Functional analysis of a biosynthetic cluster essential for production of 4-formylaminoxyvinylglycine, a germination-arrest factor from Pseudomonas fluorescens WH6

Rachel A. Okrent, Kristin M. Trippe, Maciej Maselko and Viola Manning

Abstract
Rhizosphere-associated Pseudomonas fluorescens WH6 produces the germination-arrest factor 4-formylaminoxyvinylglycine (FVG). FVG has previously been shown to both arrest the germination of weedy grasses and inhibit the growth of the bacterial plant pathogen Erwinia amylovora. Very little is known about the mechanism by which FVG is produced. Although a previous study identified a region of the genome that may be involved in FVG biosynthesis, it has not yet been determined which genes within that region are sufficient and necessary for FVG production. In the current study, we explored the role of each of the putative genes encoded in that region by constructing deletion mutations. Mutant strains were assayed for their ability to produce FVG with a combination of biological assays and TLC analyses. This work defined the core FVG biosynthetic gene cluster and revealed several interesting characteristics of FVG production. We determined that FVG biosynthesis requires two small ORFs of less than 150 nucleotides and that multiple transporters have overlapping but distinct functionality. In addition, two genes in the centre of the biosynthetic gene cluster are not required for FVG production, suggesting that additional products may be produced from the cluster. Transcriptional analysis indicated that at least three active promoters play a role in the expression of genes within this cluster. The results of this study enrich our knowledge regarding the diversity of mechanisms by which bacteria produce non-proteinogenic amino acids like vinylglycines.

INTRODUCTION
Pseudomonas fluorescens WH6, a bacterial strain originally isolated from the rhizosphere of wheat [1], has been investigated in our laboratory for its ability to produce a secondary metabolite with selective herbicidal and antimicrobial properties. WH6 culture filtrate arrests the germination of a number of weedy grass species, including annual bluegrass (Poa annua L.) [2, 3]. WH6 filtrate also inhibits the growth of the bacterial plant pathogen Erwinia amylovora [4], the causal agent of fireblight in orchard crops. Based on the biological effects of the culture filtrates on grass seed germination, the active compound was originally termed a germination-arrest factor (GAF). The molecule responsible for these activities was subsequently purified and identified as 4-amino-4-formylaminoxy-3-butenolic acid, also known as 4-formylaminoxyvinylglycine (FVG) (Fig. 1) [5, 6].

FVG is an oxyvinylglycine, a class of non-proteinogenic amino acids produced and excreted by plant- and soil-borne bacteria. Several oxyvinylglycines have been investigated for their role in plant–microbe and microbe–microbe interactions, including aminoethoxyvinylglycine from an isolate of Streptomyces sp. [7], 4-methoxyvinylglycine from Pseudomonas aeruginosa [8] and rhizobitoxine from Bradyrhizobium elkanii [9] and Burkholderia andropogonis [10] (Fig. 1). The biological activity of oxyvinylglycines is based on their ability to inhibit enzymes that require pyridoxal phosphate as a cofactor [11]. Enzymes requiring pyridoxal phosphate are necessary for important cellular processes including amino acid biosynthesis, nitrogen metabolism [12] and ethylene biosynthesis [13]. Because oxyvinylglycines target a broad spectrum of critical enzymes, they can have economically important agricultural applications, including plant growth regulation [14] and weed control [15].

Predicting the biosynthetic pathway of FVG using comparative genomics or other tools is difficult because the known biosynthetic pathways of oxyvinylglycines vary considerably...
and because FVG has an unusual aminooxy linkage (Fig. 1). Therefore, determining which genes are required for production of FVG can inform our understanding of its biosynthetic pathway. Several loci required for production of FVG were previously identified by screening Tn5 mutant libraries of the sequenced strain P. fluorescens WH6 [19, 20]. Tn5 insertions in seven regions of the chromosome led to the loss of FVG production. Of these, two occurred in regions that appeared to be specific to FVG production. One of the Tn5 insertions disrupted a gene encoding the anti-sigma factor PrtR. The role of PrtR was recently investigated and it was found to modulate negative regulation of FVG production through the ECF-type sigma factor PrtI [21]. A second 13 kb region was disrupted by multiple Tn5 insertions that led to the loss of FVG production. This chromosomal region consisted of 12 ORFs that encode putative regulatory proteins, biosynthetic enzymes and transporters [20] (Fig. 2).

In the current study, we characterize this chromosomal region, referred to here as the GAF vinylglycine (gvg) cluster. Specifically, we determined which of these genes are necessary for production of FVG through site-directed mutagenesis and complementation. Using this approach, we confirmed the importance of the four genes identified in the initial Tn5 screens [19, 20] and identified six additional genes that are involved in the production of FVG. We also investigated the transcriptional organization of the cluster and determined that, in addition to a large polycistronic transcript extending the length of the cluster, there are also several smaller transcripts. This work reveals several interesting aspects of the gvg biosynthetic gene cluster, including two small ORFs that are required for production of FVG and multiple genes encoding LysE family transporters with distinct but overlapping functionality. The presence of two genes within the gvg cluster that are not required for FVG production suggests that additional products may be produced from the gene cluster.

**METHODS**

**Bacterial strains and plasmids**

Bacterial strains are listed in Table S1 (available in the online Supplementary Material) and plasmids are listed in Table S2. The origin and characterization of P. fluorescens strain WH6 (NRRL no. B-30485) from wheat rhizosphere was described previously [1, 2].

**DNA manipulation**

DNA was isolated from bacteria using the ZR Fungal/Bacterial DNA Kit (Zymo Research) or the PowerLyzer Ultracean Microbial DNA Isolation Kit (MO BIO). Purity and concentration were determined using a Nanodrop ND1000 (Thermo Scientific) or Eon Plate Reader with Take3 attachment (BioTek). PCR was performed using high-fidelity polymerases when required [i.e. Phusion and Q5 (New England Biolabs (NEB)) and Accuprime (Invitrogen)] or OneTaq (NEB) for standard PCR. PCR products were gel purified using standard kits. Restriction enzymes, T4 DNA ligase and HiFi Assembly reagent were purchased from NEB. Sequencing was performed using an ABI 3730 capillary DNA sequence system (Applied Biosystems) by the Center for Genome Research and Biocomputing Core Facilities (Oregon State University).

**Construction of in-frame deletion mutants**

All primers used for PCR amplification in this study are listed in Table S3. Site-directed mutagenesis was performed generally as described by Okrent et al. [21]. However, a scarless method that omits the insertion of the kanamycin resistance FLP recognition target site (kan-FRT) cassettes was used in some deletions, allowing for the subsequent deletion of additional genes. The pEX-18Km vector was designed for this purpose, with a kanamycin resistance gene (kan) from pKD13 [22] replacing the tetracycline resistance gene (tet) of pEX18-Tc. The pEX-18Km vector increased transformation efficiency relative to pEX18-Tc in the scarless method (data not shown). Primers for overlap extension PCR are described in Table S4 and primers and restriction recognition sites for use in the generation of the FRT-kan-FRT fragment are described in Table S5.

Plasmids were mobilized into recipients through triparental mating using pRK2013 [23] as the mobilizing helper plasmid. Transconjugants that had undergone a recombination event were selected on 925 minimal medium agar [4] containing 50 µg ml⁻¹ ampicillin and 50 µg ml⁻¹ kanamycin. Transconjugants were transferred to plates containing 5% sucrose to counterselect for double recombinants that contain only the mutant allele. When required, eviction of kan was mediated by pBH474 [24], which encodes the FLP site-specific recombinase, and it was confirmed by replica plating on agar plates with and without kanamycin [24]. Mutants were identified by colony PCR using primers described in Table S6. Each deletion mutant was further confirmed by amplifying the region encompassing the flanking regions, ligating this region into the PCR cloning.
vector pJET1.2/blunt (CloneJET PCR cloning kit, Thermo Scientific) and sequencing the insert (Table S6).

**Genetic complementation**

The deletion mutations in *P. fluorescens* WH6 that resulted in a null or altered FVG phenotype were complemented by inserting the corresponding gene under the control of the constitutive lac promoter in the expression vector pBBR1EVM [21] (Table S7). The resulting constructs were transferred from *Escherichia coli* into the appropriate WH6 mutant strains by triparental mating as above. Transconjugants expressing resistance to gentamicin (from the plasmid) and ampicillin (from WH6) were selected. Correct transfer of plasmids was confirmed by isolation of the plasmids and digestion with the appropriate restriction enzymes. In some cases, complementation with a native promoter or inducible promoter construct was warranted. Native promoter constructs included the sequence upstream of the gene(s) of interest and were constructed in the expression vector pBBR1MCS-5 [25]. Inducible promoter constructs were under the control of the araC-pBAD system in pJN105 [26].

**Bioassays for FVG activity**

*P. fluorescens* strains were inoculated into modified *Pseudomonas* minimal salts (PMS) medium, cultured at 28 °C for 7 days and harvested as described previously [2]. Filtrates were made from duplicate or triplicate clones of each deletion mutant or complemented mutant strain and subsequently assayed to ensure reproducibility. The bioassay for germination-arrest activity in the bacterial culture filtrates was performed using the standard germination-arrest bioassay protocol and scoring system for annual bluegrass seeds [2, 27]. The agar-diffusion bioassays for antimicrobial activity against *E. amylovora* were performed as described by Halgren *et al.* [4] with the resulting zones of inhibition measured on triplicate plates using the imaging software Able Image Analyser (Mu Labs).

**TLC analyses**

Samples were prepared for TLC analysis by extracting the solids from dried bacterial culture filtrates as described previously [4, 5, 21]. TLC analysis was performed on Silica Gel GHL and microcrystalline cellulose TLC plates (250 µm thick; Analtech) with ninhydrin staining as described previously [5].

**Construction of the double transporter mutant and test for lethality**

The ΔgvgIK deletion mutation was first attempted in a wild-type *P. fluorescens* WH6 background using the suicide plasmid pEXW34. As this attempt was unsuccessful, the ΔgvgJK mutation was then constructed in the WH6-25G (ΔgvgC) background to avoid the potential for accumulation of toxic products in the cell. The pEXW34 plasmid was mobilized into WH6-25G through triparental mating, resulting in WH6-34G, a triple mutant in *gvgC*, *gvgJ* and *gvgK*.

The *gvgC* gene was cloned into the broad host range vector pJN105 [26] under the control of the pBAD arabinose-inducible promoter. This construct, pJW5251, was mobilized first into WH6-25G (ΔgvgC) and assayed for function of *gvgC* in the agar-diffusion assay and then subsequently mobilized into the triple mutant WH6-34G. The wild-type, mutant and complemented mutant strains were grown for 24 h in a modified PMS medium containing glycerol instead.
of glucose as the carbon source in the presence or absence of 100 mM l-arabinose (Ara). As an indicator of growth, the OD₆₀₀ for triplicate cultures was measured with an Eon Plate Reader (BioTek) with a path length of 0.5 cm.

**cDNA synthesis and reverse transcriptase PCR**

RNA was extracted from WH6 bacterial cultures grown to mid-exponential phase in PMS medium and stopped in RNAProtect Bacterial Reagent (Qiagen). The total RNA was extracted using an RNeasy Mini Kit (Qiagen) followed by the removal of genomic DNA with Turbo DNase treatment (Ambion). cDNA was synthesized using the SuperScript III First-Strand Synthesis System (Invitrogen) under standard conditions with random hexamer primers. A reaction without reverse transcriptase (RT) using the same RNA was performed for the no RT control sample. To detect transcripts across genes, cDNA from three pooled RNA samples was amplified in PCR reactions with overlapping primer pairs (Table S8), using cDNA as a template. Amplification reactions with no RT and gDNA controls were also included.

**5′ rapid amplification of cDNA ends**

5′ rapid amplification of cDNA ends (RACE) analysis was carried out with the GeneRacer Kit (Invitrogen) to determine the transcriptional start sites for gvgF and gvgH. RNA was purified as above and enriched for full-length transcripts by treatment with Terminator (Epicentre). Full-length transcripts with triphosphate ends were subsequently converted into monophosphate 5′ ends with tobacco acid pyrophosphatase. The GeneRacer RNA Oligo was then ligated to the treated monophosphate 5′ cDNA ends along with untreated control samples. The WH6 mRNA was reverse transcribed with random hexamers and the resulting cDNA amplified using the GeneRacer forward primer and reverse primers specific to each of the two genes examined in nested reactions (Table S9). The major PCR products were gel purified and cloned into the pCR4-TOPO cloning vector (Invitrogen). Colony PCR was performed on at least 10 clones from each PCR band to estimate insert size and the largest inserts from each set were sequenced to determine the location of 5′ ends.

**Construction of lacZ reporter fusions and assay for β-galactosidase activity**

The putative promoter regions upstream of gvgF (493 nt) and gvgH (705 nt) were cloned into the mini-Tn7 vector pXY2 for constructing fusions with lacZ to test for promoter activity [28]. The lac promoter was also amplified from pBBR1EMV and cloned into pXY2 as a positive control. Primers are described in Table S7. The promoter-lacZ fusions were transferred into the conserved attTn7 site of the wild-type WH6 genome by introduction of the pXY2-based plasmids and pTNS3 that encodes the site-specific Tn7 transposition pathway [29] via electroporation, as described by Choi et al. [30]. Ten colonies from each transformation were patched on to LB plates supplemented with 50 µg ml⁻¹ ampicillin, 15 µg ml⁻¹ gentamicin and 40 µg ml⁻¹ X-Gal. Plates were incubated at 28 ˚C for 48 h for visualization of β-galactosidase activity. Genomic DNA was extracted from three of the patched isolates per transformation and screened for genomic insertion with the PCR primers lacZ6 [28] and VM029 (Table S3).

**Bioinformatics analyses**

The locations of putative transmembrane (TM) domains were determined using the web interface of TMHMM v.2.0 from the Center for Biological Sequence Analysis at the Technical University of Denmark (http://www.cbs.dtu.dk/services/TMHMM/). Protein secondary sequence predictions were performed using The PSIPRED Protein Sequence Analysis Workbench from the Bloomsbury Centre for Bioinformatics at University College London (http://bioinf.cs.ucl.ac.uk/psipred/). Promoter sequences were predicted using BPROGRAM from Softberry (http://linux1.softberry.com/) [31]. Rho-independent transcriptional termination sequences were probed using ARNold from the Université Paris-Sud (http://rna.ignors.u-psud.fr/toolbox/arnold/index.php/) [32].

The P. fluorescens WH6 GvgF sequence was aligned with characterized O-carbamoyltransferase [33] sequences using MUSCLE v.3.8.425 implemented in CLC Main Workbench. The pairwise identity between the sequences was calculated based on the alignment using the pairwise comparisons tools in the same programme. Co-conservation of genes was investigated using the Conserved Neighborhood tool of the Integrated Microbial Genomes & Microbiomes platform from the US Department of Energy (http://img.jgi.doe.gov/) [34]. The intergenic region between gvgR and gvgA for several gvg clusters was aligned in CLC Main Workbench as above. Potential palindromic GntR binding sites were identified using the programme palindrom from EMBOSS explorer (http://emboss.bioinformatics.nl/cgi-bin/emboss/palindrome/).

**RESULTS**

**General features of the gvg biosynthetic gene cluster**

There are 12 predicted ORFs encoded within the 13 kb chromosomal region that range in length between 84 and 2211 nt. Intergenic regions vary from 13 to 307 nt. A predicted GntR family regulator, gvgR, is encoded on the opposite strand in relation to the other 11 ORFs (Fig. 2). The gvg cluster is flanked by genes encoding a putative trans-aconitate methyltransferase and a single-stranded DNA binding protein (Fig. 2). In the case of several genes (gvgB, gvgC and gvgJ), the start sites used for constructing deletion mutations and complements were different from the start sites annotated in the published genome [19] (Fig. S1).

**Effect of deletion mutations and complements on FVG production**

FVG activity is measured using a germination-arrest assay with annual bluegrass seeds [2, 3] and an agar-diffusion assay with E. amylovora [4]. Additionally, FVG reacts with ninhydrin to produce either a pink band on silica TLC plates or a purple band on cellulose TLC plates, each at a characteristic Rᵣ value [5]. The absence of one or more of
these elements in sterile or extracted culture filtrates is indicative of a null FVG phenotype. In an earlier study, Tn5 insertions in four of the gvg genes resulted in the loss of FVG production [4, 19]. These genes, gvgR, gvgC, gvgE and gvgH, are predicted to encode a GntR family transcriptional regulator, a haem-oxygenase domain-containing protein, a carbamoyltransferase (CTase) and an aminotransferase, respectively (Fig. 2). However, Tn5 insertions were not recovered in the other genes within the 13 kb region. Therefore, we used site-directed mutagenesis to determine whether these additional genes are also required for FVG production and to confirm that the phenotypes reported in earlier Tn5 studies were not due to polar effects.

Collectively, site-directed deletion of seven genes in the gvg cluster resulted in an obviously null FVG phenotype (Figs 3 and 4a, Table 1). In addition to the four genes originally identified in the screen of Tn5 insertion mutants [4, 19], a null FVG phenotype was also observed in site-directed deletions of gvgA, encoding an esterase/lipase, and two small ORFs, gvgB and gvgG. Complementation of these seven deletions with the corresponding wild-type allele in trans restored the wild-type FVG phenotype (Fig. 3, Table 1). The mutation in the gene encoding the GntR transcriptional regulator (gvgR) could not be successfully complemented with gvgR under the control of a constitutive promoter; however, it was complemented with a construct containing gvgR, gvgA and the intergenic region containing the putative promoters of both genes (Fig. 3, Table 1). The mutation in the esterase/lipase (gvgA) was not successfully complemented with gvgA or gvgB alone (data not shown). Complementation with a construct containing the gvgA-gvgB sequence under the control of a constitutive promoter restored biological activity, albeit not to wild-type levels.

Mutations in two genes within the cluster, gvgD encoding an amidinotransferase and gvgE encoding a LysE family transporter, did not affect FVG production (Figs 3 and 4a, Table 1). Mutation of the gene located downstream of gvgR, tam, did not alter the FVG phenotype (data not shown) and was thus determined to not be part of the gvg cluster.

Mutations in the remaining three genes, gvgI, gvgJ and gvgK, led to more complex phenotypes. When gvgI, encoding a formyltransferase, was deleted, the filtrate behaved similarly to wild-type in the agar-diffusion assay (Table 1). However, the activity in the germination-assay was reduced (Fig. 3) and the TLC plates were markedly different (Fig. 4a). Chromatograms of the extracted gvgI mutant filtrates lacked the FVG-specific band. Instead, a distinct bright-blue band appeared in chromatograms stained with ninhydrin (Fig. 4a). When the ΔgvgI mutation was complemented with the wild-type allele, the TLC profile characteristic of FVG (Fig. 4b) and the germination-arrest activity (Fig. 3) were both restored. Thus, it appears that the formyltransferase mutant produces a bioactive compound that is not FVG. Several attempts to identify this molecule failed as the purification process abolished its biological activity (data not shown).

The genes gvgJ and gvgK are both predicted to encode transport proteins in the LysE superfamily [35] that may facilitate the extracellular export of FVG. Mutant strains with deletions in gvgJ or gvgK produced distinct phenotypes; however, neither of these genes is absolutely required for FVG production. A mutant strain with a deletion in gvgJ was less active in the bioassay, indicating an intermediate phenotype, while mutation of gvgK resulted in wild-type FVG activity (Fig. 3, Table 1). The silica TLC plate of the extracted ΔgvgJ mutant filtrate is missing the characteristic blue band associated with FVG (Fig. 4a). Additional differences among ΔgvgJ, ΔgvgK and wild-type filtrates are apparent in the cellulose chromatograms (Fig. 4c). In the cellulose chromatograms, the bands indicative of FVG are retained in both strains, although at trace levels in the ΔgvgJ mutant strain. In both strains, there is an additional band that does not appear in wild-type samples (Fig. 4c). This band is consistently present in experimental replicates from both strains (data not shown). The complementation of the ΔgvgJ mutation with the wild-type allele restored FVG activity (Fig. 3, Table 1) and complementation of the ΔgvgJ and ΔgvgK mutations eliminated the presence of additional bands in the chromatograms (Fig. 4b and c).

**Mutation of both transporters**

The reduction, but not elimination, of FVG with the deletion of gvgJ and the similar additional bands in the TLC plates with the deletion of gvgJ or gvgK suggested that the two transporters may overlap in function. If GvgJ and GvgK are each able to export FVG, then a strain with a double ΔgvgJK deletion would be expected to display a null FVG phenotype. Our attempts to construct a strain with a ΔgvgJK genotype, however, were unsuccessful. Although transconjugants harbouring the merodiploid were isolated post-conjugation, counterselection on sucrose led to recovery of wild-type alleles only. This suggested that the double crossover event likely resulted in a lethal mutation, perhaps due to accumulation of a toxic product of the gvg cluster.

An alternative strategy for construction of the gvgJK deletion was pursued to confirm that the double gvgJK deletion is lethal (Fig. 5a). The intracellular accumulation of toxic FVG or FVG-related products was avoided by constructing the ΔgvgJK mutation in the ΔgvgC deletion strain, which does not produce FVG. The double mutation ΔgvgJK was mimicked by complementing the triple mutation in the ΔgvgC JK mutant strain with gvgC under the control of an Ara-inducible promoter. This inducible construct complements the ΔgvgC mutation similarly to the equivalent constitutive construct, as indicated by anti-Erwinia activity in the agar-diffusion assays (Fig. 52). While this complemented ΔgvgC strain grew similarly to wild-type in the presence of the inducer, the complemented triple mutant grew only slightly (Fig. 5b). This experiment confirms that the lethality of the ΔgvgJK mutation is due to product(s) from the gvg cluster, either FVG itself or its related byproducts.
Transcriptional organization of the cluster

As two genes in the centre of the cluster are not required for FVG biosynthesis, we were interested in whether the gvg cluster is transcribed in a single transcript or in multiple transcripts that bypass these intervening genes. We used several tools to investigate the transcriptional organization of the gvg cluster. RT-PCR was performed to determine whether the gvg genes are coordinately transcribed in a single mRNA transcript. Due to the difficulty in transcribing amplicons larger than 10 kb, we designed the RT-PCR experiment to detect smaller overlapping amplicons over the length of the cluster (Fig. 6a). Overlapping transcripts were detected that span the coding regions of gvgA to gvgK, but no transcript was detected that extends from gvgA into gvgR (Fig. 6b). These data suggest that a promoter upstream of gvgA drives the expression of the gvg cluster in a single transcript including gvgD and gvgE. This analysis, however, does not eliminate the possibility that smaller regions are also transcribed separately.

The possibility of additional transcription from promoters upstream of the CTase-encoding gvgF and the aminotransferase-encoding gvgH was investigated using 5’ RACE and lacZ-promoter fusions. To avoid degraded transcripts, 5’ RACE analysis was performed on mRNA samples enriched for full-length transcripts, as described in the Methods section. The longest transcript detected in the gvgF 5’ RACE analysis began 29 nt upstream of the predicted gvgF start codon (Figs 6a and S3). The longest transcript for gvgH began 203 nt upstream of the gvgH start in the intergenic region between the gvgF and gvgG (Figs 6a and S3). The presence of functional promoters was confirmed by fusing the regions upstream of these genes to a lacZ reporter in the wild-type WH6 genetic background. β-Galactosidase activity indicative of a functional promoter from the P_{gvgF}–lacZ and P_{gvgH}–lacZ fusions, along with a P_{lac}–lacZ control, was qualitatively assessed on LB agar supplemented with X-Gal (Fig. S4). A colour change in WH6 containing the putative promoter fusions but not in a negative control, was qualitatively assessed on LB agar supplemented with X-Gal (Fig. S4).
The objectives of the current study were to (1) systematically interrogate each gene within the 13 kb region to define the gvg cluster and (2) investigate the transcriptional organization of the gene cluster. Collectively, these data inform our understanding of the biosynthesis of this unusual molecule. The deletion mutagenesis and functional analysis presented here confirmed the role of genes identified in the Tn5 screen and also identified six additional genes that are required for FVG production or export.

**Two internal genes are not required for FVG production**

The gvg gene cluster is disrupted by two genes not required for FVG production. These genes encode an amidinotransferase and a LysE family transporter. Their presence suggests that there may be an additional pathway encoded by the cluster for production and export of additional compounds. While such a compound has not yet been observed using the standard methodology of detecting FVG, it would likely be hydrophilic and difficult to isolate due to the presence of an amido group. While unusual, biosynthetic gene clusters that direct the production of multiple compounds have been observed in bacteria. For example, in certain strains of *Pseudomonas chlororaphis*, phenazine-1-carboxylic acid and a hydroxylated derivative are both produced from the *phz* cluster [36]. The two compounds are thought to have different biological roles [37]. There are also several examples of superclusters in *Streptomyces* which appear to be merged from smaller clusters [38, 39].

The arrangement of gvgD and gvgE in the centre of the gvg cluster is conserved in strains of *Pseudomonas*, such as *Pseudomonas cannabina* var. *alislensis* ES4326 [20]. In contrast, *Pantoea ananatis* BRT175 encodes a cluster similar to the gvg cluster [40] but lacks the two intervening genes. The BRT175 cluster is responsible for the production of an unidentified compound which, like FVG, inhibits *E. amylovora*. Our results reported here suggest that BRT175 could produce FVG despite the absence of these two genes. Ongoing work in our laboratory is focused on the identification of the active compound from BRT175.

**Small ORFs are required for FVG production**

The requirement of two small ORFs, gvgB and gvgG, for FVG biosynthesis was surprising. Commonly used annotation pipelines do not, by default, recognize ORFs of this size. Even when annotation pipelines do identify small ORFS, they are often ignored as artefacts [41]. Indeed, the recent re-annotation of the *P. fluorescens* WH6 genome as part of the NCBI RefSeq project eliminated both of these annotations (NZ_CM001025). However, our complementation data indicate that these small ORFs act out of context of their adjacent genes; therefore, they likely encode peptides that are important for FVG production.

There is little known about the role of small peptides in the biosynthesis of secondary metabolites. Bioinformatic analyses of the sequences of GvgB and GvgC do provide some clues as to their function. The DNA sequence of gvgB is AT...
Table 1. Anti-<i>E. amylovora</i> activity of WH6 mutant and complemented strains in the agar-diffusion assay

<table>
<thead>
<tr>
<th>Genotype</th>
<th>Strain</th>
<th>Mutant</th>
<th>Zone of inhibition (cm²)</th>
<th>Complement</th>
<th>Zone of inhibition (cm²)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ΔgvgR</td>
<td>WH6-21G</td>
<td>0.0±0.00</td>
<td>gvgRA</td>
<td>12.3±0.13</td>
<td></td>
</tr>
<tr>
<td>ΔgvgA</td>
<td>WH6-23G</td>
<td>0.0±0.00</td>
<td>gvgAB</td>
<td>11.5±0.09</td>
<td></td>
</tr>
<tr>
<td>ΔgvgB</td>
<td>WH6-24G</td>
<td>0.0±0.00</td>
<td>gvgB</td>
<td>14.0±0.19</td>
<td></td>
</tr>
<tr>
<td>ΔgvgC</td>
<td>WH6-25G</td>
<td>0.0±0.00</td>
<td>gvgC</td>
<td>16.2±0.24</td>
<td></td>
</tr>
<tr>
<td>ΔgvgD</td>
<td>WH6-26G</td>
<td>18.1±0.25</td>
<td>N/A</td>
<td>N/A</td>
<td></td>
</tr>
<tr>
<td>ΔgvgE</td>
<td>WH6-27G</td>
<td>14.7±0.04</td>
<td>N/A</td>
<td>N/A</td>
<td></td>
</tr>
<tr>
<td>ΔgvgF</td>
<td>WH6-28G</td>
<td>0.0±0.00</td>
<td>gvgF</td>
<td>16.7±0.06</td>
<td></td>
</tr>
<tr>
<td>ΔgvgG</td>
<td>WH6-29G</td>
<td>0.0±0.00</td>
<td>gvgG</td>
<td>16.2±0.17</td>
<td></td>
</tr>
<tr>
<td>ΔgvgH</td>
<td>WH6-30G</td>
<td>0.0±0.00</td>
<td>gvgH</td>
<td>20.3±0.25</td>
<td></td>
</tr>
<tr>
<td>ΔgvgI</td>
<td>WH6-31G</td>
<td>18.4±0.34</td>
<td>gvgI</td>
<td>19.1±0.34</td>
<td></td>
</tr>
<tr>
<td>ΔgvgJ</td>
<td>WH6-32G</td>
<td>6.2±0.21</td>
<td>gvgJ</td>
<td>14.1±0.22</td>
<td></td>
</tr>
<tr>
<td>ΔgvgK</td>
<td>WH6-33G</td>
<td>17.4±0.20</td>
<td>gvgK</td>
<td>14.3±0.14</td>
<td></td>
</tr>
</tbody>
</table>

*Area of zone of inhibition (in cm²) shown as the average of three replicates±SEM. A typical zone of inhibition for WT <i>P. fluorescens</i> WH6 is 18.6±0.45.
†Plasmids for complementation were pNBW5248N2 (gvgRA), pEVW5249N2 (gvgAB), pEVW5250 (gvgB), pEVW5251 (gvgC), pEVW5254 (gvgF), pEVW5255 (gvgG), pEVW5256 (gvgH), pEVW5257 (gvgI), pEVW5258 (gvgJ) and pEVW5259 (gvgK). All constructs contain a constitutive lac promoter except for pNBW5248N2 which contains the native promoter.

rich (37% GC content compared with 61% for the entire cluster) and its corresponding amino acid sequence contains an unusually high proportion of lysine residues (8 of 27 total amino acids). There are no predicted TM domains within the peptide encoded by gvgB. In contrast, the DNA sequence of gvgG has a GC content (55%) that is similar to the entire gvg cluster. The putative peptide encoded by gvgG is predicted to consist of 47 amino acids with an α-helical TM domain (amino acids 5–24). TM domains are frequently found in proteins encoded by small ORFs. For example, a global characterization found that over 65% of <i>E. coli</i> proteins with fewer than 50 amino acids contain a predicted TM domain [42]. Small proteins with TM domains include membrane components that alter membrane characteristics and factors that stabilize large protein complexes [43].

Small ORFs such as gvgG are frequently co-located with an adjacent CTase-encoding gene like gvgF (data not shown). Of the CTases that have been functionally characterized, SxtI of the cyanobacterial saxitoxin cluster [44, 45] is the most similar to GvgF (with 55% identity). In <i>Cylindrospermopsis raciborskii</i> T3 and other strains of cyanobacteria, two annotated small ORFs are located immediately downstream of the CTase gene. These adjacent genes are both predicted to encode TM domains and are thought to function in combination with the SxtI protein [44, 45]. Thus, there may be a common mechanism in which TM-containing small proteins facilitate the function of CTases.

The role of multiple LysE transporters

Secondary metabolite gene clusters in <i>Pseudomonas</i> typically contain a single LysE superfamily transporter [46–49]. The gvg gene cluster in WH6 breaks this mould by encoding two functional and somewhat redundant transporters. Based on the presence of specificity-conferring residues reported by Aleshin et al. [50], the LysE superfamily transporters are members of the RhtB family (data not shown), one of three families within the LysE superfamily [51]. Although the adjacent positions of gvgI and gvgK might suggest that they arise from a recent gene duplication event and be similar in sequence, they are actually quite divergent (22% amino acid identity).

Given the phenotype of the mutant strains, it seems likely that GvgI is the primary exporter of FVG out of the cell but that GvgK also has some affinity for FVG. When either gvgI or gvgK is deleted, a compound accumulates that is not present in TLC plates from wild-type filtrates. This compound could be a precursor or byproduct of FVG biosynthesis. The double ΔgvgIK mutation is lethal, suggesting that GvgI and GvgK can compensate for each other to some extent when only one is deleted. The absence of both transporters, however, seems to result in a toxic accumulation of FVG or other products of the gvg cluster within the cell.

Transcriptional analysis and regulation

The gvg cluster from gvgA to gvgK appears to be transcribed in a single transcript which includes gvgD and gvgE, consistent with a predicted promoter upstream of gvgA and a termination signal downstream of gvgK. However, 5′ RACE analysis and lacZ fusions suggest that there are smaller transcriptional units as well. Multiple modes of transcription may be used to fine-tune the regulation of the cluster. We previously observed multiple transcripts for the prtIR genes which encode the PrtI/PrtR sigma factor/anti-sigma factor pair in <i>P. fluorescens</i> WH6. The two genes were transcribed together and in separate transcripts with prtR transcribed from its own promoter [21]. The complex regulation of gene clusters under different environmental conditions and by different sigma factors is
commonly found in global studies of transcriptional start sites in bacteria [52–54].

The gvgR gene encodes a GntR family transcriptional regulator that was previously shown to regulate the expression of gvgH and gvgJ [20]. GntR proteins regulate their own transcription and other target genes [55] by binding to palindromic sequences in promoter regions with the general pattern 5′-(N)₅GT(N)₃AC(N)₅-3′ [56]. A palindrome (5′-ATCTGT[N]₁₃ACAGAT-3′) consistent with this pattern is present in the intergenic region between gvgR and gvgA in WH6. This pattern is also conserved in orthologous gene clusters within the genus Pseudomonas. The requirement for this element or for proper dosage of GvgR and the other genes of the cluster may explain why the ΔgvgR mutation could be complemented with a native promoter construct but not with gvgR under the control of a constitutive promoter.

Complementation of the ΔgvgC mutation was not successful when either gvgA or gvgB was constitutively expressed (data not shown). However, when gvgA and gvgB were expressed in tandem, FVG activity was partially restored. Further investigation of this region is in progress and may reveal the significance of gvgA and gvgB co-expression. Specifically, RNA-seq experiments that compare global gene expression in the wild-type to global gene expression in ΔgvgA and ΔgvgR mutant strains may support our hypothesis that gene dosage plays a role in the regulation and fine-tuning of gvg gene regulation. Additionally, these experiments should provide insights into the environmental conditions that drive the expression of FVG-related genes by identifying other targets of GvgR regulation.

The FVG biosynthetic pathway

Identification of the genes within the gvg cluster required for FVG production will support future studies focused on elucidation of the biosynthetic pathway. For example, this work reveals that three transferases (aminotransferase, carbamoyltransferase and formyltransferase) are necessary for FVG production. These enzymes are predicted to add the formylamino side chain. In vitro enzymatic assays are in progress to characterize the activity of these enzymes. Identification of the bioactive molecule observed in the ΔgvgI mutant strain which lacks the formyltransferase may assist these efforts.

Previously characterized vinylglycine biosynthetic pathways use different mechanisms to form the vinyl group. Formation of the vinyl group in 1,2-amino-4-methoxy-trans-3-butenoic acid is catalysed by a pair of iron(III)/α-ketoglutarate-dependent oxygenases [18] while the vinyl group of rhizobiotoxine is formed by a membrane fatty acid desaturase-like superfamily enzyme [57]. Although the gvg cluster lacks either of these homologues, the genes that are present may function in an unknown way that leads to vinyl group formation. Alternatively, the genes required for vinyl group formation may be outside of the gvg cluster. Investigation of FVG production in species only distantly related to WH6, combined with exogenous expression of the gvg cluster, should help clarify whether the gvg cluster contains all of the necessary genes to synthesize this unusual molecule.
Conflicts of interest

No experiments involving the use of human or animal subjects were conducted during this study.

Acknowledgements

Technical assistance was provided by D. Chen (USDA-ARS FSCRU). The use of trade, firm or corporation names in this publication is for the information and convenience of the reader. Such use does not constitute an official endorsement or approval by the United States Department of Agriculture or the Agricultural Research Service of any product or service to the exclusion of others that may be suitable.

Conclusions

Functional analysis of the gvg cluster reveals a more complete picture of the genes required for FVG production than was available previously. The importance of small ORFs and the involvement of multiple LysE transporters as well as the presence of genes within the gvg cluster not required for synthesis of FVG reveal some of the complexities of FVG production. This work thus provides a foundation for subsequent studies to elucidate the biosynthetic pathway of this unique molecule and adds to our knowledge of the diverse gene clusters responsible for oxyvinylglycine biosynthesis.

Funding information

Research in the laboratory of K.T. is supported by ARS Project No. 2072-21410-004-00D. This work was supported by the Agriculture and Food Research Initiative Competitive Grant No. 2012-67012-19868 from the United States Department of Agriculture National Institute of Food and Agriculture to R.O.

References

trans-3-butenoic acid (AMB) is made from glutamate and two alanine residues via a thiotemplate-linked tripeptide precursor. Front Microbiol 2015;6:170.


Edited by: G. Preston and M. Whiteley