FEMS Microbiology Ecology, 93, 2017, fiw249

doi: 10.1093/femsec/fiw249
Advance Access Publication Date: 16 December 2016
Minireview

MINIREVIEW

The contribution of genome mining strategies to the understanding of active principles of PGPR strains

Julia Paterson[1], Ghazaleh Jahanshah[1], Yan Li[2], Qi Wang[2], Samina Mehnaz[3] and Harald Gross[1,*]

[1]Department of Pharmaceutical Biology, Pharmaceutical Institute, University of Tübingen, 72076 Tübingen, Germany, [2]Department of Plant Pathology, China Agricultural University, Beijing 100193, P. R. China and [3]Department of Biological Sciences, Forman Christian College University, Lahore 54600, Pakistan

*Corresponding author: Department of Pharmaceutical Biology, Pharmaceutical Institute, University of Tübingen, Auf der Morgenstelle 8, 72076 Tübingen, Germany. Tel: (+49)7071 29-76970; Fax: (+49)7071 29-5250; E-mail: harald.gross@uni-tuebingen.de
One sentence summary: This minireview identifies what can be achieved and the advantages of a genome analysis of a plant beneficial strain.
Editor: Gerard Muyzer

ABSTRACT

Pathogenic microorganisms and insects affecting plant health are a major and chronic threat to food production and the ecosystem worldwide. As agricultural production has intensified over the years, the use of agrochemicals has in turn increased. However, this extensive usage has had several detrimental effects, with a pervasive environmental impact and the emergence of pathogen resistance. In addition, there is an increasing tendency among consumers to give preference to pesticide-free food products. Biological control, through the employment of plant growth-promoting rhizobacteria (PGPR), is therefore considered a possible route to the reduction, even the elimination, of the use of agrochemicals. PGPR exert their beneficial influence by a multitude of mechanisms, often involving antibiotics and proteins, to defend the host plant against pathogens. To date, these key metabolites have been uncovered only by systematic investigation or by serendipity; their discovery has nevertheless been propelled by the genomic revolution of recent years, as increasing numbers of genomic studies have been integrated into this field, facilitating a holistic view of this topic and the rapid identification of ecologically important metabolites. This review surveys the highlights and advances of genome-driven compound and protein discovery in the field of bacterial PGPR strains, and aims to advocate for the benefits of this strategy.

Keywords: PGPR; biocontrol strains; rhizosphere bacteria; genome mining; *Pseudomonas*; *Bacillus*

INTRODUCTION

Plant growth-promoting rhizobacteria (PGPR) are plant-associated and beneficial soil-dwelling organisms (Kloepper, Lifshitz and Zablotowicz 1989) which competitively colonise the rhizosphere (the narrow zone of soil specifically influenced by the root system; Dobbelaere, Vanderleyden and Okon 2003), the rhizoplane (root surface) or the roots themselves (intercellularly and, in specialised nodular structures, also intracellularly; Gray and Smith 2005). PGPR and their host plants exist in a mutualistic relationship. In comparison with bulk soil, the plant provides a rich source of nutrients and energy by way of its root exudates, and offers a surface, or at least a niche, for colonisation, while the bacterium in turn promotes vegetal growth. The mechanisms by which the latter effect is mediated can be divided into direct and indirect forms. The first comprise either the facilitation of resource acquisition (e.g, by nitrogen fixation, phosphate solubilisation or iron acquisition using siderophores) or modulation of plant hormone levels (e.g, by production of indole acetic acid or by decreasing the level

Received: 11 August 2016; **Accepted:** 14 December 2016

of the stress hormone ethylene, employing the enzyme ACC deaminase). Indirect mechanisms are directed at a reduction of the inhibitory effects of various pathogens on plant growth and development. More specifically, PGPR effect biological control in this way by exerting competition for nutrients, niche exclusion, induced systemic resistance and production of antibiotic or antiviral metabolites (Lugtenberg and Kamilova 2009; Beneduzi, Ambrosini and Passaglia 2012).

Biocontrol strains have been proposed as an environmentally friendly alternative to the use of synthetic chemicals, and have drawn increasing attention due to the growing awareness of society towards issues such as environmental protection and human health. Several species of bacteria, such as *Pseudomonas, Azospirillum, Azotobacter, Klebsiella, Enterobacter, Alcaligenes, Arthrobacter, Burkholderia, Bacillus* and *Serratia,* have been reported to enhance plant growth. Of these, the genera *Bacillus* (Serenade®MAX, Quantum-400®, RhizoVital®42), *Streptomyces* (Mycostop®) and *Pseudomonas* (Cedomon® or Cerall®, BlightBan®A506, Proradix®, Liquid PSA®) have been predominantly studied and increasingly marketed as biological control agents.

Since the coining of the term PGPR in the late 1970s (Kloepper and Schroth 1978; Suslow *et al.* 1979), many research teams across the world have investigated the mechanisms, modes of action, physiology, biology and chemistry of members of the class, and thereby compiled a tremendous amount of knowledge through the application of classical biochemical and genetic experiments.

With the onset of the genomic era in bacteriology, novel strategies and research methods became available for use in the PGPR field. This era began definitively in 1995, when the first bacterial genomes were sequenced using conventional Sanger sequencing in combination with a new computerised assembly method (Fleischmann *et al.* 1995; Fraser *et al.* 1995). Despite this breakthrough, its benefits largely remained beyond the reach of the greater scientific community for a number of years, given that sequencing projects required six-figure budgets and months, if not years, of effort. However, the following decade saw the advent of the first high-throughput sequencing (*syn.* next generation sequencing) technologies (e.g. 454 pyrosequencing, Ion Torrent and Illumina sequencing), which represented a significant advance in terms of time and cost (Goodwin, McPherson and McCombie 2016). These facts completely changed the picture, as it were, and genome sequencing, though still not considered cheap, nevertheless became more affordable and therefore more widespread in its use in the scientific community. Within the next decade (2005–2015), more diverse platforms, also referred to as third (e.g. PacBio SMRT and Heliscope sequencing) and fourth generation sequencing (Oxford Nanopore sequencing) methods, were developed, which either enabled longer reads or made sequencing portable, e.g. for field work (Jain *et al.* 2016), and led to a widened distribution of the technique (Loman *et al.* 2012; Loman and Pallen 2015). As costs fell, the amount of data generated increased dramatically, with genome mining essential to its processing.

The term genome mining stems from data mining, which in turn is defined as the systematic collection, analysis and interpretation of mounds of data ('big data') with the goal of discovering significant new traits, relationships, patterns or trends. Data mining is also typically, but not necessarily, used to predict future events (Fayyad, Piatetsky-Shapiro and Smyth 1996). In the case of genome mining, whole genome sequencing data of an organism represents the big data that is interpreted by bioinformatics in order to search for new catalysts, targets or products.

Since the active principles of PGPR are mediated by enzymes and secondary metabolites, genome mining offers the opportunity to scrutinise the whole genome of a PGPR strain for genes encoding beneficial enzymes that facilitate resource acquisition and modulation of plant hormone levels, as well as for biosynthetic gene clusters (BGCs) encoding antibiotically active compounds. Since many genes are silent under standard laboratory conditions, due to the absence of appropriate natural triggers or stress signals, some functions of PGPR may be overlooked where only classical experimental methods are employed. Whole genome analyses offer the advantage that the full arsenal of a PGPR strain can be unveiled and investigated. Furthermore, as regards future applications in the field, the safety and possible virulence of a PGPR strain can also be evaluated more thoroughly, as the full potential of the strain is illuminated.

How genome mining has contributed to the understanding of the active principles of PGPR strains shall be exemplified in this minireview. In the first part of this article, currently applicable genome mining strategies will be briefly reviewed, while in the second section, for selected biocontrol strains, a summary of their biology and chemistry will be provided, before focusing on genome-driven discoveries therein. Particular emphasis will be placed on the deduced indirect biocontrol mechanisms implemented by these strains. The selected examples represent, in the authors' opinion, interesting milestones in the PGPR genome mining field, and are accompanied by results from the PGPR research programmes of our laboratories.

STRATEGIES, METHODS AND TOOLS TO IDENTIFY BIOCONTROL MECHANISMS BY GENOME MINING

Since most biocontrol mechanisms are mediated by bioactive secondary metabolites, the identification of genetic blueprints of natural products is essential to their recognition, i.e. the gene clusters coding for bioactive secondary metabolites and enzymes. This is possible because many BGCs are either already known or the rationale of their biosynthetic assembly lines is at least understood. The latter holds true particularly in the case of polyketidic or non-ribosomal peptide-based metabolites which are assembled by modular megasynthases, termed polyketide synthases (PKS) and non-ribosomal synthetases (NRPS), respectively (Fischbach and Walsh 2006). In many instances, the number of modules in an assembly line corresponds exactly to the number of building blocks (e.g. acetate units in polyketides or amino acids in peptides) incorporated into the final product. This modular architecture often allows for the prediction of the resulting compound, even those yet unknown. While in the past, elaborate manual BLAST searches were necessary for the identification of biosynthetic pathways, increasing numbers of bioinformatic software tools (e.g. NRPS-PKS by Ansari *et al.* 2004; PKS/NRPS by Bachmann and Ravel 2009; ´NP. searcher´ by Li *et al.* 2009, ´antiSMASH´ by Weber *et al.* 2015 or ´PRISM´ by Skinnider *et al.* 2015) and databases (e.g. NORINE by Caboche *et al.* 2008; ClusterMine360 by Conway and Boddy 2013; DoBISCUIT by Ichikawa *et al.* 2013; StreptomeDB by Lucas *et al.* 2013) have become available which facilitate and automate the search for BGCs. Nowadays, the easiest and most comprehensive analysis of a whole microbial genome for secondary metabolites is gained by either submitting a sequence file or the accession number of a genome of interest to the web-based analysis software antiSMASH. In contrast to the other tools, it is not only restricted to polyketides and non-ribosomal peptides but also

uncovers the genetic blueprint for ribosomal peptides, terpenoids and products of specialised pathways (e.g. homoserine lactones, phenazines, ectoine); predicts the structure of the encoded compound; and shows homologues of the cluster of interest, present in other strains. While secondary metabolites are comprehensively covered, recognition mechanisms for biocontrol enzymes and some specialised pathways (e.g. for plant hormones, selected compatible solutes or volatiles) are not yet included in antiSMASH and need to be identified by classic BLAST searches.

In order to provide proof that a certain secondary metabolite is involved in a biocontrol mechanism, the corresponding BGC must be interrupted and the resulting mutant analysed chemically (HPLC, LC/MS) or biologically (e.g. by observing the absence of an antifungal effect). Alternatively, or particularly where the resulting product is unknown or several compounds with the same biological effect are involved, the generated secondary metabolite must be obtained in a pure form and investigated by way of assays. In this case, in order to identify the products of a given BGC, five major approaches have become established (Gross 2007, 2009; Challis 2008; Zerikly and Challis 2009).

Prediction and screening approach

Upon *in silico* prediction of the resulting structure of a BGC, its physico-chemical properties, e.g. a defined mass or mass range, UV absorbance spectrum, polarity or even potential biological activity due to the presence of a certain pharmacophore, can be deduced. Consequently, extraction and fractionation processes can be optimised, particularly regarding the choice of extraction solvent, chromatographic materials and the detection method (bioassay guided, or preferably informed by UV or MS). Although conceptually simple, this approach has proved very successful (for example, see Gross 2007). It relies on as precise a bioinformatics prediction as possible. However, where an encoded product cannot be predicted, the following approaches represent further options.

Genomisotopic approach

Every complex secondary metabolite is made up of simple building blocks (*syn.* precursors, such as acetate units, amino acids or isoprene units). Even if an encoded product cannot be predicted, bioinformatics analysis at the very least facilitates the prediction of which precursor will most likely be incorporated into the final product. This putative precursor can then be fed, in a stable isotope-labelled form, to the PGPR strain during fermentation, and subsequently, the compound of interest can be tracked through the isolation process, employing sophisticated NMR experiments. It is vital to first ascertain whether the precursor is non-toxic to the PGPR strain, and that it is specifically incorporated into the compound of interest alone. This approach was initially developed for NRPS-based BGCs, employing ^{15}N-labelled amino acids (Gross *et al.* 2007), but was recently successfully extended to polyketides, meroterpenoids and fosfazinomycins using ^{13}C-labelled (Klitgaard *et al.* 2015) or ^{2}H-labelled (Gao *et al.* 2014; Saleh *et al.* 2016) precursors.

Knockout and comparative metabolic profiling approach

This approach involves inactivation of a gene within the BGC of interest, followed by comparison of the metabolites in the extracts of the wild-type (WT) organism and the non-producing mutant, by way of HPLC and/or LC/MS analysis. Metabolites present in the metabolome of the WT, but lacking in that of the mutant, are indicative of the products of the targeted BGC; accordingly, these can be isolated and structurally characterised. Prerequisites of such experiments are the amenability of the PGPR strain to genetic manipulation, and variation between the metabolic profiles of the WT and mutant strains significant enough for the identification of novel products.

All the above-mentioned approaches rely on the expression of the BGC of interest. However, empirically, the majority of the identified BGCs of a given strain are often not expressed under standard laboratory conditions and are classified as 'silent'. Where a silent BGC is encountered, the 'one strain many compounds' (OSMAC) concept can be applied (Bode *et al.* 2002). The OSMAC approach is based on the hypothesis that secondary metabolite production occurs as a specific response to an altered environment. Variation of culture conditions can mimic such a change, and lead to the activation or upregulation of the expression of a BGC. Besides this, numerous methods to activate silent gene clusters have been suggested and successfully applied, e.g. co-cultivation, addition of subinhibitory concentration of antibiotics, regulator manipulations and many more (for reviews, see Scherlach and Hertweck 2009; Chiang *et al.* 2011 and Rutledge and Challis 2015). However, if minimal expression, or its complete absence, precludes compound identification, the following two methods can be applied.

Heterologous expression

This approach includes the transfer of a complete BGC from the source PGPR strain into a suitable bacterial host organism. However, its technical realisation remains challenging, since many BGC of interest are large (>40 kb) and therefore difficult to clone into a single vector. However, an increasing number of tools, such as the phage p1-derived artificial chromosomes technique (Jones *et al.* 2013) and the transformation-associated recombination cloning technique (Yamanaka *et al.* 2014), have been developed to overcome this obstacle.

In vitro reconstitution

In this approach, the constituent biosynthetic enzymes of the BGC of interest are produced in pure recombinant form and incubated with appropriate substrates and cofactors in order to obtain the resultant compound. Since this procedure is very laborious, it is preferentially applied to smaller silent BGC, encoding up to five proteins. Successful examples exist in the field of ribosomal peptides (McClerren *et al.* 2006) and meroterpenoids (Teufel *et al.* 2014).

Pseudomonas protegens Pf-5

Pseudomonas protegens Pf-5 (formerly known as *P. fluorescens* Pf-5) inhabits root and seed surfaces and was originally isolated from soil in Texas, USA. It was first described for its capacity to suppress soil-borne diseases affecting cotton caused by *Rhizoctonia solani* (Howell and Stipanovic 1979) and *Pythium ultimum* (Howell and Stipanovic 1980). Over the two decades that followed, it could be shown that strain Pf-5 also suppresses a number of other soil- or residue-borne fungal pathogens, including *Pyrenophora tritici-repentis* (Pfender, Kraus and Loper 1993), *Sclerotinia homoeocarpa*, *Drechslera poae* (Rodriguez and Pfender 1997) and *Fusarium oxysporum* f. sp. *radicis-lycopersici* (Sharifi-Therani *et al.* 1998). Concerning bacterial pathogens, it is known to be active towards *Erwinia carotovora* (Xu and Gross 1986). Strain Pf-5 was found to produce the antibiotics pyrrolnitrin (Howell

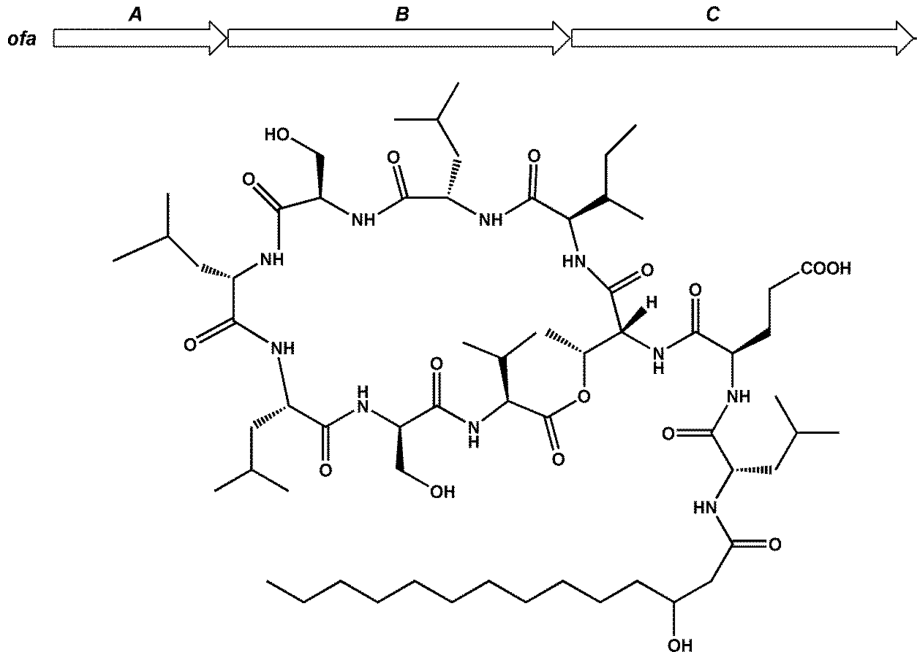

Figure 1. Diagram of the orfamide BGC found in *P. protegens* Pf-5 (top) and chemical structure of orfamide A (bottom).

and Stipanovic 1979), pyoluteorin (Howell and Stipanovic 1980; Kraus and Loper 1995), hydrogen cyanide (Kraus and Loper 1992), 2,4-diacetylphloroglucinol (Nowak-Thompson *et al.* 1994) and two siderophores: a pyoverdine of unconfirmed structure and a pyochelin-like compound.

In 2005, the whole genome sequence of strain Pf-5 (7.1 Mb) was published (Paulsen *et al.* 2005) and at the time represented the fourth and largest published *Pseudomonas* genome, and the first available for a *Pseudomonas* PGPR strain. Intriguingly, bioinformatics analyses revealed that, aside from the six expected BGCs, eight additional biosynthetic loci coding for proteins and small molecules, with potential roles in ecology and plant growth-promoting properties, were hidden in the genome, i.e. two bacteriocin-, an *mcf*-like protein, an arylpolyene-, two NRPS-, a *trans*-PKS- and a toxoflavin-like gene cluster. Six gene clusters thereof were resolved in the past decade.

Parret, Temmerman and De Mot (2005) adopted a genome mining strategy to analyse the two *llpA*-like genes which coded for the lectin-like bacteriocins $LlpA1_{Pf\text{-}5}$ and $LlpA2_{Pf\text{-}5}$. Heterologous expression of these genes in *Escherichia coli* cells demonstrated the functionality of the genes, while antimicrobial assays showed that the 31-kDa proteins are able to inhibit *P. fluorescens* strains and the mushroom pathogen *P. tolaasii*.

As regards the remaining orphan gene clusters, the large NRPS-type cluster was next to be analysed. It contained three contiguous NRPS genes (*ofaABC*) and was predicted to encode a novel lipodecapeptide. In order to obtain the resulting metabolite, a genomisotopic approach was applied and run in parallel with a traditional bioassay-guided fractionation, with which it was compared (Gross *et al.* 2007), leading to the discovery of the lipopeptides orfamides A-C. The orfamides are founding members of a new class of cyclic lipopeptides characterised by a 3-hydroxy dodecanoic or tetradecanoic acid connected to a cyclic peptide comprised of 10 amino acids. In assay studies, purified orfamide A (Fig. 1) showed no activity towards plant pathogenic fungi, but lysed zoospores of plant pathogenic oomycetes of the genus *Phytophtora* within seconds (Gross *et al.* 2007). Furthermore, it was shown that an orfamide A-deficient mutant of Pf-5 exhibits reduced swarming motility compared to the parental strain, which proved that this metabolite is physiologically decisive for the motility of the strain and growth on agar or seed surfaces (Nielsen *et al.* 2005).

Subsequently, in a continued effort to identify the products from orphan pathways of the Pf-5 genome, the Gross and Loper group, simultaneously with the Hertweck group, investigated the products synthesised from a hybrid *trans*-PKS/NRPS gene cluster (*rzxABCDEFGHI*) using a knockout and profiling strategy (Gross 2007). The application of this genome mining tool revealed that this cluster is responsible for the production of four known, and one new, rhizoxin analogue(s) (Brendel *et al.* 2007; Loper *et al.* 2008). Rhizoxins are 16-membered macrolides with potent antifungal, phytotoxic and antitumour activities (Iwasaki *et al.* 1984; Tsuruo *et al.* 1986). Of the five rhizoxin congeners, WF 1360F and its 22Z isomer (Fig. 2) were most toxic against the two agriculturally important plant pathogens *Botrytis cinerea* and *Phytophthora ramorum*. Furthermore, the Pf-5 rhizoxin derivatives caused swelling of rice roots, a symptom characteristic of rhizoxin itself (Iwasaki *et al.* 1984), but were less toxic to pea and cucumber (Loper *et al.* 2008). The latter plants exhibited only minor deformation of developing roots when exposed to the obtained rhizoxin congeners at concentrations (20 μg mL^{-1}) that were 20- to 40-fold greater than those inhibiting germination or germ tube elongation of the fungal and oomycete plant pathogens. Consequently, it is quite possible that Pf-5 could produce rhizoxin congeners on plant roots in concentrations adequate to inhibit microbial pathogens without deleterious effects on certain plant hosts. This hypothesis was very recently corroborated by Takeuchi and coworkers, who demonstrated that, in an *rzxB* mutant of the related *Pseudomonas* strain Os17 strain, which lacks rhizoxin production, the growth inhibition activity against fungal and oomycete pathogens and the overall plant protection efficacy was attenuated in comparison with the related WT bacterium (Takeuchi *et al.* 2015).

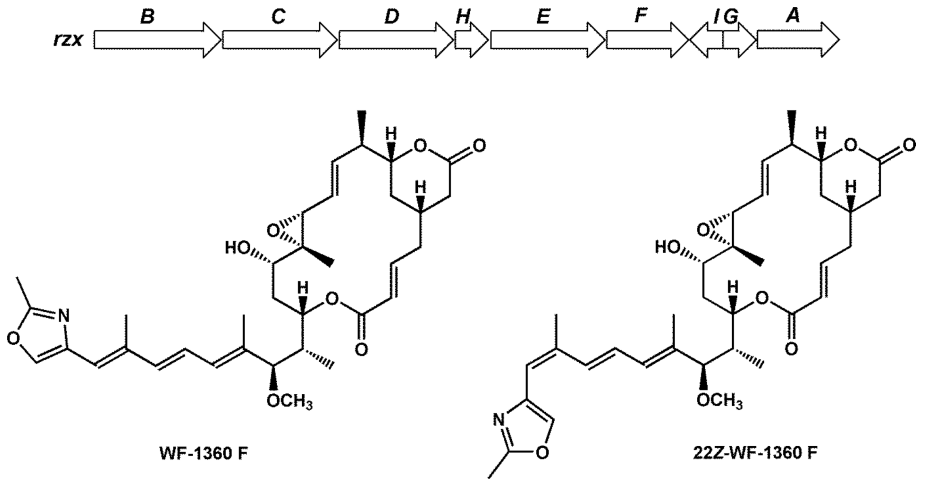

Figure 2. Diagram of the rhizoxin BGC found in *P. protegens* Pf-5 (top) and chemical structures of WF-1360 and its 22Z-isomer (bottom).

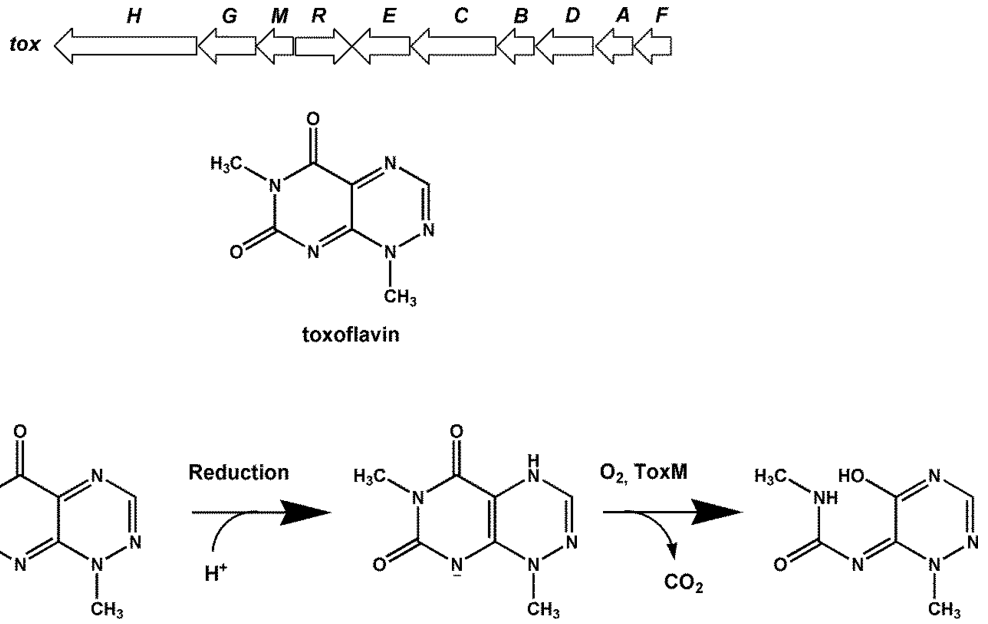

Figure 3. Diagram of the toxoflavin BGC found in *P. protegens* Pf-5 (top), chemical structure of toxoflavin (middle) and degradation pathway, catalysed by ToxM (bottom).

Intriguingly, a homologue of the *mcf* gene cluster (mcf denoting 'makes caterpillar floppy') is also present in the Pf-5 genome. Mcf is a protein, 327 kDa in size and produced by the bacterium *Photorhabdus luminescens*, that acts by inducing apoptotic cell death, ultimately causing the loss of body turgor and death of the insect (Daborn *et al.* 2002; Dowling *et al.* 2004). The teams of the Keel and the Loper lab proved, in a collaborative effort, that strain Pf-5 exhibits potent insecticidal activity towards the tobacco hornworm *Manduca sexta* and the greater wax moth *Galleria mellonella*, and, by deletion studies, that this activity was linked to the *mcf* homologous gene cluster (Péchy-Tarr *et al.* 2008). The novel insect toxin was termed 'Fit'—for *P. fluorescens* insecticidal toxin—and the gene cluster annotated *fitABCDEFGH*, accordingly.

More recently, the role of the orphan toxoflavin-like gene cluster (*toxABCDE*) was elucidated. Toxoflavin is a yellow broad-spectrum azapteridine-based toxin (Fig. 3) that possesses antibacterial properties and is best known for its role in the virulence of *Burkholderia glumae*, which causes wilt in many field crops (Jeong *et al.* 2003). As with rhizoxin, it is surprising, at first glance, that a plant-beneficial strain biosynthesises a toxin that is actually plant detrimental (i.e. towards tomato, sesame, perilla, eggplant and hot pepper at 10 μg mL^{-1}). However, the authors of the study showed that toxoflavin is produced only in traces and only under microaerobic conditions. Furthermore, employing a knockout mutant, it was demonstrated that toxoflavin production by strain Pf-5 inhibits several plant pathogenic bacteria (e.g. *P. syringae* pv. *tomato* DC3000) and that Pf-5 is capable of degrading toxoflavin by production of toxoflavin lyase ToxM (Fig. 3).

Thus, in summary, the five Pf-5 genome mining projects mentioned above led to the discovery of several new factors that contribute to the indirect biocontrol mechanisms of Pf-5, such that, nowadays, we have a much more complete picture of the full antibacterial, antifungal and anti-insecticidal armoury of the PGPR strain Pf-5. In addition, in genome-driven studies dealing with the two siderophores produced by Pf-5, stereochemical details were delineated: in comparison with the

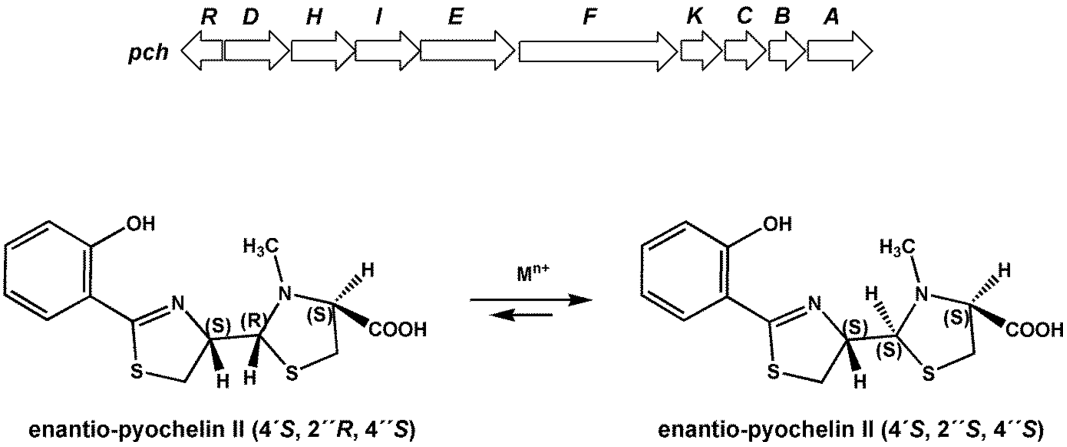

Figure 4. Diagram of the enantio-pyochelin BGC found in *P. protegens* Pf-5 (top) and chemical structure of the two interconverting isomers of enantio-pyochelin (bottom).

original pyochelin cluster of *P. aeruginosa* PAO1, it became obvious that, in the case of the pyochelin gene cluster of Pf-5, the peptide synthetase PchE lacked the epimerase domain required to generate the *R*-configuration at the chiral center C4´, which should result in the production of the optical isomer. This hypothesis was proven by subsequent isolation and chiral analysis of the encoded product, as well as by its total synthesis, and the compound was termed enantio-pyochelin (Fig. 4) (Youard *et al.* 2007), accordingly. Since these pyochelin enantiomers cannot be utilised as an iron source by regular pyochelin producer strains, it was speculated that this strategy possibly prevents potential competitors from occupying the same ecological niche. As concerns the second class of siderophores produced by Pf-5, the pyoverdines, it was originally assumed that they were identical to the pyoverdine secreted by *P. protegens* CHA0 (Wong-Lun-Sang *et al.* 1996). However, a definitive proof of this was never provided, and the absolute configuration of the two alanine residues could not be assigned. In a genome-driven study on ferric pyoverdine receptors, employing an approach combining high-resolution mass spectrometry and bioinformatics, it was possible to determine the structure as succinic acid–chromophore–L-Asp–D-FOHOrn–L-Lys-(L-Thr–L-Ala–D-Ala–D-FOHOrn–L-Lys) (FOHOrn = N5-Formyl-N5-hydroxy-ornithine), which is identical to that of the pyoverdine of *P. protegens* CHA0, with the absolute configuration of the alanine residues also clarified (Hartney *et al.* 2013).

Pseudomonas sp. SH-C52

This plant-beneficial bacterium was isolated from the rhizosphere of sugar beet plants grown in a soil suppressive of the fungal pathogen *R. solani*. It was discovered by a metagenomics analysis combined with culture-dependent functional analyses. Initially, its antifungal activity was thought to be governed by the product of an NRPS (Mendes *et al.* 2011), the structure of which was not yet determined. The application of an innovative live colony mass spectral networking approach enabled the detection and partial characterisation of what was termed thanamycin, a chlorinated lipopeptide comprising nine amino acids (Watrous *et al.* 2012). Further assay studies, employing deletion mutants, showed that thanamycin possessed a broader activity spectrum than anticipated, since it also showed activity towards the fungi *Botrytis cinerea* and *Geotrichum* sp., as well as towards the Gram-positive bacterium *Bacillus megaterium* (Van der Voort *et al.* 2015). In order to gain more insight into its biosynthetic capacity, the genome of strain SH-C52 was sequenced in parallel. The initial bioinformatics analysis of the 6.3 Mb genome clarified its uncertain taxonomic position, with phylogenetic analyses placing the strain within the *P. corrugata* clade. This finding supports the otherwise puzzling fact of the dual ecological role of *P. corrugata* strains, which can act as both biocontrol agents and plant pathogenic bacteria (Catara 2007). In addition, the genome analysis revealed, aside from the expected biosynthetic locus for thanamycin (*thaABC1C2*), the presence of BGCs for a further 11 secondary metabolites and protein toxins: hydrogen cyanide, siderophores of the achromobactin- and ornicorrugatin-type, bacteriocin, arylpolyene, two homologues for the insect toxins Tcc2 and Tcc4 and three NRPS-based lipopeptides 2 (*braABCDE*), 8 and 22 (*tnpABC*) amino acids in length; an incomplete mangotoxin gene cluster is also present (Van der Voort *et al.* 2015).

The NRPS system coding for a simple lipodipeptide included, adjacent to its structural genes, accessory genes encoding a glycosyltransferase (BraA), an unusual monooxygenase (BraC), a regulator (BraD) and an efflux transporter (BraE) (Fig. 5). Up to this point, glycosylated lipopeptides had been known to be produced by *Burkholderia* spp. (Lin *et al.* 2012), but had never been described in pseudomonads. This fact, and the aforementioned presence and arrangement of appropriate biosynthetic genes, prompted the groups of Gross and Raaijmakers to focus first on the *braABCDE* gene cluster. Metabolic profiling of a deletion mutant (Δ*braB*) readily unveiled the metabolite of interest. Its subsequent isolation and structure elucidation demonstrated that the cluster was responsible for the production of three 5,5,-bicyclic carbamate compounds, which were known under the SmithKline Beecham acronyms SB-253514, SB-253517 and SB-253518 (Fig. 5), first isolated in 2000 from *P. fluorescens* DSM 11579 (Thirkettle *et al.* 2000; Busby *et al.* 2000) and later from *P. brassicacearum* MA250 (Andersson, Levenfors and Broberg 2012). In accordance with the isolation site of the strain SH-C52, these compounds were dubbed brabantamides A, B and C. Subsequently, their biosynthesis was studied by a combination of bioinformatics, feeding experiments with isotopically labelled precursors and *in vivo* and *in vitro* analysis of the biosynthetic enzymes involved. In this way, a biosynthetic pathway was delineated that proved that brabantamides are essentially lipopeptides in disguise: their biosynthesis occurs via the formation of a linear dilipopeptide which is subsequently rearranged by BraC, a novel FAD-dependent Baeyer-Villiger monooxygenase (Schmidt *et al.* 2014). In order to gain insight into their ecological function,

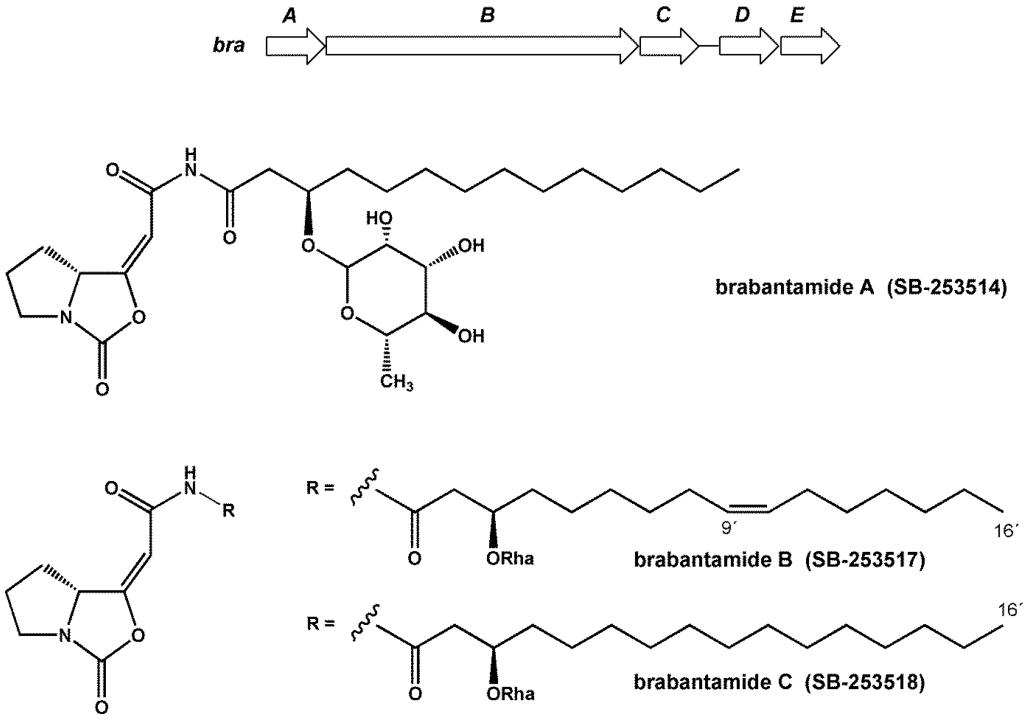

Figure 5. Diagram of the brabantamide BGC found in *P.* sp. SH-C52 (top) and chemical structures of brabantamides A-C (Rha: rhamnose).

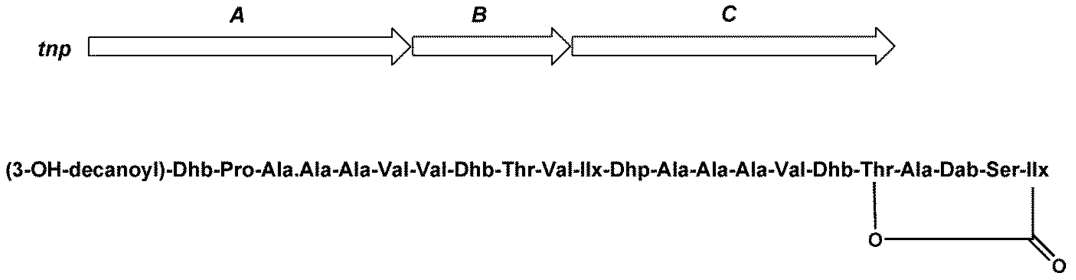

Figure 6 . Diagram of the thanapeptin BGC found in *P.* sp. SH-C52 (top) and chemical structure of thanapeptin (bottom).

several assays were performed with the purified derivatives. In antibacterial assays, brabantamides showed activity against Gram-positive bacteria, primarily towards *Bacilli* and *Arthrobacter* species (Reder-Christ *et al.* 2012; Schmidt *et al.* 2014), and moderate inhibition of the fungal plant pathogen *Microdochium nivale*, the causative agent of pink snow mould (Andersson, Levenfors and Broberg 2012). In further biological studies, Raaijmakers and coworkers demonstrated that brabantamide A possessed activity against the oomycete plant pathogens *Ph. capsici*, *Ph. infestans* and *Ph. ultimum* at a concentration of 50 μM. Since previous studies suggested that brabantamide A inhibits phospholipase A2, phosphate levels in *Ph. infestans* were also measured. Curiously, at least in the case of the *Phytophthora* system, no indication for phospholipase A2 inhibition was found. However, at a concentration of 1 μM, stimulatory effects on phospholipase D activity were observed, resulting in elevated phosphatidylbutanol levels. These findings possibly point to a new mode of action and a novel target of efforts to combat this plant pathogen (Van der Voort *et al.* 2015).

Alongside these studies on brabantamide, a large gene cluster encoding a lipopeptide comprised of 22 amino acids was analysed in depth and predicted to produce a new member of the syringopeptin family, which currently consists of syringopeptins, corpeptins and nunapeptins. The putative product was named thanapeptin, and discovered by the application of a combination of the so called ´prediction and screening´ and the ´knockout and metabolite profiling´ approach (Gross 2007). In this instance, first, from structural predictions based on the gene cluster, a mass range for the encoded compound was formulated and screened for using MALDI and live colony NanoDESI mass spectrometry. In a second step, extensive MS-based comparisons of WT strains with thanapeptin deletion mutants provided evidence that the observed compound is indeed correlated with the *tnpABC* gene cluster (Fig. 6). Further analysis of the tandem MS data enabled the deduction of the final structure (Van der Voort *et al.* 2015) (Fig. 6). The role of the thanapeptins in a PGPR context was evaluated by employing dixenic experiments and purified thanapeptins. From the results of these assays, it became apparent that thanapeptins are active towards the Gram-positive bacterium *B. megaterium* and are furthermore solely responsible for the antagonistic activity of the producing strain against a suite of oomycete pathogens, such as *Ph. infestans*, *Saprolegnia parasitica* and *Pyt. ultimum*.

Genome mining in SH-C52 represents an exemplary case of the early integration of genome mining into the study of a strain, occurring directly upon its discovery. Derived genomic

data demonstrated its huge biosynthetic capacity, and facilitated the swift and straightforward genomic-guided detection of key metabolites, which in turn clarified the cause of its plant beneficial properties.

Bacillus amyloliquefaciens subsp. *plantarum* (*methylotrophicus*) FZB42^T

Strain FZB42 was isolated as a part of a suite of *Bacillus* isolates from plant pathogen-infested soil, and shown to promote plant growth (Krebs *et al.* 1998) and inhibit that of plant-pathogenic fungi and oomycetes such as *F. oxisporum*, *Gaeumannomyces graminis*, *R. solani*, *Alternaria alternatae*, *Bo. cinerea* and *Pyt. aphanidermatum* (Grosch, Junge and Kofoet 2001). It was first identified as *B. subtilis*, later reclassified as a *B. amyloliquefaciens* strain (Idris *et al.* 2002) and, very recently, based on the latest phylogenomic analyses, subclassified as subspecies *plantarum* (Borriss *et al.* 2011; Dunlap *et al.* 2015; Wu *et al.* 2015b). Its plant beneficial effect was initially ascribed to the secretion of phytase (Idris *et al.* 2002; Makarewicz *et al.* 2006), an enzyme that makes phosphorus available to the plant by degradation of extracellular phytate (myo-inositol hexakiphosphate). Using a combined chemical and genetic approach, Borriss and coworkers were able to demonstrate that FZB42 can produce reasonable amounts of indole-3-acetic acid when tryptophan is available, and that it contributes to plant growth promotion (Idris *et al.* 2007).

As early as 2004, sample sequences of the FZB42 genome were made available through shotgun sequencing. These data provided the first insights into the metabolic capacity of the strain and subsequently disclosed, little by little, the indirect biocontrol mechanisms of FZB42. From the sampled sequence data, three NRPS gene clusters coding for the known antifungal lipopeptides surfactin, fengycin and bacillomycin D were readily identified and their expression confirmed by MALDI-TOF mass spectrometry analyses (Koumoutsi *et al.* 2004). The significance of the lipopeptides was investigated further by mutagenesis experiments. While the single mutants retained their antifungal activity, a double mutant lacking both bacillomycin D and fengycin production resulted in a strong loss of antifungal activity, which pointed to a synergistic mode of action.

From the same study, it also became apparent that three orphan *trans*-AT-PKS-based gene clusters were present in the genome. These clusters were investigated in a follow-up study, again employing mutagenesis experiments in combination with MALDI-TOF, along with LC-ESI-MS analyses. These efforts revealed that one cluster thereof, a hybrid NRPS-*trans*-AT-PKS cluster, coded for the polyene antibiotic bacillaene (*bae* cluster), while the second cluster (*dif* cluster) encoded the antibacterial macrolactone difficin and oxy-difficin (Zimmerman *et al.* 1987), respectively (Chen *et al.* 2006). In the case of both difficins and bacillaene, strong antibacterial properties were observed towards *B. megaterium* and the phytopathogenic bacterium *Erwinia carotovora*. More recently, difficin was also shown to be active against *Xanthomonas oryzae* rice pathogens (Wu *et al.* 2015a). At first, the product of the third *trans*-AT gene cluster was enigmatic. In this instance, the 'knockout and metabolite profiling' approach initially used was inconclusive, indicating no variation between the WT metabolite profile and that of the deletion mutant. In a second attempt, however, a difference in the metabolic patterns was found, and the target compounds identified as macrolactins (Schneider *et al.* 2007), which are effective against Gram-positive bacterial pathogens (Romero-Tabarez *et al.* 2006).

The whole genome sequence of FZB42 was determined in 2007, the first of a Gram-positive strain in the PGPR class (Chen *et al.* 2007). In-depth analysis revealed that a pathway for the well-established plant growth-promoting volatile compound acetoin is present, and that FZB42 harbours the *bac* and *dhb* gene clusters, responsible for the biosynthesis of the antibacterial dipeptide bacilysin (Özcengiz and Ögülür 2015) and the iron-binding siderophore bacillibactin. Bacilysin is known to be active against yeasts (Loeffler *et al.* 1986), cyanobacteria (Wu *et al.* 2014) and plant pathogenic bacteria such as xanthomonads (Wu *et al.* 2015a). In addition, seven orphan BGCs were detected in the genome. Of these, two code for terpenoid cyclases, four thereof are predicted to encode ribosomal peptides, and the remaining genetic locus (*nrs* gene cluster) directs the synthesis of a cysteine-rich hybrid NRPS/PKS compound. To date, no metabolite has been ascribed to either the *nrs* or the terpenoid sequences. However, the ribosomal gene clusters have been successfully elucidated. Two represent partial mersacidin and subtilisin gene clusters, missing the genes necessary for the synthesis of the peptides themselves (Chen *et al.* 2009). As a result, these biosynthetic loci yield no compounds, but do direct immunity against these particular lantibiotics. The products of the remaining predicted ribosomal BGCs were elucidated incrementally within the last 5 years.

One of these clusters (*pznFKGHIA-JCDBEL*) was predicted to belong to the group of thiazole/oxazole modified microcins (Lee *et al.* 2008; Haft, Basu and Mitchell 2010). Analysis of a FZB42-mutant strain deficient in non-ribosomal biosynthesis led to the detection of the corresponding products plantazolicins A and B (Fig. 7) (Kalyon *et al.* 2011) which showed antibacterial activity towards closely related *Bacillus* spp., particularly *B. anthracis* (Molohon *et al.* 2011), and a moderate nematicidal activity (Liu *et al.* 2013).

The second ribosomal gene cluster (*acnABCDEF*) was assumed to code for a circular bacteriocin. The application of a 'prediction and screening' approach, employing MALDI-TOF MS screening for a mass range of 6380–6400 Da, led to the identification and subsequent isolation of amylocyclicin, a highly hydrophobic linear peptide that consists of 64 amino acids (Scholz *et al.* 2014). Typical for this substance class, growth inhibitory activity was observed towards Gram-positive bacteria including several *Bacillus* spp., *Clavibacter michiganensis* and *Micrococcus luteus*.

In further studies, the efficacy of strain FZB42 was proven in numerous agricultural applications, and nowadays, it is commercialised and distributed as RhizoVital®42. The case of FZB42 is therefore a fine demonstration of the fact that basic science in the PGPR field can be realised as translational science, that a resulting marketed product is achievable. In this example, from the initial identification of the strain and the sequencing of its whole genome, its full biosynthetic capacity became apparent, and its further elucidation gained momentum with the classification of each biocontrol factor produced. Today, it is fully understood why and how FZB42 exerts antifungal, antibacterial and nematicidal properties. Notably, the science does not stop at this stage. Since FZB42 is genetically amenable, genetic engineering of FZB42 is envisioned as a means to enhance its plant growth-promoting properties. These strategies include regulator manipulation, optimisation of the metabolic flux and heterologous expression of further proteins and toxins (Wu *et al.* 2015b).

CONCLUSION

All the above-mentioned examples clearly demonstrate that a strain's full biosynthetic capacity for secondary metabolites and

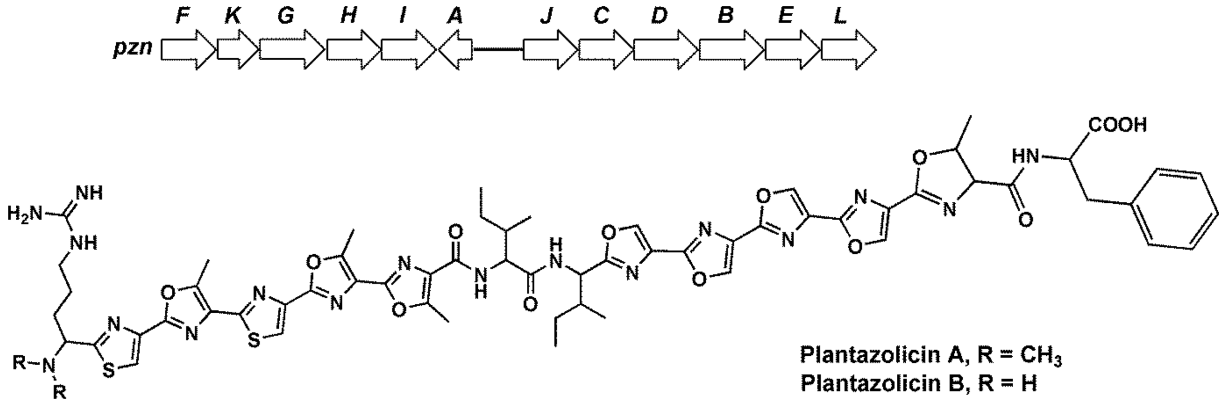

Figure 7. Diagram of the plantazolicin BGC found in *B. amyloliquefaciens* FZB42 (top) and chemical structure of plantazolicins A and B (bottom).

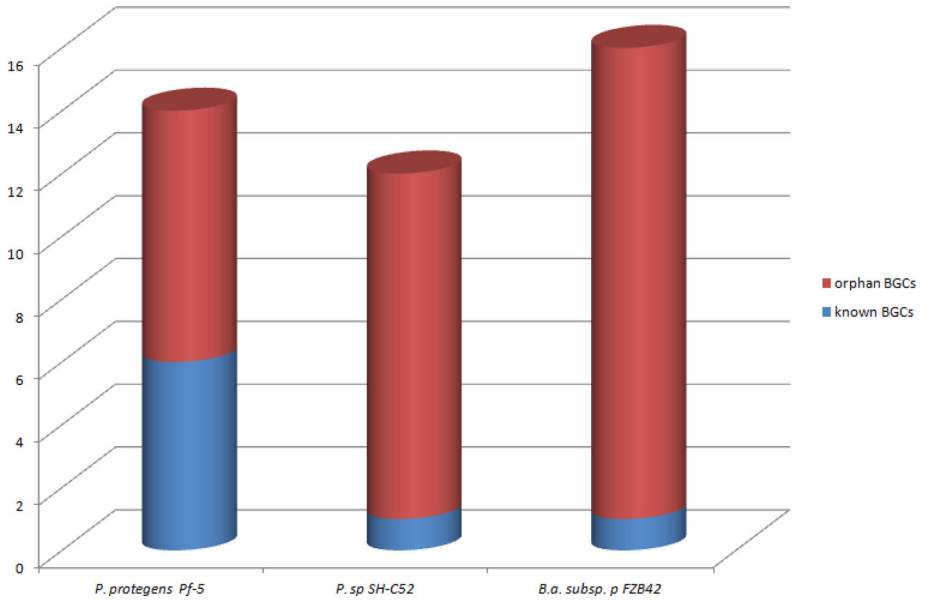

Figure 8. BGCs coding for secondary metabolites and protein toxins of strains Pf-5, SH-C52 and FZB42. Y-axis indicates the number of BGCs; blue columns show the number of BGCs known before the whole genome sequence was available, while red columns indicate the additional number of orphan BGCs that became apparent after a whole genome sequence was obtained.

proteins, i.e. the key mediators for direct and indirect PGPR mechanisms, becomes visible only after the analysis of its whole genome sequence (Fig. 8). The ability to identify encoding gene clusters has facilitated the genome-driven isolation of these compounds. In particular, as the example of strain Pf-5 shows, even though certain PGPR strains had been heavily investigated over a long period on a biological and chemical level, the majority of key biocontrol factors were still hidden in their genomes (Fig. 8). In many cases, information gleaned from orphan gene clusters concerning their predicted products subsequently informed cultivation strategies and, most crucially, appropriate extraction and screening procedures, these being adapted or optimised to ultimately lead to the isolation of targeted compounds. In turn, this allowed the testing of pure compounds and revealed their ecological functions in a PGPR context. In addition, iterative genome mining studies won new and valuable insights into the biosynthesis and the preceding metabolic flux of these compounds, facilitating potential future optimisation of the PGPR strains.

Additionally, concerning the application of PGPR strains in the field, genome mining can be used to support classification in terms of their safety. The absence of so-called pathogenicity islands (Dobrindt *et al.* 2004; Schmidt and Hensel 2004; Che, Hasan and Chen 2014) or genes encoding possible virulence factors or toxins within a given genome, for example, can provide grounds to catalogue a bacterial strain as a safe organism and therefore facilitate the commercial registration process.

In summary, it can be noted that genome mining has tremendously advanced and enriched the scientific field of PGPR research. As a direct consequence, genomic analysis has become an established tool in the characterisation of plant-beneficial microbes. In addition, since sequencing is expected to become more and more affordable, more precise and even portable, this development will continue to contribute to the identification, investigation and development of highly effective PGPR strains, which should in turn bolster sustainable agriculture and serve to decrease the use of agrochemicals worldwide.

ACKNOWLEDGEMENTS

Owing to space limitations, the authors apologise for not being able to present further interesting PGPR strains and to cite every relevant paper.

FUNDING

This work was supported by the German Research Foundation (grant GR2673/2-1 to HG) and the National Natural Science Foundation of China (grant 31672074 to YL). GJ is a recipient of a doctoral fellowship from the DAAD.

Conflict of interest. None declared.

REFERENCES

Andersson PF, Levenfors J, Broberg A. Metabolites from *Pseudomonas brassicacearum* with activity against the pink snow mould causing pathogen *Microdochium nivale*. *BioControl* 2012;**57**:463–9.

Ansari MZ, Yadav G, Gokhale RS *et al.* NRPS-PKS: a knowledge-based resource for analysis of NRPS/PKS megasynthases. *Nucleic Acids Res* 2004;**32**:W405–13.

Bachmann BO, Ravel J. In silico prediction of microbial secondary metabolic pathways from DNA sequence data. *Method Enzymol* 2009;**458**:181–217.

Beneduzi A, Ambrosini A, Passaglia LMP. Plant growth-promoting rhizobacteria (PGPR): their potential as antagonists and biocontrol agents. *Gen Mol Biol* 2012;**35**:1044–51.

Bode HB, Bethe B, Höfs R *et al.* Big effects from small changes: possible ways to explore nature´s chemical diversity. *ChemBioChem* 2002;**3**:619–27.

Borriss R, Chen XH, Rueckert C *et al.* Relationship of *Bacillus amyloliquefaciens* clades associated with strains DSM7^{T} and FZB42^{T}: a proposal for *Bacillus amyloliquefaciens* subsp. *amyloliquefaciens* subsp. nov. and *Bacillus amyloliquefaciens* subsp. *plantarum* subsp. nov. based on complete genome sequence comparisons. *Int J Syst Evol Micr* 2011;**61**:1768–801.

Brendel N, Partida-Martinez LP, Scherlach K *et al.* A cryptic PKS-NRPS gene locus in the plant commensal *Pseudomonas fluorescens* Pf-5 codes for the biosynthesis of an antimitotic rhizoxin complex. *Org Biomol Chem* 2007;**5**:2211–3.

Busby DJ, Copley RC, Hueso JA *et al.* SB-253514 and analogues: novel inhibitors of lipoprotein associated phospholipase A2 produced by *Pseudomonas fluorescens* DSM 11579. II. Physicochemical properties and structure elucidation. *J Antibiot* 2000;**53**:670–6.

Caboche S, Pupin M, Leclère V *et al.* NORINE: a database of nonribosomal peptides. *Nucleic Acids Res* 2008;**36**:D326–31.

Catara V. *Pseudomonas corrugata*: plant pathogen and/or biological resource? *Mol Plant Pathol* 2007;**8**:233–44.

Challis GL. Mining microbial genomes for new natural products and biosynthetic pathways. *Microbiology* 2008;**154**:1555–69.

Che D, Hasan MS, Chen B. Identifying pathogenicity islands in bacterial pathogenomics using computational approaches. *Pathogens* 2014;**3**:36–56.

Chen XH, Koumoutsi A, Scholz R *et al.* Comparative analysis of the complete genome sequence of the plant growth-promoting bacterium *Bacillus amyloliquefaciens* FZB42. *Nat Biotechnol* 2007;**25**:1007–14.

Chen XH, Koumoutsi A, Scholz R *et al.* Genome analysis of *Bacillus amyloliquefaciens* FZB42 reveals its potential for biocontrol of plant pathogens. *J Biotechnol* 2009;**140**:27–37.

Chen XH, Vater J, Piel J *et al.* Structural and functional characterization of three polyketide synthase gene clusters in *Bacillus amyloliquefaciens* FZB42. *J Bacteriol* 2006;**188**:4024–36.

Chiang YM, Chang SL, Oakley BR *et al.* Recent advances in awakening silent biosynthetic gene clusters and linking orphan clusters to natural products in microorganisms. *Curr Opin Chem Biol* 2011;**15**:137–43.

Conway KR, Boddy CN. ClusterMine360: a database of microbial PKS/NRPS biosynthesis. *Nucleic Acids Res* 2013;**41**:D402–7.

Daborn PJ, Waterfield N, Silva CP *et al.* A single *Photorhabdus* gene, makes caterpillars floppy (*mcf*), allows *Escherichia coli* to persist within and kill insects. *P Natl Acad Sci USA* 2002;**99**:10742–7.

Dobbelaere S, Vanderleyden J, Okon Y. Plant growth-promoting effects of diazotrophs in the rhizosphere. *Crit Rev Plant Sci* 2003;**22**:107–49.

Dobrindt U, Hochhut B, Hentschel U *et al.* Genomic islands in pathogenic and environmental microorganisms. *Nat Rev Microbiol* 2004;**2**:414–24.

Dowling AJ, Daborn PJ, Waterfield NR *et al.* The insecticidal toxin Makes caterpillar floppy (Mcf) promotes apoptosis in mammalian cells. *Cell Microbiol* 2004;**6**:345–53.

Dunlap CA, Kim SJ, Kwon SW *et al.* Phylogenomic analysis shows that *Bacillus amyloliquefaciens* subsp. *plantarum* is a later heterotypic synonym of *Bacillus methylotrophicus*. *Int J Syst Evol Micr* 2015;**65**:2104–9.

Fayyad UM, Piatetsky-Shapiro G, Smyth P. From data mining to knowledge discovery in databases. *AI Mag* 1996;**17**:37–54

Fleischmann RD, Adams MD, White O *et al.* Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science* 1995;**269**:496–512.

Fischbach MA, Walsh CT. Assembly-line enzymology for polyketide and nonribosomal peptide antibiotics: logic, machinery, and mechanisms. *Chem Rev* 2006;**106**:3468–96.

Fraser CM, Gocayne JD, White O *et al.* The minimal gene complement of *Mycoplasma genitalium*. *Science* 1995;**270**:397–403.

Gao J, Ju KS, Yu X *et al.* Use of a phosphonate methyltransferase in the identification of the Fosfazinomycin bioynthetic gene cluster. *Angew Chem Int Edit* 2014;**53**:1334–7.

Goodwin S, McPherson JD, McCombie WR. Coming of age: ten years of next-generation sequencing technologies. *Nat Rev Genet* 2016;**17**:333–51.

Gray EJ, Smith DL. Intracellular and extracellular PGPR: commonalities and distinctions in the plant-bacterium signaling processes. *Soil Biol Biochem* 2005;**37**:395–412.

Grosch R, Junge H, Kofoet A. Evaluation of isolates of *Bacillus*-group against *Fusarium oxysporum* f. sp. *radicis-lycopersici* on tomato plants in soilless culture. *Gartenbauwissenschaft* 2001;**66**:262–9.

Gross H. Strategies to unravel the function of orphan biosynthesis pathways: recent examples and future prospects. *Appl Microbiol Biot* 2007;**75**:267–77.

Gross H. Genomic mining—a concept for the discovery of new bioactive natural products. *Curr Opin Drug Di De* 2009;**12**:207–19.

Gross H, Stockwell VO, Henkels MD *et al.* The genomisotopic approach: a systematic method to isolate products of orphan biosynthetic gene clusters. *Chem Biol* 2007;**14**:53–63.

Haft DH, Basu MK, Mitchell DA. Expansion of ribosomally produced natural products: a nitrile hydratase- and Nif11-related precursor family. *BMC Biol* 2010;**8**:70.

Hartney SL, Mazurier S, Girard MK *et al.* Ferric-pyoverdine recognition by Fpv outer membrane proteins of *Pseudomonas protegens* Pf-5. *J Bacteriol* 2013;**195**:765–76.

Howell CR, Stipanovic RD. Control of *Rhizoctonia solani* in cotton seedlings with *Pseudomonas fluorescens* and with an antibiotic produced by the bacterium. *Phytopathology* 1979;**69**:480–2.

Howell CR, Stipanovic RD. Suppression of *Phythium ultimum* induced damping-off of cotton seedlings by *Pseudomonas fluorescens* and its antibiotic pyoluteorin. *Phytopathology* 1980;**70**:712–5.

Ichikawa N, Sasagawa M, Yamamoto M *et al.* DoBISCUIT: a database of secondary metabolite biosynthetic gene clusters. *Nucleic Acids Res* 2013;**41**:D408–14.

Idris EE, Iglesias DJ, Talon M *et al.* Tryptophan-dependent production of indole-3-acetic acid (IAA) affects level of plant growth promotion by *Bacillus amyloliquefaciens* FZB42. *Mol Plant-Microbe In* 2007;**20**:619–26.

Idris EE, Makarewicz O, Farouk A *et al.* Extracellular phytase activity of *Bacillus amyloliquefaciens* FZB45 contributes to its plant-growth-promoting effect. *Microbiology* 2002;**148**: 2097–109.

Iwasaki S, Kobayashi H, Furukawa J *et al.* Studies on macrocyclic lactone antibiotics. VII Structure of a phytotoxin "Rhizoxin" produced by *Rhizopus chinensis*. *J Antibiot* 1984;**37**:354–62.

Jain M, Olsen HE, Paten B *et al.* The Oxford Nanopore MinION: delivery of nanopore sequencing to the genomics community. *Genome Biol* 2016;**17**:239.

Jeong Y, Kim J, Kim S *et al.* Toxoflavin produced by *Burkholderia glumae* causing rice grain rot is responsible for inducing bacterial wilt in many field crops. *Plant Dis* 2003;**87**:890–5.

Jones AC, Gust B, kulik A *et al.* Phage p1-derived artificial chromosomes facilitate heterologous expression of the FK506 gene cluster. *PLoS One* 2013;**8**:e69319.

Kalyon B, Helaly SE, Scholz R *et al.* Plantazolicin A and B: structure elucidation of ribosomally synthesized thiazole/oxazole peptides from *Bacillus amyloliquefaciens* FZB42. *Org Lett* 2011;**13**:2996–9.

Klitgaard A, Frandsen RJN, Holm DK *et al.* Combining UHPLC-high resolution MS and feeding of stable isotope labeled polyketide intermediates for linking precursors to end-products. *J Nat Prod* 2015;**78**:1518–25.

Kloepper JW, Lifshitz R, Zablotowicz RM. Free-living bacterial inocula for enhancing crop productivity. *Trends Biotechnol* 1989;**7**:39–43.

Kloepper JW, Schroth MN. Plant growth promoting rhizobacteria on radishes. *Proceedings of the Fourth International Conference on Plant Pathogen Bacteria, INRA*, Vol. 2. Gilbert-Clarey, Tours, France 1978, 879–82.

Koumoutsi A, Chen XH, Henne A *et al.* Structural and functional characterization of gene clusters directing nonribosomal synthesis of bioactive cyclic lipopeptides in *Bacillus amyloliquefaciens* strain FZB42. *J Bacteriol* 2004;**186**: 1084–96.

Kraus J, Loper JE. Lack of evidence for a role of antifungal metabolite production by *Pseudomonas fluorescens* Pf-5 in biological control of *Phythium* damping-off of cucumber. *Phytopathology* 1992;**82**:264–71.

Kraus J, Loper JE. Characterization of a genomic region required for production of the antibiotic pyoluteorin by the biological control agent *Pseudomonas fluorescens* Pf-5. *Appl Environ Microb* 1995;**61**:849–54.

Krebs B, Höding B, Kübart SM *et al.* Use of *Bacillus subtilis* as biocontrol agent. 1. Activities and characterization of *Bacillus subtilis* strains. *J Plant Dis Protect* 1998;**105**:181–97.

Lee SW, Mitchell DA, Markley AL *et al.* Discovery of a widely distributed toxin biosynthetic gene cluster. *P Natl Acad Sci USA* 2008;**105**:5879–84.

Li MHT, Ung PMU, Zajkowski J *et al.* Automated genome mining for natural products. *BMC Bioinformatics* 2009;**10**:185.

Lin Z, Falkinham JO III, Tawfik KA *et al.* Burkholdines from *Burkholderia ambifaria*: antifungal agents and possible virulence factors. *J Nat Prod* 2012;**75**:1518–23.

Liu Z, Budiharjo A, Wang P *et al.* The highly modified microcin peptide plantazolicin is associated with nematicidal activity of *Bacillus amyloliquefaciens* FZB42. *Appl Microbiol Biot* 2013;**97**:10081–90.

Loeffler W, Tschen SM, Vanittanakom N *et al.* Antifungal effect of bacilysin and fengymycin from *Bacillus subtilis* F-29-3. A comparison with activities of other *Bacillus* antibiotics. *J Phytopathol* 1986;**155**:204–13.

Loman NJ, Constantinidou C, Chan JZM *et al.* High-throughput bacterial genome sequencing: an embarrassment of choice, a world of opportunity. *Nat Rev Microbiol* 2012;**10**: 599–606.

Loman NJ, Pallen MJ. Twenty years of bacterial genome sequencing. *Nat Rev Microbiol* 2015;**13**:787–94.

Loper JE, Henkels MD, Shaffer BT *et al.* Isolation and identification of rhizoxin analogs from *Pseudomonas fluorescens* Pf-5 by using a genomic mining strategy. *Appl Environ Microb* 2008;**74**:3085–93.

Lucas X, Senger C, Erxleben A *et al.* StreptomeDB: a resource for natural compounds isolated from *Streptomyces* species. *Nucleic Acids Res* 2013;**41**:D1130–6.

Lugtenberg B, Kamilova F. Plant-growth-promoting rhizobacteria. *Annu Rev Microbiol* 2009;**63**:541–56.

McClerren AL, Cooper LE, Quan C *et al.* Discovery and *in vitro* biosynthesis of haloduracin, a two-component lantibiotic. *P Natl Acad Sci USA* 2006;**103**:17243–8.

Makarewicz O, Dubrac S, Msadek T *et al.* Dual role of the PhoP∼P response regulator: *Bacillus amyloliquefaciens* FZB45 phytase gene transcription is directed by positive and negative interaction with the *phyC* promoter. *J Bacteriol* 2006;**188**:6953–65.

Mendes R, Kruijt M, de Bruijn I *et al.* Deciphering the rhizosphere microbiome for disease-suppressive bacteria. *Science* 2011;**332**:1097–100.

Molohon KJ, Melby JO, Lee J *et al.* Structure determination and interception of biosynthetic intermediates for the plantazolicin class of highly discriminating antibiotics. *ACS Chem Biol* 2011;**16**:1307–13.

Nielsen TH, Nybroe O, Koch B *et al.* Genes involved in cyclic lipopeptide production are important for seed and straw colonization by *Pseudomonas* sp. strain DSS73. *Appl Environ Microb* 2005;**71**:4112–6.

Nowak-Thompson B, Gould SJ, Kraus J *et al.* Production of 2,4-diacetylphloroglucinol by the biocontrol agent *Pseudomonas fluorescens* Pf-5. *Can J Microbiol* 1994;**40**:1064–6.

Özcengiz G, Ögülür I. Biochemistry, genetics and regulation of bacilysin biosynthesis and its significance more than an antibiotic. *New Biotechnol* 2015;**32**:612–9.

Parret AHA, Temmerman K, De Mot R. Novel lectin-like bacteriocins of biocontrol strain *Pseudomonas fluorescens* Pf-5. *Appl Environ Microb* 2005;**71**:5197–207.

Paulsen IT, Press CM, Ravel J *et al.* Complete genome sequence of the plant commensal *Pseudomonas fluorescens* Pf-5. *Nat Biotechnol* 2005;**23**:873–8.

Péchy-Tarr M, Bruck DJ, Maurhofer M *et al.* Molecular analysis of a novel gene cluster encoding an insect toxin in plant-associated strains of *Pseudomonas fluorescens*. *Environ Microbiol* 2008;**10**:2368–86.

Pfender WF, Kraus J, Loper JE. A genomic region from *Pseudomonas fluorescens* Pf-5 required for pyrrolnitrin production

and inhibition of *Pyrenophora tritici-repentis* in wheat straw. *Phytopathology* 1993;**83**:1223–8.

Reder-Christ K, Schmidt Y, Dörr M *et al*. Model membrane studies for characterization of different antibiotic activities of lipopeptides from *Pseudomonas*. *BBA Biomembranes* 2012;**1818**:566–73.

Rodriguez F, Pfender WF. Antibiosis and antagonism of *Sclerotia homoeocarpa* and *Drechslera poae* by *Pseudomonas fluorescens* Pf-5 *in vitro* and *in planta*. *Phytopathology* 1997;**87**: 614–21.

Romero-Tabarez M, Jansen R, Sylla M *et al*. 7-O-malonyl macrolactin A, a new macrolactin antibiotic from *Bacillus subtilis* active against methicillin resistant *Staphylococcus aureus*, vancomycin-resistant enterococci, and small-colony variant of *Burkholderia cepacia*. *Antimicrob Agents Ch* 2006;**50**: 1701–9.

Rutledge PJ, Challis GL. Discovery of microbial natural products by activation of silent biosynthetic gene clusters. *Nat Rev Microbiol* 2015;**13**:509–23.

Saleh H, Petras D, Mainz A *et al*. Deuterium-labeled precursor feeding reveals a new *p*ABA-containing meroterpenoid from the mango pathogen *Xanthomonas citri* pv *mangiferaeindicae*. *J Nat Prod* 2016;**79**:1532–7.

Scherlach K, Hertweck C. Triggering cryptic natural product biosynthesis in microorganisms. *Org Biomol Chem* 2009;**7**:1753–60.

Schmidt H, Hensel M. Pathogenicity Islands in Bacterial Pathogenesis. *Clin Microbiol Rev* 2004;**17**:14–56.

Schmidt Y, van der Voort M, Crüsemann M *et al*. Biosynthetic origin of the antibiotic cyclocarbamate brabantamide A (SB-253514) in plant-associated *Pseudomonas*. *ChemBioChem* 2014;**15**:259–66.

Schneider K, Chen XH, Vater J *et al*. Macrolactin is the polyketide biosynthesis product of the pks2 cluster of *Bacillus amyloliquefaciens* FZB42. *J Nat Prod* 2007;**70**:1417–23.

Scholz R, Vater J, Budiharjo A *et al*. Amylocyclicin, a novel circular bacteriocin produced by *Bacillus amyloliquefaciens* FZB42. *J Bacteriol* 2014;**196**:1842–52.

Sharifi-Therani A, Zala M, Natsch A *et al*. Biocontrol of soilborne fungal plant diseases by 2,4-diacetylphloroglucinol-producing fluorescent pseudomonads with different restriction profiles of amplified 16S rDNA. *Eur J Plant Pathol* 1998;**104**:631–43.

Skinnider MA, Dejong CA, Rees PN *et al*. Genomes to natural products Prediction Informatics for Secondary Metabolomes (PRISM). *Nucleic Acids Res* 2015;**43**:9645–62.

Suslow TV, Kloepper JW, Schroth MN *et al*. Beneficial bacteria enhance plant growth. *Calif Agr* 1979;**33**:15–7.

Takeuchi K, Noda N, Katayose Y *et al*. Rhizoxin analogs contribute to the biocontrol activity of a newly isolated *Pseudomonas* strain. *Mol Plant-Microbe In* 2015;**28**: 333–42.

Teufel R, Kaysser L, Vilaume MT *et al*. One-pot enzymatic synthesis of merochlorin A and B. *Angew Chem Int Edit* 2014;**53**:11019–22.

Thirkettle J, Alvarez E, Boyd H *et al*. SB-253514 and analogues; novel inhibitors of lipoprotein-associated phopholipase A2 produced by *Pseudomonas fluorescens* DSM 11579. I. Fermentation of producing strain, isolation and biological activity. *J Antibiot* 2000;**53**:664–9.

Tsuruo T, Oh-Hara T, Iida H *et al*. Rhizoxin, a macrocyclic lactone antibiotic, as a new antitumor agent against human and murine tumor cells and their vincristine-resistant sublines. *Cancer Res* 1986;**46**:381–5.

Van der Voort M, Meijer HJG, Schmidt Y *et al*. Genome mining and metabolic profiling of the rhizosphere bacterium *Pseudomonas* sp. SH-C52 for antimicrobial compounds. *Front Microbiol* 2015;**6**:693.

Watrous J, Roach P, Alexandrov T *et al*. Mass spectral molecular networking of living microbial colonies. *P Nat Acad Sci USA* 2012;**109**:E1743–52.

Weber T, Blin K, Duddela S *et al*. antiSMASH 3.0—a comprehensive resource for the genome mining of biosynthetic gene clusters. *Nucleic Acids Res* 2015;**43**:W237–43.

Wong-Lun-Sang S, Bernardini Hennard JJC, Kyslik P *et al*. Bacterial siderophores: structure elucidation, 2D ^{1}H and ^{13}C NMR assignments of pyoverdines produced by *Pseudomonas fluorescens* CHA0. *Tetrahderon Lett* 1996;**37**:3329–32.

Wu L, Wu H, Chen L *et al*. Bacilysin from *Bacillus amyloliquefaciens* FZB42 has specific bactericidal activity against harmful algal bloom species. *Appl Environ Microb* 2014;**80**:7512–20.

Wu L, Wu H, Chen L *et al*. Difficidin and bacilysin from *Bacillus amyloliquefaciens* FZB42 have antibacterial activity against *Xanthomonas oryzae* rice pathogens. *Sci Rep* 2015a;**5**:12975.

Wu L, Wu HJ, Qiao J *et al*. Novel routes for improving biocontrol activity of *Bacillus* based bioinoculants. *Front Microbiol* 2015b;**6**:1395.

Xu GW, Gross DC. Selection of fluorescent pseudomonads antagonistic to *Erwinia carotovora* and suppressive of potato seed piece decay. *Phytopathology* 1986;**76**:414–22.

Yamanaka K, Reynolds KA, Kersten RD *et al*. Direct cloning and refactoring of a silent lipopeptide biosynthetic gene cluster yields the antibiotic taromycin A. *P Natl Acad Sci USA* 2014;**111**:1957–62.

Youard ZA, Mislin GLA, Majcherczyk PA *et al*. *Pseudomonas fluorescens* CHA0 produces enantio-pyochelin, the optical antipode of the *Pseudomonas aeruginosa* siderophore pyochelin. *J Biol Chem* 2007;**282**:35546–53.

Zerikly M, Challis GL. Strategies for the discovery of new natural products by genome mining. *ChemBioChem* 2009;**10**:625–33.

Zimmerman SB, Schwartz CD, Monaghan RL *et al*. Difficin and oxydifficin: novel broad spectrum antibacterial antibiotics produced by *Bacillus subtilis*. I. Production, taxonomy and antibacterial activity. *J Antibiot* 1987;**40**:1677–81.