Brockhurst, 2015; Maxwell, Sepulveda, Turissini, Goldman, & Matute, 2018), and there is evidence that hybrid parasites may be able to infect a wider range of host species (Volf et al., 2007) or exhibit increased virulence (Farrer et al., 2011).

Here, we use genomic analyses to investigate short and long-term adaptive evolution in the monogenean parasite, *Gyrodactylus bullatarudis.* Monogeneans are economically important fish pathogens and provide ideal model systems for studying host-pathogen coevolution. They have a direct life cycle, simplifying theoretical predictions about coevolution and virulence evolution. In addition, some species can reproduce asexually by mitotic division, by automictic parthenogenesis, and also by sexual reproduction, allowing us to explore the role of different reproductive strategies in coevolution and the potential role of hybridization in shaping patterns of genetic variation (Cable & Harris, 2002; Schelkle, Faria, Johnson, van Oosterhout, & Cable, 2012). Except for *Gyrodactylus salaris* (see Hahn et al. 2014), a significant pathogen of Atlantic salmon, relatively little is known about the molecular basis of macro- and microevolution of monogeneans.

We thus explore the genome of *G. bullatarudis*, parasite of the guppy (*Poecilia reticulata*) – model species in eco-evolutionary research – with particular emphasis on identifying candidate genes involved in host-pathogen coevolution. Interactions between gyrodactylids and guppies have been a subject of research for more than 50 years. A large part of this investigation stems from the role of guppies as an important model species in evolutionary ecology (Magurran, 2005). The impact of gyrodactylids on guppy behaviour (e.g. van Oosterhout et al. 2003; Jacquin et al. 2016; Reynolds et al. 2018), fitness (Houde & Torio, 1992), phenotype and population/community dynamics (Pérez-Jvostov, Hendry, Fussmann, & Scott, 2016; Stephenson, van Oosterhout, & Cable, 2015), as well as the biology of the parasite itself (reviewed in Bakke et al. 2007), have been well documented. Mitochondrial markers suggest considerable population structure in wild *G. bullatarudis*, and the presence of cryptic species in the natural range of these parasites (Xavier et al., 2015), and there is also evidence of adaptation of gyrodactylids to local guppy immunity genes (MHC; Phillips et al. 2018). However, nothing is known about geographic structuring of gyrodactylid genes likely important in their host adaptation. We thus investigated *G. bullatarudis* genomic variation within, and differentiation among, natural populations in Trinidad and Tobago. Through whole-genome sequencing, we aimed to (i) identify genes that have been evolving adaptively since divergence from other taxa, thus likely contributing to co-evolution with the guppy host; (ii) describe genes differentiated between local populations of Trinidad, thus potentially involved in adaptation to local host populations; and (iii) validate previous findings about cryptic species suggested for *G. bullatarudis* and their potential impact on coevolution.

Materials and Methods

Data collection

The origin of gyrodactylid individuals for reference genome assembly was Roxborough River on Tobago. The culture was set up from a single worm collected in 2016, and 2000-3000 individuals were obtained by infecting parasite-naïve guppies from the mesocosm populations established by Phillips et al. (2018) at our field station in Charlotteville, Tobago. Fish with sufficiently large numbers of worms were euthanized with an overdose of tricaine methane-sulfonate (MS-222), preserved in 97% analytical ethanol and transferred to Adam Mickiewicz University, Poland, where parasites were isolated from their hosts.

During two subsequent sampling trips (2017 and 2018), guppies were collected from five streams on Trinidad. These fish were transported to the field station, where each population was kept in separate aquaria. Fish (anesthetized with MS-222) were screened for the presence of *Gyrodactylus* spp. under a dissecting microscope. If parasites were identified, a single worm was allowed to move to a naïve, anesthetized fish, with the transfer closely monitored to ensure movement of just a single worm. Infected fish were screened every 2-3 days for the presence of gyrodactylids. Number of parasites and their location on hosts were recorded. After 9-12 days, infected fish were euthanized with an overdose of MS-222, number of worms were counted, preserved in 97% analytical ethanol and transported to Poland. Samples with more than 10 individuals of *Gyrodactylus* spp. were used for DNA extraction and species identification. *G. bullatarudis* was identified in three streams/populations (Caura River, Lopinot River, Santa Cruz River) and samples from these sites were used for genome resequencing.

For RNA sequencing we used *G. bullatarudis* individuals farmed in the Cardiff University parasitology laboratory from a culture isolated from ornamental guppies in 2017 and maintained for approx. 3 months. Heavily infected fish were euthanized and preserved in RNAlater, and 5,000 individual worms later separated from their hosts in fresh RNAlater for transport to Poland for RNA extraction.

The project, including collection of wild guppies, was conducted with the permission from the Tobago House of Assembly (permit number 004/2014). All national guidelines for the care and use of animals were followed. Procedures and protocols were conducted under UK Home Office license (PPL 302876) with approval by the Cardiff University Animal Ethics Committee.

DNA extraction, library preparation and sequencing

All DNA extraction was from pools of individuals, each derived from a single worm. Extraction was by Proteinase K digestion (3 h) and MagJET Genomic DNA kit (Thermo Scientific™). DNA concentration was measured with Qubit High Sensitivity reagents and DNA quality was assessed on agarose gels. For the reference genome, a PCR-free library was prepared and sequenced by the CRG Sequencing Unit in Barcelona. Sequencing was performed on an Illumina HiSeq4000 in Rapid Mode and yielded 88.4 million 2 x 250 bp reads. Two mate-pair libraries (approx. 3 kb and 10 kb insert size) were constructed from the same DNA samples and were sequenced on a half lane of a HiSeq2500 machine.

RNA was extracted with RNAzol, and quality was assessed by TapeStation. Because of the low RNA yield, we used SMARTer Ultra Low RNA kit and TruSeq RNA stranded library construction. We then sequenced ca. 10 Gb on HiSeq2500 with 2 x 100 bp mode. Library construction and sequencing were performed by Macrogen Korea.

DNA from samples collected for genome resequencing was extracted with MagJET reagents as described above. Species ID was determined by sequenced COII fragment of mtDNA (Xavier et al., 2015). Sequences were aligned with records downloaded from the NCBI Genbank, and a neighbor joining tree was constructed with MEGA-X software (version 10.0.5; Kumar, Stecher, Li, Knyaz, & Tamura, 2018) (with 500 bootstraps; Supplementary Figure S1). Based on the DNA quality and quantity, *G. bullatarudis* samples were then selected to prepare libraries using Nextera Flex kit, and were sequenced on an Illumina HiSeq2500 (Macrogen Korea).

Genome assembly and annotation

The 2 x 250 pair-end reads were assembled with the shovil pipeline (version 1.0-pre1; https://github.com/tseemann/shovill; with default parameters), which uses SPAdes assembler (Bankevich et al., 2012) but reduces fastq files to manageable depth. Contigs shorter than 200 bp were removed from the assembly. The assembly was then screened for contamination using Blast against UniRef90, which was later visualized with MEGAN software (version 6.13.1; Huson, Auch, Qi, & Schuster, 2007). Aggregate properties of the assembly (GC content vs. coverage) were visualized using blobtools (version v1.0; Laetsch & Blaxter, 2017). Putative contaminant contigs (coverage < 100x; GC content > 50%) were removed after examination of blobplot outputs (Supplementary Figure S2). The remaining contigs were then subjected to scaffolding with BESST software (version 2.2.8; Sahlin, Vezzi, Nystedt, Lundeberg, & Arvestad, 2014). Prior to scaffolding, mate-pair reads were mapped to contigs with nxtrim (version v0.4.3-778bea9) and bwa mem (version 0.7.10-r789; Li & Durbin, 2010; O'Connell et al., 2015). Scaffolds shorter than 500 bp were removed from the genome draft. Finally, gaps were filled with the GapCloser software (version 1.12; Luo et al., 2012). Genome quality was assessed with QUAST software (Gurevich et al. 2013). Detailed description of functional and structural annotation of the nuclear genome is provided in the Supplementary Materials and Methods.

The mitochondrial genome was assembled with MITObim (version 1.9; Hahn, Bachmann, & Chevreux, 2013) *de novo*, using a subset of 20 million sequenced reads and a COII mtDNA fragment (Genbank accession KP168347) as initial bait. Results were manually inspected, and annotation performed with MITOS (Bernt et al., 2013).

The assembled genome was submitted to the GenBank database (accession no. PRJNA532341). During submission, 77 short scaffolds were identified as contaminated (derived either from *P. reticulata* or from adapters). Of these, 30 scaffolds containing 8 predicted protein coding genes were removed. Other scaffolds were trimmed or masked.

Secretome

To define the secretome we applied a strategy similar to Cuesta-Astroz et al. (2017). Briefly, SignalP (version 4.1; Petersen, Brunak, von Heijne, & Nielsen, 2011) was used to identify classical secretory proteins. The proteins without signal peptide were analyzed with SecretomeP (version 1.0; Bendtsen, Jensen, Blom, Von Heijne, & Brunak, 2004) to predict non-classical secretor proteins (only records with neural network score >0.9 were assigned as secreted proteins). TargetP (version 1.1; Emanuelsson, Nielsen, Brunak, & Von Heijne, 2000) was used to exclude mitochondrial proteins and TMHMM (version 2.0c) to identify transmembrane helicases.

Comparative genomics

To identify orthologous sequences between *G. bullatarudis* and *G. salaris* we ran reciprocal blastp (version 2.2.31; -evalue 0.001, -num_alignments 1). For subsequent analyses, we selected only such pairs in which identity was >30% across an alignment length of >70% of the *G. bullatarudis* sequence. TranslatorX (version v1.1; Abascal, Zardoya, & Telford, 2010) was used for nucleotide sequence alignment based on amino acid information. Stop codons were changed for gap sequences and the yn00 program from PAML (version 4.9h; Yang, 2007) was used to estimated dN and dS for all pairs of orthologs. The same analyses, except for the alignment length filtering, were performed to determine orthologous sequences between *G. bullatarudis* and the draft genome of *G. turnbulli* (another guppy gyrodactylid, unpublished).

We used 16 genomes of Platyhelminthes, downloaded from the WormBase ParaSite (Howe, Bolt, Shafie, Kersey, & Berriman, 2017) in November 2018, to assess the phylogenetic relationships of *G. bullatarudis*. OMA software (Altenhoff et al., 2018) was used for classifying protein sequences of orthologous groups. We first ran analyses with automatic species tree prediction, and selected 472 Orthologous Groups with maximum two species missing per cluster. Muscle (version v3.8.31; Edgar, 2004) was then used for sequence alignment, and trimal (version v1.4.rev22; Capella-Gutiérrez, Silla-Martínez, & Gabaldón, 2009) for alignment cleaning (with -gappyout parameter). Sequences were concatenated with FASconCAT (version v1.11; Kück & Meusemann, 2010). RAxML (version 8.2.12c; Stamatakis, 2006) was used to reconstruct phylogenetic relationships, with the GAMMA model of rate heterogeneity and automatically selected substitution model. We performed this separately for each genome partition (gene). *Schmidtea mediterranea* and *Macrostomum lignano* were defined as an outgroup, and 100 alternative runs on distinct starting trees were initialized. Best ML tree was then used to rerun the OMA pipeline (version 2.3.0) and to identify evolutionary events and orthogroups. We searched for genes duplicated between the common ancestor of all Neodermata and the *G. bullatarudis* genome. To explore orthology groups and to identify the most dynamic gene families, we used pyham (Train, Pignatelli, Altenhoff, & Dessimoz, 2018) and custom scripts, which were run on the OMA output, to summarize results and select orthology groups with the largest number of duplications. Gene Ontology terms associated with *G. bullatarudis* genes were merged within orthology groups, and gene ontology enrichment was calculated with topGO package in R. Functional analyses are based on the Gene Ontology annotated with Pannzer2 software (Törönen, Medlar, & Holm, 2018). To confirm findings, we repeated enrichment analyses with Gene Ontologies annotated with GOA-Uniprot approach (details in Supplementary Information).

Population genetics

Raw read quality was inspected with FastQC (Andrews & Babraham Bioinformatics, 2010), and low quality reads were trimmed with Trimmomatic, with default trimming parameters recommended within the software manual (version 0.36; Bolger, Lohse, & Usadel, 2014). Reads were then mapped to the reference genome with bwa mem (version 0.7.10-r789), and duplicates were marked with picard tools (version 2.18.5-6). Files were then inspected with qualimap (García-Alcalde et al., 2012). SNPs and indels were called with samtools mpileup (version 1.6.0, options -R -C50 -t DP,ADF,ADR) and bcftools (version 1.6, options -f GQ -vmO v). We filtered out SNPs within 5 bp of an indel, with quality below 15, and, based on empirical distribution, with sequencing depth summed across all samples smaller than 50 or larger than 400. Using SNPs that remained after filtering, we performed principal component analyses (PCA) with plink (version 1.90; Purcell et al., 2007) and default parameters. Genetic variation (π), and differentiation between populations (Weir and Cockerham F_{ST} estimator) were calculated with vcftools (version v0.1.12b; Danecek et al., 2011) in 25 kb windows. Using the PopGenome Package (version 2.6.1; Pfeifer, Wittelsbürger, Ramos-Onsins, & Lercher, 2014) in R, we also calculated genetic variation (π) and differentiation (d_{XY}) per gene. F_{ST} outlier analysis was performed for each polymorphic site with BayeScan software (v2.1; Foll and Gaggiotti 2008). In all cases, analyses were performed after excluding indel polymorphisms. For each individual, we calculated divergence from the reference genome in the 25 kb non-overlapping windows, by counting the number of non-reference variants (adding 1 if heterozygous and 2 if alternative homozygous site). Windows of <12.5 kb were excluded from the analyses (i.e.

ends of the scaffolds or entire scaffolds/contigs shorter than 12.5 kb). Based on the empirical distribution of divergence (number of variants divided by two times the window length), we classified each window in each sample to be A) Gb1-like (divergence smaller than 0.25% from the reference genome), B) Gb2-like (divergence >0.35% and <0.8%), or C) of undetermined origin (divergence >0.25% and <0.35% or >0.8%). The entire reference sequence was determined as Gb1-like. If a given gene was localized in either Gb1-like or Gb2-like regions in different samples, one Gb1-like and one Gb2-like sequence per sample were randomly chosen for downstream analyses. The rate of synonymous and non-synonymous substitutions between these two sequences were then calculated with yn00 program from PAML (Yang 2007). Numbers of non-synonymous and synonymous substitutions were then summed for genes localized in regions fixed for different haplotypes in different populations. The sums were divided by the sum of non-synonymous and synonymous sites respectively, giving a final rate of non-synonymous to synonymous substitutions (dN/dS). The rate was then calculated for the same number of genes randomly selected from the genes having dN/dS calculated between haplotypes. This procedure was repeated 1000 times to produce genome-wide random expectations against which we compared observed values.

To investigate sensitivity of population genetic parameter estimates to particular variant calling protocols, we also called SNPs using GATK (version 4.1.4.0), following the best practices workflow (DePristo et al. 2011). Using this dataset, we calculated F_{ST} among populations and divergence from reference genome. The results were compared with the analyses calculated based on SNPs called with samtools.

Results

Reference genome

A total of 44 gigabases of sequencing data were used to generate the draft assembly. Contigs were scaffolded with mate-pair libraries generating the final assembly, with the assembly size of 84.4 Mb and scaffold N50 size of 0.31 Mb (Table 1, Supplementary Table S1, Supplementary Figure S3). Combination of several *ab initio,* RNA-Seq and orthology based strategies were used to generate 10,749 protein coding gene predictions (Table 1, Supplementary Table S1). Average genes span 4,691 bp, containing 6 exons. Quality controls support high quality of gene predictions and confirm the absence of bacteria, fish or human contamination (Supplementary Figure S4).

Comparative genomics

Using proteomes available from 13 other Platyhelminthes, we investigated the evolution of gene families and long-term evolution in the lineage leading to *G. bullatarudis*, in a phylogeny reconstructed with 472 highly conserved genes. Our analysis placed the Monogenea as a fast-evolving sister lineage to Cestoda and Trematoda (Figure 1A, 1B). The divergence between the two *Gyrodactylus* (*G. salaris* and *G. bullatarudis*) species is much deeper than that between any other congeneric pair of species in the phylogeny, suggesting rapid molecular diversification within monogeneans.

We inferred homologs among Platyhelminth species to show general patterns of gene birth, loss and duplication (Figure 1A). The fraction of genes without orthologs in other species was highest in monogeneans, which may reflect an increased rate of gene births in this lineage. Alternatively, this finding may be a consequence of rapid divergence in the clade, hampering identification of orthologous genes. Since the split from the most common ancestor of *Gyrodactylus* spp., more genes appear to have been lost in the *G. salaris* lineage (1,772) compared to *G. bullatarudis* (1,364) and fewer genes have been retained in *G. salaris* (5,581) than in *G. bullatarudis* (5,916), suggesting the difference in the genomes' completeness. However, the overall high rate of gene gain (Figure 1A) is consistent with the rapid evolutionary rate we inferred for this clade. Such a high rate of diversification can potentially hamper investigations of long-term adaptive evolution due to signal loss, which may account for the low number of homologous sequences (3873) between *G. bullatarudis* and *G. salaris* that met the minimum criteria for inferring homology (blast e-value < 10^{-3}, alignment length >70% of *G. bullatarudis* sequence). Furthermore, non-synonymous site divergence was high for these genes (average dN = 0.35, Supplementary Figure S5), as was synonymous site divergence (dS >> 1; Supplementary Figure S6). The same analyses with the draft genome of another *Gyrodactylus* species infecting guppies (*G. turnbulli,* Ch. Hahn et al. unpublished data) similarly showed high divergence of synonymous and non-synonymous sites (Supplementary Figures S5 and S6).

Analysis of duplicated genes can reveal important historical adaptive events. We identified 522 gene families with putative duplications in *G. bullatarudis* as compared to the common ancestor of all Neodermata. In these genes, several Gene Ontology terms were significantly enriched (Figures 1C, S7 and S8). The gene family with the largest number of duplications was 'serine proteases', showing homology to *S. mansoni* cercarial elastase genes (Hierarchical Orthologous Group HOG01193 in our OMA analysis, Figure 1D). For this gene family most of the species have 1-3 paralogs, whereas *S. mansoni* has 7, *G. salaris* 13 and *G. bullatarudis* 15 paralogs (Figure 1D). Other gene families with pronounced expansion in *G. bullatarudis* include those with homology to venom allergen-like proteins (HOG1412) and to dynein light-like proteins (HOG1668, Supplementary Figures S9 and S10).

Population genomics

We explored micro-evolutionary genomic changes in three local populations of *G. bullatarudis* from Trinidad (Caura River, Lopinot River and Santa Cruz River), by analyzing 11 samples sequenced to 23x coverage on average (Supplementary Table S3, Supplementary Figure S11). Across the Trinidadian populations we identified 77,162 Single Nucleotide Polymorphisms (SNPs) and 18,305 variable indels, including 6,793 SNPs localized within protein coding sequences of 1,963 genes. In addition, 193,420 single nucleotide positions and 36,193 indels were fixed for the alternative alleles when compared to the reference genome from Tobago. The PCA on SNP genotypes showed that samples were clustered by population and that the genomes diverged between rivers (Figure 2A). Per gene nucleotide diversity was low and very similar among the three populations. Genes without orthologous sequences in the *G. salaris* genome showed, however, higher nucleotide diversity than genes for which orthologous sequences were found (Supplementary Figure S12). Similarly, per gene divergence (calculated as d_{XY}) did not differ between the three inter-population comparisons, but the divergence was higher for genes without orthologous sequences in the *G. salaris* genome (Figure 2B), suggesting faster evolution than other genes in the genome (high divergence can lead to difficulties in orthologs identification). The gene with the highest divergence between populations was Gbulla1a000092, elastase, a member of the previously mentioned family of serine proteases (HOG01193, Supplementary Figure S13). We found high non-synonymous divergence in this gene ($d_{XY} = 0.02$ in two out of three comparisons).

Most genes showing high rate of non-synonymous substitutions did not have orthologs in *G. salaris*, and the genes without identified orthologs differentiated faster in non-synonymous, but not in synonymous sites, compared with genes with identified orthologs (Figure 2B). Given the lack of homology between the species, it is hard to infer molecular functions of these genes. However, we were able to predict the secretome of *G. bullatarudis* bioinformatically, based on the presence of signaling peptides. In parasitic Platyhelminths, secretory/excretory genes might be primarily involved in the host-pathogen dialogue (Garg & Ranganathan, 2012; Hewitson, Grainger, & Maizels, 2009), and thus such genes are likely involved in the host-pathogen coevolution. We observed that secretory/excretory genes are significantly over-represented among *G. bullatarudis* genes without orthologs in *G. salaris* ($p < 10^{-5}$, Fisher Exact Test). However, genes predicted as secretory/excretory do not differentiate faster between local populations than other genes in the genome (Supplementary Figure S14).

Genome-wide genetic differentiation between populations offers another way to reveal genomic regions that might be associated with phenotypic divergence. We therefore calculated Weir and Cockerham estimates of F_{ST} between the three populations in 25 kb windows across the genome. For all three comparisons (Caura vs Lopinot, Caura vs Santa Cruz, Lopinot vs Santa Cruz), median F_{ST} fell close to zero, but in all three cases, we also found windows with F_{ST} values close to 1 (Figure 2C-E**,** Supplementary Figures S15-S17). While no single SNP reached statistical significance in F_{ST} outlier tests (FDR=0.05) due to the small number of individuals sequenced per population, genes localized in windows with extreme F_{ST} values (F_{ST} > 0.98) contain several excretory/secretory genes with considerable non-synonymous divergence - interesting candidates for future research (Table 2).

Signatures of hybridization in G. bullatarudis *genomes*

When we manually investigated random regions in the genome, we found that divergence from the reference genome fell into one of two categories: relatively high or low divergence from the reference genome in different genomic locations. Indeed, when we calculated divergence from the reference genome within 25 kb non-overlapping windows genome-wide, we found a pronounced bimodal distribution, with one peak around 0.55% divergence and the second equal or smaller to 0.1% divergence (Figure 3A). The pattern was very similar for all samples from the three Trinidadian populations (Figure 3B and Supplementary Figure S18), and consistent even if we excluded indels or heterozygotic genotypes, or if we used different software for SNP calling. We interpret this pattern as a signature of hybridization and subsequent recombination between two divergent lineages, one lineage similar to the reference genome and the other lineage with ~0.5% divergence from the reference, which occurred before parasites colonized the three sites studied here.

The divergence from the reference genome calculated in 25 kb non-overlapping windows was used to determine nGb1 (similar to the Tobagonian reference genome) or nGb2 origin (diverged from the reference genome by about 0.5%) for each sample (Figure 3A). Windows of nGb1 and nGb2 origin did not differ in heterozygosity (Supplementary Figure S19), but number of non-reference homozygotic positions was larger in the nGb2 windows (Supplementary Figure S20), showing that the high divergence from the reference genome in nGb2 regions is not driven by their elevated heterozygosity. More than 25% of scaffolds/contigs contained both nGb1 and nGb2 windows (Supplementary Figure S21). Despite heterozygosity being low across genomes in all populations (median number of heterozygotic genotypes per 25 kb window = 4, median ratio of heterozygotic to homozygotic genotypes = 0.06), the ratio of heterozygotic to homozygotic sites was elevated in several windows in samples from the Santa Cruz population (Supplementary Figure S22). These are likely the genomic regions heterozygotic for nGb1/nGb2 origins, which could have been associated either with a recent expansion (such that there was not enough time for fixation of one of the haplotypes), or with heterozygote advantage associated with genes localized in these regions.

Hybridization and subsequent recombination could have fixed advantageous combinations of alleles at the island level, but some combinations could give an advantage only in the context of local host populations. That process could produce elevated F_{ST} values in some of the genomic locations. To explore this possibility, we selected genomic windows polymorphic for nGb1/nGb2 origins in Trinidad and investigated F_{ST} distribution in these loci. These distributions (Supplementary Figure 23), showed the same peaks of extreme F_{ST} values as observe in analysis of all windows (Figure 2). Thus, differentiation between populations appeared to be mostly driven by local fixations of fragments of the two highly divergent haplotypes. We further searched for signatures of adaptive evolution among genes showing high inter-population differentiation. We identified 326 genes localized in regions which were fixed for different genomes of origin (nGb1 or nGb2) in different Trinidadian rivers. We tested whether the rate of non-synonymous to synonymous substitutions among these genes was higher compared to the randomly sampled genes from the genome, for which we could calculate dN/dS between haplotypes. We found significantly larger dN/dS for locally-fixed set of genes only in nGb2 haplotype of the Lopinot population (Supplementary Figure 24). The function of most of these genes (n=17; Supplementary Table S4) is unknown.

Discussion

Testing scenarios of host-parasite coevolution requires an understanding of how parasites adapt to their hosts at the molecular level (Schmid-Hempel, 2011; Woolhouse et al., 2002). We explored the evolution of *Gyrodactylus bullatarudis* by analyzing its genome, comparing genomic composition with a related parasite species, and comparing genomic variation of parasites derived from different local populations. Our assembly of the *G. bullatarudis* genome shows substantially increased contiguity and completeness (Supplementary Table S1) compared to the only other monogenean genome published so far (*G. salaris*; see Hahn et al. 2014). *G. bullatarudis* gene size is almost twice as long (4.7 vs. 2.7 kb), genes contain significantly more exons (6 vs. 4), while have only slightly longer introns (769 vs. 659 bp) and similar size of exons (288 vs. 289 bp). Most likely the differences between species in these properties result from better contiguity of the *G. bullatarudis* genome and from availability of transcriptomic data, which improved gene predictions in the present study.

Our predicted phylogenetic relationships between platyhelminths were generally consistent with previous studies (Hahn et al. 2014; Egger et al. 2015, but see Laumer et al. 2015), placing the Monogenea as a fast-evolving sister lineage to Cestoda and Trematoda (Figure 1A-B). Similar to results reported for other flatworms (Coghlan et al., 2019), we found high fractions of clade-specific gene families, suggesting fast molecular evolution despite considerable morphological conservatism. All these results demonstrate that the *G. bullatarudis* genome provides a valuable source of information to mine the molecular basis of adaptation in the context of host-pathogen coevolution.

Molecular basis of adaptation

Selection on coding sequences is typically measured by the rate of non-synonymous to synonymous substitutions (dN/dS), but with dS > 0.4 the test loses more than 50% of its power (Gharib & Robinson-Rechavi, 2013). Given that dS was 3.7 since the split of *Gyrodactylus* species, we did not perform this classical test, and instead based our comparative inference of past adaptive evolution on patterns of gene duplication. Gene duplications that persist in an evolving lineage can be beneficial from the time of their origin, e.g. due to protein dosage effect, or can confer advantage in a later phase of evolution due to neofunctionalization (Kondrashov, Rogozin, Wolf, & Koonin, 2002). Experimental studies confirmed that organisms often evolve duplications in response to environmental challenge (Kondrashov, 2012), and such events have previously been documented for genes relevant to parasitism and drug resistance evolution in flatworms (Coghlan et al., 2019).

Among genes duplicated in the *G. bullatarudis* lineage, the G-protein coupled receptor signaling pathway was the most abundant group of Gene Ontology terms. G proteins are involved in transmitting signals from a variety of stimuli outside a cell to its interior. For example, in the amoebozoan parasite *Entamoeba histolytica,* G proteins modulate attachment to and killing of host cells, regulate invasion, phagocytosis and evasion of the host immune response by surface receptor capping (Bosch & Siderovski, 2013). Among helminths, it has been suggested that *Schistosoma mansoni* G-receptors likely play key roles in pathogenesis (Zamanian et al., 2011). It may thus be the case that these proteins are particularly important in the coevolution of monogenean parasites. Many other genes associated with such enriched terms as biological regulation, response to external stimulus, detoxification and behaviour could have played a role in coevolution as well (Figure 1C, Supplementary Figure S7 and S8**,** Supplementary Table S2).

The gene family with the highest number of duplications was 'serine proteases', with homology to *S. mansoni* cercarial elastase genes – an enzyme that plays a pivotal role in the penetration of host skin by cercariae to initiate infection (Salter et al., 2002). Several paralogs found in the *S. mansoni* genome show high similarity, indicating selection for increased gene expression of cercarial elastase gene via a dosage effect (Ingram et al., 2012). In contrast, monogenean paralogs are considerably diverged, suggesting evolutionary pressure for neo- or subfunctionalization in the monogenean lineage, as well as in the individual *Gyrodactylus* lineages (Supplementary Figure S13). The function of these genes in monogeneans is unknown, but given their homology to cercarial elastases and enzymatic activity in other species, these genes might play a crucial role in digesting host tissue. This inference is consistent with the fact that all gyrodactylids are epidermal browsers that occasionally eat dermal cells (Bakke, Cable, & Harris, 2007).

Interestingly, all but one (Gbulla1a000092) member of the serine proteases gene family showed almost complete conservation between local populations. The high non-synonymous divergence in Gbulla1a000092 might be thus interpreted as a signature of recent diversifying selection acting on this gene. Inspection of reads that mapped to the contig containing this gene revealed patterns suggesting a small inversion at the end of the gene (Supplementary Figure S25). This might have caused open reading frame shifts, followed by rapid neofunctionalization.

The highly expanded gene families also included those with homology to venom allergen-like proteins and to dynein light-like proteins. The expression of venom allergen-like proteins is specifically up-regulated during parasitic phases of the life cycles of helminths, and these proteins are abundantly secreted during several stages of parasitism, causing extensive damage to host tissues (Wilbers et al., 2018). Dynein light-like proteins, a helminth-specific group of proteins binding calcium ions, have been linked to host immune stimulation (Jones, Gobert, Zhang, Sunderland, & McManus, 2004), and it therefore seems plausible that these genes could have evolved in gyrodactylids under evolutionary pressure from the host's immune system.

Overall, our comparative analyses show that the Monogenea is a dynamic, fast-evolving clade of parasites, and that many evolutionary events of gene duplication could have been related to interactions with their hosts. Some of these genes show differentiation between populations, suggesting strong diversifying selection by host populations. Elucidation of the specific functions of the candidate genes we identified, and their potential role in the host-parasite coevolution, could be the focus of future hypothesis-driven work.

Hybridization dominated population history of Trinidadian G. bullatarudis

Many parasitic organisms are capable of parthenogenetic reproduction, which facilitates colonization of hosts from just a single individual. However, occasional sexual reproduction appears essential for purging deleterious mutations and restoring evolutionary potential (Heitman, 2010). *Gyrodactylus* species are capable of asexual, parthenogenetic and sexual reproduction (Cable & Harris, 2002; Schelkle et al., 2012), although relative frequencies of these reproductive modes are unknown and likely vary between species.

Our data support the role of recombination in the genus *Gyrodactylus*, which is increasingly used as a model for host-parasite coevolution (Hutson, Cable, Grutter, Paziewska-Harris, & Barber, 2018; Phillips et al., 2018; Robertson, Bradley, & MacColl, 2017). In addition to recombination, sexual reproduction enables hybridization between individuals from previously reproductively isolated populations, or even species. In recent years, the potential for such events has increased due to human activity and global changes which breakdown barriers in species distribution. For example, Tihon et al. (Tihon, Imamura, Dujardin, Van Den Abbeele, & Van den Broeck, 2017) demonstrated extensive hybridization between phylogenetically distinct lineages of *Trypanosoma congolense,* challenging the traditional view of predominantly clonal evolution in this genus (Tibayrenc & Ayala, 2012). Likewise, schistosomiasis that reached southern Europe in 2013 was caused by a hybrid species (Kincaid-Smith et al., 2019), and variation that arose in another medically-important parasite – *Leishmania* – was caused by a recombination event between two previously diverse strains (Rogers et al., 2014). These examples highlight the importance of hybridization in parasite evolution beyond the study of adaptive evolution, into epidemiology and public health (King et al., 2015).

A previous study of *G. bullatarudis*, which included samples collected from some of the same rivers as ours 13 years previously, reported two very divergent mtDNA lineages (mtGb1 and mtGb2) present in Trinidadian populations of *G. bullatarudis*. The level of divergence (11.8-13%), led the authors to suggest cryptic speciation (Xavier et al., 2015). However, this study did not sequence any nuclear loci, which might have detected hybridization between the two lineages. Consistent with those previous findings, our genome-wide analyses revealed that nuclear genomes of Trinidadian *G. bullatarudis* is built from two divergent types. One type was very similar to the Tobagonian reference genome (nGb1), and the other diverged from the reference by about 0.5% (nGb2). We interpret this as evidence for hybridization between two divergent lineages of *G. bullatarudis.* It seems more likely that hybridization has occurred only once in the lineages' history, as the contributions of both lineages to the genome is approximately equal (repeated backcrossing would be expected to reduce the share of one of the lineages). We found the mtGb1 haplotype to be more closely related to mtDNA sequences from samples collected by us on Tobago (divergence 3.1-3.8%) than to mtGb2 (Supplementary Figure S26). This suggests that the split between the two mtDNA strains reflects the same isolation event between Trinidad and Tobago that caused divergence of the nuclear genome of the Trinidadian lineage (nGb2) from the genome inferred to be closely related to Tobagonian reference (nGb1). Thus, in the populations we studied, a recombinant between an indigenous *G. bullatarudis* genome and the genome related to our Tobagonian reference genome apparently replaced that indigenous population. Given that mtDNA evolves much faster than nuclear DNA (Allio, Donega, Galtier, & Nabholz, 2017), the 11.8-13% divergence in mtDNA corresponds well to the 0.55% nuclear divergence in genomic DNA found in our study.

Mitochondrial divergence of 3.8% (mtGb1 vs Tobagonian references) and 13% (mtGb1 vs mtGb2) suggests the following scenario of events: Trinidad and Tobago populations of *G. bullatarudis* diverged, and evolved independently; , and later on Tobagonian worms, carrying mtGb1, were introduced to Trinidad, where a hybridization event occurred with a local lineage carrying nGb2 genome. That recombinant genotype then appears to have replaced indigenous nGb2/mtGb2 strain, at least in the population we studied. Indeed, using the same primers to amplify mtDNA COII gene as Xavier et al. (2015), we found that all 44 samples from Trinidad harbored the Gb1 mitochondrial haplotype, supporting the replacement scenario (Supplementary Figure 26). The high invasion success of the resulting recombinant genome is in line with growing appreciation of the role of hybridization in parasite adaptation, and its association with increased virulence (King et al., 2015). Indeed, sexual reproduction between inbred *G. turnbulli* strains (Schelkle et al. 2012) demonstrated that mixed populations have significantly increased virulence.

Our results demonstrate that hybridization between divergent *G. bullatarudis* lineages had dramatic consequences for the parasite evolution, resulting in the emergence of a stable “hybrid” species. The invasion success of a recombinant strain might arise from synergistic epistasis between recombined genomic regions, or from advantage stemming from heterozygosity in some genomic regions. The elevated heterozygosity for nGb1 and nGb2 that we found in the Santa Cruz population may be the result of heterozygosity maintained by such overdominance. However, such signals were not detected in other populations, suggesting that nGb1/nGb2 heterozygotic regions (containing 70 genes) played a role in local adaptation in Santa Cruz, rather than in island-wide invasion success of a recombinant strain.

Hybridization and subsequent recombination can also increase the scope for local adaptation by providing genetic variation on which selection can act, such that alternative variants of diverged alleles could fix in different populations, depending on the selection pressure from local populations of hosts. At 326 genes, we found alternative variants fixed in different populations, indicating their possible role in adaptation to local host populations. While molecular functions and potential adaptive advantage of these genes need future validation, these are undoubtedly worthwhile candidates for future investigations.

Conclusions

Our study has revealed signatures of adaptive evolution in *G. bullatarudis* at different timescales. At the macroevolutionary timescale, we observed duplications of genes whose functions strongly suggest their involvement in the host-parasite arms race, such as G-proteins or serine proteases. Divergence within the latter gene family suggests that they have undergone evolution by subfunctionalisation, although, interestingly, a member of the serine proteases family was identified as a top candidate for diversifying selection at the microevolutionary scale. Our findings indicate fast adaptive evolution, resulting in the rapid loss of orthology to other *Gyrodactylus* species, in the excretory-secretory protein gene group that is likely important to host-parasite interactions in Platyhelminths. A number of these genes, including the serine protease gene, showed extreme inter-population differentiation, indicative of local adaptation. By identifying a number of strong candidate genes likely involved in host adaptation, our study opens the way to investigate host-parasite coevolution in a complex system of vertebrate host and parasite flatworm. Finally, our analyses revealed that hybridization and a consequent recombination event involving two divergent *Gyrodactylus* genomes has dramatically affected the genetic composition of Trinidadian *G. bullatarudis* populations. Consistent with the notion that recombination between divergent pathogen genomes can result in particularly successful parasites, the recombinant genome apparently managed to invade Trinidad and completely replace local populations.

Acknowledgments

We thank W. Babik and two anonymous reviewers for their help in improving this manuscript; N. Cook, A. Szubert-Kruszyńska, A. Sadowska-Konczal and staff of the Environmental Research Institute Charlotteville (ERIC), Tobago, for support in the field; P. Turpin of Man O War Bay Cottages for renting us the field station; Tobago House of Assembly (THA) for granting permission to conduct field surveys; K. Dudek and J. Raubic for help in molecular laboratory and J. Hecht for sequencing. The research was funded by Polish National Science Center Fuga Grant UMO-509-2016/20/S/NZ8/00208. This work was supported by the INB ("Instituto Nacional de Bioinformatica") Project PT13/0001/0021 (ISCIII -FEDER). We also acknowledge support of the Spanish Ministry of Economy and Competitiveness, 'Centro de Excelencia Severo Ochoa 2013-2017', SEV-2012-0208. The computations were performed at the Poznan Supercomputing and Networking Center.

References

Abascal, F., Zardoya, R., & Telford, M. J. (2010). TranslatorX: Multiple alignment of nucleotide sequences guided by amino acid translations. *Nucleic Acids Research*. doi:

10.1093/nar/gkq291

Allio, R., Donega, S., Galtier, N., & Nabholz, B. (2017). Large variation in the ratio of mitochondrial to nuclear mutation rate across animals: Implications for genetic diversity and the use of mitochondrial DNA as a molecular marker. *Molecular Biology and Evolution*. doi: 10.1093/molbev/msx197

Altenhoff, A. M., Glover, N. M., Train, C. M., Kaleb, K., Warwick Vesztrocy, A., Dylus, D., … Dessimoz, C. (2018). The OMA orthology database in 2018: Retrieving evolutionary relationships among all domains of life through richer web and programmatic interfaces. *Nucleic Acids Research*. doi: 10.1093/nar/gkx1019

Andrews, S., & Babraham Bioinformatics. (2010). FastQC: A quality control tool for high throughput sequence data. *Manual*. doi: citeulike-article-id:11583827

Bakke, T. A., Cable, J., & Harris, P. D. (2007). The Biology of Gyrodactylid Monogeneans: The "Russian-Doll Killers." *Advances in Parasitology*. doi: 10.1016/S0065-308X(06)64003-7

Bankevich A., Nurk S., Antipov D., Gurevich A., Dvorkin M., Kulikov A. S., … Pevzner P. A. (2012) SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. Journal of Computational Biology. doi: 10.1089/cmb.2012.0021.

Bendtsen, J. D., Jensen, L. J., Blom, N., Von Heijne, G., & Brunak, S. (2004). Feature-based prediction of non-classical and leaderless protein secretion. *Protein Engineering, Design and Selection*. doi: 10.1093/protein/gzh037

Bernt, M., Donath, A., Jühling, F., Externbrink, F., Florentz, C., Fritzsch, G., … Stadler, P. F. (2013). MITOS: Improved de novo metazoan mitochondrial genome annotation. *Molecular Phylogenetics and Evolution*. doi: 10.1016/j.ympev.2012.08.023

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*. doi: 10.1093/bioinformatics/btu170

Bosch, D. E., & Siderovski, D. P. (2013). G protein signaling in the parasite *Entamoeba histolytica*. *Experimental and Molecular Medicine*. doi: 10.1038/emm.2013.30

Cable, J., & Harris, P. D. (2002). Gyrodactylid developmental biology: Historical review, current status and future trends. *International Journal for Parasitology*. doi: 10.1016/S0020-7519(01)00330-7

Capella-Gutiérrez, S., Silla-Martínez, J. M., & Gabaldón, T. (2009). trimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*. doi: 10.1093/bioinformatics/btp348

Choi, Y. J., Tyagi, R., McNulty, S. N., Rosa, B. A., Ozersky, P., Martin, J., … Mitreva, M. (2016). Genomic diversity in *Onchocerca volvulus* and its Wolbachia endosymbiont. *Nature Microbiology*. doi: 10.1038/nmicrobiol.2016.207

Coghlan, A., Tyagi, R., Cotton, J. A., Holroyd, N., Rosa, B. A., Tsai, I. J., … Berriman, M. (2019). Comparative genomics of the major parasitic worms. *Nature Genetics*. doi: 10.1038/s41588-018-0262-1

Cole, R., & Viney, M. (2018). The population genetics of parasitic nematodes of wild animals. *Parasites & Vectors*. doi: 10.1186/s13071-018-3137-5

Criscione, C. D., Poulin, R., & Blouin, M. S. (2005). Molecular ecology of parasites: Elucidating ecological and microevolutionary processes. *Molecular Ecology*. doi: 10.1111/j.1365-294X.2005.02587.x

Cuesta-Astroz, Y., Oliveira, F. S. de, Nahum, L. A., & Oliveira, G. (2017). Helminth secretomes reflect different lifestyles and parasitized hosts. *International Journal for Parasitology*. doi: 10.1016/j.ijpara.2017.01.007

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., … Durbin, R. (2011). The variant call format and VCFtools. *Bioinformatics*. doi: 10.1093/bioinformatics/btr330

Denver, D. R., Morris, K., Lynch, M., Vassilieva, L. L., & Thomas, W. K. (2000). High direct estimate of the mutation rate in the mitochondrial genome of *Caenorhabditis elegans*. *Science*. doi: 10.1126/science.289.5488.2342

DePristo, M., Banks, E., Poplin, R., Garimella, K., Maguire, J., Hartl, C., ... Daly, M., (2011). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Nature Genetics 43:491-498

Dobson, A., Lafferty, K. D., Kuris, A. M., Hechinger, R. F., & Jetz, W. (2008). Homage to Linnaeus: How many parasites? How many hosts? *Proceedings of the National Academy of Sciences*. doi: 10.1073/pnas.0803232105

Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*. doi: 10.1093/nar/gkh340

Egger, B., Lapraz, F., Tomiczek, B., Müller, S., Dessimoz, C., Girstmair, J., … Telford, M. J. (2015). A transcriptomic-phylogenomic analysis of the evolutionary relationships of flatworms. *Current Biology*. doi: 10.1016/j.cub.2015.03.034

El Nagar, A., & Maccoll, A. D. C. (2016). Parasites contribute to ecologically dependent postmating isolation in the adaptive radiation of three-spined stickleback. *Proceedings of the Royal Society B: Biological Sciences*, *283*(1836). doi: 10.1098/rspb.2016.0691

Emanuelsson, O., Nielsen, H., Brunak, S., & Von Heijne, G. (2000). Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. *Journal of Molecular Biology*. doi: 10.1006/jmbi.2000.3903

Farrer, R. A., Weinert, L. A., Bielby, J., Garner, T. W. J., Balloux, F., Clare, F., … Fisher, M. C. (2011). Multiple emergences of genetically diverse amphibian-infecting chytrids include a globalized hypervirulent recombinant lineage. *Proceedings of the National Academy of Sciences*. doi: 10.1073/pnas.1111915108

Foll, M., & Gaggiotti, O. (2008). A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. Genetics, 180(2), 977-993.

Frainer, A., McKie, B. G., Amundsen, P. A., Knudsen, R., & Lafferty, K. D. (2018). Parasitism and the Biodiversity-Functioning Relationship. *Trends in Ecology and Evolution*. doi: 10.1016/j.tree.2018.01.011

García-Alcalde, F., Okonechnikov, K., Carbonell, J., Cruz, L. M., Götz, S., Tarazona, S., … Conesa, A. (2012). Qualimap: Evaluating next-generation sequencing alignment data. *Bioinformatics*. doi: 10.1093/bioinformatics/bts503

Garg, G., & Ranganathan, S. (2012). Helminth secretome database (HSD): A collection of helminth excretory/secretory proteins predicted from expressed sequence tags (ESTs). *BMC Genomics*. doi: 10.1186/1471-2164-13-S7-S8

Gharib, W. H., & Robinson-Rechavi, M. (2013). The branch-site test of positive selection is surprisingly robust but lacks power under synonymous substitution saturation and variation in GC. *Molecular Biology and Evolution*. doi: 10.1093/molbev/mst062

Gurevich, A., Saveliev, V., Vyahhi, N., & Tesler, G. (2013). QUAST: quality assessment tool for genome assemblies. Bioinformatics. doi: 10.1093/bioinformatics/btt086.

Hahn, C., Bachmann, L., & Chevreux, B. (2013). Reconstructing mitochondrial genomes directly from genomic next-generation sequencing reads - A baiting and iterative mapping approach. *Nucleic Acids Research*. doi: 10.1093/nar/gkt371

Hahn, C., Fromm, B., & Bachmann, L. (2014). Comparative genomics of flatworms (Platyhelminthes) reveals shared genomic features of ecto- and endoparastic neodermata. *Genome Biology and Evolution*. doi: 10.1093/gbe/evu078

Hamilton, W. D. (2006). Sex versus Non-Sex versus Parasite. *Oikos*. doi: 10.2307/3544435

Hamilton, W. D., & Zuk, M. (1982). Heritable true fitness and bright birds: A role for parasites? *Science*. doi: 10.1126/science.7123238

Hatcher, M. J., Dick, J. T. A., & Dunn, A. M. (2012). Diverse effects of parasites in ecosystems: Linking interdependent processes. *Frontiers in Ecology and the Environment*. doi: 10.1890/110016

Heitman, J. (2010). Evolution of eukaryotic microbial pathogens via covert sexual reproduction. *Cell Host and Microbe*. doi: 10.1016/j.chom.2010.06.011

Hewitson, J. P., Grainger, J. R., & Maizels, R. M. (2009). Helminth immunoregulation: The role of parasite secreted proteins in modulating host immunity. *Molecular and Biochemical Parasitology*. doi: 10.1016/j.molbiopara.2009.04.008

Houde, A. E., & Torio, A. J. (1992). Effect of parasitic infection on male color pattern and female choice in guppies. *Behavioral Ecology*. doi: 10.1093/beheco/3.4.346

Howe, K. L., Bolt, B. J., Shafie, M., Kersey, P., & Berriman, M. (2017). WormBase ParaSite − a comprehensive resource for helminth genomics. *Molecular and Biochemical Parasitology*. doi: 10.1016/j.molbiopara.2016.11.005

Hupalo, D. N., Luo, Z., Melnikov, A., Sutton, P. L., Rogov, P., Escalante, A., … Carlton, J. M. (2016). Population genomics studies identify signatures of global dispersal and drug resistance in *Plasmodium vivax*. *Nature Genetics*. doi: 10.1038/ng.3588

Huson, D. H., Auch, A. F., Qi, J., & Schuster, S. C. (2007). MEGAN analysis of metagenomic data. *Genome Research*. doi: 10.1101/gr.5969107

Hutson, K. S., Cable, J., Grutter, A. S., Paziewska-Harris, A., & Barber, I. (2018). Aquatic Parasite Cultures and Their Applications. *Trends in Parasitology*. doi: 10.1016/j.pt.2018.09.007

Ingram, J. R., Rafi, S. B., Eroy-Reveles, A. A., Ray, M., Lambeth, L., Hsieh, I., … McKerrow, J. H. (2012). Investigation of the proteolytic functions of an expanded cercarial elastase gene family in *Schistosoma mansoni*. *PLoS Neglected Tropical Diseases*. doi: 10.1371/journal.pntd.0001589

Jacquin, L., Reader, S. M., Boniface, A., Mateluna, J., Patalas, I., Pérez-Jvostov, F., & Hendry, A. P. (2016). Parallel and nonparallel behavioural evolution in response to parasitism and predation in Trinidadian guppies. *Journal of Evolutionary Biology*. doi: 10.1111/jeb.12880

Jones, M. K., Gobert, G. N., Zhang, L., Sunderland, P., & McManus, D. P. (2004). The cytoskeleton and motor proteins of human schistosomes and their roles in surface maintenance and host-parasite interactions. *BioEssays*. doi: 10.1002/bies.20058

Kincaid-Smith, J., Tracey, A., Augusto, R. de C., Bulla, I., Holroyd, N., Rognon, A., … Toulza, E. (2019). Morphological and Genomic Characterisation of the Hybrid Schistosome Infecting Humans In Europe Reveals a Complex Admixture Between *Schistosoma haematobium* and *Schistosoma bovis* Parasites. *BioRxiv*. doi: 10.1101/387969

King, K. C., Stelkens, R. B., Webster, J. P., Smith, D. F., & Brockhurst, M. A. (2015). Hybridization in Parasites: Consequences for Adaptive Evolution, Pathogenesis, and Public Health in a Changing World. *PLOS Pathogens*. doi: 10.1371/journal.ppat.1005098

Kondrashov, F. A. (2012). Gene duplication as a mechanism of genomic adaptation to a changing environment. *Proceedings of the Royal Society B: Biological Sciences*. doi: 10.1098/rspb.2012.1108

Kondrashov, F. A., Rogozin, I. B., Wolf, Y. I., & Koonin, E. V. (2002). Selection in the evolution of gene duplications. *Genome Biology*.

Kück, P., & Meusemann, K. (2010). FASconCAT: Convenient handling of data matrices. *Molecular Phylogenetics and Evolution*. doi: 10.1016/j.ympev.2010.04.024

Kumar, S., Stecher, G., Li, M., Knyaz, C., & Tamura, K. (2018). MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Molecular Biology and Evolution*. doi: 10.1093/molbev/msy096

Laetsch, D. R., & Blaxter, M. L. (2017). BlobTools: Interrogation of genome assemblies. *F1000Research*. doi: 10.12688/f1000research.12232.1

Laumer, C. E., Hejnol, A., & Giribet, G. (2015). Nuclear genomic signals of the ‘microturbellarian’ roots of platyhelminth evolutionary innovation. *ELife*. doi: 10.7554/elife.05503

Li, H., & Durbin, R. (2010). Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics*. doi: 10.1093/bioinformatics/btp698

Luo, R., Liu, B., Xie, Y., Li, Z., Huang, W., Yuan, J., … Wang, J. (2012). SOAPdenovo2: An empirically improved memory-efficient short-read de novo assembler. *GigaScience*. doi: 10.1186/2047-217X-1-18

Magurran, A. E. (2005). Evolutionary ecology: the Trinidadian guppy. In *Oxford Series in Ecology and Evolution*. doi: 10.1093/acprof

Maxwell, C. S., Sepulveda, V. E., Turissini, D. A., Goldman, W. E., & Matute, D. R. (2018). Recent admixture between species of the fungal pathogen Histoplasma . *Evolution Letters*. doi: 10.1002/evl3.59

Morran, L. T., Schmidt, O. G., Gelarden, I. A., Parrish, R. C., & Lively, C. M. (2011). Running with the Red Queen: Host-parasite coevolution selects for biparental sex. *Science*. doi: 10.1126/science.1206360

O'Connell, J., Schulz-Trieglaff, O., Carlson, E., Hims, M. M., Gormley, N. A., & Cox, A. J. (2015). NxTrim: Optimized trimming of Illumina mate pair reads. *Bioinformatics*. doi: 10.1093/bioinformatics/btv057

Papkou, A., Guzella, T., Yang, W., Koepper, S., Pees, B., Schalkowski, R., … Schulenburg, H. (2019). The genomic basis of Red Queen dynamics during rapid reciprocal host–pathogen coevolution. *Proceedings of the National Academy of Sciences*. doi: 10.1073/pnas.1810402116

Pérez-Jvostov, F., Hendry, A. P., Fussmann, G. F., & Scott, M. E. (2016). An experimental test of antagonistic effects of competition and parasitism on host performance in semi-natural mesocosms. *Oikos*. doi: 10.1111/oik.02499

Petersen, T. N., Brunak, S., von Heijne, G., & Nielsen, H. (2011). SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nature Methods*. doi: 10.1038/nmeth.1701

Pfeifer, B., Wittelsbürger, U., Ramos-Onsins, S. E., & Lercher, M. J. (2014). PopGenome: An efficient swiss army knife for population genomic analyses in R. *Molecular Biology and Evolution*. doi: 10.1093/molbev/msu136

Phillips, K. P., Cable, J., Mohammed, R. S., Herdegen-Radwan, M., Raubic, J., Przesmycka, K. J., … Radwan, J. (2018). Immunogenetic novelty confers a selective advantage in host–pathogen coevolution. *Proceedings of the National Academy of Sciences*. doi: 10.1073/pnas.1708597115

Poulin, R., & Morand, S. (2004). The Diversity of Parasites. *The Quarterly Review of Biology*. doi: 10.1086/393500

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., … Sham, P. C. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *American Journal of Human Genetics*. doi: 10.1086/519795

Reynolds, M., Arapi, E. A., & Cable, J. (2018). Parasite-mediated host behavioural modifications: *Gyrodactylus turnbulli* infected Trinidadian guppies increase contact rates with uninfected conspecifics. *Parasitology*. doi: 10.1017/S0031182017001950

Robertson, S., Bradley, J. E., & MacColl, A. D. C. (2017). No evidence of local adaptation of immune responses to *Gyrodactylus* in three-spined stickleback (Gasterosteus aculeatus). *Fish and Shellfish Immunology*. doi: 10.1016/j.fsi.2016.11.058

Rogers, M. B., Downing, T., Smith, B. A., Imamura, H., Sanders, M., Svobodova, M., … Smith, D. F. (2014). Genomic Confirmation of Hybridisation and Recent Inbreeding in a Vector-Isolated *Leishmania* Population. *PLoS Genetics*. doi: 10.1371/journal.pgen.1004092

Sahlin, K., Vezzi, F., Nystedt, B., Lundeberg, J., & Arvestad, L. (2014). BESST - Efficient scaffolding of large fragmented assemblies. *BMC Bioinformatics*. doi: 10.1186/1471-2105-15-281

Salter, J. P., Choe, Y., Albrecht, H., Franklin, C., Lim, K. C., Craik, C. S., & McKerrow, J. H. (2002). Cercarial elastase is encoded by a functionally conserved gene family across multiple species of schistosomes. *Journal of Biological Chemistry*. doi: 10.1074/jbc.M202364200

Schelkle, B., Faria, P. J., Johnson, M. B., van Oosterhout, C., & Cable, J. (2012). Mixed infections and hybridisation in monogenean parasites. *PLoS ONE*. doi: 10.1371/journal.pone.0039506

Schmid-Hempel, P. (2011). The integrated study of infections, immunology, ecology and genetics. *Evolutionary Parasitology, Oxford University Press*.

Small, S. T., Reimer, L. J., Tisch, D. J., King, C. L., Christensen, B. M., Siba, P. M., … Zimmerman, P. A. (2016). Population genomics of the filarial nematode parasite *Wuchereria bancrofti* from mosquitoes. *Molecular Ecology*. doi: 10.1111/mec.13574

Stamatakis, A. (2006). RAxML-VI-HPC: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics*. doi: 10.1093/bioinformatics/btl446

Stephenson, J. F., Van Oosterhout, C., & Cable, J. (2015). Pace of life, predators and parasites: Predator-induced life-history evolution in Trinidadian guppies predicts decrease in parasite tolerance. *Biology Letters*. doi: 10.1098/rsbl.2015.0806

Tibayrenc, M., & Ayala, F. J. (2012). Reproductive clonality of pathogens: A perspective on pathogenic viruses, bacteria, fungi, and parasitic protozoa. *Proceedings of the National Academy of Sciences*. doi: 10.1073/pnas.1212452109

Tihon, E., Imamura, H., Dujardin, J. C., Van Den Abbeele, J., & Van den Broeck, F. (2017). Discovery and genomic analyses of hybridization between divergent lineages of *Trypanosoma congolense*, causative agent of Animal African Trypanosomiasis. *Molecular Ecology*. doi: 10.1111/mec.14271

Törönen, P., Medlar, A., & Holm, L. (2018). PANNZER2: A rapid functional annotation web server. *Nucleic Acids Research*. doi: 10.1093/nar/gky350

Train, C.-M., Pignatelli, M., Altenhoff, A., & Dessimoz, C. (2018). iHam and pyHam: visualizing and processing hierarchical orthologous groups. *Bioinformatics*. doi:

10.1093/bioinformatics/bty994

Van Oosterhout, C., Harris, P. D., & Cable, J. (2003). Marked variation in parasite resistance between two wild populations of the Trinidadian guppy, *Poecilia reticulata* (Pisces: *Poeciliidae*). *Biological Journal of the Linnean Society*. doi: 10.1046/j.1095-8312.2003.00203.x

Venditti, C., Meade, A., & Pagel, M. (2010). Phylogenies reveal new interpretation of speciation and the Red Queen. *Nature*. doi: 10.1038/nature08630

Volf, P., Benkova, I., Myskova, J., Sadlova, J., Campino, L., & Ravel, C. (2007). Increased transmission potential of *Leishmania major/Leishmania infantum* hybrids. *International Journal for Parasitology*. doi: 10.1016/j.ijpara.2007.02.002

Wilbers, R. H. P., Schneiter, R., Holterman, M. H. M., Drurey, C., Smant, G., Asojo, O. A., … Lozano-Torres, J. L. (2018). Secreted venom allergen-like proteins of helminths: Conserved modulators of host responses in animals and plants. *PLoS Pathogens*. doi: 10.1371/journal.ppat.1007300

Windsor, D. A. (2002). Controversies in parasitology, Most of the species on Earth are parasites. *International Journal for Parasitology*. doi: 10.1016/s0020-7519(98)00153-2

Wit, J., & Gilleard, J. S. (2017). Resequencing Helminth Genomes for Population and Genetic Studies. *Trends in Parasitology*. doi: 10.1016/j.pt.2017.01.009

Woolhouse, M. E. J., Webster, J. P., Domingo, E., Charlesworth, B., & Levin, B. R. (2002). Biological and biomedical implications of the co-evolution of pathogens and their hosts. *Nature Genetics*. doi: 10.1038/ng1202-569

Xavier, R., Faria, P. J., Paladini, G., Van Oosterhout, C., Johnson, M., & Cable, J. (2015). Evidence for cryptic speciation in directly transmitted gyrodactylid parasites of trinidadian guppies. *PLoS ONE*. doi: 10.1371/journal.pone.0117096

Yang, Z. (2007). PAML 4: Phylogenetic analysis by maximum likelihood. *Molecular Biology and Evolution*. doi: 10.1093/molbev/msm088

Zamanian, M., Kimber, M. J., McVeigh, P., Carlson, S. A., Maule, A. G., & Day, T. A. (2011). The repertoire of G protein-coupled receptors in the human parasite *Schistosoma mansoni* and the model organism *Schmidtea mediterranea*. *BMC Genomics*. doi: 10.1186/1471-2164-12-596

Data Accessibility Statement

The raw sequences will be available as FASTQ files and the final reference genome as a FASTA file in the GeneBank (BioProject accession no. PRJNA532341).

Authors Contributions

M.K. and J.R. designed research, M.K., K.J.P., R.S.M., K.P.P and S.C. collected samples; F.C., and R.G contributed new analytical tools; M.K. analyzed data with contribution from K.P.P. and C.H.; M.K. drafted the manuscript and J.R., J.C. and K.P.P. contributed to the MS writing. All authors read and approved the final manuscript.

Tables and Figures

Table 1. Genome assembly completeness (based on BUSCO Eukaryota dataset) and annotation overview

Genome assembly	
Genome size	84.40 Mb
Number of scaffolds	4,362
Longest scaffold	2.03 Mb
Scaffold N50	0.31 Mb
L50	75
Number of contigs	5,049
Contig N50	0.12 Mb
Contig L50	188
GC content	31%
Genome completeness	
Complete BUSCOs (single copy)	221 (73%)
Complete BUSCOs (duplicated)	4 (1%)
Fragmented BUSCOs	26 (7%)
Missing BUSCOs	52 (17%)
Genome annotation	
Number of genes	10749
Number of transcripts	15919
Intron GC content	24.3%
Exon GC content	37.7%
Avg. gene length	4691 bp
Avg. exon length (single exon genes)	758 bp
Avg. exon length (multiple exon genes)	270 bp
Avg. intron length	769 bp

Table 2. Secretory genes with non-synonymous divergence found within 25 kb windows with extreme values of F_{ST} ($F_{ST} > 0.98$).

#	Gene ID	Annotation/comments on homology	dxy(nsyn)	dxy(syn)
	Lopinot vs Caura			
1	Gbulla1a000092	Elastase	0.0220	0.0089
2	Gbulla1a003378	Cysteine ase inhibitor	0.0114	0.0144
3	Gbulla1a000016	Ribonuclease T2	0.0033	0
4	Gbulla1a008623	Uncharacterized protein	0.0026	0
5	Gbulla1a004344	disulfide-isomerase	0.0016	0.0102
6	Gbulla1a004942	F-actin-capping subunit	0.0014	0
7	Gbulla1a010110	Uncharacterized protein	0.0005	0
8	Gbulla1a010751	LOW QUALITY PROTEIN	0.0002	0
	Lopinot vs Santa Cruz			
-	-		-	-
	Santa Cruz vs Caura			
1	Gbulla1a000092	Elastase	0.0225	0.0094
2	Gbulla1a003378	Cysteine ase inhibitor	0.0114	0.0144
3	Gbulla1a008623	Uncharacterized protein	0.0026	0
4	Gbulla1a004344	disulfide-isomerase	0.0016	0.0102
5	Gbulla1a004942	F-actin-capping subunit	0.0014	0
6	Gbulla1a004662	Uncharacterized protein	0.0007	0.0022
7	Gbulla1a010751	LOW QUALITY PROTEIN	0.0002	0

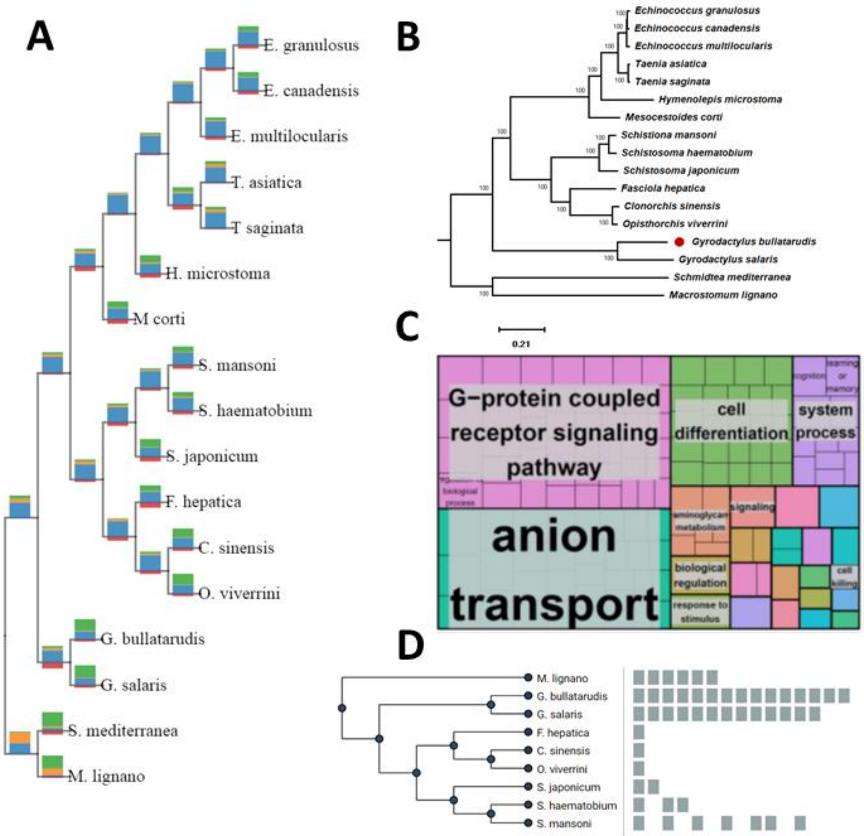

Figure 1. Phylogenetic relationship and gene duplications in the flatworms. **A**: Relative number of genes retained (blue), duplicated (yellow), gained (green) and lost (red) during flatworm evolution, as predicted by the OMA analyses. The topology is based on the phylogeny inferred with RAxML. **B**: Phylogenetic relationship calculated with RAxML based on 217,373-long amino-acid alignment built from 472 orthology groups. **C**: Biological processes (Gene Ontology terms) enriched in the orthology groups duplicated in the lineage between common ancestor of all Neodermata and *G. bullatarudis*. The list of GO terms were summarized and visualized with ReviGO software. **D**: Number of genes in the orthology group HOG01193, i.e. genes with homology to cercarial elastase genes in *Schistoma mansoni*, according to their hierarchical orhologous groups (columns).

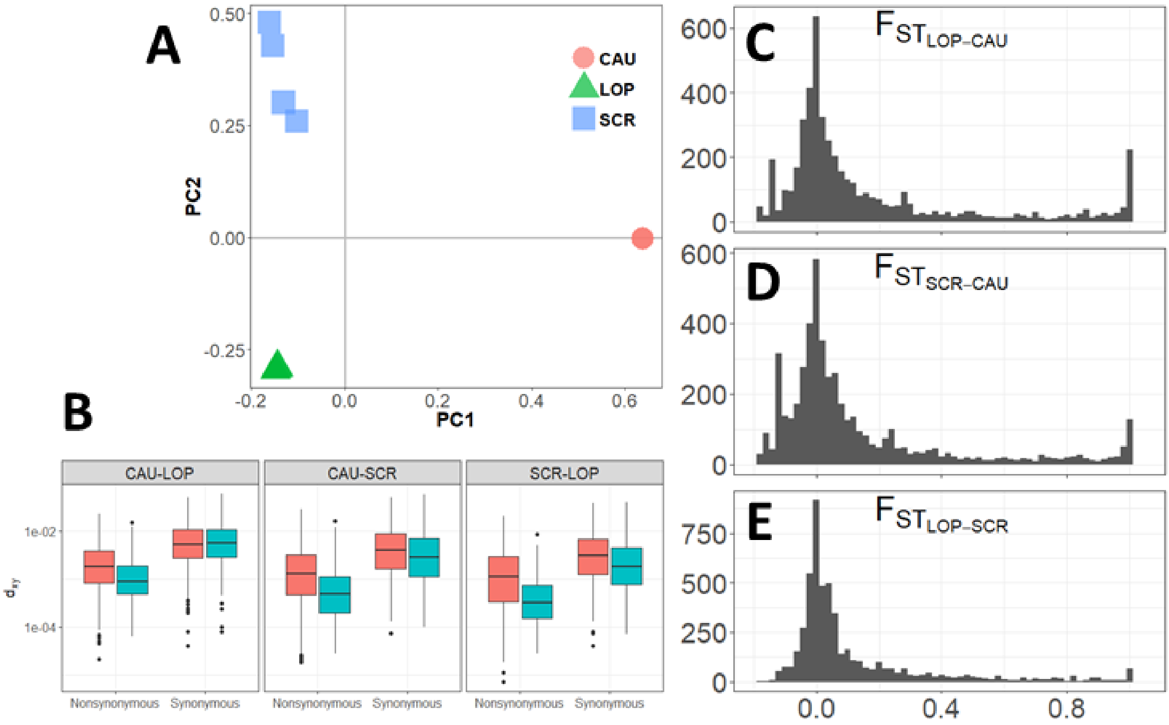

Figure 2. Genetic differentiation between local populations (Lopinot: LOP, Caura: CAU and Santa Cruz: SCR) of *Gyrodactylus bullatarudis*. **A**: Genome wide genetic differentiation between samples represented by Principal Component Analyses plot calculated based on genotypes from all identified SNPs. **B:** Per gene genetic differentiation (d_{XY}) calculated for non-synonymous and synonymous sites. Genes are divided for those for which orthologous sequences were identified in the *G. salaris* genome (green), and genes without such orthology (red). **C-E**: Histograms of Weir and Cockerham F_{ST} estimator values calculated in the 25,000 bp windows.

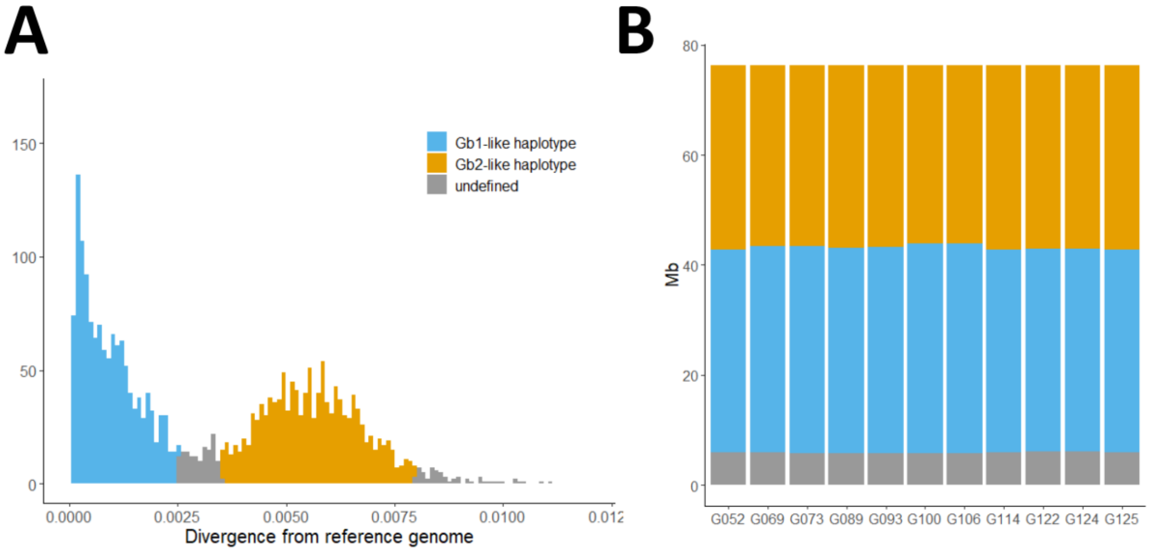

Figure 3. Recombination between two diverged lineages of *Gyrodactylus bullatarudis*. **A:** Haplotypes, defined by the divergence from the reference genome in 25 kb non-overlapping windows. Data shown for scaffolds longer than 100 kb (80% assembled genome). **B:** Fraction of genome assigned to Gb1 and Gb2 haplotypes in 25 kb windows.