environmental
microbiology

Environmental Microbiology (2018) **20**(1), 62–74 doi:10.1111/1462-2920.13957

Tropical soils are a reservoir for fluorescent *Pseudomonas* spp. biodiversity

Lucas Dantas Lopes,[1,2] Edward W. Davis II,[2,3]
Michele de C. Pereira e Silva,[1]
Alexandra J. Weisberg,[2] Luana Bresciani,[1]
Jeff H. Chang,[2,3] Joyce E. Loper[2,3] and
Fernando D. Andreote[1]*
[1] *Department of Soil Science, "Luiz de Queiroz" College of Agriculture, University of São Paulo, Piracicaba, SP, Brazil.*
[2] *Department of Botany and Plant Pathology, Oregon State University, Corvallis, OR 97331, USA.*
[3] *Molecular and Cellular Biology Program, Oregon State University, Corvallis, OR 97331, USA.*

Summary

Fluorescent *Pseudomonas* spp. are widely studied for their beneficial activities to plants. To explore the genetic diversity of *Pseudomonas* spp. in tropical regions, we collected 76 isolates from a Brazilian soil. Genomes were sequenced and compared to known strains, mostly collected from temperate regions. Phylogenetic analyses classified the isolates in the *P. fluorescens* (57) and *P. putida* (19) groups. Among the isolates in the *P. fluorescens* group, most (37) were classified in the *P. koreensis* subgroup and two in the *P. jessenii* subgroup. The remaining 18 isolates fell into two phylogenetic subclades distinct from currently recognized *P. fluorescens* subgroups, and probably represent new subgroups. Consistent with their phylogenetic distance from described subgroups, the genome sequences of strains in these subclades are asyntenous to the genome sequences of members of their neighbour subgroups. The tropical isolates have several functional genes also present in known fluorescent *Pseudomonas* spp. strains. However, members of the new subclades share exclusive genes not detected in other subgroups, pointing to the potential for novel functions. Additionally, we identified 12 potential new species among the 76 isolates from the tropical soil. The unexplored diversity found in the tropical soil is possibly related to biogeographical patterns.

Received 16 May, 2017; revised 7 October, 2017; accepted 8 October, 2017. *For correspondence. E-mail fdandreo@gmail.com; Tel. +55(19)34172123; Fax (19) 3417-2100/3417-2110.

Introduction

The bacterial genus *Pseudomonas* contains multiple species of bacteria (Palleroni, 1992; Silby *et al.*, 2011; Loper *et al.*, 2012). These species are genetically, ecologically and functionally diverse and can be found in many terrestrial and aquatic habitats (Spiers *et al.*, 2000). *Pseudomonas* species can be soil saprophytes; degrade pollutants; be mutualists or pathogens associated to plants, insects, animals and humans; and produce several compounds of industrial interest (Silby *et al.*, 2011; Loper *et al.*, 2012). Currently, the *Pseudomonas* species are classified in 10 groups, *P. fluorescens*, *P. aeruginosa*, *P. putida*, *P. syringae*, *P. anguilliseptica*, *P. oryzihabitans*, *P. stutzeri*, *P. oleovorans*, *P. straminae* and *P. pertucinogena* (Gomila *et al.*, 2015; Garrido-Sanz *et al.*, 2016). The *P. fluorescens* group is one of the most diverse and is further subclassified into ten subgroups (Gomila *et al.*, 2015; Garrido-Sanz *et al.*, 2016). As a consequence of its high diversity, the classification of species of *Pseudomonas* is challenging and is constantly being amended as new species are described and new methods are developed.

In agricultural systems, there is particular interest in fluorescent *Pseudomonas* spp. that live in the soil/plant environments because of their potential to improve plant productivity. Strains of fluorescent *Pseudomonas* characterized as plant growth promoting rhizobacteria (PGPR) are most commonly within the *P. putida* and *P. fluorescens* groups (Silby *et al.*, 2011). PGPR strains produce iron-chelating siderophores (Trapet *et al.*, 2016), synthesize phytohormones (Patten and Glick, 2002), solubilize phosphorus (de Werra *et al.*, 2009), protect against abiotic stresses (Cho *et al.*, 2015) or control against insects or microbial pathogens (Hofte and Altier, 2010; Loper *et al.*, 2012; Hol *et al.*, 2013).

The tropics occupy approximately one-third of Earth's land surface. The climate, vegetation, geomorphology, lithology and consequently the soils within these regions vary more than those of temperate and arctic regions (Kalpage, 1974). Such variation provides expanded habitats and niches highly contributing to the evolution and diversification of organisms. Therefore, tropical ecosystems are major reservoirs of biological diversity on our planet. The diversity of macroscopic species decreases from the tropics to the poles but the existence of a latitudinal gradient in microbial diversity is controversial. A latitudinal gradient has been described for *Streptomyces* (Andam *et al*., 2016) and geographical distance was shown to be the main factor for the structuring of soil fungal communities (Gumiere *et al*., 2016). On the other hand, edaphic variables were suggested to be the main factors controlling biogeography of soil bacterial communities (Fierer and Jackson, 2006). Fluorescent *Pseudomonas* spp. seem to have some degree of endemicity, that is, specificity of some species/strains to different geographical locations, and were suggested to be unequally distributed around the planet following biogeographical patterns (Cho and Tiedje, 2000).

Understanding the diversity of microorganisms in the tropics is key to resolving these fundamental concepts of global biogeography. Although *Pseudomonas* spp. are important in agricultural systems, knowledge of these bacteria in tropical regions has lagged behind those from temperate regions. The few reported examples of tropical strains include the PGPR strain *P. fluorescens* Ps006, isolated from banana roots in Colombia (Gamez *et al*., 2016); the biocontrol and PGPR strain *P. fluorescens* UM270 isolated from the rhizosphere of wild *Medicago* sp. in Mexico (Hernández-Salmerón *et al*., 2016); strains *P. fluorescens* BRIP34879 and *P. fluorescens* SRM1 obtained from cereal crops and spoiled raw milk, respectively, in Queensland, Australia, (Gardiner *et al*., 2013; Lo *et al*., 2015); and strain *P. fluorescens* RP47, isolated from the maize rhizosphere in Brazil (Araujo *et al*., 1994).

Most studies in microbial biogeography have used molecular biomarkers, which are unable to accurately analyse the intraspecies diversity. The goal of this study was to expand knowledge on the diversity of fluorescent *Pseudomonas* spp. in a tropical soil using phylogenetic and genomic approaches. We discovered that tropical soils harbour novel members of the *P. fluorescens* group that fall outside of previously defined subgroups and with gene inventories that are distinct from known strains in this group of bacteria. The genomic approach allowed the intraspecies comparison of fluorescent *Pseudomonas* spp., and the high degree of phylogenetic and species novelty in the tropical soil argue for an unequal distribution of soil fluorescent *Pseudomonas* throughout the world and support the possibility of endemism.

Results and discussion

Soil characterization and selection of fluorescent colonies

The soil samples used in this study were previously assessed in a microbiome analysis comparing rhizosphere and bulk soil habitats, showing that bacterial community structure and physiological profiles are different between sugarcane rhizosphere and bulk soil (Lopes *et al*., 2016). Based on these differences, we targeted both habitats for sampling and used an equitable number of isolates from each, aiming to access as much genetic diversity as possible.

Soil texture was dominated by clay (45%), followed by sand (29%) and silt (26%) fractions. Soil samples had low average content of organic matter (2.7%) and low pH (5.0), typical of tropical soils. March, when samples were collected, has average temperature of 23°C and approximately 135 mm of precipitation (https://pt.climate-data.org/location/748/, 07/19/2017). Fluorescent *Pseudomonas* spp. isolation resulted in an average of 1.9×10^4 CFU g soil^{-1} for bulk soil samples and 3.1×10^4 CFU g soil^{-1} for rhizosphere samples. Approximately 5% of the colonies in each plate showed fluorescence when exposed to UV light. These colonies were purified and stored. A total of 76 isolates – 40 from bulk soil and 36 from rhizosphere – were further characterized in this study.

Genome sequencing and assembly

A total of 560 million reads were generated from the multiplexed 76 libraries. The reads were processed, and independently *de novo* assembled for each genome sample. The genome assemblies had an average of 43 contigs (from 15 to 94 contigs) (Supporting Information Table S2). The total within-assembly sizes for each genome were comparable to sizes of finished genome sequences of related *Pseudomonas* spp. (Supporting Information Table S2). We thus, estimate the average coverage was approximately 70X for each sequenced isolate. The assemblies were annotated and yielded an average of 5559 coding sequences (CDSs) per genome sequence, consistent with the expected density of most bacterial genomes.

Most of the tropical isolates are members of the P. fluorescens *and* P. putida *groups*

To determine the identity and phylogeny of the isolates, a phylogenetic tree based on sequences of seven housekeeping genes was constructed (Loper *et al*., 2012). Seventy-four of the 76 isolates clustered with strains that belong to one of two recognized groups of *Pseudomonas*, *P. fluorescens* and *P. putida* (Fig. 1). The 55 isolates that clustered to the *P. fluorescens* group formed 18 nodes. The 19 isolates that clustered to *P. putida* group formed nine nodes. None of the included reference strains were positioned in the 27 nodes formed by the isolates collected and analysed in this study. However, the reference strains *P. fluorescens* Pf0–1 (Compeau *et al*., 1988) and *Pseudomonas* sp. MS586 (unpublished) were closely related to the

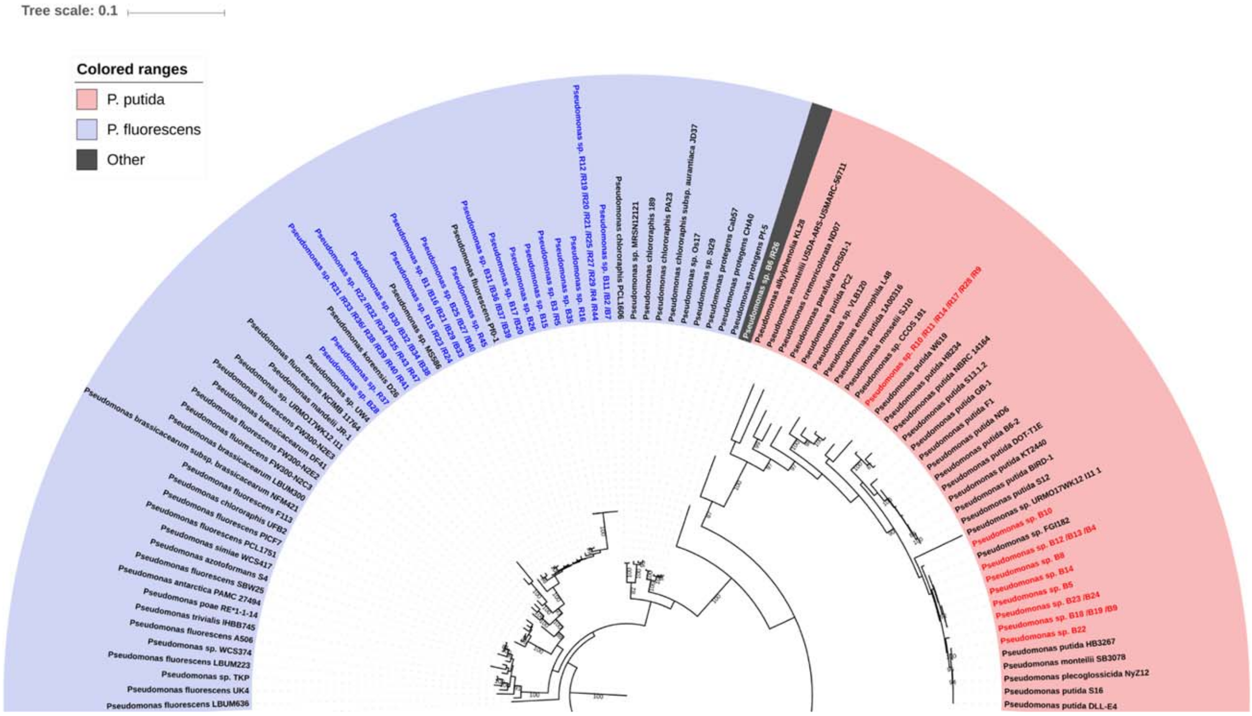

Fig. 1. Maximum likelihood phylogenetic tree of *Pseudomonas* spp. inferred by multilocus sequence analysis. Translated sequences for *dnaE*, *guaA*, *mutL*, *ppsA*, *pyrC*, *recA* and *rpoB* from 76 sequenced isolates and reference strains downloaded from NCBI were identified, aligned, concatenated and used to generate a phylogenetic tree. The light blue region indicates the *P. fluorescens* group, while the light red region indicates the *P. putida* group. The dark grey region comprises the isolates B6 and R26 that were not classified in the known subgroups in this tree (Other). Labels in black are the reference strains, blue and red labels are the ones isolated in this study. Branch lengths are supported by bootstrap values (numbers below each clade). [Colour figure can be viewed at wileyonlinelibrary.com]

isolates in the *P. fluorescens* group, while *Pseudomonas sp.* FGI182 (unpublished), *P. putida* HB3267 (Molina *et al.*, 2014) and *P. putida* W619 (Taghavi *et al.*, 2009) were closely related to isolates in the *P. putida* group (Fig. 1).

The remaining two isolates (B6 and R26), although belonging to *Pseudomonas*, did not associate to either of the aforementioned groups, or any other established group of *Pseudomonas* (Fig. 1). Although Fig. 1 highlights the *P. fluorescens* and *P. putida* groups, strains of all other *Pseudomonas* groups were analysed in the MLSA, and these two isolates did not cluster to any of those groups. The two formed their own node, which shared a common ancestor to the *P. putida* group based on this phylogenetic tree (Fig. 1).

Tropical soils harbour isolates belonging to new subclades of P. fluorescens *group*

To test for subgroup associations, we constructed a second phylogenetic tree using just four housekeeping genes (16S rRNA, *rpoB*, *rpoD* and *gyrB*), which allowed for inclusion of a greater diversity and quantity of type strains and for comparison with previous studies that focused on *P. fluorescens* subgrouping classification (Gomila *et al.*, 2015; Garrido-Sanz *et al.*, 2016). This tree was consistent with the previously described tree in assigning 55 isolates to the *P. fluorescens* group and 19 isolates to the *P. putida* group, and placing isolates B6 and R26 in a separate subclade (Fig. 2). Of the 55 isolates in the *P. fluorescens* group, 37 associated to the *P. koreensis* subgroup. Isolates R37 and B28 were classified in the *P. jessenii* subgroup (Fig. 2). Those from the tropical soil that associated to the *P. putida* group were intermingled with reference strains, but represent new haplotypes. This second phylogenetic tree agreed with the previous one in suggesting that the newly sequenced isolates are new haplotypes. The number of haplotypes was slightly higher using this approach, increasing to 10 and 24 in the *P. putida* and *P. fluorescens* group respectively (Fig. 2).

One subclade (X) containing 16 isolates within the *P. fluorescens* group did not cluster to any previously recognized subgroup. The separation of subclade X from its closest subgroup, *P. koreensis*, is well-supported (bootstrap value = 100). These results suggest that some of the isolates collected from the tropical soil are distinct from strains available in the databases, which were mostly identified from temperate regions. Additionally, B6 and R26 did not associate to any currently recognized subgroup.

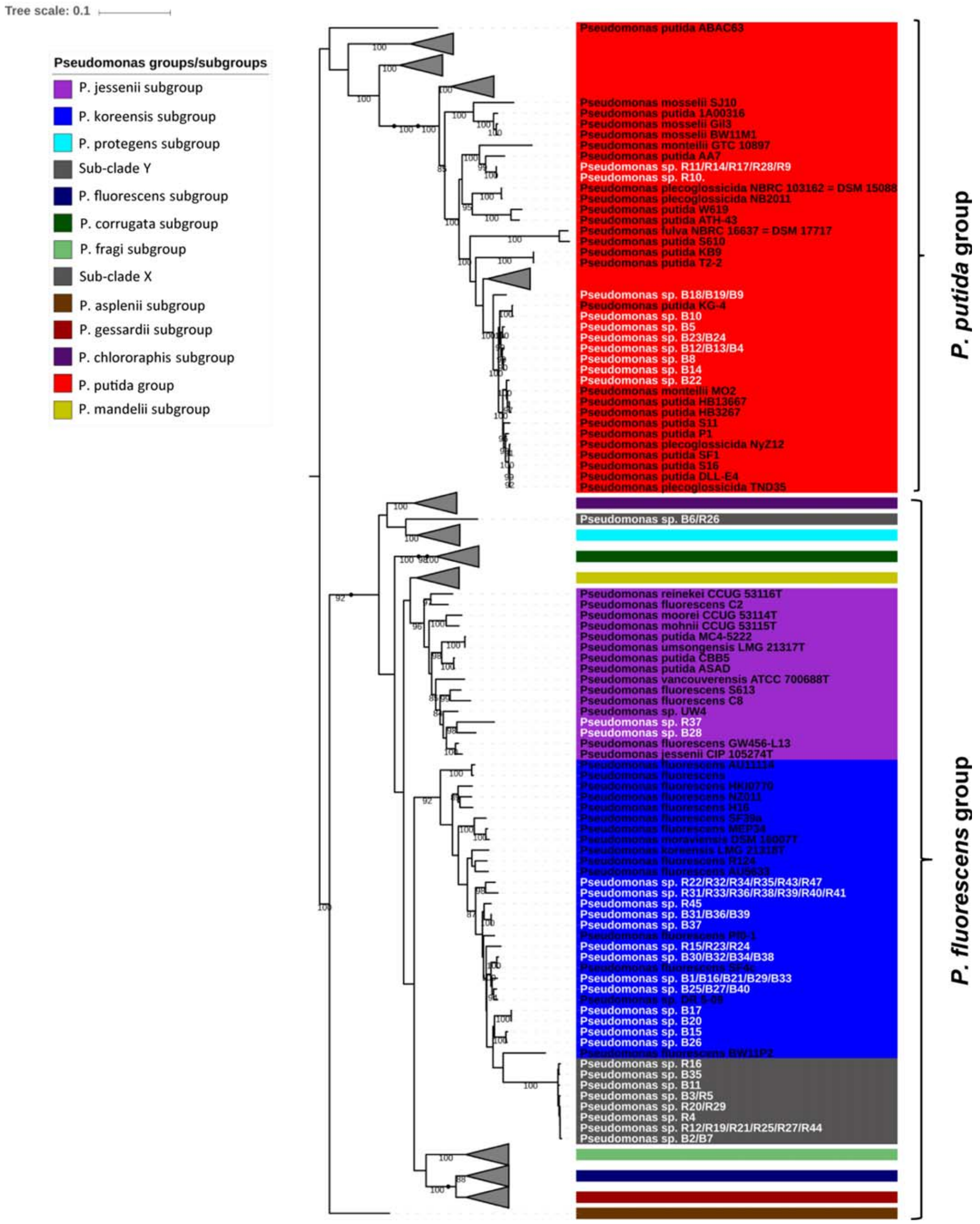

Fig. 2. Maximum likelihood phylogenetic tree inferred by multilocus sequence analysis.
Translated sequences of housekeeping genes *gyrB*, *rpoD, rpoB* and 16S rRNA from 76 sequenced isolates; all reference strains classified in the *P. putida* and *P. fluorescens* groups downloaded from NCBI; and informative *Pseudomonas* type strains of the *P. putida* group and all *P. fluorescens* subgroups were aligned, concatenated and used to generate a phylogenetic tree. Colours refer to the *Pseudomonas* groups and *P. fluorescens* subgroups. The subclades representing potential new subgroups are coloured in lighter (X) and darker (Y) grey. Labels in black are the reference and type strains, while labels in white are the isolates of this study. Branch lengths are supported by bootstrap values (numbers below each clade). [Colour figure can be viewed at wileyonlinelibrary.com]

However, unlike the previous analysis, these two isolates associated with the *P. fluorescens* group, and compose a distinct subclade (Y) located between the *P. protegens* and *P. chlororaphis* subgroups (Fig. 2). These distinct subclades may represent new subgroups of *P. fluorescens*.

In order to assess the possibility that known strains not included in the phylogenies fit within the new subclades, a BLASTn approach was used. The genes *rpoD*, *rpoB* and *gyrB* from isolates representing the new subgroups were used as queries. There were no hits with higher identity

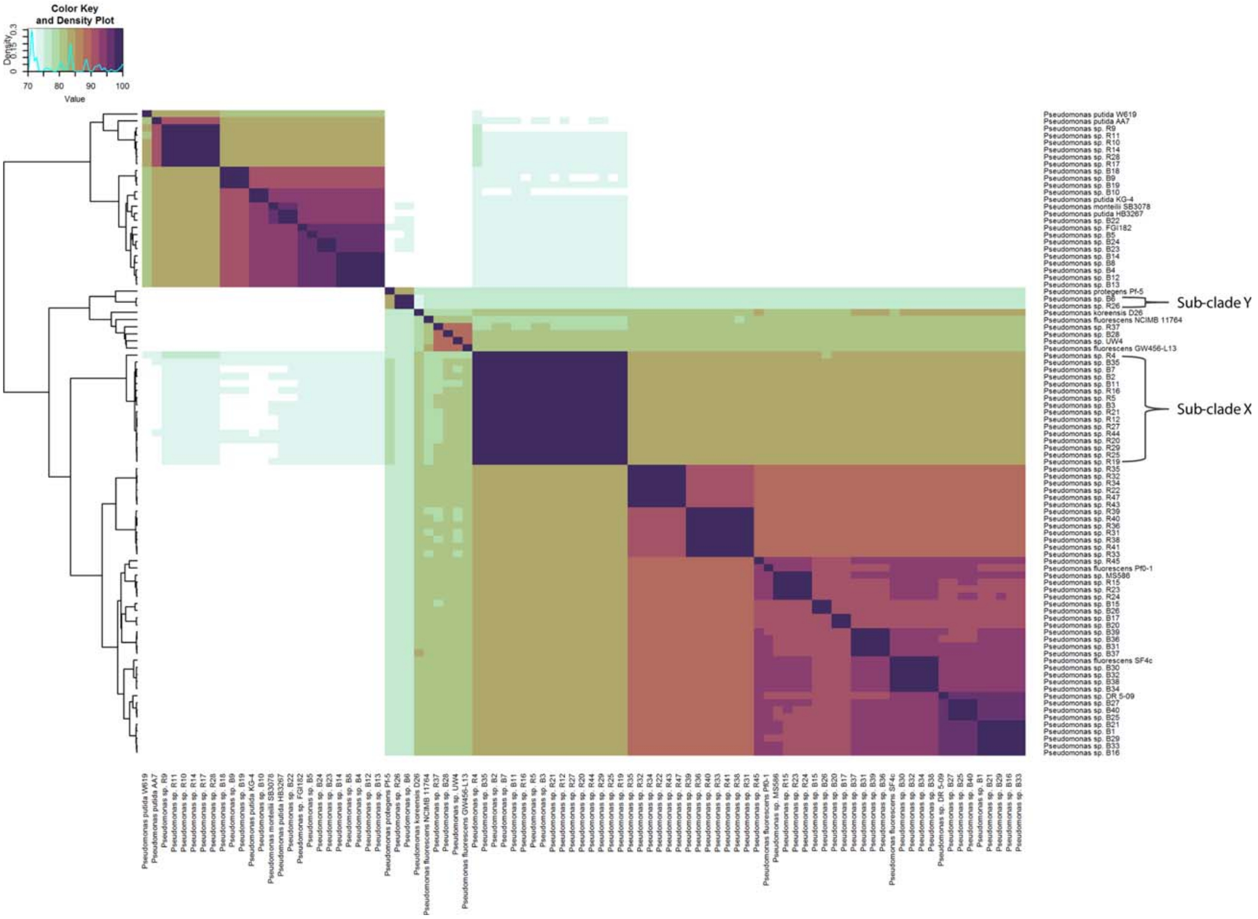

Fig. 3. Heatmap of all possible pairwise comparisons of average nucleotide identity (ANI).
Segments from the genome sequences from 76 tropical soil isolates and the nearest phylogenetically related reference strains were compared in all possible pairwise combinations and the ANI values were computed. The values are shown as a heatmap; the darker the colours, the higher the ANI percentages. The clusters of the new subclades are highlighted. [Colour figure can be viewed at wileyonlinelibrary.com]

and similarity than the strains already included in our phylogenetic trees. Thus, we concluded that there are no other strains available in the databases that fit in the new subclades detected in this study.

ANI supports the phylogenetic classification and the distinction of new subclades

A whole-genome based approach was used to support conclusions derived from phylogenetic analyses. ANI was calculated for all possible pairwise comparisons between the genome sequences of the 76 isolates and those from their most phylogenetically related reference strains, revealed by MLSA (Fig. 3). Type strains were not included in ANI analysis, since other studies observed a high correlation between MLSA and ANI (Gomila *et al.*, 2015), and type strains were less related to our isolates than reference strains. The pattern based on ANI clustering supported the conclusions drawn from the analyses of phylogenetic trees based on only seven or four marker genes (Figs 1–3). Use of ANI clustered isolates into groups that mirrored the clades defined from analyses of phylogenetic trees. More importantly, use of ANI confirmed the distinction of the new subclades X and Y to the recognized subgroups. Intergroup comparisons between members of the new subclades and previously defined subgroups had ANI values of less than 85%. Analysis using NMDS of the ANI values was consistent in supporting subgroup assignments and the separate clustering of the subclades X and Y (Supporting Information Figure S1).

ANI was also used to infer the number of possible new species found in our study. ANI values of 94%–96% correspond to 70% similarity in DNA-DNA hybridization and this threshold can be used to operationally classify bacteria into species (Konstantinidis and Tiedje, 2005; Kampfer, 2012; Oren and Garrity, 2014). Using a 95% ANI cutoff to define species, we identified 18 species among our 76 isolates, 13 classified in the *P. fluorescens* group and five in

Table 1. Number of homologous genes (gene clusters) among genome sequences analysed.

	All isolates (76)	All isolates + 18 references (94)	Isolates in the *P. fluorescens* group (57)	Isolates in the *P. putida* group (19)
Shared gene clusters	2839	2064	3580	3816
Accessory gene clusters	12299	21497	7640	4313
Total gene clusters	15138	23561	11220	8127

the *P. putida* group. In the *P. fluorescens* group, eight isolates are potentially the same species as the reference strain *Pseudomonas* sp. DR 5–09 (unpublished); four isolates are the same species as the reference strain *P. fluorescens* SF4c (Ly *et al*., 2015); and three isolates are the same species as the reference strain *Pseudomonas* sp. MS586 (unpublished). The other 42 isolates in this group comprised 10 individual clusters showing $<$ 95% ANI to any phylogenetically related reference strain and thus are potential new species (Fig. 3).

In the *P. putida* group, eight isolates are potentially the same species as the reference strain *Pseudomonas* sp. FGI182 (unpublished); one isolate is the same species as the reference strain *P. putida* HB3267 (Molina *et al*., 2014); and one isolate is the same species as the reference strain *P. putida* KG-4 (Dawar and Aggarwal, 2015). The remaining nine isolates formed two clusters with $<$ 95% ANI to any strain used in this analysis; these isolates could also represent new species. Therefore, using the 95% ANI cutoff, we identified ten putative new species in the *P. fluorescens* group and two putative new species in the *P. putida* group (Fig. 3). However, to get the accurate number of new species, we need to follow the taxonomy protocol, meaning that other analyses are required and will be performed in the future.

Specific gene clusters and genome organization of isolates in the new subclades

With the 76 isolates clustered into groups, subgroups and potential new species, we next turned our attention to identifying discriminating functions that could be used to distinguish strains belonging to the new subclades from those in the known subgroups of *P. fluorescens* group. The Get Homologues software was used to identify orthologous clusters. From this, a total of 15 138 gene clusters were found, of which 2839 clusters were present and had an orthologue in each of the 76 genome sequences (Table 1). The remaining 12 299 were polymorphic as presence/absence. Inclusion of 18 reference genome sequences reduced the number of conserved clusters to 2064 and significantly increased the total number of clusters to 23 561 (Table 1).

When analysing each *Pseudomonas* group separately, we observed 3580/3816 shared gene clusters (core genome) and 11 220/8127 total gene clusters (pangenome) for the *P. fluorescens* and *P. putida* groups respectively (Table 1). Analysis of the core and pan genome of 93 strains encompassing all the *P. fluorescens* subgroups revealed a lower core genome (1334) and larger pangenome (30 848) than we observed in the tropical isolates, which was expected since a higher number of genome sequences and phylogenetic diversity was considered in that study (Garrido-Sanz *et al*., 2016). Similarly, it was detected a lower core genome (3386) and larger pangenome (10 297) size in an analysis of nine *P. putida* genome sequences representing a high species diversity of this group (Udaondo *et al*., 2016). Despite our study considered a higher number of genome sequences (19) classified in the *P. putida* group, they seem to represent only five species. Therefore, we also covered a lower phylogenetic diversity in this group (Udaondo *et al*., 2016). These results show that the tropical soil isolates analysed in our study represent a subset of the genetic diversity in the *P. fluorescens* and *P. putida* groups. Nevertheless, the phylogenetic and ANI results suggested a high degree of novelty in the tropical isolates, which can potentially represent an increase in the pangenome size of these *Pseudomonas* groups.

To determine whether the isolates in the new subclades of *P. fluorescens* group have distinguishing functions, we focused on clusters that are present in all members of the new subclades and absent in all other subgroups. For the subclade X (related to *P. koreensis*), 105 gene clusters were exclusive to its members. The 105 gene loci are unlinked and distributed throughout each genome, arguing against their acquisition through a limited number of horizontal gene transfer (HGT) events. Sixty-one of the 105 genes were annotated as ‘hypothetical proteins’. The small number of genes with annotated functions makes it difficult to test for enriched functions that may distinguish subclade X from species of established subgroups. Results from synteny analyses of representative genome sequences also supported that strains in the subclade X are distinct from strains in the *P. koreensis* subgroup, which is consistent with our phylogenetic and ANI analyses (Fig. 4A). The genome sequence of Brazilian isolate B34 is represented by 15 contigs, and they are co-linear to the genome sequence of strain Pf0–1. In contrast, the genome sequence of isolate R21, which was assigned to subclade X and is represented by 57 contigs, is

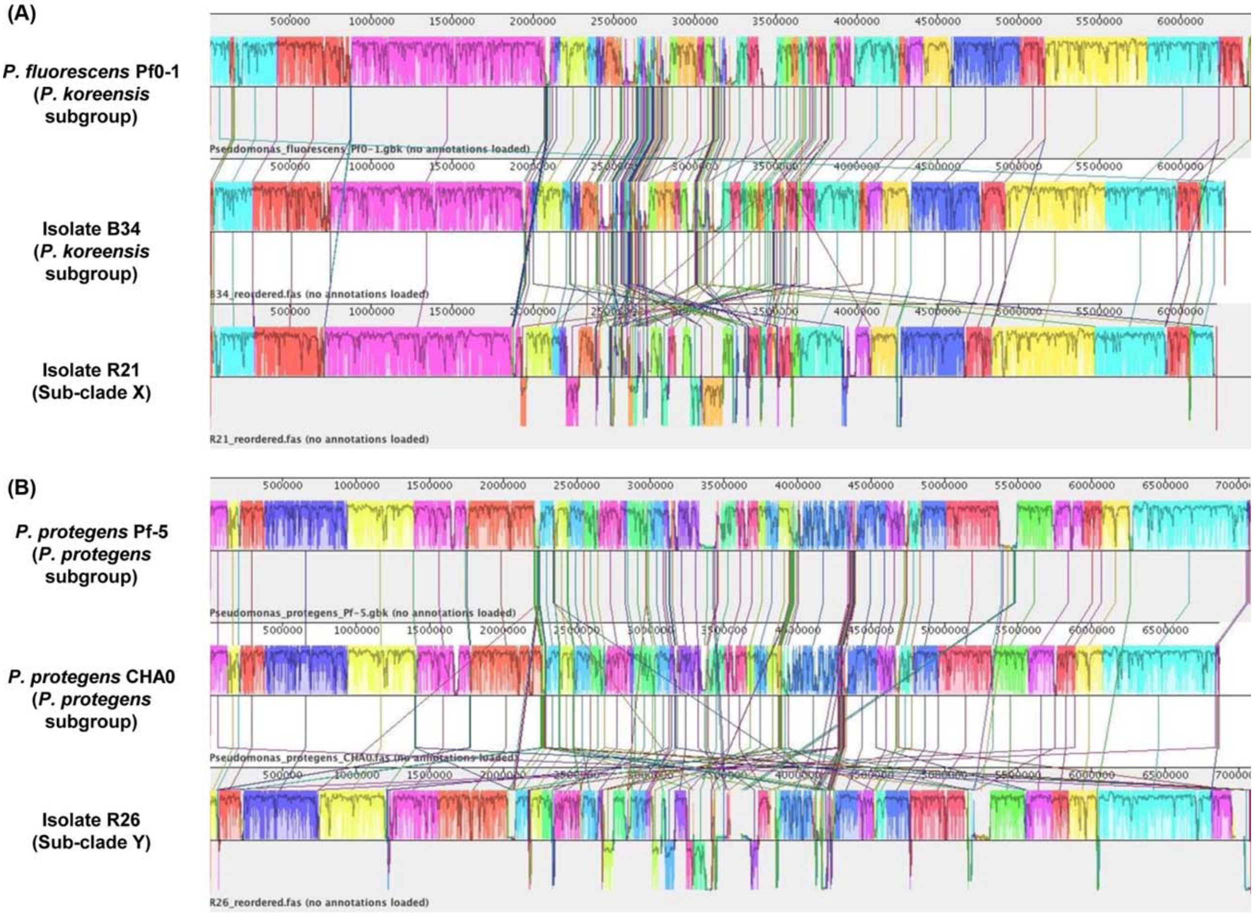

Fig. 4. Multiple genome alignments. Selected genome sequences were aligned, using Mauve.
Different colours correspond to genome regions with high percentage alignment. Lines indicate rearranged regions and show their new position in the compared genomes. Origin of replication is in the extreme left of each genome sequence, based on the reference genome sequences used to reorder the contigs of the other genomes analysed, that is A) *P. fluorescens* Pf0–1 for the comparison between the *P. koreensis* subgroup with the new subclade X; and B) *P. protegens* Pf-5 for the comparison between the *P. protegens* subgroup and the new subclade Y. [Colour figure can be viewed at wileyonlinelibrary.com]

slightly different from the B34 or Pf0–1 genomes in organization and has an inversion event that includes the region of termination. Additionally, Mauve alignment of genome sequences from all isolates that belong to subclade X reveal them to be largely syntenic (Supporting Information Figure S2A).

Similarly, 527 gene clusters were exclusively identified in the subclade Y (neighbour to *P. protegens*). Of these, 315 were annotated as ‘hypothetical proteins’. As observed for the subclade X, these exclusive gene clusters are dispersed throughout the genome, which argues against their acquisition through a few HGT events. Genome organization is also quite divergent compared to strains in the *P. protegens* subgroup (Fig. 4B). As was the case above, the two strains of the subclade Y are co-linear (Supporting Information Figure S2A and B). Therefore, our results support the conclusion that some of the tropical isolates may represent new subgroups of the *P. fluorescens* group.

Genetic content of the isolates covers several known functions performed by fluorescent Pseudomonas *spp.*

Several functions, such as antibiotic production, protein secretion systems and mechanisms, for interacting with plants have been identified as important for the plant-associated lifestyles of fluorescent *Pseudomonas* spp. (Rezzonico *et al.*, 2005; Kidarsa *et al.*, 2013). We therefore mined the 76 genome sequences to determine whether isolates inhabiting tropical soils could potentially have similar functions and exhibit similar lifestyles as their temperate counterparts (Supporting Information Table S3). Two sets of genes were found in all 57 genome sequences of the isolates in the *P. fluorescens* group (100%). One set is associated with pyoverdine biosynthesis and the other is associated with Gamma-aminobutyric acid (GABA) catabolism. Almost all the other function-encoding genes analysed were found in most, but not all of the genome sequences of the *P. fluorescens* group. Of the functions that were examined, only ACC deaminase was not found in any of the 57 isolates in this group (0%).

Pyoverdine biosynthesis genes are perhaps the most outstanding of the analysed functions (Supporting Information Table S3). Pyoverdines are a class of structurally related siderophores that are produced by fluorescent *Pseudomonas* spp., function in iron chelation and uptake, and are characterized by their green fluorescence (Visca *et al*., 2007). It is assumed that pyoverdine is a public good and that *Pseudomonas* spp. populations have a balance between pyoverdine producers and pyoverdine defective genotypes (Lujan *et al*., 2015). In alkaline temperate soils, pyoverdine depletes the amount of iron available to pathogens and helps control their populations in the rhizosphere (Visca *et al*., 2007). However, the tropical soil used in this study is an acidic (pH = 5.0) Oxisol, which is characterized by a high availability of iron due to its chemical characteristics (Brady and Weil, 2002). In principle, the increased iron availability should reduce the beneficial effect of pyoverdine production for plant protection. Nonetheless, pyoverdine may also be involved in biofilm formation and could be advantageous for survival of bacteria in such soils with lower levels of organic matter (Ross, 1993; Visca *et al*., 2007; Guimaraes *et al*., 2013). We therefore speculate that the tropical isolates could synthesize pyoverdine for this latter benefit.

GABA is a non-protein amino acid that functions in stress tolerance of plants as well as in plant-microbe communication (Dagorn *et al*., 2013). The presence of GABA caused increase accumulation of LPS and affected biofilm maturation in *P. fluorescens* (Dagorn *et al*., 2013) and could also be related to this process in tropical soils.

Other sets of genes found in $>$ 70% of the genome sequences of the *P. fluorescens* group were ones related to bacteriocins (S-type + cytotoxic domain proteins and carocin-like); the exoprotease AprA; type II secretion systems (T2SS Hxc and T2SS Hxc-2) and type VI secretion systems (T6SS HSI-II); chitinase; and hydrogen cyanide (HCN) biosynthesis (Supporting Information Table S3). In general, these genes function in intermicrobial competition. The extracellular protease AprA is related to both control of protist grazing and entomopathogenic activity (Jousset *et al*., 2006; Chen *et al*., 2014). Bacteriocins play a role in *P. fluorescens* competitiveness with other bacteria (Bruce *et al*., 2017), and the T6SS delivers toxic effector proteins into competing bacteria, providing a benefit to the producer strain in complex bacterial communities (Ho *et al*., 2014). Chitinase degrades chitin, which is present in arthropods and fungal cell walls, and could be related to antifungal activity (Manikandan and Raguchander, 2015). Finally, HCN has both antimicrobial and insecticidal activity (Flury *et al*., 2017).

Strains B6 and R26, assigned to the new subclade Y, have genes for the biosynthesis of the antibiotics hydrogen cyanide, pyoluteorin, 2,4-diacetylphloroglucinol (DAPG), pyrrolnitrin and rhizoxin (Supporting Information Table S3). The closest subgroup of this subclade is *P. protegens* (Garrido-Sanz *et al.* 2016), which has strains that also contain genes for the biosynthesis of these antibiotics (Loper *et al*., 2012).Thus, although the strains in subclade Y and in *P. protegens* subgroup are phylogenetically different, they share common features for interacting in the microbial community. All the screened functional genes that were found in the genome sequences of tropical strains assigned to the *P. koreensis* subgroup were also found in the genome sequences of strains in subclade X (Supporting Information Table S3). Both results indicate that genes related to biotic interactions are shared by strains in the new subclades and in their most-closely related subgroups, despite the phylogenetic distinctions revealed by MLSA and the presence of gene clusters exclusive to subclades X and Y. These results suggest a considerable degree of functional redundancy between isolates inhabiting tropical soils and previously characterized *P. fluorescens* strains primarily from temperate regions.

The ability of fluorescent *Pseudomonas* spp. to produce chemicals for competing in the soil/plant environments is widely known, and several of these functions can help plants to defend against pathogens and/or pests (Chen *et al*., 2014; Manikandan and Raguchander, 2015; Flury *et al*., 2017). It is noteworthy that the rhizosphere isolates of the *P. putida* group have some degree of phylogenetic relatedness to *P. putida* W619 (Figs 1 and 2), which is a strain isolated from tissues of the plant *Populus trichocarpa* x *deltoide*s cv. ‘Hoogvorst, and showed plant growth promotion activity (Taghavi *et al*., 2009). A close relative of the isolates classified in the *P. jessenii* subgroup (R37, B28) is the reference strain *Pseudomonas* sp. UW4 (Figs 1 and 2), which was isolated from the rhizosphere of common reeds and promotes plant growth (Duan *et al*., 2013). The strains isolated from tropical soils in this study have not yet been tested for their effects on plant growth or health, but the close relationship of some of these tropical isolates to well-characterized PGPRs suggests that they have potential to be beneficial to plants.

New diversity found may be related to geographical distribution

The results of this study showed that the *Pseudomonas* genus is a plentiful source of new strains and species, reflecting its tremendous diversity (Spiers *et al*., 2000). Fluorescent *Pseudomonas* spp. are commonly isolated from agroecosystems and are well-known for their activities in plant growth promotion and biocontrol (Hofte and Altier, 2010). Even so, our study pointed to the presence of several new haplotypes and 12 possible new species from the 18 (66.7%) identified among our 76 isolates using a 95% ANI cutoff; and also found distinct subclades indicating the possibility of two new subgroups of *P. fluorescens*. Thus, based on phylogenetic analyses and operational criteria based on ANI, the 76 isolates that we collected from tropical soils represent a substantial new diversity.

Despite the attention given to *Pseudomonas* in the past and the ease with which species in this genus are cultured, we isolated several new haplotypes and possible new species. We speculate that this reflects our focus on a relatively unexplored habitat and geographical location. Currently, there are 104 genome sequences of strains named as *P. fluorescens* in the NCBI Genome database. Only eight – *P. fluorescens* Pt14 (Rani *et al*., 2017), *P. fluorescens* BRIP34879 (Gardiner *et al*., 2013), *P. fluorescens* SRM1 (Lo *et al*., 2015), *P. fluorescens* EGD-AQ6 (unpublished), *P. fluorescens* S12 (unpublished), *P. fluorescens* UM270 (Hernández-Salmerón *et al*., 2016), *P. fluorescens* et76 (Aarab *et al*., 2016) and *P. fluorescens* Ps006 (Gamez *et al*., 2016) – were submitted by institutions from tropical countries (https://www.ncbi.nlm.nih.gov/genome/genomes/150, searched in 02/10/2017). Even considering that the taxonomic classification of these strains may be ambiguous, these data nonetheless exemplify how the diversity of fluorescent *Pseudomonas* spp. is less explored in tropical than temperate regions. Thus, the lack of available data and studies from tropical countries can explain why we found several possible new species of fluorescent *Pseudomonas* in this study.

Microbial biogeography has shown that soil microbial communities change across the world and neutral or selective factors, including pH, carbon sources, plant species, moisture and temperature, can influence their composition and structure (Fierer and Jackson, 2006; Fierer *et al*., 2012). In addition to shaping communities, geographic distance also influences the evolution and diversification of microbial species, since both genetic drift and natural selection take place independently for populations facing geographical separation (Hanson *et al*., 2012). Thus, another recurring issue in microbial biogeography is the existence of endemism. Cho and Tiedje (2000) suggested the possibility of endemism of soil fluorescent *Pseudomonas* spp.*,* arguing against the omnipresent distribution of these species in the world. Our results strengthen this idea by finding several putative new species not described in any other place. The existence of novel species and subclades in the tropical soil analysed could represent a case of endemism, until proven otherwise. Even considering the high dispersion and migration potential of prokaryotes – amplified by human interference – the Earth might be big enough to prevent exogenous bacteria from rapidly conquering new ecosystems; while environmental filtering can limit the easy establishment of new isolates elsewhere. In addition, as environment differs with distance, it is possible to suggest a rapid and distinct bacterial evolution in populations inhabiting very distant regions (Hanson *et al*., 2012). Hence, the biogeographical patterns added to the lack of available genome sequences and strains from tropical soils possibly resulted in our findings, a description of not only new haplotypes and possible species, but also putative new *P. fluorescens* subgroups.

The similarities found in gene content indicate that the biogeographical effects may have affected the genome arrangement and phylogenetic diversification of tropical soil isolates, but not their functionality, at least with respect to activities associated with plant interaction of *P. fluorescens*. Notwithstanding, the set of specific gene clusters found in the new subclades suggests that its members may perform some different activities compared to strains in the neighbouring subgroups (*P. koreensis and P. protegens*). Biochemical and phenotypic analyses are required to assess this issue more deeply.

Conclusions

The tropical soil analysed in our study harbours new haplotypes and 12 possible new species of the *P. fluorescens* and *P. putida* groups. Two subclades of isolates have phylogenetic and genomic distinctions compared to the known subgroups of *P. fluorescens*, and therefore can possibly represent new subgroups. Biogeography may explain our finding of many putative new species and subclades, since most studies and available data on fluorescent *Pseudomonas* spp. are from temperate regions. Although phylogenetically distinct, strains isolated from the tropical soil share genes key to plant-microbe and other biotic interactions with fluorescent *Pseudomonas* spp. isolated from temperate environments. Nevertheless, hundreds of gene clusters were found exclusively in the tropical isolates assigned to subclades X and Y, relative to the most closely related strains of *P. fluorescens* previously described. This study demonstrates that the tremendous diversity of *Pseudomonas* spp. is not yet fully known and that tropical soils provide an untapped source of microbial diversity.

Experimental procedures

Soil sampling and isolation of fluorescent Pseudomonas *spp.*

Isolates were obtained in March 2014 from a soil (Oxisol) in Piracicaba-SP, Brazil, managed for the past 10 years using green-harvest system for sugarcane cultivation. Samples were collected from six different points of the field and from two soil compartments: bulk soil, the root-free soil located between the crop rows and at a depth of 0–10 cm; and rhizosphere, collected by removing the plants from soil and scraping the 1–2 mm of soil adhering to the roots. A total of 12 samples (6 from bulk soil and 6 from rhizosphere) were obtained. All bulk soil or rhizosphere samples were homogenized, making two composite samples. Soil samples were serially diluted and aliquots were plated on the selective media Pseudomonas Agar Base (Oxoid, Basingstoke, Hampshire, UK) supplemented with the antibiotics cetrimide, fucidin and cephalosporin (10, 10 and 50 mg L^{-1} respectively). Plates were incubated at 28°C and colonies of *Pseudomonas* were periodically visualized under UV light over the span of two days. A total of 76 single fluorescent colonies were isolated and stored in a 35% glycerol solution at −80°C.

DNA isolation and genome sequencing

Genomic DNA from the 76 isolates was extracted using the Wizard Kit (Promega, Madison, Wisconsin, USA) after growing cultures overnight with the same (broth) medium and conditions described above. Nextera libraries were constructed and sequenced by the Center for Genome Research and Biocomputing (CGRB) at Oregon State University. The 76 barcoded libraries were sequenced using the 150mer paired end sequencing kit, in a single lane on an Illumina HiSeq 3000.

Genome assemblies and annotations

FastQC (Andrews, 2010) was used to assess the quality of the sequencing reads and the Cutadapt 1.9.1 software (Martin, 2011) was used to trim off adapter sequences. Paired-end sequencing reads were *de novo* assembled, using SPAdes 3.7.0 (Bankevich *et al*., 2012), with a phred-offset of 33 and multiple kmers (21, 33, 55, 77, 99). Small contigs ($<$ 500 bp) and palindromic sequences were removed from the assemblies. The scaffold genomes were annotated using the Prokka software (Seeman, 2014), which also generated the .gbk files. This Whole Genome Shotgun project (BioProject PRJNA383709) has been deposited at DDBJ/ENA/GenBank databases under the accession numbers NEHD00000000-NEKA00000000. The versions described here are NEHD01000000-NEKA01000000. The accession numbers of all 76 isolates are available in the Supporting Information Table S1.

Phylogenetic and genomic analyses

Using sequences of the housekeeping genes *dnaE*, *guaA*, *mutL*, *ppsA*, *pyrC*, *recA* and *rpoB* of *P. aeruginosa* PA01 as queries and autoMLSA v. 1.0 (Davis *et al*., 2016), homologous sequences from only *Pseudomonas* genome sequences having all seven housekeeping genes were retrieved from the Genome database of NCBI (Loper *et al*., 2012). We refer to these strains with available genome sequences in the NCBI Genome database – but not necessarily type strain of a species – as reference strains. Sequences of the housekeeping genes were also extracted from the genome assemblies of each isolate sequenced herein. The gene sequences corresponding to each genome were aligned (MAFFT v. 7.271, L-insi-i algorithm), concatenated and RAxML v. 8.2.8 was used to generate a maximum likelihood multilocus sequence analysis (MLSA) phylogenetic tree with bootstrap support (1000 bootstrap replicates) (Katoh and Standley, 2013; Stamatakis, 2014). The ItoL platform was used to visualize results and produce the phylogenetic trees (Letunic and Bork, 2016). Analysis was performed on 06/2016.

A second MLSA tree, based on the analysis of four housekeeping genes (16S rRNA, *rpoB*, *rpoD* and *gyrB*), was constructed to infer subgroup classifications. All strains with sequences available in the NCBI Genome database and classified in the *P. fluorescens* or *P. putida* groups (reference strains), as well as several type strains of all *P. fluorescens* subgroups and the *P. putida* group, were included in this analysis. Analysis was performed on 03/2017.

The software autoANI v. 1.0 was used to calculate all possible pairwise average nucleotide identity values (ANI; Davis *et al*., 2016). Each genome sequence was chunked into 1020 nt fragments and used as queries in all possible reciprocal pairwise BLAST searches. BLAST hits above the default 30% identity and 70% coverage were averaged to calculate the pairwise ANI values. The R-based software program ggplot2 was used to generate a heatmap of the ANI matrix (Wickham, 2009). The Past software (Hammer *et al*., 2001) was used to normalize the ANI data and a non-metric multidimensional scaling (NMDS) was performed using the Bray-Curtis index.

The orthoMCL algorithm in the Get Homologues software (Contreras-Moreira and Vinuesa, 2013) was used to identify clusters of homologue genes from the coding sequences (CDSs). The CDSs from 18 reference strains: *P. fluorescens* Pf0–1 (Compeau *et al*., 1988), *Pseudomonas* sp. MS586 (unpublished), *P. aeruginosa* DK2 (Yang *et al*., 2011), *P. alkylphenolia* KL28 (Mulet *et al*., 2015), *P. chlororaphis* PCL1606 (Pliego *et al*., 2012), *P. fluorescens* FW300-N2E3 (unpublished), *P. fluorescens* NCIMB 11 764 (Vilo *et al*., 2012), *P. mandelii* JR-1 (Jang *et al*., 2012), *P. monteilii* SB 3078 (Dueholm *et al*., 2014), *P. mosselii* SJ10 (Park *et al*., 2014), *P. protegens* Pf-5 (Paulsen *et al*., 2005), *P. putida* H8234 (Molina *et al*., 2013), *P. putida* HB3267 (Molina *et al*., 2014), *P. putida* S12 (Tao *et al*., 2012), *P. putida* W619 (Taghavi *et al*., 2009), *Pseudomonas* sp. FGI182 (unpublished), *Pseudomonas* sp. UW4 (Duan *et al*., 2013) and *P. syringae* CC1557 (Hockett *et al*., 2014) – representing phylogenetically related and distant strains – were used for comparisons. Mauve 2.4.0 software, using the progressiveMauve algorithm, was used to align multiple genome sequences for synteny analyses (Darling *et al*., 2010).

A database of translated sequences that exemplify key functions of fluorescent *Pseudomonas* spp. was developed and used as queries in BLAST searches (tBLASTn) against the genome assemblies of the 76 isolates from Brazilian soil. The amino acid sequences were extracted from the .faa files corresponding to: *P. fluorescens* Q8r1–96 (ACC deaminase; T3SS-1; T3SS-2; T6SS HSI-I) (Loper *et al*., 2012), *P. fluorescens* Pf0–1 (AprA; chitinase; GABA catabolism; T2SS Hxc; T6SS HSI-II; T6SS HSI-III; Tcc4; Tcc5; carocin-like; S-type-+cytotoxic domain proteins; prototypic S-type) (Compeau *et al*., 1988), *P. protegens* Pf-5 (DAPG; HCN; pyoluteorin; pyrrolnitrin; rhizoxin) (Paulsen *et al*., 2005), *P. chlororaphis* O-6 (IAA biosynthesis) (Loper *et al*., 2012), *P. fluorescens* SBW25 (T2SS Hxc-2) (Rainey and Bailey, 1996) and *P. fluorescens* Q2–87 (T3SS-3) (Loper *et al*., 2012).

Acknowledgements

We acknowledge CAPES for the scholarship support provided to LDL and MCPS; FAPESP and CNPq for the financial support to FDA, project 2013/50353-7 and grant no. 302168/2014-1 respectively. This work was also supported by the National Institute of Food and Agriculture, US Department of Agriculture award 2014–51181-22384 to JHC and 2011–67019-30192 to JEL; and by the National Science Foundation Graduate Research Fellowship grant no. DGE-1314109 to EWD. The authors declare no conflicts of interest.

Author contributions

LDL wrote the manuscript and performed the molecular, phylogenetic and genomic analyses; EWD and AJW performed the phylogenetic and genomic analyses; MCPS and LB performed the isolation of soil *Pseudomonas* strains; FDA, JEL and JHC contributed to the conception, writing and editing of the manuscript.

References

Aarab, S., Arakrak, A., Ollero, F.J., Megías, M., Gomes, D.F., Ribeiro, R.A., and Hungria, M. (2016) Draft genome sequence of *Pseudomonas fluorescens* strain ET76, isolated from rice rhizosphere in northwestern Morocco. *Genome Announc* **4:** e00356–e00316.

Andam, C.P., Doroghazi, J.R., Campbell, A.N., Kelly, P.J., Choudoir, M.J., and Buckley, D.H. (2016) A latitudinal diversity gradient in terrestrial bacteria of the genus *Streptomyces*. *mBio* **7:** e02200–e02215.

Andrews, S. (2010) *FastQC: a quality control tool for high throughput sequence data*. [WWW document]. URL http://www.bioinformatics.babraham.ac.uk/projects/fastqc

Araujo, M.A.V., MendonÃ§A-Hagler, L.C., Hagler, A.N., and Elsas, J.D. (1994) Survival of genetically modified *Pseudomonas fluorescens* introduced into subtropical soil microcosms. *FEMS Microbiol Ecol* **13:** 205–216.

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikov, A.S., *et al*. (2012) SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* **19:** 455–477.

Brady, N.C., and Weil, R.R. (2002) *The Nature and Properties of Soils*. Upper Saddle River, NJ: Prentice Hall.

Bruce, J.B., West, S.A., and Griffin, A.S. (2017) Bacteriocins and the assembly of natural *Pseudomonas fluorescens* populations. *J Evol Biol* **30:** 352–360.

Chen, W.J., Hsieh, F.C., Hsu, F.C., Tasy, Y.F., Liu, J.R., and Shih, M.C. (2014) Characterization of an insecticidal toxin and pathogenicity of *Pseudomonas taiwanensis* against insects. *PLoS Pathog* **10:** e1004288. doi:10.1371/journal.ppat.1004288

Cho, J.C., and Tiedje, J.M. (2000) Biogeography and degree of endemicity of fluorescent *Pseudomonas* strains in soil. *Appl Environ Microbiol* **66:** 5448–5456.

Cho, S.-T., Chang, H.-H., Egamberdieva, D., Kamilova, F., Lugtenberg, B., and Kuo, C.-H. (2015) Genome analysis of *Pseudomonas fluorescens* PCL1751: a rhizobacterium that controls root diseases and alleviates salt stress for its plant host. *PLoS One* **10:** e0140231. doi:10.1371/journal.pone.0140231.

Compeau, G., Al-Achi, B.J., Platsouka, E., and Levy, S.B. (1988) Survival of rifampin-resistant mutants of *Pseudomonas fluorescens* and *Pseudomonas putida* in soil systems. *Appl Environ Microbiol* **54:** 2432–2438.

Contreras-Moreira, B., and Vinuesa, P. (2013) GET_HOMOLOGUES, a versatile software package for scalable and robust microbial pangenome analysis. *Appl Environ Microbiol* **79:** 7696–7701.

Dagorn, A., Chapalain, A., Mijouin, L., Hillion, M., Duclairoir-Poc, C., Chevalier, S., *et al*. (2013) Effect of GABA, a bacterial metabolite, on *Pseudomonas fluorescens* surface properties and cytotoxicity. *Int J Mol Sci* **14:** 12186–12204. doi:10.3390/ijms140612186.

Darling, A.E., Mau, B., and Perna, N.T. (2010) progressiveMauve: multiple genome alignments with gene gain, loss and rearrangement. *PLoS One* **5:** e11147. doi:10.1371/journal.pone.0011147.

Davis, E.W., II, Weisberg, A.J., Tabima, J.F., Grunwald, N.J., and Chang, J.H. (2016) Gall-ID: tools for genotyping gall-causing phytopathogenic bacteria. *PeerJ* **4:** e2222.

Dawar, C., and Aggarwal, R.K. (2015) Draft genome sequence of hydrocarbon-degrading *Pseudomonas putida* strain KG-4, isolated from soil samples collected from Krishna-Godavari Basin in India. *Genome Announc* **3:** e00590–e00515.

Duan, J., Jiang, W., Cheng, Z., Heikkila, J.J., Glick, B.R., and van Passel, M.W.J. (2013) The complete genome sequence of the plant growth-promoting bacterium *Pseudomonas* sp. UW4. *PLoS One* **8:** e58640.

Dueholm, M.S., Albertsen, M., D'Imperio, S., Tale, V.P., Lewis, D., Nielsen, P.H., and Nielsen, J.L. (2014) Complete genome sequences of *Pseudomonas monteilii* SB3078 and SB3101, two benzene-, toluene-, and ethylbenzene-degrading bacteria used for bioaugmentation. *Genome Announc* **2:** e00524–e00514.

Fierer, N., and Jackson, R.B. (2006) The diversity and biogeography of soil bacterial communities. *Proc Natl Acad Sci USA* **103:** 626–631.

Fierer, N., Leff, J.W., Adams, B.J., Nielsen, U.N., Bates, S.T., Lauber, C.L., *et al*. (2012) Cross-biome metagenomic analyses of soil microbial communities and their functional attributes. *Proc Natl Acad Sci USA* **109:** 21390–21395.

Flury, P., Vesga, P., Pechy-Tarr, M., Aellen, N., Dennert, F., Hofer, N., *et al*. (2017) Antimicrobial and insecticidal: cyclic lipopeptides and hydrogen cyanide produced by plant-beneficial *Pseudomonas* strains CHA0, CMR12a, and PCL1391 contribute to insect killing. *Front Microbiol* **8:** 100. doi:10.3389/fmicb.2017.00100.

Gamez, R.M., Rodriguez, F., Ramirez, S., Gomez, Y., Agarwala, R., Landsman, D., and Marino, R.L. (2016) Genome sequence of the banana plant growth-promoting rhizobacterium *Pseudomonas fluorescens* PS006. *Genome Announc* **4:** pii: e00329–16. doi:10.1128/genomeA.00329-16.

Gardiner, D.M., Stiller, J., Covarelli, L., Lindeberg, M., Shivas, R., and Manners, J.M. (2013) Genome sequences of *Pseudomonas* spp. isolated from cereal crops. *Genome Announc* **1:** e00209–e00213. doi:10.1128/genomeA.00209-13.

Garrido-Sanz, D., Meier-Kolthoff, J.P., Göker, M., Martín, M., Rivilla, R., and Redondo-Nieto, M. (2016) Genomic and genetic diversity within the *Pseudomonas fluorescens* complex. *PLoS One* **11:** e0150183. doi:10.1371/journal.pone.0150183.

Gomila, M., Peña, A., Mulet, M., Lalucat, J., and García-Valdés, E. (2015) Phylogenomics and systematics in *Pseudomonas*. *Front Microbiol* **6:** 214. doi:10.3389/fmicb.2015.00214.

Guimaraes, D.V., Gonzaga, M.I.S., Silva, T.O., Silva, T.L., Dias, N.S., and Matias, M.I.S. (2013) Soil organic matter pools and carbon fractions in soil under different land uses. *Soil. Till Res* **126:** 177–182.

Gumiere, T., Durrer, A., Bohannan, B., and Andreote, F.D. (2016) Biogeographical patterns in fungal communities from soils cultivated with sugarcane. *J Biogeogr* **43:** 2016–2026.

Hammer, Ø., Harper, D.A.T., and Ryan, P.D. (2001) PAST: paleontological statistics software package for education and data analysis. *Paleontol Electron* **4:** 1–9.

Hanson, C.A., Fuhrman, J.A., Horner-Devine, M.C., and Martiny, J.B. (2012) Beyond biogeographic patterns: processes shaping the microbial landscape. *Nat Rev Microbiol* **10:** 497–506.

Hernández-Salmerón, J.E., Hernandez-Leon, R., Orozco-Mosqueda, M.D.C., Valencia-Cantero, E., Moreno-Hagelsieb, G., and Santoyo, G. (2016) Genome sequence of the biocontrol and plant growth-promoting rhizobacteria *Pseudomonas fluorescens* strain UM270. *Stand Genomic Sci* **11:** 5. doi:10.1186/s40793-015-0123-9.

Ho, B.T., Dong, T.G., and Mekalanos, J.J. (2014) A view to a kill: the bacterial type VI secretion system. *Cell Host Microb* **15:** 9–21.

Hockett, K.L., Nishimura, M.T., Karlsrud, E., Dougherty, K., and Baltrus, D.A. (2014) *Pseudomonas syringae* CC1557: a highly virulent strain with an unusually small type III effector repertoire that includes a novel effector. *Mol Plant Microbe Interact* **27:** 923–932.

Hofte, M., and Altier, N. (2010) Fluorescent pseudomonads as biocontrol agents for sustainable agricultural systems. *Res Microbiol* **161:** 464–471.

Hol, W.H.G., Bezemer, T.M., and Biere, A. (2013) Getting the ecology into interactions between plants and the plant growth-promoting bacterium *Pseudomonas fluorescens*. *Front Plant Sci* **4:** 81. doi:10.3389/fpls.2013.00081.

Jang, S.H., Kim, J., Kim, J., Hong, S., and Lee, C. (2012) Genome sequence of cold-adapted *Pseudomonas mandelii* strain JR-1. *J Bacteriol* **194:** 3263.

Jousset, A., Lara, E., Wall, L.G., and Valverde, C. (2006) Secondary metabolites help biocontrol strain *Pseudomonas fluorescens* CHA0 to escape protozoan grazing. *Appl Environ Microbiol* **72:** 7083–7090.

Kalpage, F.S.C.P. (1974) *Tropical Soils: classification, Fertility and Management*. New York, NY: St. Martin's Press, 294 p.

Kampfer, P. (2012) Systematics of prokaryotes: the state of the art. *Antonie Van Leeuwenhoek* **101:** 3–11.

Katoh, K., and Standley, D.M. (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* **30:** 772–780. doi:10.1093/molbev/mst010.

Kidarsa, T.A., Shaffer, B.T., Goebel, N.C., Roberts, D.P., Buyer, J.S., Johnson, A., *et al*. (2013) Genes expressed by the biological control bacterium *Pseudomonas protegens* Pf-5 on seed surfaces under the control of the global regulators GacA and RpoS. *Environ Microbiol* **15:** 716–735.

Konstantinidis, K.T., and Tiedje, J.M. (2005) Genomic insights that advance the species definition for prokaryotes. *Proc Natl Acad Sci USA* **102:** 2567–2572.

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

Lo, R., Stanton-Cook, M.J., Beatson, S.A., Turner, M.S., and Bansal, N. (2015) Draft genome sequence of *Pseudomonas fluorescens* SRM1, an isolate from spoiled raw milk. *Genome Announc* **3:** e00138–e00115. doi:10.1128/genomeA.00138-15.

Loper, J.E., Hassan, K.A., Mavrodi, D.V., Davis, E.W., Lim, C.K., Shaffer, B.T., *et al*. (2012) Comparative genomics of plant-associated *Pseudomonas* spp.: Insights into diversity and inheritance of traits involved in multitrophic interactions. *PLoS Genet* **8:** e1002784.

Lopes, L.D., Pereira e Silva, M.C., and Andreote, F.D. (2016) Bacterial abilities and adaptation toward the rhizosphere colonization. *Front Microbiol* **7:** 1341. doi:10.3389/fmicb.2016.01341.

Lujan, A.M., Gomez, P., and Buckling, A. (2015) Siderophore cooperation of the bacterium *Pseudomonas fluorescens* in soil. *Biol Lett* **11:** 20140934.

Ly, L.K., Underwood, G.E., McCully, L.M., Bitzer, A.S., Godino, A., Bucci, V., *et al*. (2015) Draft genome sequences of *Pseudomonas fluorescens* strains SF39a and SF4c, potential plant growth promotion and biocontrol agents. *Genome Announc* **3:** e00219–e00215.

Manikandan, R., and Raguchander, T. (2015) *Pseudomonas fluorescens* (Pf1) mediated chitinolytic activity in tomato plants against *Fusarium oxysporum* f. sp. *lycopersici*. *Afr J Microbiol Res* **9:** 1331–1337.

Martin, M. (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* **17:** 10–12.

Molina, L., Bernal, P., Udaondo, Z., Segura, A., and Ramos, J.L. (2013) Complete genome sequence of a *Pseudomonas putida* clinical isolate, strain H8234. *Genome Announc* **1:** e00496–e00413.

Molina, L., Udaondo, Z., Duque, E., Fernández, M., Molina-Santiago, C., Roca, A., *et al*. (2014) Antibiotic resistance determinants in a *Pseudomonas putida* strain isolated from a hospital. *PLoS One* **9:** e81604.

Mulet, M., Sánchez, D., Lalucat, J., Lee, K., and García-Valdés, E. (2015) *Pseudomonas alkylphenolia* sp. nov., a bacterial species able to form special aerial structures when grown on p-cresol. *Int J Syst Evo Microbiol* **65:** 4013–4018.

Oren, A., and Garrity, G.M. (2014) Then and now: a systematic review of the systematics of prokaryotes in the last 80 years. *Antonie Van Leeuwenhoek* **106:** 43–56.

Palleroni, N.J. (1992) Introduction to the Pseudomonadaceae. In *The Prokaryotes, a Handbook on the Biology of Bacteria, Ecophysiology, Isolation, Identification and Applications*, Vol.**III**, 2nd ed. Balows, A., Truper, H.G., Dworkin, M., Harder, W., and Schlepper, K.H. (eds). New York, NY, USA: Springer, pp. 3071–3085.

Park, G.S., Chu, J.H., Hong, S.J., Kwak, Y., Khan, A.R., Jung, B.K., *et al*. (2014) Complete genome sequence of the caprolactam-degrading bacterium *Pseudomonas mosselii* SJ10 isolated from wastewater of a nylon 6 production plant. *J Biotechnol* **192:** 263–264.

Patten, C.L., and Glick, B.R. (2002) Role of *Pseudomonas putida* indoleacetic acid in development of the host plant root system. *Appl Environ Microbiol* **68:** 3795–3801.

Paulsen, I.T., Press, C.M., Ravel, J., Kobayashi, D.Y., Myers, G.S., Mavrodi, D.V., *et al*. (2005) Complete genome sequence of the plant commensal *Pseudomonas fluorescens* Pf-5. *Nat Biotechnol* **23:** 873–878.

Pliego, C., López-Herrera, C., Ramos, C., and Cazorla, F.M. (2012) Developing tools to unravel the biological secrets of *Rosellinia necatrix*, and emergent threat to woody crops. *Mol Plant Pathol* **13:** 226–239.

Rainey, P.B., and Bailey, M.J. (1996) Physical and genetic map of the *Pseudomonas fluorescens* SBW25 chromosome. *Mol Microbiol* **19:** 521–533.

Rani, P., Mahato, N.K., Sharma, A., Rao, D.L.N., Kamra, K., and Lal, R. (2017) Genome mining and predictive functional profiling of acidophilic rhizobacterium *Pseudomonas fluorescens* Pt14. *Indian J Microbiol* **57:** 155–161.

Rezzonico, F., Binder, C., Défago, G., and Moënne-Loccoz, Y. (2005) The type III secretion system of biocontrol *Pseudomonas fluorescens* KD targets the phytopathogenic Chromista *Pythium ultimum* and promotes cucumber protection. *Mol Plant Microbe Interact* **18:** 991–1001.

Ross, S.M. (1993) Organic matter in tropical soils: current conditions, concerns and prospects for conservation. *Prog Phys Geogr* **17:** 265–305.

Seeman, T. (2014) Prokka: rapid prokaryotic genome annotation. *Bioinformatics* **30:** 2068–2069.

Silby, M.W., Winstanley, C., Godfrey, S.A.C., Levy, S.B., and Jackson, R.W. (2011) *Pseudomonas* genomes: diverse and adaptable. *FEMS Microbiol Rev* **35:** 652–680.

Spiers, A.J., Buckling, A., and Rainey, P.B. (2000) The causes of *Pseudomonas* diversity. *Microbiology* **146:** 2345–2350.

Stamatakis, A. (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30:** 1312–1313.

Taghavi, S., Garafola, C., Monchy, S., Newman, L., Hoffman, A., Weyens, N., *et al*. (2009) Genome survey and characterization of endophytic bacteria exhibiting a beneficial effect on growth and development of poplar trees. *Appl Environ Microbiol* **75:** 748–757.

Tao, F., Shen, Y., Fan, Z., Tang, H., and Xu, P. (2012) Genome sequence of *Pseudomonas putida* S12, a potential platform strain for industrial production of valuable chemicals. *J Bacteriol* **194:** 5985–5986.

Trapet, P., Avoscan, L., Klinguer, A., Pateyron, S., Citerne, S., Chervin, C., *et al*. (2016) The *Pseudomonas fluorescens* siderophore pyoverdine weakens *Arabidopsis thaliana* defense in favour of growth in iron-deficient conditions. *Plant Physiol* **171:** 675–693. doi:10.1104/pp.15.01537.

Udaondo, Z., Molina, L., Segura, A., Duque, E., and Ramos, J.L. (2016) Analysis of the core genome and pangenome of *Pseudomonas putida*. *Environ Microbiol* **18:** 3268–3283.

Vilo, C.A., Benedik, M.J., Kunz, D.A., and Dong, Q. (2012) Draft genome sequence of the cyanide-utilizing bacterium *Pseudomonas fluorescens* strain NCIMB 11764. *J Bacteriol* **194:** 6618–6619.

Visca, P., Imperi, F., and Lamont, I.L. (2007) Pyoverdine siderophores: from biogenesis to biosignificance. *Trends Microbiol* **15:** 22–30.

de Werra, P., Pechy-Tarr, M., Keel, C., and Maurhofer, M. (2009) Role of gluconic acid production in the regulation of biocontrol traits of *Pseudomonas fluorescens* CHA0. *Appl Environ Microbiol* **75:** 4162–4174.

Wickham, H. (2009) *ggplot2: Elegant Graphics for Data Analysis*. New York, NY, USA: Springer-Verlag.

Yang, L., Jelsbak, L., Marvig, R.L., Damkiær, S., Workman, C.T., Rau, M.H., *et al*. (2011) Evolutionary dynamics of bacteria in a human host environment. *Proc Natl Acad Sci USA* **108:** 7481–7486.

Supporting information

Additional Supporting Information may be found in the online version of this article at the publisher's web-site:

Fig. S1. Non-metric Multidimensional Scaling (NMDS) using the Bray-Curtis similarity index for ordination of samples (isolates and references) according to the ANI values. Black dots are the reference strains; blue dots are isolates in the *P. koreensis* subgroup; purple dots are isolates in the *P. jessenii* subgroup; red dots are isolates in the *P. putida* group; and grey dots are isolates in the new subclades X and Y.

Fig. S2. Synteny analyses performed by multiple genome alignments using the mauve software. Different colours in the genome sequences correspond to genome regions with high percentage alignment. Lines indicate rearranged regions and show their position in the compared genomes. The origin of replication is approximately in the middle of each genome sequence, based on the reference genome sequences used to reorder the contigs of the other assemblies analysed, that is, A) isolate R12 ($\sim$2962 Mbp) for the comparison inside the subclade X; and B) isolate R26 ($\sim$3148 Mbp) for the comparison inside the subclade Y.

Table S1. Accession numbers of the 76 genome sequences available in the Genbank/DDBJ/ENA databases.

Table S2. Main quality parameters of genome assemblies output.

Table S3. BLAST searches for the set of genes related to important functions performed by known *P. fluorescens* strains. Functions were considered present in a genome only if the complete set of genes were found with high similarity and identity hits in the tBLASTn. Abbreviations: DAPG, 2,4-diacetylphloroglucinol; GABA, Gamma-aminobutyric acid; IAA, indole acetic acid (*iaaMH*); Tcc4 and Tcc5, toxin complex clusters; AprA, exoprotease; T2SS, type 2 secretion system; T3SS, type 3 secretion system; T6SS, type VI secretion system. Toxin complex clusters and secretion systems are designated according to Loper and colleagues (2012).