















# Discovery of the *Pseudomonas* Polyne Protegencin by a Phylogeny-Guided Study of Polyne Biosynthetic Gene Cluster Diversity

 Alex J. Mullins,<sup>a</sup>
 Gordon Webster,<sup>a</sup>
 Hak Joong Kim,<sup>b</sup>
 Jinlian Zhao,<sup>c</sup>
 Yoana D. Petrova,<sup>a</sup>
 Christina E. Ramming,<sup>c</sup>
 Matthew Jenner,<sup>c,d</sup>
 James A. H. Murray,<sup>e</sup>
 Thomas R. Connor,<sup>a</sup>
 Christian Hertweck,<sup>b,f</sup>
 Gregory L. Challis,<sup>c,d,g,h</sup>
 Eshwar Mahenthiralingam<sup>a</sup>

<sup>a</sup>Microbiomes, Microbes and Informatics Group, Organisms and Environment Division, School of Biosciences, Cardiff University, Cardiff, United Kingdom

<sup>b</sup>Department of Biomolecular Chemistry, Leibniz Institute for Natural Product Research and Infection Biology, Hans Knöll Institute, Jena, Germany

<sup>c</sup>Department of Chemistry, University of Warwick, Coventry, United Kingdom

<sup>d</sup>Warwick Integrative Synthetic Biology Centre, University of Warwick, Coventry, United Kingdom

<sup>e</sup>Molecular Biosciences Division, School of Biosciences, Cardiff University, Cardiff, United Kingdom

<sup>f</sup>Faculty of Biological Sciences, Friedrich Schiller University Jena, Jena, Germany

<sup>g</sup>Department of Biochemistry and Molecular Biology, Biomedicine Discovery Institute, Monash University, Clayton, Victoria, Australia

<sup>h</sup>ARC Centre of Excellence for Innovations in Peptide and Protein Science, Monash University, Clayton, Victoria, Australia

Alex J. Mullins, Gordon Webster, and Hak Joong Kim made equal contributions to this article. The study was initiated and led by Cardiff University with Alex J. Mullins and Gordon Webster in the first and second positions, and bioinformatic analysis carried out by Alex J. Mullins in the first position. Hak Joong Kim made a parallel characterization of protegencin and demonstrated the essentiality of the conserved desaturases.

**ABSTRACT** Natural products that possess alkyne or polyne moieties have been isolated from a variety of biological sources and possess a broad range of bioactivities. In bacteria, the basic biosynthesis of polyynes is known, but their biosynthetic gene cluster (BGC) distribution and evolutionary relationship to alkyne biosynthesis have not been addressed. Through comprehensive genomic and phylogenetic analyses, the distribution of alkyne biosynthesis gene cassettes throughout bacteria was explored, revealing evidence of multiple horizontal gene transfer events. After investigation of the evolutionary connection between alkyne and polyne biosynthesis, a monophyletic clade was identified that possessed a conserved seven-gene cassette for polyne biosynthesis that built upon the conserved three-gene cassette for alkyne biosynthesis. Further diversity mapping of the conserved polyne gene cassette revealed a phylogenetic subclade for an uncharacterized polyne BGC present in several *Pseudomonas* species, designated *pgn*. Pathway mutagenesis and high-resolution analytical chemistry showed the *Pseudomonas protegens pgn* BGC directed the biosynthesis of a novel polyne, protegencin. Exploration of the biosynthetic logic behind polyne production, through BGC mutagenesis and analytical chemistry, highlighted the essentiality of a triad of desaturase proteins and a thioesterase in both the *P. protegens pgn* and *Trinickia caryophylli* (formerly *Burkholderia caryophylli*) caryonencin pathways. We have unified and expanded knowledge of polyne diversity and uniquely demonstrated that alkyne and polyne biosynthetic gene clusters are evolutionarily related and widely distributed within bacteria. The systematic mapping of conserved biosynthetic genes across the available bacterial genomic diversity proved to be a fruitful method for discovering new natural products and better understanding polyne biosynthesis.

**IMPORTANCE** Natural products bearing alkyne (triple carbon bond) or polyne (multiple alternating single and triple carbon bonds) moieties exhibit a broad range of important biological activities. Polyne metabolites have been implicated in important ecological roles such as cepacin mediating biological control of plant pathogens and

**Citation** Mullins AJ, Webster G, Kim HJ, Zhao J, Petrova YD, Ramming CE, Jenner M, Murray JAH, Connor TR, Hertweck C, Challis GL, Mahenthiralingam E. 2021. Discovery of the *Pseudomonas* polyne protegencin by a phylogeny-guided study of polyne biosynthetic gene cluster diversity. *mBio* 12: e00715-21. <https://doi.org/10.1128/mBio.00715-21>.

**Editor** Vaughn S. Cooper, University of Pittsburgh

**Copyright** © 2021 Mullins et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to Alex J. Mullins, MullinsA@cardiff.ac.uk, or Eshwar Mahenthiralingam, MahenthiralingamE@cardiff.ac.uk.

**Received** 17 March 2021

**Accepted** 30 June 2021

**Published** 3 August 2021

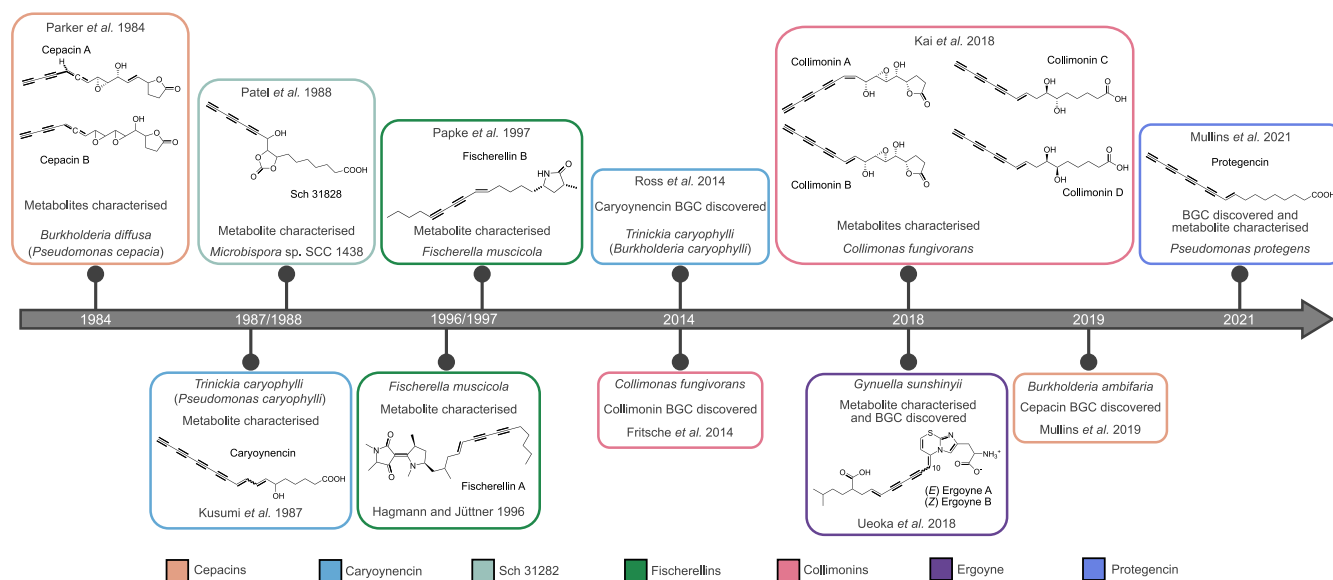
caryoynencin protecting Lagriinae beetle eggs against pathogenic fungi. After further phylogenetic exploration of polyynes diversity, we identified a novel gene cluster in *Pseudomonas* bacteria with known biological control abilities and proved it was responsible for synthesizing a new polyynes metabolite, protegencin. The evolutionary analysis of polyynes pathways showed that multiple biosynthetic genes were conserved, and using mutagenesis, their essentiality was demonstrated. Our research provides a foundation for the future modification of polyynes metabolites and has identified a novel polyynes, protegencin, with potential bioactive roles of ecological and agricultural importance.

**KEYWORDS** *Pseudomonas*, biosynthetic gene clusters, natural products, phylogenetics, polyynes

**B**acteria and fungi are an unparalleled source of structurally and functionally diverse metabolites with important applications in medicine and agriculture. Different classes of natural products can possess common structural features. One such moiety is the carbon-carbon triple (alkyne) bond. More than 65 alkyne-containing natural products have been isolated from marine bacteria and possess biotechnologically exploitable spectra of biological activity (1). Other metabolites possess elongated chains of alternating carbon-carbon single and triple bonds (polyynes). Polyynes have been isolated from diverse sources, including plants, fungi, bacteria, and even insects (2). The first bacterial polyynes, cepacins A and B, were discovered from the bacterium *Burkholderia diffusa* (formerly *Pseudomonas cepacia*) (3). However, the biosynthetic origin of the cepacins was only defined recently in the closely related species *Burkholderia ambifaria*, where these metabolites were shown to function in the biocontrol of damping off disease caused by the oomycete *Globisporangium ultimum* (4). The timeline of bacterial polyynes discovery is interesting, with multiple studies characterizing molecular diversity and different ecological roles (Fig. 1). Following the discovery of cepacins A and B in 1984 (3), several other polyynes were identified in *Proteobacteria*. Caryoynencin was isolated from *Trinickia caryophylli* (formerly *Burkholderia caryophylli*) (5) and *Burkholderia gladioli* (6). Alongside other antifungal compounds biosynthesized by *B. gladioli*, Lagriinae beetles exploit caryoynencin in a symbiotic relationship to protect their eggs from fungal attack (7). Collimonins were discovered from *Collimonas fungivorans* and displayed antifungal activity (8, 9), and ergoynes were found in the marine grass endophyte *Gynerium juncea* (10) (Fig. 1). For the polyynes Sch 31828, isolated from *Actinobacteria* (11), and fischerellins A and B, isolated from *Cyanobacteria* (12, 13), the associated biosynthetic gene clusters (BGCs) remain unknown. While alkyne (14) and polyynes (6) biosynthetic mechanisms have been investigated, the evolution of polyynes biosynthesis, its relationship to alkyne biosynthesis, and overall polyynes diversity have yet to be established.

The influx of bacterial genomic assemblies over the last decade has revolutionized our understanding of bacterial evolution and enhanced our ability to discover natural products through multiple genome mining techniques (15). Common approaches for identifying the metabolic products of novel BGCs discovered by genome mining include comparative metabolic profiling following mutagenesis of target BGCs, activation/inactivation of cluster-situated regulators, and heterologous expression (15, 16). Alternative methods fueled by the increasing availability of genomic data include analyzing the evolutionary diversity of bacteria to identify lineages talented in specialized metabolite biosynthesis (15). A second, phylogeny-based mining strategy exploits the diversity of biosynthetic genes to discover natural product derivatives of known metabolites (15). Such an approach has the advantage of gleaning insight into the horizontal transfer of genes from BGCs by comparing biosynthetic gene trees to evolutionary phylogenies.

Considering the limited insights into polyynes evolution despite evidence of an evolutionarily broad distribution (4, 11) (Fig. 1), we sought to integrate existing knowledge and expand our understanding of the distribution of these structurally intriguing moieties. Here, we show their evolutionary history, by examining the co-occurrence of



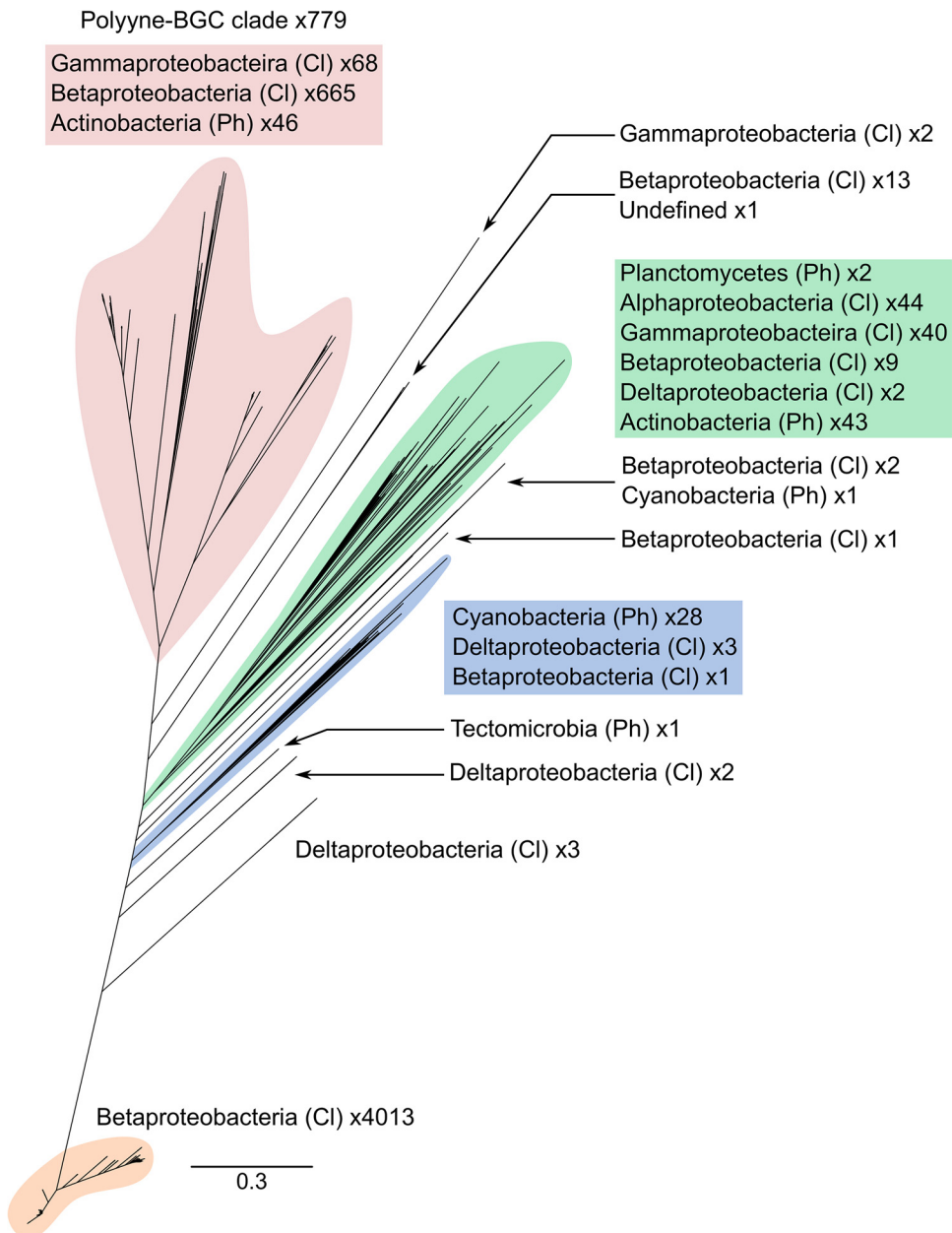
**FIG 1** Timeline highlighting the discovery of polyene metabolites and their biosynthetic gene clusters. The history of seven polyenes is displayed, highlighting the interval between the discovery of the metabolites and their BGCs.

alkyne and polyene biosynthetic cassettes (a minimum gene collection to biosynthesize a specific structural moiety), and their distribution, through a phylogeny-guided genome mining approach. This approach involved constructing a phylogeny based on genes associated with the alkyne and polyene cassettes, identifying phylogenetic clades of interest, and mining representative genomes from these clades for uncharacterized polyene biosynthetic gene clusters. Mixed evolutionary lineages within the alkyne phylogeny provided further evidence of their highly promiscuous nature. A distinct, monophyletic clade composed of polyene biosynthetic gene clusters was observed within the broader alkyne gene cassette distribution. By examining subclade architecture, we identified a previously unexplored *Pseudomonas* polyene clade that resulted in the characterization of a novel polyene BGC, *pgn*, and its associated metabolite, protegenicin.

## RESULTS

**Distribution of alkyne biosynthesis and emergence of polyene biosynthesis.** A phylogenetic tree based on 4,990 protein sequences of the alkyne biosynthetic fatty acyl-AMP ligase, *JamA*, was constructed to assess the distribution of alkyne biosynthesis in bacteria (Fig. 2). Phylogenies were also constructed based on the corresponding gene, *jamA*, alongside the protein and gene sequences of the alkyne fatty acid desaturase *JamB/jamB*, and acyl carrier protein *JamC/jamC* (see Fig. S1 in the supplemental material).

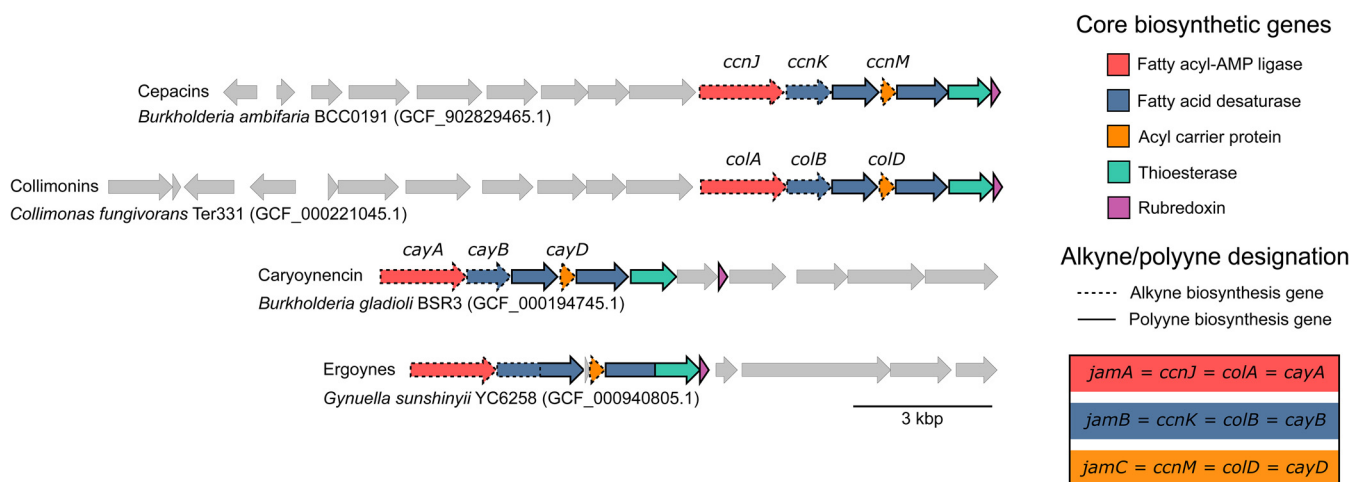
The ability to biosynthesize alkynes was widely distributed across *Proteobacteria*, occurring in the *Alpha*-, *Beta*-, *Delta*-, and *Gammaproteobacteria*, and represented 95.5% of available sequences (4,868 of 4,990). Within the *Proteobacteria*, *Betaproteobacteria* were the most dominant representatives at 96.6% (4,704 of 4,868 *Proteobacteria*) and occurred in multiple deep-branching lineages, potentially indicating several acquisition events into the phylum (Fig. 2), which is also supported by the additional phylogenies of alkyne biosynthetic genes and proteins (Fig. S1). However, the rearrangement of the branchpoints observed in the *JamABC/jamABC* protein and gene phylogenies confounds the ability to determine the number of horizontal gene transfer events that have occurred (Fig. S1). Despite these phylogenetic limitations, all six phylogenies (Fig. 2; Fig. S1) supported a similar overarching topology. Most sequences (80% [4,013 of 4,990]) occurred in a basal clade composed entirely of *Burkholderia* species, including *B. pseudomallei*, *B. thailandensis*, and *B. ubonensis* (Fig. 2), while the opposing end of the unrooted phylogeny



**FIG 2** Fatty acyl-AMP ligase protein (JamA) phylogeny of potential alkyne-synthesizing bacteria. The phylogeny was constructed with 4,990 sequences using FastTree. The composition of each clade is indicated along with the number of representatives. Ph, phylum; CI, class. Clades that contain multiple classes and phyla are highlighted with different colors.

consistently encompassed 779 sequences with a congruent topology (Fig. 2; Fig. S1). Outside of the *Proteobacteria*, examples of the alkyne cassette were found in members of the *Cyanobacteria* (29 genomes), *Planctomycetes* (2 genomes), and the candidate phylum *Tectomicrobia* uncultivated sponge symbiont "Candidatus Entotheonella" (1 genome).

Construction of the phylogeny of the biosynthetic fatty acyl-AMP ligase JamA also highlighted a discrepancy in the literature regarding the previously characterized *B. pseudomallei* alkyne biosynthetic locus (14). Inclusion of the purported JamA homologue alongside the JamA homologue identified during this analysis confirmed the latter to be the genuine JamA homologue (see Fig. S2 in the supplemental material). Annotation of the biosynthetic locus revealed the genuine fatty acyl-AMP ligase was encoded downstream of the previously characterized JamA protein (Fig. S2).

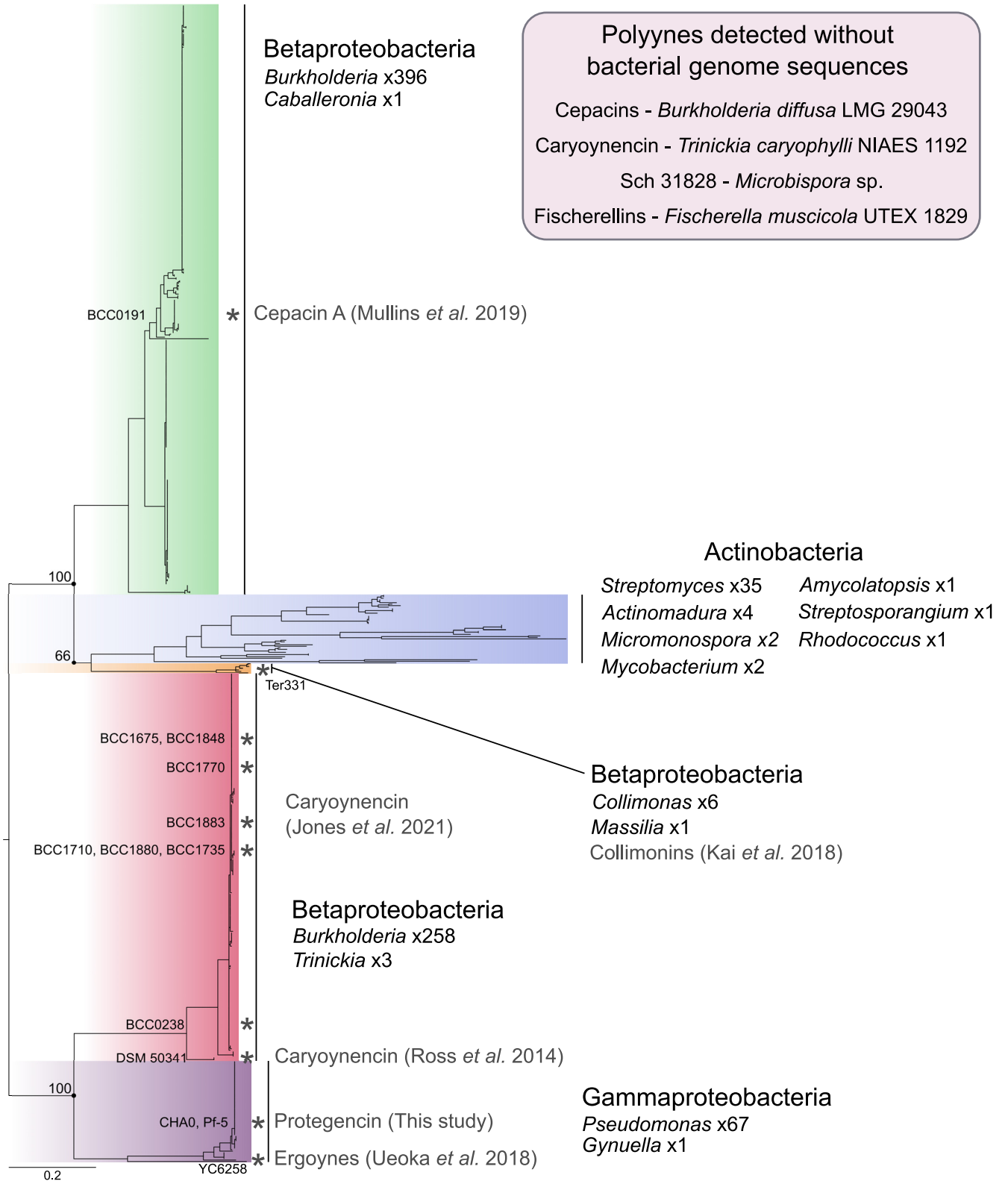


**FIG 3** Comparison of gene organization between characterized polyene biosynthetic gene clusters. Genes associated with alkyne biosynthesis are indicated by a bold outline: fatty acyl-AMP ligase, *jamA*; desaturase, *jamB*; and the acyl carrier protein, *jamC*. Genes identified as polyene biosynthesis-specific genes are indicated by a dashed outline: two further desaturase genes, a thioesterase gene, and a rubredoxin gene. Biosynthetic gene cluster (BGC)-specific nomenclature for *jamABC* homologues is included for cepacin, caryoynencin, and collimonin BGCs. The gene nomenclature of the ergoyne BGC is unavailable. The NCBI locus tags for the polyene biosynthetic gene clusters in the representative RefSeq genomes are as follows: *B. ambifaria* BCC0191 (GCF\_902829465.1), HWW27\_RS03890 to HWW27\_RS03965; *C. fungivorans* Ter331 (GFA\_000221045.1), CFU\_RS05585 to CFU\_RS05660; *B. gladioli* BSR3 (GCF\_000194745.1), BGLA\_RS09975 to BGLA\_RS10025; and *G. sunshinyii* YC6258 (GCF\_000940805.1), YC6258\_RS21350 to YC6258\_RS27625.

To understand the broader relationship between bacterial alkyne and polyene biosynthesis, a comparison of characterized polyene biosynthetic gene clusters was performed. Analysis of the gene content and architecture of four characterized/published polyene BGCs (for cepacins, collimonins, caryoynencin, and ergoynes) identified seven common genes (Fig. 3). In addition to the three genes encoding the alkyne biosynthetic cassette, *jamABC* (14), genes encoding two additional fatty acid desaturases, a thioesterase, and rubredoxin were found in all BGCs (Fig. 3). Using this knowledge, we screened DNA sequences flanking the *jamABC* alkyne biosynthetic cassettes for the presence of the remaining four genes. This revealed a monophyletic clade in the alkyne phylogenies (Fig. 2; Fig. S1) where the 779 corresponding genomes possessed the conserved polyene gene cassette (Fig. 3), with a few exceptions. Three discrepancies were observed within the monophyletic polyene clade: *B. gladioli* strain 3848s-5 and three *Streptomyces* strains appeared to lack the colocalized thioesterase and rubredoxin genes with the remaining polyene core biosynthetic genes, but manual inspection of these genomes revealed the BGCs were split across two contigs. A subset of 10 actinobacterial genomes appeared to have the thioesterase- and rubredoxin-encoding genes replaced by a gene encoding a cytochrome P450. These 10 genomes represented three genera (*Streptomyces*, *Micromonospora*, and *Amycolatopsis*) and were confined to a single subclade in the monophyletic polyene clade. The final discrepancy included two representatives of the family *Mycobacteriaceae* that lacked the rubredoxin gene.

To investigate the diversity of the monophyletic clade, a separate phylogeny was constructed based on one of the polyene-associated desaturase proteins (Fig. 4). This phylogeny was rooted using the basal branches of the clade of interest from both the *JamA* and *JamB* phylogenies (Fig. 2): a *Gammaproteobacteria* subclade and *Betaproteobacteria* subclade. Within the resulting phylogeny, we defined five major clades representing three *Betaproteobacteria* clades, one *Gammaproteobacteria* clade, and an *Actinobacteria* clade (Fig. 4). Each of the four previously characterized polyenes corresponded to a different clade, with collimonins, caryoynencin, and cepacins localized to the three distinct *Betaproteobacteria* clades (Fig. 4). The ergoynes, biosynthesized by *G. sunshinyii*, were in the *Gammaproteobacteria* clade, but with deep branching separating *G. sunshinyii* from the remainder of the clade members (Fig. 4). Each *Proteobacteria* clade was dominated by a single genus and mainly structured with relatively shallow branching. In comparison, the *Actinobacteria* clade possessed deep branching and contained representatives





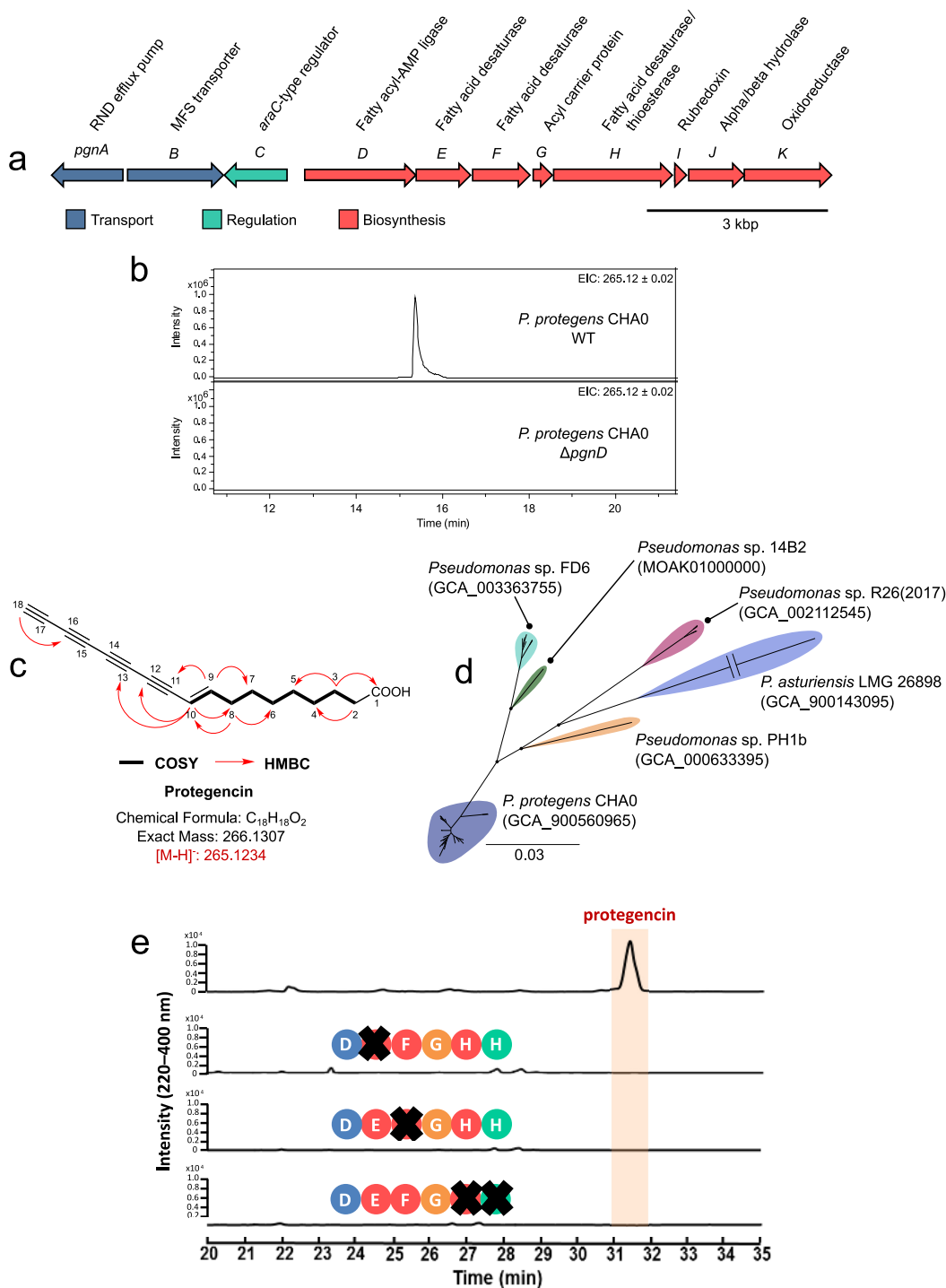
**FIG 4** Desaturase protein-based phylogeny of poly-yne-producing bacteria. Homologues of the cepacin desaturase CnnN (protegenicin PgnH) were extracted from bacterial genomes represented in the monophyletic alkyne clade as poly-yne producers. The four *Proteobacteria* clades, their composite genera, and associated poly-yne metabolites are indicated, in addition to the *Actinobacteria* phylum clade. The *Gammaproteobacteria* clade was used as the root based on the topologies of alkyne gene phylogenies. Known poly-yne producers are indicated with asterisks, and the specific strains are labeled. Bootstrap values are indicated for splits between the 5 major clades. The scale bar represents the number of substitutions per position.

of seven genera, including *Micromonospora*, *Actinomadura*, and *Rhodococcus*, but was dominated by *Streptomyces* species. This analysis identified the cepacin BGC in several species that were previously not known to carry the gene cluster (Fig. 4), including *B. contaminans*, *B. vietnamiensis* and *Caballeronia peredens*.

**Exploration of the *Gammaproteobacteria* clade reveals an uncharacterized polyynes.** Aside from the single representative of the *Gynuella* genus, the *Gammaproteobacteria* clade was dominated by *Pseudomonas*. However, this genus is not known to produce polyynes. Evidence of a *Pseudomonas* polyynes BGC has been alluded to as a homologous gene cluster of the collimonin (8) and caryoynencin (6) BGCs during the discovery of these polyynes. As such, we sought to investigate the production of an uncharacterized polyynes in *Pseudomonas* (Fig. 5a), focusing on *Pseudomonas protegens* (formerly *P. fluorescens*) strains Pf-5 and CHA0 as model systems (see Table S1 in the supplemental material). High-performance liquid chromatography (HPLC) analysis of these two strains revealed a small chromatographic peak with a characteristic UV absorbance spectrum as observed for other polyynes (6, 8). Comparative negative-ion-mode high-resolution electrospray ionization quadrupole time of flight mass spectrometry (HR-ESI-Q-TOF MS) analysis of the wild-type *P. protegens* Pf-5 and CHA0 strains and mutants with in-frame deletions in the fatty acyl-AMP ligase gene (Pf-5  $\Delta$ *pgnD* and CHA0  $\Delta$ *pgnD*, respectively) identified a compound, which we named protegencin, with the molecular formula  $C_{18}H_{18}O_2$  (Calculated for  $C_{18}H_{17}O_2^-$ : 265.1234. Found: 265.1239) as the product of the polyynes BGC (Fig. 5b and c; see Fig. S3a and b in the supplemental material).

**NMR spectroscopy confirms protegencin is a novel *Pseudomonas* polyynes.** Polyynes are notorious for being unstable and difficult to isolate, with recent studies requiring derivatization by click chemistry prior to spectroscopic analysis (6). The isolation of protegencin required careful optimization to enable spectroscopic characterization of the compound without derivatization. Purified fractions of protegencin were dried under vacuum for 2 to 3 h, with the addition of small volumes of MeCN to promote the removal of water from the sample. Freeze-drying of protegencin-containing fractions resulted in a polymerized brown oil. Using this procedure, protegencin was isolated as a brownish, amorphous powder. Its  $^1H$ ,  $^{13}C$ , correlation spectroscopy (COSY), heteronuclear single quantum coherence (HSQC), and heteronuclear multiple-bond correlation (HMBC) spectra were acquired in deuterated dimethyl sulfoxide ( $DMSO-d_6$ ) (see Table S2 and Fig. S3c to g in the supplemental material). The  $^1H$  NMR spectroscopic data displayed two olefinic protons ( $\delta_H$  6.65, 1H, dt,  $J = 16.0, 6.5$ , H-9;  $\delta_H$  5.79, 1H, d,  $J = 16.0$  H-10), a methine proton ( $\delta_H$  4.06, 1H, H-18), and seven pairs of methylene protons. The  $^{13}C$  NMR and HSQC spectroscopic data (Table S2) indicated 18 carbons, including three methine carbons ( $\delta_C$  155.4, 107.3, and 74.7), seven methylene carbons ( $\delta_C$  34.1, 33.4,  $28.9 \times 2$ , 28.8, 28.0, and 24.9), one carbonyl carbon ( $\delta_C$  175.0), and seven quaternary carbons. The above data suggested a similar polyynes structure to caryoynencin (5, 6), but lacking a pair of olefinic protons and an oxymethine proton. The structure was further established by COSY and HMBC spectroscopic data analysis (Fig. S3f and g). The HMBC correlations of H-9 with C-11, C-8, and C-7, along with the couplings of H-10 to C-9, C-11, C-12, C-8, and C-13, confirmed a double bond was located at C-9/C-10 next to the polyynes scaffold, as observed in caryoynencin. The double bond at C-7/C-8 and hydroxyl group at C-6 in caryoynencin were missing in protegencin, as evidenced by HMBC correlations from a methylene (H<sub>2</sub>-8) to two methine carbons (C-9 and C-10) and two methylene carbons (C-6 and C-7), and from a methylene (H<sub>2</sub>-4) to two methylene carbons (C-6 and C-5), as well as COSY couplings of H<sub>2</sub>-8 to H-9 and H<sub>2</sub>-7. The other COSY correlations of H<sub>2</sub>-3 to H<sub>2</sub>-4 and H<sub>2</sub>-2, and of H<sub>2</sub>-4 to H<sub>2</sub>-5, together with HMBC correlations of H-2 to C-1, C-3, and C-4, and of H-3 with C-1, C-2, C-4, and C-5, confirmed the structure of the saturated region of this metabolite. Therefore, the structure of protegencin was elucidated as a novel polyynes natural product, as shown in Fig. 5c.

**Distribution of protegencin (*pgn*) BGC within *Pseudomonas*.** Following the discovery of the previously uncharacterized polyynes metabolite protegencin, we sought to fully understand the species distribution of the *pgn* locus. The *Pseudomonas*



**FIG 5** Organization and distribution of the protegencin (*pgn*) BGC and analysis of protegencin production. (a) Organization and putative function of genes within the *pgn* BGC. (b) Extracted-ion chromatograms at  $m/z = 265.12 \pm 0.02$ , corresponding to  $[M - H]^-$  for protegencin, from LC-MS analyses of crude extracts made from agar-grown cultures of *P. protegens* CHA0 (top) and the *P. protegens* CHA0  $\Delta pgnD$  mutant (bottom). (c) Structure of protegencin, determined by a combination of high-resolution mass spectrometry and NMR spectroscopy (see Table S2 and Fig. S3a to g). (d) Core gene-based phylogeny, using 1,487 genes, of 67 *Pseudomonas* genomes carrying the *pgn* BGC. The main nodes that demarcate the *Pseudomonas* species are highlighted, and all possess bootstrap values of 100. Representative strains and genome assembly accession numbers are included for each defined species. The scale bar represents the number of substitutions per site. The *P. asturiensis* branch was shortened (indicated by a break), and as such, the scale bar does not apply. (e) HPLC chromatograms (220 to 400 nm) of *P. protegens* Pf-5 wild-type and in-frame insertional mutant cultures. Only in the presence of all three desaturase genes (*pgnE*, *pgnF*, and *pgnH*) is protegencin produced. No polyene precursors can be detected in the mutant strains.



branches of the *Gammaproteobacteria* clade represented 67 *Pseudomonas* genomes. Subsequent average nucleotide identity analysis (ANI) of these genomes indicated the presence of multiple species. Based on the established 95% species delineation threshold for ANI (17, 18), six species were identified: these included two named species, *Pseudomonas protegens* (*P. fluorescens* group) and *Pseudomonas asturiensis* (*P. syringae* group) (19), and four unnamed species. The relatedness of these two species to one another is highlighted in the core-gene-based phylogeny (Fig. 5d). *P. protegens* was the dominant species possessing the *pgn* BGC, representing approximately 75% of genomes. A wider search for genome representatives of these six species in the European Nucleotide Archive (ENA) revealed that all genomes available of these species possess the protegencin (*pgn*) BGC, except for *P. asturiensis*. Of the two available *P. asturiensis* genomes, only the type strain LMG 26898<sup>T</sup> contained the *pgn* BGC. It was absent from *Pseudomonas* sp. strain 286 (98.9% ANI to LMG 26898). The *pgn* locus is present in five out of six *Pseudomonas* species examined in this study.

**A conserved desaturase triad is essential for polyne formation.** The high conservation of the three desaturase genes and the thioesterase gene across all orthologous polyne BGCs is notable (Fig. 3). To elucidate their roles, we performed targeted gene replacements. Specifically, we individually replaced the desaturase and thioesterase genes with a kanamycin and apramycin resistance cassette in the *P. protegens pgn* and *T. caryophylli cay* BGCs, respectively (Fig. 5e; see Fig. S4 in the supplemental material). Sequence analyses indicated that pairs of desaturase genes (*pgnE/cayB* and *pgnF/cayC*) would have similar functions. The deduced product of *pgnH* is a didomain enzyme with putative desaturase and thioesterase functions that corresponds to *cayE* and *cayF*, respectively. The metabolic profiles of the mutant strains were compared by HPLC (220 to 400 nm) with those of the wild-type strains, with or without the empty pGL42a or pJET1.2/blunt vector used for mutagenesis (Fig. 5e; Fig. S4). Whereas *P. protegens* Pf-5 (with or without the empty vector) produces protegencin, in the  $\Delta pgnE$  Kan<sup>r</sup>,  $\Delta pgnF$  Kan<sup>r</sup>, and  $\Delta pgnH$  Kan<sup>r</sup> mutant strains, no polyne precursor could be identified (Fig. 5e). Deletions of the desaturase genes *cayB*, *cayC*, and *cayE* and the thioesterase gene *cayF* in *T. caryophylli* abolished the production of caryonencin. The wild type (with or without an empty vector) generates the 7E/Z-isomers of caryonencin, but the mutant strains ( $\Delta cayB$  Apr<sup>r</sup>,  $\Delta cayC$  Apr<sup>r</sup>,  $\Delta cayE$  Apr<sup>r</sup>, and  $\Delta cayF$  Apr<sup>r</sup>) produce neither polyynes nor pathway intermediates (Fig. S4). These data indicate that the three desaturases and the thioesterase synergize in the production of polyynes. Interestingly, the same multienzyme system that gives rise to a tetrayne in the protegencin and caryonencin BGCs appears to form a triyne in the collimonin pathway and a diynyl allene in the cepacin pathway (Fig. 1).

## DISCUSSION

**Highly transmissible alkyne and polyne cassettes.** Our results identify evidence of a single point of evolution of polyne biosynthesis within bacteria and demarcate its evolution from alkyne biosynthesis (Fig. 2). The basal positioning of *Proteobacteria* within the polyne phylogeny hints at a potential origin of this biosynthetic ability (Fig. 4), followed by horizontal gene transfer into *Actinobacteria* and other *Proteobacteria* classes. Additionally, the occurrence of alkyne biosynthetic genes across diverse bacterial lineages was also indicative of multiple horizontal gene transfer events. Few other fatty acid synthase-based biosynthetic capabilities appear to occur across a spectrum of bacterial lineages.

While examples of polyne biosynthesis exist across plants, fungi, and insects, they appear to have different biosynthetic origins compared to bacteria (2). In contrast to the biosynthetic mechanism for multiple carbon-carbon triple bond formation defined in this study, there is no evidence of other biosynthetic pathways evolving from an alkyne precursor biosynthetic gene cassette. Within bacteria, a separate, evolutionarily independent, mechanism exists for the biosynthesis of multiple carbon-carbon triple bonds in the form of enediynes (20). In contrast to the seven-gene cassette required for polyne biosynthesis, a minimal five-gene cassette was defined by comparing 10 biosynthetic pathways associated with production of enediyne-containing natural products (20). Mining of bacterial genomes revealed comparably fewer examples of

the enediynes gene cassette (20, 21); however, there is evidence of horizontal gene transfer across several phyla (20) similar to the alkyne and polyynes gene cassettes.

**Phylogeny-driven metabolite discovery.** Mapping the diversity of polyynes biosynthetic gene clusters through functional gene and protein phylogenies permitted the discovery of an uncharacterized *Pseudomonas* polyynes BGC, *pgn*, and metabolite, protegencin. Hotter et al. (22) have recently demonstrated that this *P. protegens* polyynes, protegencin, acts as an algicidal toxin of the green alga *Chlamydomonas reinhardtii*. In parallel to these studies characterizing protegencin, Murata et al. (23) identified the same polyynes biosynthetic gene cluster in the biocontrol strain *P. protegens* Cab57, designating the molecules produced as protegenins.

Function-based phylogenies have been exploited previously to gain insight into natural product diversity. For example, ketosynthase (KS) and condensation (C) domains have been used to identify polyketide synthase (PKS) and nonribosomal peptide synthetase (NRPS) BGCs, respectively (24). Mining for genes known to encode enzymes that biosynthesize specific structural moieties also enables discovery and comparison to other structurally related metabolites. A novel glutarimide, gladiostatin, was recently discovered in *Burkholderia gladioli* by identifying a BGC possessing genes similar to those associated with the biosynthesis of glutarimide antibiotics in *Streptomyces* species (25, 26).

The deep branching observed within the *Actinobacteria* clade of the polyynes phylogeny represents evidence of sequence divergence and may translate into structural diversity of the resulting polyynes natural products. No *Actinobacteria* polyynes has been associated with a biosynthetic gene cluster to date, and the only published *Actinobacteria* polyynes, Sch 31828, originated from a strain that lacks a genome sequence and has not been characterized at the species level, *Microbispora* sp. strain SCC 1438 (11). In *Cyanobacteria*, many alkyne-containing natural products have been characterized (1); in contrast, only two polyynes have been discovered to date (12, 13). The lack of a genome sequence for the *Fischerella muscicola* strains that produce fischerellins also impedes our mapping of their phylogenetic relationship to other polyynes biosynthetic gene clusters, and they potentially represent an uncharacterized *Cyanobacteria* clade.

**Evidence for an uncharacterized polyynes in *P. protegens*.** We identified and characterized a novel *Pseudomonas* polyynes metabolite produced by the widely studied *P. protegens* strains Pf-5 and CHA0 (Table S1). Both strains have an extensive history of biopesticidal properties (27, 28), indicative of the array of potent antimicrobial natural products biosynthesized by this species, such as the antifungal metabolites 2,4-diacetylphloroglucinol and pyoluteorin (27, 28). Previous sequence comparisons had highlighted the existence of a polyynes BGC in *P. protegens* with similarities to the caryoyencin (6) and collimonin (8) BGCs. However, homology to only the core biosynthetic region was defined in these studies (6, 8) (Fig. 2), and the metabolic product was not identified. Additionally, a transcriptomics analysis of the Gac global regulatory system highlighted a locus possessing similarities to those in *Burkholderia* (29), with a gene organization and putative gene functions like those found in the cepacin BGC (4).

Overall, we sought to understand the evolution and diversity of polyynes biosynthesis following emergence from the alkyne biosynthetic gene cassette. This study exploited functional gene phylogenetics alongside evolutionary analyses to explore polyynes biosynthetic diversity. Bioinformatics analyses supported by molecular biology and analytical chemistry led to the discovery of a *Pseudomonas*-derived polyynes BGC, *pgn*, and its metabolic product, protegencin. The conserved multienzyme system was proven to be essential for polyynes formation in both protegencin and caryoyencin biosynthesis (Fig. 5e). Discovering novel polyynes and investigating their biosynthetic mechanisms will support future endeavors to better understand these unusual biologically active metabolites.

## MATERIALS AND METHODS

**Detection of alkyne and polyynes biosynthetic gene clusters.** A BLASTp (30) search of NCBI genomes, excluding *Burkholderia* (taxid: 32008) and a local database of *Burkholderia* assemblies (3,002 downloaded genomes and 4,434 genomes assembled from publicly available Illumina read data) was performed with the cepacin homologue (CcnK) (4) of the desaturase JamB as the query. *Burkholderia* genomic assemblies were downloaded from the European Nucleotide Archive (ENA) using a script from

enaBrowserTools (<https://github.com/enasequence/enaBrowserTools>). The local assemblies constructed from publicly available Illumina paired-end fastq data were assembled via Shovill v0.9.0 (<https://github.com/tseemann/shovill>). The top 5,000 genus and species hits from NCBI were dereplicated, and their associated genomes were downloaded and combined with the local collection. The flanking 30-kbp sequence of the protein hit (E value of  $<1.00e-50$ ) was extracted, and the encoded protein domains were predicted using Interproscan v5.38-76.0 (31). Each sequence was screened for the presence of three domains corresponding to the presence of a fatty acyl-AMP ligase (IPR040097), fatty acid desaturase (IPR005804), and acyl carrier protein (IPR009081). The presence of these three homologues was considered evidence of alkyne biosynthesis potential. These sequence fragments were further screened for the presence of four additional protein homologues—two desaturases, a thioesterase, and a rubredoxin protein—via BLASTp, to determine the potential of polyynes biosynthesis. A threshold of  $1.00e-100$  was used to determine the presence of the additional desaturase proteins based on a noticeable change in E value between protein presence and absence. Manual analysis of the sequence fragments for the presence or absence of alkyne- and polyynes-associated genes was necessary to define the thioesterase and rubredoxin thresholds due to an indistinct change in E value and BGCs occurring near contig edges.

**Phylogenetic and phylogenomic analyses of alkyne and polyynes BGCs.** Protein and nucleotide alignments were generated using MAFFT v7.455 (32), with the exception of core gene alignments, which were generated with Roary v3.13.0 (33). Alkyne-related phylogenies were constructed using multi-threaded FastTree v2.1.10 with a general time-reversible model and gamma distribution for nucleotide alignments (34). The remaining phylogenies were constructed using RAxML v8.2.12 (35) with a general time-reversible model and gamma distribution supported by 100 bootstraps. In cases where the protein or gene sequence of interest occurred as a fusion, the region of interest was extracted for use in the alignment. Bacterial genomes were annotated with Prokka v1.14.5 (36). Average nucleotide identity (ANI) analyses were initially performed with fastANI v1.2 (17) and supported by PyANI v0.2.9 (mummer) (37). A comparison of the annotated sequences was visualized using Easyfig (38).

**Mutagenesis of polyynes biosynthetic gene clusters.** A range of in-frame, gene replacement, and insertional inactivation mutants were constructed in *P. protegens* and *T. caryophylli* (Table S1) to link polyynes biosynthesis to gene clusters and cassette function as described in the supplemental material.

**Metabolite extraction and LC-MS analysis of *P. protegens* wild types and  $\Delta$ pgnD mutants.** *P. protegens* wild-type strains (CHA0 and Pf-5) and mutants (CHA0  $\Delta$ pgnD and Pf-5  $\Delta$ pgnD) were grown in LB broth at 30°C overnight with agitation and then inoculated onto pea exudate medium (PEM) agar plates (see supplemental material for PEM constituents). After incubation on PEM agar at 22°C for 3 days, the medium in a single plate was cut into approximately 1- by 1- by 0.5-cm pieces after removing surface growth and extracted with 10 ml of ethyl acetate (EtOAc), submerging the agar pieces, for 2 h static with periodic agitation. The crude extract was then filtered, followed by rotary evaporation and redissolving in 1 ml of 50% acetonitrile in water. The crude extracts were then analyzed by ultrahigh-performance (UHPLC)-ESI-Q-TOF MS after centrifugation to remove debris. UHPLC-ESI-Q-TOF MS analysis was performed using a Dionex UltiMate 3000 UHPLC device connected to a Zorbax Eclipse Plus  $C_{18}$  column (100 by 2.1 mm, 1.8  $\mu$ m) coupled to a Bruker MaXis Impact mass spectrometer. The mobile phases consisted of water and acetonitrile (MeCN), each supplemented with 0.1% formic acid. After 5 min of isocratic elution at 5% MeCN, a gradient of 5 to 100% MeCN in 12 min was employed with a flow rate 0.2 ml min<sup>-1</sup>, followed by isocratic elution for a further 5 min and then returning to the initial conditions within 3 min. The mass spectrometer was operated in positive-ion or negative-ion mode with a scan range of 50 to 3,000 *m/z*. The source conditions were end-plate offset at -500 V, capillary at -4,500 V, nebulizer gas (N<sub>2</sub>) at 1.6 bars, dry gas (N<sub>2</sub>) at 81 min<sup>-1</sup>, and dry temperature at 180°C. The ion transfer conditions were ion funnel radio frequency (RF) at 200 Vpp, multiple RF at 200 Vpp, quadrupole low mass at 55 *m/z*, collision energy at 5.0 eV, collision RF at 600 Vpp, ion cooler RF at 50 to 350 Vpp, transfer time at 121  $\mu$ s, and prepulse storage time at 1  $\mu$ s. Calibration was performed with 1 mM sodium formate through a loop injection of 15  $\mu$ l at the start of each run. Additional LC-MS methods are described in the supplemental material.

**Preparative HPLC purification and structure elucidation by NMR spectroscopy.** *P. protegens* Pf-5 metabolite production was scaled up by growth on 53 PEM agar plates (1.5 liters of medium in total). After growth at 22°C for 3 days, the medium was processed as described for the LC-MS analyses. The purification was performed on an Agilent 1200 series HPLC instrument equipped with a diode array detector and an Agilent Zorbax  $C_{18}$  column (100 by 21.1 mm, 5  $\mu$ m), and the crude EtOAc extract was separated with an MeCN-H<sub>2</sub>O gradient (0 min, 5% MeCN; 5 min, 30% MeCN; 50 min, 30% MeCN; 80 min, 100% MeCN; 90 min, 100% MeCN) at a flow rate of 9 ml/min and monitoring absorbance at 260 nm. This resulted in the isolation of a putative polyynes metabolite (1.5 mg, *t<sub>R</sub>* = 76.8 min). The structure of this compound was elucidated using NMR spectroscopy. The sample was dissolved in 0.6 ml of deuterated DMSO in a Norell standard series 5-mm NMR tube, and 1D/2D spectra (<sup>1</sup>H, <sup>13</sup>C, COSY, HSQC, and HMBC) were obtained at 500 MHz for <sup>1</sup>H NMR and 125 MHz for <sup>13</sup>C NMR on a Bruker Avance III HD 500-MHz spectrometer. Chemical shifts ( $\delta$ ) are given in ppm, and coupling constants (*J*) are given in hertz (Hz). Additional HPLC methods are described in the supplemental material.

**Data availability.** All bacterial genome assemblies and Illumina reads analyzed during this study were downloaded from the National Center for Biotechnology Information (NCBI) or European Nucleotide Archive public databases.

## SUPPLEMENTAL MATERIAL

Supplemental material is available online only.

**TEXT S1**, PDF file, 0.1 MB.

**FIG S1**, PDF file, 0.2 MB.

**FIG S2**, PDF file, 0.1 MB.

**FIG S3**, PDF file, 0.5 MB.

**FIG S4**, PDF file, 0.1 MB.

**TABLE S1**, PDF file, 0.1 MB.

**TABLE S2**, PDF file, 0.1 MB.

**TABLE S3**, PDF file, 0.1 MB.

**TABLE S4**, PDF file, 0.04 MB.

## ACKNOWLEDGMENTS

A.J.M., G.W., J.Z., E.M., G.L.C., J.A.H.M., and T.R.C. acknowledge funding from the Biotechnology and Biological Sciences Research Council (BBSRC) grant references BB/S007652/1 and BB/S008020/1. A.J.M. acknowledges previous funding by the BBSRC South West Doctoral Training Partnership (BB/M009122/1). T.R.C. acknowledges funding from Medical Research Council award MR/L015080/1, which funded the Cloud Infrastructure for Microbial Bioinformatics (CLIMB) used for data analysis. The Dionex 3000RS/Bruker MaXis Impact instrument used in this work was purchased with a grant to G.L.C. from the BBSRC (BB/K002341/1). M.J. is supported by a BBSRC Discovery Fellowship (BB/R012121/1). C.H. and H.J.K. acknowledge funding by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation)—SFB 1127/2, ChemBioSys—239748522 and the Pakt für Forschung und Innovation.

We thank Gail Preston (Department of Plant Sciences, Oxford) for supplying the strain *P. protegens* Pf-5 and George O'Toole (Dartmouth College, Hanover, NH, USA) for providing the pMQ30 mutagenesis strains and constructs. We thank the School of Biosciences Genomics Research Hub at Cardiff University for genome sequencing services and project student George Mears for laboratory technical assistance during his B.Sc. (Honors) final year project.

G.L.C. is a codirector of Erebagen, Ltd. The other authors declare no competing interests.

Conceptualization, A.J.M., G.W., E.M., H.J.K., and C.H.; Data curation, A.J.M., J.Z., M.J., and H.J.K.; Formal analysis, A.J.M., G.W., Y.D.P., J.Z., M.J., and H.J.K.; Funding acquisition, E.M., G.L.C., J.A.H.M., T.R.C., and C.H.; Investigation, A.J.M., G.W., Y.D.P., J.Z., C.E.R., M.J., and H.J.K.; Methodology, A.J.M., Y.D.P., J.Z., M.J., E.M., and H.J.K.; Project administration, A.J.M., J.Z., G.L.C., and E.M.; Resources, E.M., G.L.C., J.A.H.M., and C.H.; Software, A.J.M.; Supervision, E.M., G.L.C., and C.H.; Validation, A.J.M., Y.D.P., J.Z., M.J., and H.J.K.; Visualization, A.J.M., J.Z., M.J., and H.J.K.; Writing—original draft, A.J.M., G.W., Y.D.P., J.Z., M.J., E.M., and H.J.K.; Writing—review & editing, A.J.M., E.M., G.W., J.Z., T.R.C., M.J., G.L.C., H.J.K., C.H.

## REFERENCES

- Chai QY, Yang Z, Lin HW, Han BN. 2016. Alkynyl-containing peptides of marine origin: a review. *Mar Drugs* 14:216. <https://doi.org/10.3390/md14110216>.
- Shi Shun ALK, Tykwinski RR. 2006. Synthesis of naturally occurring polyynes. *Angew Chem Int Ed Engl* 45:1034–1057. <https://doi.org/10.1002/anie.200502071>.
- Parker WL, Rathnum ML, Seiner V, Trejo WH, Principe PA, Sykes RB. 1984. Cepacin A and cepacin B, two new antibiotics produced by *Pseudomonas cepacia*. *J Antibiot (Tokyo)* 37:431–440. <https://doi.org/10.7164/antibiotics.37.431>.
- Mullins AJ, Murray JAH, Bull MJ, Jenner M, Jones C, Webster G, Green AE, Neill DR, Connor TR, Parkhill J, Challis GL, Mahenthiralingam E. 2019. Genome mining identifies cepacin as a plant-protective metabolite of the biopesticidal bacterium *Burkholderia ambifaria*. *Nat Microbiol* 4:996–1005. <https://doi.org/10.1038/s41564-019-0383-z>.
- Kusumi T, Ohtani I, Nishiyama K, Kakisawa H. 1987. Caryoynencins, potent antibiotics from a plant pathogen *Pseudomonas caryophylli*. *Tetrahedron Lett* 28:3981–3984. [https://doi.org/10.1016/S0040-4039\(00\)96437-2](https://doi.org/10.1016/S0040-4039(00)96437-2).
- Ross C, Scherlach K, Kloss F, Hertweck C. 2014. The molecular basis of conjugated polyene biosynthesis in phytopathogenic bacteria. *Angew Chem Int Ed Engl* 53:7794–7798. <https://doi.org/10.1002/anie.201403344>.
- Flórez LV, Scherlach K, Gaube P, Ross C, Sitte E, Hermes C, Rodrigues A, Hertweck C, Kaltenpoth M. 2017. Antibiotic-producing symbionts dynamically transition between plant pathogenicity and insect-defensive mutualism. *Nat Commun* 8:15172. <https://doi.org/10.1038/ncomms15172>.
- Fritsche K, van den Berg M, de Boer W, van Beek TA, Raaijmakers JM, van Veen JA, Leveau JHJ. 2014. Biosynthetic genes and activity spectrum of antifungal polyynes from *Collimonas fungivorans* Ter331. *Environ Microbiol* 16:1334–1345. <https://doi.org/10.1111/1462-2920.12440>.
- Kai K, Sogame M, Sakurai F, Nasu N, Fujita M. 2018. Collimonins A–D, unstable polyynes with antifungal or pigmentation activities from the fungus-feeding bacterium *Collimonas fungivorans* Ter331. *Org Lett* 20:3536–3540. <https://doi.org/10.1021/acs.orglett.8b01311>.
- Ueoka R, Bhushan A, Probst SI, Bray WM, Lokey RS, Linington RG, Piel J. 2018. Genome-based identification of a plant-associated marine bacterium as a rich natural product source. *Angew Chem Int Ed Engl* 57:14519–14523. <https://doi.org/10.1002/anie.201805673>.
- Patel M, Conover M, Horan A, Loebeberg D, Marquez J, Mierzwa R, Puar MS, Yarborough R, Waitz JA. 1988. Sch 31828, a novel antibiotic from a *Microbispora* sp. Taxonomy, fermentation, isolation and biological properties. *J Antibiot* 41:794–797. <https://doi.org/10.7164/antibiotics.41.794>.



12. Hagmann L, Jüttner F. 1996. Fischerellin A, a novel photosystem-II-inhibiting allelochemical of the cyanobacterium *Fischerella muscicola* with antifungal and herbicidal activity. *Tetrahedron Lett* 37:6539–6542. [https://doi.org/10.1016/0040-4039\(96\)01445-1](https://doi.org/10.1016/0040-4039(96)01445-1).
13. Papke U, Gross EM, Francke W. 1997. Isolation, identification and determination of the absolute configuration of fischerella B. A new algicide from the freshwater cyanobacterium *Fischerella muscicola* (Thuret). *Tetrahedron Lett* 38:379–382. [https://doi.org/10.1016/S0040-4039\(96\)02284-8](https://doi.org/10.1016/S0040-4039(96)02284-8).
14. Zhu X, Su M, Manickam K, Zhang W. 2015. Bacterial genome mining of enzymatic tools for alkyne biosynthesis. *ACS Chem Biol* 10:2785–2793. <https://doi.org/10.1021/acschembio.5b00641>.
15. Ziemert N, Alanjary M, Weber T. 2016. The evolution of genome mining in microbes—a review. *Nat Prod Rep* 33:988–1005. <https://doi.org/10.1039/c6np00025h>.
16. Zerikly M, Challis GL. 2009. Strategies for the discovery of new natural products by genome mining. *ChemBiochem* 10:625–633. <https://doi.org/10.1002/cbic.200800389>.
17. Jain C, Rodriguez-R LM, Phillippy AM, Konstantinidis KT, Aluru S. 2018. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nat Commun* 9:5114. <https://doi.org/10.1038/s41467-018-07641-9>.
18. Goris J, Klappenbach JA, Vandamme P, Coenye T, Konstantinidis KT, Tiedje JM. 2007. DNA–DNA hybridization values and their relationship to whole-genome sequence similarities. *Int J Syst Evol Microbiol* 57:81–91. <https://doi.org/10.1099/ijs.0.64483-0>.
19. Peix A, Ramírez-Bahena MH, Velázquez E. 2018. The current status on the taxonomy of *Pseudomonas* revisited: an update. *Infect Genet Evol* 57:106–116. <https://doi.org/10.1016/j.meegid.2017.10.026>.
20. Shen B, Hindra, Yan X, Huang T, Ge H, Yang D, Teng Q, Rudolf JD, Lohman JR. 2015. Enehydines: exploration of microbial genomics to discover new anticancer drug leads. *Bioorganic Med Chem Lett* 46. <https://doi.org/10.1002/chin.201506306>.
21. Yan X, Ge H, Huang T, Hindra, Yang D, Teng Q, Crnovčić I, Li X, Rudolf JD, Lohman JR, Gansemans Y, Zhu X, Huang Y, Zhao LX, Jiang Y, van Nieuwerburgh F, Rader C, Duan Y, Shen B. 2016. Strain prioritization and genome mining for enehydine natural products. *mBio* 7:e02104-16. <https://doi.org/10.1128/mBio.02104-16>.
22. Hotter V, Zopf D, Kim HJ, Silge A, Schmitt M, Aiyar P, Fleck J, Matthäus C, Hniopek J, Yan Q, Loper J, Sasso S, Hertweck C, Popp J, Mittag M. 2021. A polyynes toxin produced by an antagonistic bacterium blinds and lyses a green microalga. *bioRxiv* 2021.03.24.436739. <https://www.biorxiv.org/content/10.1101/2021.03.24.436739v1>.
23. Murata K, Suenaga M, Kai K. 2021. Genome mining discovery of protegenins A–D, bacterial polyynes involved in the antioomycete and biocontrol activities of *Pseudomonas protegens*. *ACS Chem Biol* <https://doi.org/10.1021/acschembio.1c00276>.
24. Ziemert N, Podell S, Penn K, Badger JH, Allen E, Jensen PR. 2012. The natural product domain seeker NaPDoS: a phylogeny based bioinformatic tool to classify secondary metabolite gene diversity. *PLoS One* 7:e34064. <https://doi.org/10.1371/journal.pone.0034064>.
25. Nakou IT, Jenner M, Dashti Y, Romero-Canelón I, Masschelein J, Mahenthalingam E, Challis GL. 2020. Genomics-driven discovery of a novel glutarimide antibiotic from *Burkholderia gladioli* reveals an unusual polyketide synthase chain release mechanism. *Angew Chem Int Ed Engl* 59:23145–23153. <https://doi.org/10.1002/anie.202009007>.
26. Niehs SP, Kumpfmüller J, Dose B, Little RF, Ishida K, Flórez LV, Kaltenpoth M, Hertweck C. 2020. Insect-associated bacteria assemble the antifungal butenolide gladiofungin by non-canonical polyketide chain termination. *Angew Chem Int Ed Engl* 59:23122–23126. <https://doi.org/10.1002/anie.202005711>.
27. Paulsen IT, Press CM, Ravel J, Kobayashi DY, Myers GSA, Mavrodi DV, DeBoy RT, Seshadri R, Ren Q, Madupu R, Dodson RJ, Durkin AS, Brinkac LM, Daugherty SC, Sullivan SA, Rosovitz MJ, Gwinn ML, Zhou L, Schneider DJ, Cartinhour SW, Nelson WC, Weidman J, Watkins K, Tran K, Khouri H, Pierson EA, Pierson LS, Thomashow LS, Loper JE. 2005. Complete genome sequence of the plant commensal *Pseudomonas fluorescens* Pf-5. *Nat Biotechnol* 23:873–878. <https://doi.org/10.1038/nbt1110>.
28. Haas D, Défago G. 2005. Biological control of soil-borne pathogens by fluorescent pseudomonads. *Nat Rev Microbiol* 3:307–319. <https://doi.org/10.1038/nrmicro1129>.
29. Hassan KA, Johnson A, Shaffer BT, Ren Q, Kidarsa TA, Elbourne LDH, Hartney S, Duboy R, Goebel NC, Zabriske TM, Paulsen IT, Loper JE. 2010. Inactivation of the GacA response regulator in *Pseudomonas fluorescens* Pf-5 has far-reaching transcriptomic consequences. *Environ Microbiol* 12:899–915. <https://doi.org/10.1111/j.1462-2920.2009.02134.x>.
30. Morgulis A, Coulouris G, Raytselis Y, Madden TL, Agarwala R, Schaffer AA, Madden TL, Ainscough R, Alexandersson M, An P. 2008. BLAST+: architecture and applications. *Bioinformatics* 24:1757–1764. <https://doi.org/10.1093/bioinformatics/btn322>.
31. Jones P, Binns D, Chang H-Y, Fraser M, Li W, McAnulla C, McWilliam H, Maslen J, Mitchell A, Nuka G, Pesseat S, Quinn AF, Sangrador-Vegas A, Scheremetjov M, Yong S-Y, Lopez R, Hunter S. 2014. InterProScan 5: genome-scale protein function classification. *Bioinformatics* 30:1236–1240. <https://doi.org/10.1093/bioinformatics/btu031>.
32. Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* 30:772–780. <https://doi.org/10.1093/molbev/mst010>.
33. Page AJ, Cummins CA, Hunt M, Wong VK, Reuter S, Holden MTG, Fookes M, Falush D, Keane JA, Parkhill J. 2015. Roary: rapid large-scale prokaryote pan genome analysis. *Bioinformatics* 31:3691–3693. <https://doi.org/10.1093/bioinformatics/btv421>.
34. Price MN, Dehal PS, Arkin AP. 2010. FastTree 2—approximately maximum-likelihood trees for large alignments. *PLoS One* 5:e9490. <https://doi.org/10.1371/journal.pone.0009490>.
35. Stamatakis A. 2014. RAXML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30:1312–1313. <https://doi.org/10.1093/bioinformatics/btu033>.
36. Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30:2068–2069. <https://doi.org/10.1093/bioinformatics/btu153>.
37. Pritchard L, Glover RH, Humphris S, Elphinstone JG, Toth IK, Thomas CM, Pallen MJ, Moiemens NS, Bamford A, Oppenheim B, Loman NJ, Foster JT, Wagner DM, Okinaka RT, Sim SH, Pearson O, Wu Z, Chang J, Kaul R, Hoffmaster AR, Brettin TS, Robison RA, Mayo M, Gee JE, Tan P, Currie BJ, Keim P, He Y, Hines HM, Ibrahim N, Jackson LM, Jaiswal P, James-Zorn C, Kohler S, Lecointre G, Lapp H, Lawrence CJ, Le Novere N, Lundberg JG, Macklin J, Mast AR, Midford PE, Miko I, Mungall CJ, Oellrich A, Osumi-Sutherland D, Parkinson H, Ramirez MJ, Richter S, Robinson PN, et al. 2016. Genomics and taxonomy in diagnostics for food security: soft-rotting enterobacterial plant pathogens. *Anal Methods* 8:12–24. <https://doi.org/10.1039/C5AY02550H>.
38. Sullivan MJ, Petty NK, Beatson SA. 2011. Easyfig: a genome comparison visualizer. *Bioinformatics* 27:1009–1010. <https://doi.org/10.1093/bioinformatics/btr039>.