Unexpected distribution of the 4-formylaminooxyvinylglycine (FVG) biosynthetic pathway in *Pseudomonas* and beyond

Edward W. Davis, II, Rachel A. Okrent, Viola A. Manning, Kristin M. Trippe

1 Center for Genome Research and Biocomputing, Oregon State University, Corvallis, OR, United States of America, 2 Forage Seed and Cereal Research Unit, United States Department of Agriculture, Agricultural Research Service, Corvallis, OR, United States of America, 3 Department of Crop and Soil Sciences, Oregon State University, Corvallis, OR, United States of America

Abstract

The biological herbicide and antibiotic 4-formylaminooxyvinylglycine (FVG) was originally isolated from several rhizosphere-associated strains of *Pseudomonas fluorescens*. Biosynthesis of FVG is dependent on the *gvg* biosynthetic gene cluster in *P. fluorescens*. In this investigation, we used comparative genomics to identify strains with the genetic potential to produce FVG due to presence of a *gvg* gene cluster. These strains primarily belong to two groups of *Pseudomonas*, *P. fluorescens* and *P. syringae*, however, a few strains with the *gvg* cluster were found outside of *Pseudomonas*. Mass spectrometry confirmed that all tested strains of the *P. fluorescens* species group produced FVG. However, *P. syringae* strains did not produce FVG under standard conditions. Several lines of evidence regarding the transmission of the *gvg* cluster including a robust phylogenetic analysis suggest that it was introduced multiple times through horizontal gene transfer within the *Pseudomonas* lineage as well as in select lineages of *Thiomonas*, *Burkholderia* and *Pantoea*. Together, these data broaden our understanding of the evolution and diversity of FVG biosynthesis. In the course of this investigation, additional gene clusters containing only a subset of the genes required to produce FVG were identified in a broad range of bacteria, including many non-pseudomonads.

Introduction

Members of the genus *Pseudomonas* are prolific producers of specialized metabolites that facilitate life in their environment. These chemicals contribute to many processes, including nutrient acquisition, host colonization, manipulation of host physiology and competition with other microbes. One example of a specialized metabolite produced by some strains of *Pseudomonas fluorescens* is 4-formylaminooxyvinylglycine (FVG, Fig 1A), a non-canonical amino acid (NCA) that is not found in the usual genetic code.
Acid [1, 2]. FVG arrests the germination of weedy grasses, and thus was originally termed a germination-arrest factor (GAF) [3]. FVG also displays activity against the bacterial plant pathogen *Erwinia amylovora*, the causal agent of fireblight [4]. As with other vinylglycines, the activity of FVG is attributed to inhibition of enzymes that utilize pyridoxal phosphate as a co-factor, such as 1-aminocyclopropane-1-carboxylate synthase in the ethylene biosynthesis pathway of plants and several enzymes involved in amino acid metabolism in bacteria [5, 6]. The relatively recent discovery of FVG contrasts with the identification of other vinylglycines of bacterial origin that occurred decades earlier, e.g. rhizobitoxine [7, 8], aminoethoxyvinylglycine (AVG) [9], and 4-methoxyvinylglycine (L-2-amino-4-methoxy-trans-3-butenoic acid) (MVG/AMB) [10].

In earlier studies, FVG production was restricted to a small number of *P. fluorescens* isolated from grass and wheat rhizosphere in the Willamette Valley of Oregon [11]. FVG production was not detected in a phylogenetically diverse group of other *P. fluorescens* strains, including *P. fluorescens* A506, PFO-1, Pf-5, and SBW25 [2, 3, 12]. The narrow host range and geographic origin of FVG-producers coupled with the relatively recent identification of FVG might suggest that it is produced by few strains in a common, but specialized, ecological niche. However, our understanding of the relationship between these strains is incomplete. Previous analyses included a limited number of strains, either due to lack of available genomic or gene sequences [13], or lack of available cultures for bioassays [3]. As the available genome sequences in the *Pseudomonas fluorescens* group has greatly expanded, along with the sequences of the genus as a whole, a more robust analysis using phylogenomic and targeted locus scrutinization is now possible [14].

The 13-kb GAF vinylglycine (gvg) biosynthetic gene cluster (Fig 1B) is essential for FVG production in *P. fluorescens* WH6 [15–17], and the presence of this cluster is correlated with the ability of other *P. fluorescens* strains to produce FVG [2]. Although the biosynthetic pathway for FVG has not yet been elucidated, the genes within the gvg cluster required for FVG production have been determined in Pfl WH6 [17]. The gvg cluster has also been reported in genomes of a few other species. For example, several genes similar to those of the gvg cluster...
were sequenced and characterized in *P. chlororaphis* (*Pch*) PCL1391, though in relation to the regulation of phenazine rather than FVG [18]. *P. syringae* (*Psy*) pv. *maculicola* ES4326, also known as *P. cannabina* (*Pca*) pv. *alislensis* ES4326, was reported to contain the gvg cluster but not to produce FVG under standard conditions [16]. Screens for novel antibiotics recently revealed that the non-pseudomonad *Pantoea ananatis* (Pan) BRT175 encodes the gvg biosynthetic gene cluster [19]. These examples suggest a more complex history of FVG biosynthesis but is not a systematic analysis.

In this study, we combine the analysis of the original collection of Willamette Valley FVG-producers with mining of publicly available sequence databases to define the phylogenetic, host, and geographic distributions of the gvg cluster. We also determine if the presence of the gvg cluster is sufficient for FVG production in distinct genetic backgrounds. The origin of the gvg cluster in various strains, through vertical descent or horizontal gene transfer, is explored. We use a robust phylogenetic analysis to test the hypothesis that the gvg locus is restricted to a monophyletic clade of *Pseudomonas* spp. In the course of the analysis, additional “orphan” biosynthetic gene clusters are identified that share similarities with the gvg cluster but are predicted to produce distinct compounds.

**Methods**

**Search for strains with gvg clusters**

The formyltransferase gene was used as a probe in two general strategies to search for sequenced strains containing the gvg biosynthetic gene cluster. The NCBI database was searched for homologs of the formyltransferases encoded by *gvgI* from *Pfl* WH6 (PFWH6_5257; WP_003176394.1) and *Pan* BRT175 (L585_04235; ERM15483.1) using BLAST+ 2.3.0 (BLASTP and TBLASTN) under default parameters [20]. Locus tags were retrieved from positive hits and used to search the NCBI Nucleotide database to confirm the presence of the other genes in the cluster. The Gene Neighborhood tool in Integrated Microbial Genome (IMG) v4 from the Joint Genome Institute was also used to search for complete gene clusters, again with the *gvgI* as a probe [21]. Comparisons of the gene content or genetic context were made using the Gene Neighborhood tool or manually based on gene annotations. Sequences were downloaded from NCBI or IMG and imported into CLC Main Workbench v. 6 (Qiagen) for additional analysis. For synteny comparisons, chromosome regions were cropped approximately 25 kb upstream and downstream of the Carbamoyltransferase (CTase) encoding gene and colored based on the synteny of genetic contexts.

Variations of the gvg cluster that may lack homologs of *gvgI*, termed gvg-like clusters, were identified using the *gvgA* (PFWH6_5249), *gvgC* (PFWH6_5251), and *gvgF* (PFWH6_5254) genes. Additional gvg amino acid sequences and accessions can be found in S2 File. The NCBI database was searched for homologs of the amino acid sequences corresponding to these genes from *Pfl* WH6 by BLASTP. Additional homologs were retrieved by Genome BLASTP in IMG. The Gene Neighborhood tool in IMG was used to facilitate identification of gene clusters.

**Phylogenetic tree construction of the *P. fluorescens* and *P. syringae* groups**

The multi-locus sequence analysis based phylogenetic tree was generated as previously described [22]. The code for the automated multi-locus sequence analysis pipeline (automlsa2) is available from the python package index (pypi) at https://pypi.org/project/automlsa2/ and builds upon autoMLSA.pl as previously reported [23]. Briefly, TBLASTN+ (v. 2.10.0) was used to identify and extract sequences for 99 of the 100 housekeeping genes as identified in Hesse et al. 2018 [14]; PA4393 was excluded from our analysis as it was unable to be routinely recovered using TBLASTN+ from all selected *Pseudomonas* spp. Included genome sequences are
shown in S1 Table. As P. aeruginosa has previously been shown to be a suitable outgroup for the P. fluorescens and P. syringae groups, three selected P. aeruginosa isolates were used as an outgroup [24]. All genes that passed filter were aligned using MAFFT (v. 7.427) [25, 26].

Phylogenetic tree construction was done using IQ-TREE2 [27–30] with a partitioned nexus file containing references to each aligned sequence as input. Model selection and partition finding was done using the ‘-m MFP—merge rclusterf’ mode in IQ-TREE2, with protein models limited with the ‘—msub nuclear’ flag. Branch supports were determined using 1000 ultra fast bootstraps, and 1000 SH-aLRT test replicates (-B 1000, -alrt 1000). Nodes with > = 95% and > = 80% supports, respectively, were considered as well supported, and are so marked on the trees. Nodes with less support are unlabeled. Trees were visualized using ggtree v2.4.1 in R v4.0.1 [31–34].

In addition, the P. fluorescens and P. gessardii subgroups were trimmed from the larger Pseudomonas phylogeny and annotated using ggtree, as above, and inkscape (https://inkscape.org/). The genes present at a particular locus, between the genes encoding a methyl transferase and single-stranded DNA binding protein (tam and ssb), were identified for these strains. Identification was performed by searching for the sequence encoded by the P. fluorescens tam gene using IMG BLASTP (v4), cross-referencing results with P. fluorescens and P. gessardii subgroups in the phylogeny, and manually examining gene annotations.

A phylogenetic tree including Pseudomonas strains encoding the gvg-gene cluster was constructed generally as above, but with Pan BRT175 as the outgroup.

### Gene diagram for gvg locus
The gvg locus in each genome was identified using BLAST with the protein sequences from WH6 used as queries against the predicted protein sequences from each of the target genomes as the databases. Subsetted genbank files (accessions listed in the figure) with 10kb regions upstream and downstream the putative gvg loci were extracted for each representative genome (20kb in total). Subset genbank files were used as input to clinker for visualization [35]. Unannotated genes were identified using TBLASTN+ v 2.10.0 using the Gvg protein sequences from WH6 as the query and each individual genome sequence as the target (S2 File; S6 Table).

### Comparison of tree topologies
One strain from each clade in the Pseudomonas phylogeny containing the gvg cluster was selected for comparison of the topologies of gvg and species trees with Pan BRT175 as the outgroup. Seven of the Gvg proteins (GvgR, GvgA, GvgC, GvgF, GvgI, and GvgJ) were used in construction of the gvg phylogeny. Tree topology tests, including the approximately unbiased (-au) and Shimodaira-Hasegawa (SH) tests were done using iqtree2 v2.1.2 with 10,000 replicates (-zb 10000) [36, 37].

### Phylogenetic tree of carbamoyltransferase proteins
Carbamoyltransferase (CTase) sequences were retrieved using BLASTP (v2.3.0) of the NCBI nt database with gvgF (PFWH6_5254) as the query sequence. Additional sequences were identified by a genome BLASTP search of the IMG database (v4.540). Sequences were confirmed to be present within clusters containing other homologs to gvg cluster genes. CTase sequences from the in-house genomes were added and all sequences were aligned using MAFFT (v. 7.244 [25]) and the percent identities calculated in CLC Main Workbench. A representative sequence of groups of sequences with greater than 95% identity was chosen for further analysis. These sequences were aligned again using MAFFT with the L-INS-i setting and trimmed using Gblocks (v. 0.91b) with half gapped positions allowed (-b5 = h setting). The LG
substitution model was selected based on the PROTGAMMAAUTO option in RAxML (v8 [38]). Bootstrap analysis was performed using the autoMRE bootstrapping criterion, resulting in 350 bootstrap replicates. The resulting phylogenetic tree was visualized and annotated in iTOL [39].

### Bacterial isolation

Bacterial strains used in this study are listed in Table 1. The strains in the ARS collection of FVG producers were isolated from wheat and grass rhizosphere in the central Willamette Valley of OR, USA as described in [11]. *Pseudomonas fluorescens* P5A was isolated from the surface of a basidiomycete fungus, *Marasmius oreades*, at the edge of a Douglas fir plantation in the Willamette Valley. The sample was transferred to sterile water (100 ml) with glass beads and shaken. The resulting suspension was serially diluted, plated on Fluorescent Pseudomonad Media (FPM) agar with cycloheximide (75 mg/L), and incubated at 28°C for 24 h. Colonies were re-streaked on FPM agar and screened for germination arrest and anti-*Erwinia* activity consistent with FVG production as in [3, 4]. Initial taxonomic identification of *P. fluorescens* P5A was performed by amplifying and sequencing 16S rDNA using standard protocols.

The sequencing, assembly and annotation of the genomes is described in Okrent et al. [43]. The contigs were ordered based on whole genome alignment with closely related complete genomes and the presence of the gvg gene cluster was confirmed.

### Preparation of culture filtrates and biological assays

Bacteria were inoculated into either 60 mL or 5 mL of *Pseudomonas* Minimal Salts Medium [3] in a 125 mL Wheaton bottle or 20 mL test tube, respectively. The cultures were grown at 28°C with shaking for either seven (60 mL cultures) or two days (5 mL cultures). Subsequently, cells were removed from the filtrate by centrifugation (3000 x g for 15 min) and filter

---

**Table 1. Strains used in this study.**

| Strain* | Origin or description | Location | Reference | Genome Reference |
|---------|----------------------|----------|-----------|------------------|
| *Pch* 30–84 | Wheat rhizosphere | WA, USA | [40] | [41] |
| Pch O6 | Soil | UT, USA | [42] | [41] |
| *Pfl* A3422A | Unknown | OR, USA | [2] | [43] |
| *Pfl* AH4 | Bluegrass rhizosphere | OR, USA | [3] | [43] |
| *Pfl* E24 | Bluegrass rhizosphere | OR, USA | [2] | [43] |
| *Pfl* G2Y | Ryegrass rhizosphere | OR, USA | [2] | [43] |
| *Pfl* P5A | Fungal surface | OR, USA | This study | [43] |
| *Pfl* WH6 | Wheat rhizosphere | OR, USA | [11] | [15] |
| *Pfl* WH6-30G | ∆gvgH mutant | N/A | [17] | N/A |
| *Pfl* WH6-31G | ∆gvgI mutant | N/A | [17] | N/A |
| *Pfl* TDH40 | Bluegrass rhizosphere | OR, USA | [13] | [43] |
| *Pss* A342 | Unknown | OR, USA | [13] | [43] |
| *Pss* BG33R | Peach rhizosphere | SC, USA | [44] | [41] |
| *Psa* BS91 | Broccoli raab | CA, USA | [45] | [46] |
| *Psy* ES4326 | Radish | WI, USA | [47] | [48] |
| *Pan* BRT175 | Strawberry | Unknown | [19] | [49] |
| *Eam* 153 | Apple | OR, USA | [50] | N/A |

*Eam*, *Erwinia amylovora*; *Pan*, *Pantoea ananatis*; *Pch*, *Pseudomonas pv. alisalensis*; *Pfl*, *Pseudomonas chlororaphis*; *Pss*, *Pseudomonas fluorescens*; *Psp*, *Pseudomonas sp.*; *Psa*, *Pseudomonas synxantha*; *Psy*, *Pseudomonas syringae*.

https://doi.org/10.1371/journal.pone.0247348.t001
sterilization (Millipore GP express Steritop, 0.22-μm pore size (EMD Millipore) or 25 mm Pall Acrodisc with 0.22 μm pore size Supor membrane (Pall Corporation), respectively. For initial laser ablation electrospray ionization mass spectrometry (LAESI-MS) experiments, culture filtrate was extracted with 90% ethanol as described previously [1].

The agar diffusion bioassays for antimicrobial activity against *Erwinia amylovora* were performed generally as described in [4], however smaller volumes were used than reported previously. Aliquots of *E. amylovora* cultures grown overnight and diluted to an OD600 of 0.2 (300 μL) were spread on a 100x100x15 mm square plate containing 35 mL of 925 agar media [4]. Four wells per plate were cut using a #2 cork borer and the plugs removed. Samples of filtrate (40 μL) were added per well and plates incubated at 28˚C for 2 days. Plates were scanned and the areas of the zones of inhibition were measured on triplicate wells for each sample using ImageJ from the National Institutes of Health [51].

**Laser ablation electrospray ionization mass spectrometry analysis**

LAESI-MS analysis was performed by Protea Biosciences, Inc. Samples of crude culture filtrate, non-inoculated filtrate, or extracted culture filtrate (20 μl) were aliquoted into individual wells of a 96-well plate and analyzed with a LAESI DP-1000 (Protea Biosciences) coupled to a Q Exactive Orbitrap mass spectrometer (Thermo Scientific) according to the methods described in [52].

Conditions were slightly altered for analysis of FVG in crude culture filtrates. Settings for the LAESI DP-1000 were electrospray (+ 4000 V, 1.0 μl/min flow, 50% methanol, 0.1% acetic acid), 100 pulses per well at 20 Hz laser repetition rate, and laser energy ~600 μJ. The samples were analyzed in selected-ion monitoring mode at a resolution of 140,000 with scans centered on the sodium adduct of FVG, [M+Na]^+ m/z 183.037. The average scan signal intensities were extracted on each well using LAESI Bridge software. Samples were analyzed by direct analysis using the LAESI-MS interface to the Q Exactive mass spectrometer without any sample preparation. Samples were analyzed in SIM mode at resolution of 140,000 for greatest sensitivity, specificity, and accuracy. Mass accuracy was maintained through analyses at < 0.5 ppm error with the sodium adduct detected. After optimization on positive and negative controls, all remaining samples were run using the same ablation and scanning conditions. Measurements were also performed on a second set of biological replicates for comparison.

**Results and discussion**

**Identification of potential FVG-producers**

The *gvg* gene cluster in *Pfl WH6* contains 12 genes (*gvgR*, *gvgA*-*gvgK*), encoding putative regulatory, transferase or transport activity, and two small ORFs encoding peptides of less than 50 amino acids (Fig 1B) [16, 17]. Three genes, *gvgD*, *gvgE* and *gvgK*, are dispensable for production of FVG [17]. In preliminary analyses, the *gvgI* gene, encoding a formyltransferase, was proposed as a specific marker for FVG biosynthesis [13]. Based on previous studies, we hypothesized that the *gvgI* gene would be found in a limited number of isolates within the three previously reported groups of *Pseudomonas* (i.e. WH6-like strains, *P. chlororaphis*, and few *P. syringae* isolates), in addition to *Pantoea ananatis* isolate known to encode the *gvg* locus.

Our strategy for prediction of additional FVG-producing strains was to mine genome databases for the *gvgI* gene and confirm its presence within a *gvg* cluster. This strategy identified over 30 additional sequenced strains containing a *gvg* cluster in NCBI and IMG databases as of December 2020. Counter to our expected result, the strains originated from a broad spectrum of geographic locations and varied taxonomic units (*S2 Table*). In general, strains encoding the *gvg* cluster were collected from rhizosphere or soil samples and belong to the genus...
Pseudomonas. However, the gvg cluster was also found in other genera including Thiomonas, Pantoea and Burkholderia, and strains from environmental or clinical samples (S2 Table).

Distribution of gvg-cluster containing strains

The distribution of strains containing the gvg cluster within the genus Pseudomonas is not readily apparent by species name alone. The Pseudomonas genus contains over 230 recognized species names (www.bacterio.net/pseudomonas.html). The P. fluorescens and P. syringae groups or species complexes are diverse and heterogeneous, and are each further divided into subgroups. Recent phylogenies have grouped P. fluorescens into 9 subgroups [24, 53] and P. syringae into at least 9 genomospecies or 11 phylogroups with multiple subgroups [54]. As a further complication, several of the Pseudomonas strains containing a gvg cluster have not been conclusively assigned to species (S2 Table). Additionally, researchers have suggested that some species designations in Pseudomonas may be incorrect based on newer tools, namely comparisons to type strains using multilocus or whole genome-based methods [14, 24, 55].

As the species labels on these isolates do not allow us to make conclusions about phylogenetic relationships between them, and in order to investigate the distribution of strains containing the gvg cluster within the genus Pseudomonas, a maximum likelihood phylogenetic tree was constructed from the amino acid sequences of 99 housekeeping genes [14, 41], focusing on groups containing strains with gvg-like clusters. The resulting phylogenetic tree, shown in Fig 2A, is largely consistent with the phylogenetic tree published in Hesse et al. 2018 [14, 24, 53, 56] and clustering of average amino acid identities from whole genomes [14, 57]. Strains encoding the gvg cluster are shown using blue or green circles at the tips for the P. fluorescens and P. syringae groups, respectively.

Despite the known limitations of accurate taxonomic designations within the Pseudomonas genus, analysis of the Pseudomonas phylogeny revealed that strains with homologous gvg clusters are present in the P. fluorescens and P. syringae groups only. Further, and contrary to our original hypothesis, the gene clusters appear in multiple subgroups within the P. fluorescens group, i.e. the P. fluorescens and P. chlororaphis subgroups, as well as Pfl FW300-N2E3, not yet placed in a subgroup. Strains containing the cluster within the P. fluorescens subgroup are not monophyletic, but instead are dispersed throughout the subgroup (Fig 2B; bold). The strains from the original Willamette Valley collection of FVG producers are found in multiple clades within the P. fluorescens subgroup (Fig 2B; bold). In contrast, gvg clusters are restricted to two paraphyletic groups in a small subclade of P. syringae (Fig 2C; bold). These strains are variously annotated as P. cannabina and P. syringae and include the model Arabidopsis pathogen Psy ES4326. The full phylogenetic tree is included as S1 File.

The gvg cluster was also found in several non-pseudomonads, including one strain of Pantoea ananatis, Thiomonas sp. and several strains of Burkholderia. In each case, the presence of a gvg-like cluster is not typical of the species. For example, a PCR-based survey of 117 strains of Pantoea failed to detect the cluster in any strain other than Pan BRT175 [19]. Similarly, of the 22 sequenced strains of Pantoea ananatis in the IMG database as of December 2020, Pan BRT175 and the closely related Pan PANS 99–36 are the only strains with the cluster. The gvg cluster is also unique to Thiomonas sp. (Tsp) CB2 of the 11 strains of Thiomonas that have been sequenced. Likewise, although there were approximately 40 sequenced genomes in the Burkholderia cepacia complex, only three contain complete gvg clusters (S2 Table).

Gene content of gvg clusters

As gvg gene clusters are present in a variety of bacterial strains, we were interested in determining whether there are variations in the gene content. To this end, the clusters were examined
for gene loss, gain or rearrangement. Rearrangement was not evident in these clusters, as genes consistently appear in the same order. However, gene loss and gain were both evident and consistent with the phylogeny, as shown in Fig 3. The three genes not essential for FVG production, gvgD, gvgE and gvgK, are absent in some clusters. While most strains within the P. fluorescens subgroup have the same set of genes as Pfl WH6, the two strains in the E24 clade, Pfl E24 and Pfl ATCC 13525, as well as Pfl FW300-N2E3, lack the partially-redundant LysE transporter gene, gvgK. The gene cluster in Pfl FW300-N2E3 lacks segments of the non-essential amidinotransferase encoded by gvgD, indicative of gene decay. The strains within the P. syringae consistently lack gvgK but contain all other genes (Fig 3).

Outside of Pseudomonas, the gvg-like clusters are more variable. Pan BRT175 lacks orthologs of gvgD, gvgE and gvgK (Fig 3 and S1 Fig). Tsp CB2 lacks gvgK only and Burkholderia
strains lack both extra LysE genes, \textit{gvgD} and \textit{gvgK} (S1 Fig). Uniquely, \textit{Thiomonas} and \textit{Burkholderia} clusters have gene sequences inserted between the \textit{gvgR} and \textit{gvgA} orthologs and between the \textit{gvgH} and \textit{gvgI} orthologs (S1 Fig). These genes encode putative amino acid transporters, a metallohydrolase, and proteins of unknown function. What effect the presence of these genes may have on the metabolites produced by these strains is unknown; they may produce a variant of FVG rather than the compound itself.

Formylaminooxyvinylglycine chemical analysis

The routine determination of an FVG phenotype typically requires an ethanol-based extraction of crude culture filtrate followed by thin-layer chromatography [1] as well as two biological assays [3, 4]. However, this protocol is not amenable to high-throughput analysis and is sometimes ambiguous. We recently developed a LAESI-MS-based method to directly detect FVG by mass in crude culture filtrates without extraction [5]. This protocol was used here to analyze a diverse selection of strains to determine if the presence of a \textit{gvg}-like cluster corresponds to the ability to produce FVG. The culture filtrates were analyzed by LAESI-MS and, using \textit{Erwinia amylovora} as an indicator strain, inhibition using the agar diffusion assay. All tested strains within the \textit{P. fluorescens} group produced FVG, including strains in the \textit{P. fluorescens} and \textit{P. chlororaphis} subgroups (Table 2). As previously reported, the non-pseudomonad...
Pan BRT175 also produced FVG (Table 2) [51]. However, neither biological activity consistent with FVG production nor the mass associated with FVG were detected from the three strains within the *P. syringae* group (Table 2). The absence of FVG production from these strains is consistent with previous analyses of *Psy* ES4326 [16, 19, 51]. We investigated multiple strains from the *P. syringae* group in the current study, due to the possibility that a mutation in the laboratory strain *Psy* ES4326 was responsible for the lack of FVG production previously observed. This was not the case, as the closely related isolates of *Pca* also did not produce FVG.

Although we tested several possible hypotheses for the lack of FVG production in *P. syringae* strains, we were not able to conclusively demonstrate why these strains do not produce FVG under standard conditions. However, the most likely explanation is related to differences in gene regulation. *P. syringae* lack the sigma factor/anti-sigma factor pair involved in regulation of FVG production in WH6 encoded by the *prtIR* locus [58]. This suggests that regulation of FVG production differs in these strains relative to *P. fluorescens*. In *P. fluorescens* WH6, transcription of the *gvg* cluster occurs from multiple promoters, with one located in the intergenic region between *gvgR* and *gvgA* as well as additional promoters internal to the cluster [17]. We thus examined the promoter regions in *P. syringae* and compared them to those of several strains in the *P. fluorescens* group. The predicted promoter sequences located between *gvgR* and *gvgA* are conserved within *P. fluorescens* group strains (including *P. chlororaphis*) but substantially different in *P. syringae* strains (S2 Fig). Therefore, *P. syringae* may require different conditions for induction of the *gvg* cluster, which may reflect the different ecological niches or nutrient acquisition strategies of the two species.

Evidence of horizontal gene transfer

The observed pattern of *gvg* clusters, present in a given phylogenetic group but not closely related strains, is consistent with horizontal gene transfer (HGT) and/or the inclusion of clusters in genomic islands. Horizontally transferred genes typically demonstrate a lack of correspondence between phylogenetic trees representing gene evolution and those representing species evolution [59]. We therefore compared the phylogeny of the *gvg* genes and the housekeeping genes for strains within *Pseudomonas*. One representative from each clade that contains a *gvg* cluster was chosen for construction of maximum likelihood phylogenetic trees for amino acid sequences encoded by *gvg* and housekeeping genes. A comparison of the species tree and a partitioned Gvg protein maximum likelihood tree is shown in Fig 4. The most striking finding is that the placement of *Psy* isolates in the Gvg phylogeny is inconsistent with the species tree. Additionally, using the approximately unbiased (AU) and Shimodaira-Hasegawa (SH) tests with 10,000 resamplings, we determined that these two phylogenies had significantly different topologies (*p* < 0.01 for both tests). While we are unable to determine the origin of these gene loci in the *Pseudomonas* spp., this pattern is indicative of two or more HGT events within and/or between *gvg*-encoding *Pseudomonas* spp.

Genomic context of the *gvg* locus

The *gvg* cluster appears in several distinct genetic contexts (Fig 3). These different genetic contexts correspond to both phylogeny and gene content of clusters, i.e. clusters in closely related strains have the same gene content and are located in the same genomic context (Fig 3). The genetic context is the same for all strains in the *P. syringae* group, while there are six observed in the *P. fluorescens* group (Fig 3). Strains that are closely related to *gvg*-cluster-containing strains but that do not contain *gvg* clusters maintain synteny of the neighboring genes. This suggests specific insertion of *gvg* clusters rather than larger genetic rearrangements. We
propose that each example of a cluster located in a distinct genetic context represents an independent insertion of the gvg cluster in that lineage.

There is one example where gvg clusters are present in the same genetic context even though the strains are in multiple clades of the phylogeny. These strains consist of members of the Pfl WH6 clade, the Psn A342 clade and Pfl1112. In all of these strains, the gvg cluster is located in the same chromosomal location, between tam and ssb genes. The tam gene encodes a methyl transferase and ssb encodes a single stranded DNA-binding protein. Remarkably, this common location of the gvg cluster is true despite the many related strains lacking the cluster (S3 Fig).

Two possible explanations for this observation are that 1) the gene cluster has been inserted in at least three lineages in the same location or 2) the common ancestor to these three clades contained the gvg cluster at this location but that it has been lost in multiple lineages. In strains without the gvg cluster at this location, alternative genes present between tam and ssb encode a LysR transcription factor, chloroperoxidase, chemotaxis protein, and various hypothetical proteins. In some strains, an integrase, transposase or restriction enzyme is also apparent. Like the presence of the gvg cluster, these variations are only partially consistent with phylogeny (S3 Fig). For example, chloroperoxidase and lysR regulatory proteins are present in this location in several disparate clades (S3 Fig). This analysis reveals that the position between the tam and ssb genes is highly variable in the P. fluorescens subgroups, perhaps indicating a hotspot for recombination or insertion.
These observations regarding HGT of the FVG biosynthetic cluster are consistent with the evidence of HGT for other specialized metabolites produced by plant-associated bacteria, including several metabolites commonly associated with *Pseudomonas* (reviewed in [60, 61]). For example, the gene clusters responsible for biosynthesis of 2,4-diacylphloroglucinol (DAPG), phenazines and pyochelin are observed in *Pseudomonas* as well as disparate, unrelated taxa [62–65]. As we observed with the *gvg* cluster, biosynthetic gene clusters frequently are present in both *Pseudomonas* and *Burkholderia* in particular, which may be explained by the presence of these bacteria in the same habitats [60].

Similarly, the vinylglycine rhizobitoxine is produced by bacteria from multiple genera, including the rhizobial *Bradyrhizobium elkanii* and the plant pathogen *Burkholderia andropogonis*. Rhizobitoxine biosynthesis (*rtx*) genes flanked by insertion sequences have been reported in *Xanthomonas oryzae* pv. *oryzae* KACC10331 [66], though rhizobitoxine production has not been reported in this strain. AMB production has only been reported in *Pseudomonas aeruginosa* and AVG in one strain of *Streptomyces* sp., though systematic analyses have not been performed for other vinylglycines.

**Characteristics of gvg-like gene clusters**

In addition to the strains containing *gvg* gene clusters, there are many strains with only some of the genes of the cluster, sometimes in combination with different genes. These variations are present in a broader range of bacteria than is the complete *gvg* cluster (S4 Table). Genera with strains containing these *gvg*-like clusters include the Alphaproteobacterium *Salinarimonas*; Betaproteobacteria *Burkholderia*, *Ralstonia* and *Thiomonas*; Gammaproteobacteria *Pseudomonas*, *Pantoea*, *Acinetobacter* and *Serratia*; and the Actinobacteria *Streptomyces* and *Saccharopolyspora* (S4 Table). These clusters contain genes in various arrangements and do not contain all of the genes shown to be necessary for FVG production [17], and thus would not be predicted to produce FVG. The core genes of these cluster variants are *gvgA*, *gvgB*, *gvgC*, and *gvgF*. The vast majority of these gene clusters are uncharacterized.

The evolutionary relationship of the *gvg* and *gvg*-like clusters was examined by constructing a phylogeny of the carbamoyltransferase (CTase) encoded by *gvgF*, which is common to the majority of the gene clusters, including the complete *gvg* clusters discussed above. The CTases from gene clusters with the same gene content tend to be in the same clades, regardless of taxonomy (Fig 5). Additionally, the CTases from the complete *gvg* clusters form a separate clade from those in *gvg*-like clusters, irrespective of taxonomy. (The exception is CTases encoded by truncated clusters from the *Burkholderia mimosarum* strains, which are also found in this clade, suggesting a relatively recent origin from a complete *gvg* cluster.) (Fig 5).

At least eight cluster variations are present (Fig 5), leading to the hypothesis that each contributes to the production of a unique compound, perhaps an amino acid or similar product. One variant is located within a known cluster for biosynthesis of the macrolide antibiotics aldgamycin and chalcomycin in *Streptomyces* sp. HK2006-1 [67]. This cluster variation contains homologs of *gvgA*, *gvgC*, *gvgF* and *gvgG*, annotated as *almUII*, *almULIII*, *almDIV* and *almUVI*, respectively. However, exactly how these genes may function in the synthesis of the cluster products is unknown. These *gvg*-like loci are promising candidates for future molecular characterization. An important question is whether these gene clusters lead to the production of as yet undiscovered compounds or to previously identified compounds for which the biosynthetic genes are not yet known.

The presence of subsets of core biosynthetic genes in unrelated strains has also been observed for the phenazine *phz* and DAPG *phl* clusters encoded in other *Pseudomonas* spp. [49, 50]. This recycling of biosynthetic sub-clusters into new clusters may be a key method for evolution of biosynthetic diversity [68, 69].
Historically, secondary metabolites were identified in single organisms or groups of organisms. The range of organisms capable of producing the metabolites were not known. With the advent of genome sequencing and the linking of genes clusters to metabolites, it is now possible to obtain a deeper understanding of their distribution and abundance.

**Conclusion**

This is a maximum likelihood phylogenetic tree of the carbamoyltransferase encoded by gvg clusters and other gene clusters. Clades are colored and labeled based on the genes present in the gene cluster. The red A clade corresponds to a complete gvg cluster with all of the genes present for production of FVG. In the key, the genes present in each cluster are listed, with parentheses around genes that are optional and curly braces indicating insertions. Sequence accessions are listed in S5 Table.

https://doi.org/10.1371/journal.pone.0247348.g005
Analysis of sequence databases reveals that the gvg cluster is widespread within the genus *Pseudomonas*. It appears in strains commonly studied and utilized for biological control of a number of pathogens/pests, including all sequenced strains of *P. chlororaphis*. FVG was produced and excreted by both strains of *P. chlororaphis* tested and its activity may be complementary to other useful properties of these strains. Production of FVG is not a defining characteristic of the *P. fluorescens* group, as only some of the strains are capable of producing the compound. The gvg cluster is not restricted to single clade or subgroup of the *P. fluorescens* group, but instead is dispersed throughout the group in multiple subgroups. This varied pattern indicates that FVG provides a function that is beneficial in specific ecological contexts and niches but may be a burden in others, leading to loss of the gvg locus even in closely related isolates.

In contrast, although the gvg cluster is present in several pathogenic strains of the *P. syringae* group, FVG was not produced under the conditions tested for three of the strains. Additional cues for the induction of FVG production may be required in these strains. Experiments to understand different modes of regulation of FVG production in these strains are ongoing. However, FVG production is not restricted to *Pseudomonas*. The strain *Pan* BRT175 was confirmed to produce FVG. In this case, FVG biosynthesis must be embedded in different regulatory networks than in *Pseudomonas*. The unexpected, diverse pattern of gvg encoding strains indicates that FVG and related compounds may have additional benefits and outcomes that have yet to be elucidated.

The pattern of distribution of gvg cluster, non-congruence between gene and species trees, and variation in genetic context suggest that the cluster has been horizontally transferred multiple times. In some cases, indicators of integration consistent with a “classic” genomic island, such as an integrase, transposase or tRNA genes, are located adjacent to the cluster. However, the lack of these indicators in many strains may be reflective of an ancient integration event, for which those signs are not typically found. The frequency of which the cluster appears in unrelated strains suggest a useful function for a variety of bacteria living in a variety of plant or soil-related environments.

Finally, gvg-like gene clusters containing some of the genes of the gvg clusters are found in a wide range of bacteria. Study of these gene clusters may help us to understand the biosynthetic pathway of FVG as well as discover novel compounds or result in new sources of known compounds.

**Supporting information**

**S1 Fig. The gvg gene cluster maintains co-linearity, including outside of the *Pseudomonas* genus.** Comparison of gvg gene clusters in *Pseudomonas fluorescens* WH6, *Pantoea ananatis* BRT175, *Thiomonas* sp. CB2 and *Burkholderia cenocepacia*. Gene arrows are colored by function and are to scale. Gray arrows indicate genes not present in gvg clusters from *P. fluorescens*. (TIF)

**S2 Fig. Promoter regions are different between *Pseudomonas fluorescens* and *P. syringae*.** Comparison of the promoters in the intergenic region between gvgR and gvgA of the gvg clusters from *Pseudomonas fluorescens*, *P. chlororaphis* and *P. syringae* strains. The locations of the -35 and -10 promoter sites for *P. fluorescens* WH6 are indicated by gray arrows above the sequences. (TIF)

**S3 Fig. The gvg locus inserted in a putative hotspot for horizontal gene transfer in some isolates.** This is a subset of Fig 2 focusing on the *Pseudomonas fluorescens* subgroup. The
putative identity of genes present between the *tam* and *ssb* genes in a given strain are indicated by color (red = gvg locus; turquoise = *lysR* protein; green = chloroperoxidase protein; orange = chemotaxis protein; purple = other genes; gray = no insert). Strains containing a gvg locus are shown in bold.

**S1 Table.** Strains used to construct multi-locus sequence phylogeny of pseudomonas.
(XLSX)

**S2 Table.** Strains containing gvg biosynthetic gene cluster.
(XLSX)

**S3 Table.** LAESI-MS ablation values for FVG (m/z 183.037).
(XLSX)

**S4 Table.** Bacterial strains with gvg-like clusters.
(XLSX)

**S5 Table.** Carboxymethyltransferase protein sequences used to construct phylogeny.
(XLSX)

**S6 Table.** Identification of unannotated gvg genes using tblastn.
(XLSX)

**S1 File.** Full maximum likelihood phylogenetic tree.
(NWK)

**S2 File.** FASTA format amino acid sequences & accession numbers for Gvg proteins.
(FAA)

**Acknowledgments**

We thank Anne Anderson, Carolee Bull, Jeff Chang, Daniel Kluepfel, Joyce Loper, Leland Pearson, John Stavrinides, and Pat Wechter for generously providing bacterial strains. Strain P5A was isolated during a classroom-based research experience by students in the class of Alice Eldridge at Philomath High School. The LAESI-MS analysis was performed by Callee Walsh and Gregory Boyce at Protea Biosciences, Inc.

**Author Contributions**

**Conceptualization:** Rachel A. Okrent, Viola A. Manning, Kristin M. Trippe.

**Data curation:** Edward W. Davis, II, Rachel A. Okrent, Viola A. Manning.

**Formal analysis:** Edward W. Davis, II, Rachel A. Okrent, Viola A. Manning.

**Funding acquisition:** Rachel A. Okrent, Kristin M. Trippe.

**Investigation:** Edward W. Davis, II, Rachel A. Okrent.

**Methodology:** Edward W. Davis, II, Rachel A. Okrent, Viola A. Manning.

**Project administration:** Kristin M. Trippe.

**Resources:** Rachel A. Okrent, Kristin M. Trippe.

**Software:** Edward W. Davis, II.

**Visualization:** Viola A. Manning, Kristin M. Trippe.
Writing – original draft: Edward W. Davis, II, Rachel A. Okrent, Viola A. Manning, Kristin M. Trippe.

Writing – review & editing: Edward W. Davis, II, Rachel A. Okrent, Viola A. Manning, Kristin M. Trippe.

References

1. Armstrong D, Azevedo M, Mills D, Bailey B, Russell B, Groenig A, et al. Germination-Arrest Factor (GAF): 3. Determination that the herbicidal activity of GAF is associated with a ninhydrin-reactive compound and counteracted by selected amino acids. Biol Control. 2009; 51: 181–190. https://doi.org/10.10121/wp1004856 PMID: 20979386

2. McPhail KL, Armstrong DJ, Azevedo MD, Banowetz GM, Mills DI. 4-Formylaminoxy vinylglycine, an herbicidal germination-arrest factor from Pseudomonas rhizosphere bacteria. J Nat Prod. 2010; 73: 1853–1857. https://doi.org/10.1021/np1004856 PMID: 20979386

3. Banowetz GM, Azevedo MD, Armstrong DJ, Halgren AB, Mills DI. Germination-Arrest Factor (GAF): biological properties of a novel, naturally-occurring herbicide produced by selected isolates of rhizosphere bacteria. Biol Control. 2008; 46: 380–390.

4. Halgren A, Azevedo M, Mills D, Armstrong D, Thimmaiah M, McPhail K, et al. Selective inhibition of Erwinia amylovora by the herbicidally active germination-arrest factor (GAF) produced by Pseudomonas bacteria. J Appl Microbiol. 2011; 111: 949–959. https://doi.org/10.1111/j.1365-2672.2011.05098.x PMID: 21726360

5. Berkowitz DB, Charette BD, Karukurichi KR, McFadden JM. α-Vinylic amino acids: occurrence, asymmetric synthesis, and biochemical mechanisms. Tetrahedron: Asymmetry. 2006; 17: 869–882.

6. Percudani R, Peracchi A. A genomic overview of pyridoxal-phosphate-dependent enzymes. EMBO Rep. 2003; 4: 850–854. https://doi.org/10.1038/sj.embor.embor914 PMID: 12949584

7. Owens LD, Wright DA. Production of the soybean-chlorosistoxin by Rhizobium japonicum in pure culture. Plant Physiol. 1965; 40: 931–933. https://doi.org/10.1104/pp.40.5.931 PMID: 16656177

8. Owens LD, Thompson TF, Pitcher R, Williams T. Structure of rhizobitoxine, an antimitabolite enol-ether amino-acid from Rhizobium japonicum. J Chem Soc, Chem Commun. 1972; 714.

9. Pruess DL, Scannell JP, Kellett M, Ax HA, Janecek J, Williams TH, et al. Antimetabolites produced by microorganisms. X. L-2-amino-4-(2-aminoethoxy)-trans-3-butenonic acid. J Antibiot (Tokyo). 1974; 27: 229–33. https://doi.org/10.7164/antibiotics.27.229 PMID: 4850608

10. Scannell JP, Pruess D, Blount JF, Ax HA, Kellett M, Weiss F, et al. Antimetabolites produced by microorganisms. XII. (S)-alanyl-3-[α-(S)-chloro-3-(S)-hydroxy-2-oxo-3-azetidinylmethyl]-(S)-alanine, a new beta-lactam containing natural product. J Antibiot (Tokyo). 1975; 28: 1–6.

11. Elliott LF, Azevedo MD, Mueller-Warrant GW, Horwath WR. Weed control with rhizobacteria. Soil Sci Agrochem Ecol. 1998; 33: 3–7.

12. Trippe K, McPhail K, Armstrong D, Azevedo M, Banowetz G. Pseudomonas fluorescens SBW25 produces furanomycin, a non-proteinogenic amino acid with selective antimicrobial properties. BMC Microbiol. 2013; 13: 1–10. https://doi.org/10.1186/1471-2180-13-1 PMID: 23286760

13. Armstrong DJ, Azevedo MD, Mcphail KL, Jackson MD, Mills DI, Banowetz G, et al. Control of grassy weeds with vinylglycines and vinylglycine-producing organisms. Google Patents; 2010.

14. Hesse C, Schulz F, Bull CT, Shaffer BT, Yan Q, Shapiro N, et al. Genome-based evolutionary history of Pseudomonas spp. Environ Microbiol. 2018; 20: 2142–2159. https://doi.org/10.1111/1462-2920.14130 PMID: 29633519

15. Kimbrel JA, Givan SA, Halgren AB, Creason AL, Mills DI, Banowetz GM, et al. An improved, high-quality draft genome sequence of the Germination-Arrest Factor-producing Pseudomonas fluorescens WH6. BMC Genomics. 2010; 11: 522. https://doi.org/10.1186/1471-2164-11-522 PMID: 20920191

16. Halgren A, Maselko M, Azevedo M, Mills D, Armstrong D, Banowetz G. Genetics of germination-arrest factor (GAF) production by Pseudomonas fluorescens WH6: identification of a gene cluster essential for GAF biosynthesis. Microbiology, 2013; 159: 36–45. https://doi.org/10.1099/mic.0.062166-0 PMID: 23125119

17. Okrent RARA Trippe KMKM, Maselko M, Manning V. Functional analysis of a biosynthetic cluster essential for production of 4-formylaminooxyvinylglycine, a germination-arrest factor from Pseudomonas fluorescens WH6. Microbiology. 2017; 163: 207–217. https://doi.org/10.1099/mic.0.000418 PMID: 28270265

18. van Rij ET. Environmental and molecular regulation of phenazine-1-carboxamide biosynthesis in Pseudomonas chlororaphis strain PCL1391. 2006.
19. Walterson AM, Smith DDN, Stavrinides J. Identification of a Pantoea biosynthetic cluster that directs the synthesis of an antimicrobial natural product. PLoS One. 2014; 9: e96208. https://doi.org/10.1371/journal.pone.0096208 PMID: 24796857

20. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, et al. BLAST+: architecture and applications. BMC Bioinformatics. 2009; 10: 1–9. https://doi.org/10.1186/1471-2105-10-1 PMID: 19118496

21. Chen I-MA, Chu K, Palaniappan K, Ratner A, Huang J, Huntemann M, et al. The IMG/M data management and analysis system v. 6.0: new tools and advanced capabilities. Nucleic Acids Res. 2021; 49: D751–D763. https://doi.org/10.1093/nar/gkaa939 PMID: 33119741

22. Creason AL, Davis EW, Putnam ML, Vandeputte OM, Chang JH. Use of whole genome sequences to develop a molecular phylogenetic framework for Rhodococcus fascians and the Rhodococcus genus. Front Plant Sci. 2014; 5: 406. https://doi.org/10.3389/fpls.2014.00406 PMID: 25237311

23. Davis E, II, Weisberg A, Tabima J, Gruenwald N, Chang J. Gall-ID: tools for genotyping gall-causing phytopathogenic bacteria. PeerJ. 2016;e2222. https://doi.org/10.7717/peerj.2222 PMID: 27547538

24. Gomila M, Peña A, Mulet M, Lalucat J, García-Valdés E. Phylogenomics and systematics in Pseudomonas. Front Microbiol. 2015; 6: 214. https://doi.org/10.3389/fmicb.2015.00214 PMID: 26074881

25. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 2013; 30: 772–780. https://doi.org/10.1093/molbev/mst010 PMID: 23329690

26. Castresana J. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol Biol Evol. 2000; 17: 540–552. https://doi.org/10.1093/oxfordjournals.molbev.a026334 PMID: 10742046

27. Kalyaanamorthy S, Minh BQ, Wong TKF, Von Haeseler A, Jermiin LS. ModelFinder: fast model selection for accurate phylogenetic estimates. Nat Methods. 2017; 14: 587–589. https://doi.org/10.1038/nmeth.4285 PMID: 28481363

28. Nguyen L-T, Schmidt HA, Von Haeseler A, Minh BQ. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol. 2015; 32: 268–274. https://doi.org/10.1093/molbev/msu300 PMID: 25371430

29. Hoang DT, Chennorom O, Von Haeseler A, Minh BQ, Vinh LS. UFBoot2: improving the ultrafast bootstrap approximation. Mol Biol Evol. 2018; 35: 518–522. https://doi.org/10.1093/molbev/msy281 PMID: 29077904

30. Chennorom O, Von Haeseler A, Minh BQ. Terrace aware data structure for phylogenomic inference from supermatrices. Syst Biol. 2016; 65: 997–1008. https://doi.org/10.1093/sysbio/syw037 PMID: 27121966

31. Yu G. Using ggtree to visualize data on tree-like structures. Curr Protoc Bioinforma. 2020; 69: e96. https://doi.org/10.1002/cpbi.96 PMID: 32162851

32. Yu G, Lam TT-Y, Zhu H, Guan Y. Two methods for mapping and visualizing associated data on phylogeny using ggtree. Mol Biol Evol. 2018; 35: 3041–3043. https://doi.org/10.1093/molbev/msy194 PMID: 30351396

33. Yu G, Smith DK, Zhu H, Guan Y, Lam TT. ggtree: an R package for visualization and annotation of phylogenetic trees with their covariates and other associated data. Methods Ecol Evol. 2017; 8: 28–36.

34. Wang L-G, Lam TT-Y, Xu S, Dai Z, Zhou L, Feng T, et al. Treeio: an R package for phylogenetic tree input and output with richly annotated and associated data. Mol Biol Evol. 2020; 37: 599–603. https://doi.org/10.1093/molbev/msz240 PMID: 31633786

35. Gilchrist CLM, Chooi Y-HH. clinker & clustermap. js: Automatic generation of gene cluster comparison figures. bioRxiv. 2020.

36. Shimodaira H. An approximately unbiased test of phylogenetic tree selection. Syst Biol. 2002; 51: 492–508. https://doi.org/10.1080/10635150290069913 PMID: 12079646

37. Shimodaira H, Hasegawa M. Multiple comparisons of log-likelihoods with applications to phylogenetic inference. Mol Biol Evol. 1999; 16: 1114.

38. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. 2014; 30: 1312–1313. https://doi.org/10.1093/bioinformatics/btu033 PMID: 24451623

39. Letunic I, Bork P. Interactive tree of life (iTOP) v3: an online tool for the display and annotation of phylogenetic and other trees. Nucleic Acids Res. 2016; 44: W242–W245. https://doi.org/10.1093/nar/gkw290 PMID: 27095192

40. Thomashow LS, Weiller DM, Bonsall RF, Pierson LS. Production of the antibiotic phenazine-1-carboxylic acid by fluorescent Pseudomonas species in the rhizosphere of wheat. Appl Environ Microbiol. 1990; 56: 908–912. https://doi.org/10.1128/AEM.56.4.908-912.1990 PMID: 16348176
41. Loper JE, Hassan KA, Mavrodi DV, Davis EW, Lim CK, Shaffer BT, et al. Comparative genomics of plant-associated Pseudomonas spp.: insights into diversity and inheritance of traits involved in multi-trophic interactions. PLoS Genet. 2012; 8: e1002784. https://doi.org/10.1371/journal.pgen.1002784 PMID: 22792073

42. Radtke C, Cook WS, Anderson A. Factors affecting antagonism of the growth of Phanerochaete chrysosporium by bacteria isolated from soils. Appl Microbiol Biotechnol. 1994; 41: 274–280.

43. Okrent RA, Manning VA, Trippe KM. Draft genome sequences of seven 4-formylaminoxyvinylglycine producers belonging to the Pseudomonas fluorescens species complex. Genome Announc. 2017; 5: 5–7. https://doi.org/10.1128/genomeA.00277-17 PMID: 28473382

44. Kluepfel DA, McInnis TM, Zehr EI. Involvement of root-colonizing bacteria in peach orchard soils suppressive of the nematode Criconemella xenoplax. Phytopathol YORK Balt THEN ST PAUL-. 1993; 83: 1240.

45. Cintas NA, Koike ST, Bull CT. A new pathovar, Pseudomonas syringae pv. alisalensis pv. nov., proposed for the causal agent of bacterial blight of broccoli and broccoli raab. Plant Dis. 2002; 86: 992–998. https://doi.org/10.1094/PDIS.2002.86.9.992 PMID: 3081561

46. Sarris PF, Trantas EA, Baltrus DA, Bull CT, Wechter WP, Yan S, et al. Comparative genomics of multiple strains of Pseudomonas cannabina pv. alisalensis, a potential model pathogen of both monocots and dicots. PLoS One. 2013; 8: e59366. https://doi.org/10.1371/journal.pone.0059366 PMID: 23555661

47. Williams PH, Keen NT. Bacterial blight of radish. Plant Dis. 1966; 50: 192–195.

48. Baltrus DA, Nishimura MT, Romanchuk A, Chang JH, Mukhtar MS, Cherkis K, et al. Dynamic evolution of pathogenicity revealed by sequencing and comparative genomics of 19 Pseudomonas syringae isolates. PLoS pathog. 2011; 7: e1002132. https://doi.org/10.1371/journal.ppat.1002132 PMID: 21796644

49. Smith DDN, Kirzinger MWB, Stavrinides J. Draft genome sequence of the antibiotic-producing epiphytic isolate Pantoea ananatis BRT175. Genome Announc. 2013; 1. https://doi.org/10.1128/genomeA.00902-13 PMID: 24201193

50. Johnson KB, Stockwell VO, McLaughlin RJ, Sugar D, Loper JE, Roberts RG. Effect of antagonistic bacteria on establishment of honey bee-dispersed Erwinia amylovora in pear blossoms and on fire blight control. Phytopathol. 1993.

51. Schneider CA, Rasband WS, Eliceiri KW. NIH Image to ImageJ: 25 years of image analysis. Nat Methods. 2012; 9: 671–675. https://doi.org/10.1038/nmeth.2089 PMID: 22930834

52. Okrent RA, Trippe KM, Manning VA, Walsh CM. Detection of 4-formylaminoxyvinylglycine in culture filtrates of Pseudomonas fluorescens WH6 and Pantoea ananatis BRT175 by laser ablation electrospray ionization-mass spectrometry. PLoS One. 2018; 13: e0200481. https://doi.org/10.1371/journal.pone.0200481 PMID: 29990341

53. Garrido-Sanz D, Meier-Kolthoff JP, Göker M, Martín M, Rivilla R, Redondo-Nieto M. Genomic and genetic diversity within the Pseudomonas fluorescens complex. PLoS One. 2016; 11: e0150183. https://doi.org/10.1371/journal.pone.0150183 PMID: 26915094

54. Bull CT, Koike ST. Practical benefits of knowing the enemy: modern molecular tools for diagnosing the etiology of bacterial diseases and understanding the taxonomy and diversity of plant-pathogenic bacteria. Annu Rev Phytopathol. 2015; 53: 157–180. https://doi.org/10.1146/annurev-phyto-080614-120122 PMID: 26002289

55. Tran PN, Savka MA, Gan HM. In-silico taxonomic classification of 373 genomes reveals species misidentification and new genospecies within the genus Pseudomonas. Front Microbiol. 2017; 8: 1296. https://doi.org/10.3389/fmicb.2017.01296 PMID: 28747902

56. Rangel LI, Henkels MD, Shaffer BT, Walker FL, Davis EW II, Stockwell VO, et al. Characterization of Toxin Complex Gene Clusters and Insect Toxicity of Bacteria Representing Four Subgroups of Pseudomonas fluorescens. PLoS One. 2016; 11: e0161120. Available: https://doi.org/10.1371/journal.pone.0161120 PMID: 27580176

57. Jun S-R, Wassenaar TM, Nookaew I, Hauser L, Wanchai V, Land M, et al. Comparative genome analysis of Pseudomonas genomes including Populus-associated isolates. Appl Environ Microbiol. 2015. https://doi.org/10.1128/AEM.02612-15 PMID: 26519390

58. Okrent RA, Halgren AB, Azevedo MD, Chang JH, Mills DI, Maselko M, et al. Negative regulation of germination-arrest factor production in Pseudomonas fluorescens WH6 by a putative extracytoplasmic function sigma factor. Microbiology. 2014; 160: 2432–42. https://doi.org/10.1099/mic.0.080317-0 PMID: 25165126

59. Planet PJ. Tree disagreement: measuring and testing incongruence in phylogenies. J Biomed Inform. 2006; 39: 86–102. https://doi.org/10.1016/j.jbi.2005.08.008 PMID: 16243006
60. Gross H, Loper JE. Genomics of secondary metabolite production by Pseudomonas spp. Nat Prod Rep. 2009; 26: 1408–1446. https://doi.org/10.1039/b817075b PMID: 19844639

61. Mousa WK, Raizada MN. Biodiversity of genes encoding anti-microbial traits within plant associated microbes. Front Plant Sci. 2015; 6: 231. https://doi.org/10.3389/fpls.2015.00231 PMID: 25914708

62. Almario J, Bruto M, Vacheron J, Prigent-Combaret C, Moënne-Loccoz Y, Muller D. Distribution of 2, 4-diacetylphloroglucinol biosynthetic genes among the Pseudomonas spp. reveals unexpected polyphyletism. Front Microbiol. 2017; 8: 1218. https://doi.org/10.3389/fmicb.2017.01218 PMID: 28713346

63. Mavrodi DV, Peever TL, Mavrodi OV, Pareiko JA, RaaijmakersJM, Lemanceau P, et al. Diversity and evolution of the phenazine biosynthesis pathway. Appl Environ Microbiol. 2010; 76: 866–879. https://doi.org/10.1128/AEM.02009-09 PMID: 20008172

64. Fitzpatrick DA. Lines of evidence for horizontal gene transfer of a phenazine producing operon into multiple bacterial species. J Mol Evol. 2009; 68: 171–185. https://doi.org/10.1007/s00239-009-9198-5 PMID: 19189039

65. Castignetti D. Probing of Pseudomonas aeruginosa, Pseudomonas aureofaciens, Burkholderia (Pseudomonas) cepacia, Pseudomonas fluorescens, and Pseudomonas putida with the ferrirypochelin receptor A gene and the synthesis of pyochelin in Pseudomonas aureofaciens, Pseudomonas. Curr Microbiol. 1997; 34: 250–257. https://doi.org/10.1007/s002849900178 PMID: 9058547

66. Sugawara M, Okazaki S, Nukui N, Ezura H, Mitsui H, Minamisawa K. Rhizobitoxine modulates plant–microbe interactions by ethylene inhibition. Biotechnol Adv. 2006; 24: 382–388. https://doi.org/10.1016/j.biotechadv.2006.01.004 PMID: 16516430

67. Tang X, Dai P, Gao H, Wang C, Chen G, Hong K, et al. A Single Gene Cluster for Chalcomycins and Aldgamycins: Genetic Basis for Bifurcation of Their Biosynthesis. ChemBioChem. 2016; 17: 1241–1249. https://doi.org/10.1002/cbic.201600118 PMID: 27191535

68. Fischbach MA, Walsh CT, Clardy J. The evolution of gene collectives: How natural selection drives chemical innovation. Proc Natl Acad Sci. 2008; 105: 4601–4608. https://doi.org/10.1073/pnas.0709132105 PMID: 18216259

69. Medema MH, Cimermanic P, Sali A, Takano E, Fischbach MA. A systematic computational analysis of biosynthetic gene cluster evolution: lessons for engineering biosynthesis. PLoS Comput Biol. 2014; 10: e1004016. https://doi.org/10.1371/journal.pcbi.1004016 PMID: 25474254