Check for updates

Genomic diversity in *Paenibacillus polymyxa*: unveiling distinct species groups and functional variability

Adrian Wallner[1*], Livio Antonielli[1], Ouiza Mesguida[2,3], Patrice Rey[2] and Stéphane Compant[1]

Abstract

Background *Paenibacillus polymyxa* is a bacterial species of high interest, as suggested by the increased number of publications on its functions in the past years. Accordingly, the number of described strains and sequenced genomes is also on the rise. While functional diversity of *P. polymyxa* has been suggested before, the available genomic data is now sufficient for robust comparative genomics analyses.

Results Using 157 genomes, we found significant disparities among strains currently affiliated to *P. polymyxa*. Multiple taxonomic groups were identified with conserved predicted functions putatively impacting their respective ecology. As strains of this species have been reported to exhibit considerable potential in agriculture, medicine, and bioremediation, it is preferable to clarify their taxonomic organization to facilitate reliable and durable approval as active ingredients.

Conclusions Strains currently affiliated to *P. polymyxa* can be separated into two major species groups with differential potential in nitrogen fixation, plant interaction, secondary metabolism, and antimicrobial resistance, as inferred from genomic data.

Keywords *Paenibacillus polymyxa*, Taxonomy, Species groups, Plant growth promotion

Background

Paenibacillus polymyxa is a frequent inhabitant of various niches including soil, plant rhizospheres and plant tissues but also digestive tracts of different animals [1, 2]. It has further been occasionally found in samples of seawater and fermented foods, but is also an inhabitant of the International Space Station [3–5]. The species has undergone repeated taxonomic shifts and was reclassified as the type species of the novel genus *Paenibacillus* in 1994 [6, 7].

Due to its prolific secondary metabolism, manifold plant growth promoting potentials, probiotic status and bioremediation activities, *P. polymyxa* has been described as a model organism for sustainable plant, human and animal health management within the One Health framework [8].

Strains of *P. polymyxa* are reported to provide multiple benefits for plant health in the shape of direct growth promotion. They increase nutrient availability through the conversion of atmospheric nitrogen to plant available ammonia, phosphate solubilization and iron scavenging [9–12]. *P. polymyxa* is also used as a biocontrol agent,

*Correspondence:
Adrian Wallner
adrian.wallner@ait.ac.at
[1]Center for Health & Bioresources, Bioresources Unit, AIT Austrian Institute of Technology, Konrad Lorenz Str. 24, Tulln 3430, Austria
[2]E2S UPPA, CNRS, IPREM, Université de Pau et des Pays de l'Adour, Pau 64000, France
[3]GreenCell, Biopôle Clermont-Limagne, Saint Beauzire 63360, France

known for its ability to produce volatile compounds that inhibit fungal plant pathogens [13, 14]. For example, the production of antifungal compounds by *P. polymyxa* is highly effective against *Fusarium* species, and this bacterium can also induce systemic resistance to pathogens in plants [14–16]. The overall beneficial traits of *P. polymyxa*, i.e. its potential for biofertilization, biocontrol and protection against abiotic stresses, have been recently reviewed [4].

High interest in this species and simplified application of sequencing technologies have led to a leap in genome sequence depositions on public repositories. The first comparative genomics analysis for *P. polymyxa* is a decade old and included only four strains [17]. The pangenome of 14 strains was explored more recently [18]. In 2024, about 160 strains are affiliated to *P. polymyxa* when cross-referencing NCBI and GTDB databases. This considerable amount of genomic information coupled with the latest analysis tools enable a robust comprehension of this high-interest species.

Here we provide a species-wide comparative genomics analysis underlining the molecular oppositions between the different species groups currently affiliated to *P. polymyxa*. We also describe the putative functionalities of these groups on a basis of plant growth promotion traits, antimicrobial resistance, and secondary metabolism to obtain a comprehensive picture of their host interaction capacities.

Methods
Genomic data
Genome sequences belonging to *P. polymyxa* were downloaded from NCBI's RefSeq database. Additional genome sequences referenced as *P. polymyxa* on GTDB were downloaded from NCBI's GenBank database (Table S1). The homogenous affiliation to *P. polymyxa* was verified using the GTDB-Tk v2.1.1 classify workflow [19–25]. Genome completeness levels were assessed using the checkM v1.2.2 lineage workflow [26]. All assemblies were annotated *de novo* using Bakta v1.9.2 with baktaDB5.1 to standardize protein annotation between strains [27].

Overall genome relatedness index calculations
ANI calculations were conducted with Pyani v0.2.12 [28] using either BlastN (ANIb) or MUMmer (ANIm) supported alignments [29, 30]. Standard threshold for species delimitation were considered i.e., 95% identity with 70% sequence coverage. ANIb was used to compare the nucleotide sequences from the 157 strains affiliated to *P. polymyxa* and ANIm to compare the type strains of 93 *Paenibacillus* species. ANIm is stricter in detecting matches between more divergent genomes and was thus preferred to ANIb for the intra-genus comparison [31].

Exact ANIb identity scores and coverage levels are available in the supplementary data (Table S2-S3).

AAI was inferred with EzAAI v1.2.3 [32] using the default parameters for protein similarity detection based on MMseqs2 [33]. Exact AAI identity scores and coverage levels are available in the supplementary data (Table S4-S5).

Digital DDH was calculated using the GGDC 3.0 web application [34]. As all-against-all comparisons are impractical with this tool, only one representative of each subspecies group was used for distance calculation to all *P. polymyxa* genomes. The type strain genome, GCF_022811565.1, was selected for species group A2 and genomes with a good completeness status were chosen for the remaining species groups namely GCF_000237325.1, GCF_000146875.3, GCF_035323085.1 and GCF_001719045.1 for species groups A1, B1, B2 and B3 respectively. Given the genomes completeness of the input genomes we used and report the results from formula 3, summing all identities found in HSPs and dividing them by total genome length (Table S6).

Phylogenetic reconstructions
Detection of conserved gene orthologues was conducted separately on the complete set of 157 genomes affiliated to *P. polymyxa* and to the same set with five additional *P. ottowii* genomes using OrthoFinder v2.5.5 [35]. OrthoFinder was run with MSA, MAFFT and FastTree parameters. The aligned amino acid sequences of single copy orthologous genes were used for approximately-maximum-likelihood phylogenetic tree inference using FastTree v2.1.11 [23]. The JTT model of amino acid evolution with CAT approximation and a Gamma20-based likelihood branch rescaling was used. Local support values are provided through the Shimodaira-Hasegawa test [36].

Pangenome analyses
Gene conservation across strains was calculated using PPanGGOLiN v2.0.2 on bakta-generated gbff files [37]. Separate analyses were run for the total of 157 *P. polymyxa* affiliated strains, strains belonging to species group A and strains belonging to species group B (Table S7). For all analyses, gene clustering thresholds at 80% identity and 80% coverage were used.

Functional screening
Protein sequences were annotated and analyzed with the PGPT_BASE_nr_Aug2021n_ul_1 database from PLaBAse and the AnnoTree Bacteria database (version of 25 August 2020) using the DIAMOND+MEGAN pipeline with DIAMOND v2.1.8 and MEGAN v6.24.23 [38–41]. DIAMOND BLASTp was run with a threshold of 60% sequence similarity and an E-value cutoff of $<10^{-5}$.

Further hit filtering occurred in MEGAN6 (Naïve Lowest Common Ancestor (LCA) method: Minimum-Score=50, Top-percent filter=10%, Minimum-Support=50). Functional cluster analyses were generated using PCoA based on Bray-Curtis distances. Complete ontologies of detected functions in the pgpt-db, SEED and EC classifications are available in the supplementary data (Table S8-S10). The distribution of the *nifHDKENB* operon was confirmed by a blast screening of the 10,485-nucleotide long sequence from *P. polymyxa* strain CR1. Detection thresholds were set to 80% similarity on 80% sequence coverage.

GenBank format annotation and sequence files were used as input to screen for secondary metabolite synthesis pathways using antiSMASH v.6.1.1 [42]. Only contigs with a length of 500 nt or more were considered for the analysis consisting of a whole-genome HMMer scan, active site finder and comparison to antiSMASH's database of known cluster (Table S11).

For antimicrobial resistance screening, the CARD database v3.2.9 was used with the resistance gene identifier tool v 6.0.3 on protein data files using the DIAMOND alignment tool [43]. Additionally, resistance genes that were not detected by alignment to the CARD databases but mentioned in a study on the metabolic potential of *P. polymyxa* OSY-DF i.e., penicillin, tellurium, daunorubicin, fosmidomycin, erythromycin, kanamycin, teicoplanin and chloramphenicol resistance genes, were screened manually using the reference sequences provided by the authors [3].

Macromolecular systems and CRISPR-Cas systems were detected using MacSyFinder v2.1.3 with TXSScan v1.1.3 and CasFinder v3.1.0 models [44–46]. The accumulation of CRISPR/Cas systems can be a sign of adaptation to environments with high phage pressure [2].

Synteny analyses

Genomic regions of interest were detected by applying a differential abundance analysis on the pangenome output using LEfSe v1.1.2 [36]. Species groups were used as class, Species subgroups were used as subclass and individual IDs were used as subject for the analysis.

Synteny comparison figures were drawn using Easyfig v2.2.5 [47]. For every region of interest, one representative of each species subgroup in which the region is fully or partially present is depicted. The similarity scores were inferred through blastn [29].

Results

Strains affiliated to *P. polymyxa* segregate into distinct species groups

Assemblies for 110 *P. polymyxa* genomes are available in the NCBI RefSeq database. Their affiliation to *P. polymyxa* was confirmed using the GTDB toolkit [19]. The GTDB references 47 additional genomes belonging to *P. polymyxa* which in NCBI are either affiliated to alternative or unnamed *Paenibacillus* species. A total of 157 whole genome assemblies, representing 152 different strains are compared here. We kept the multiple entries from the same strains in our comparison as they have different genome assemblies. The contamination levels inferred through CheckM [26] range from very low (<1% contamination, 147 genomes) to medium (1–6% contamination, 9 genomes) and only one genome has a contamination level of 20.7% at a completeness level of 98% (Table S1). This genome was kept throughout the different analyses as it displayed no particular divergence, and its singular status prevents any influence on the orthology approaches.

Average Nucleotide Identity (ANI) is a whole-genome similarity metric allowing robust comparisons of taxonomic identities in prokaryotes [48]. The ANI analyses conducted in a previous pangenome approach including 14 *P. polymyxa* strains concluded the existence of three species candidates [18]. Seven were considered part of the "true" *P. polymyxa* while six were attributed to *Paenibacillus peoriae* and one to a new species. All *P. polymyxa* strains included in the present analysis are genetically distinct from the *P. peoriae* type strain, as well as from any other type-strain of the 74 most closely affiliated *Paenibacillus* species (Figure S1).

Strains affiliated to *P. polymyxa* are separated in two major ANI groups, within which further subgroups can be distinguished (Fig. 1A). All species groups have an average ANIb identity score below 90% similarity while the subgroups share an average internal similarity of at least 97.5% identity (Table S2). The cluster containing the *P. polymyxa* type-strain ATCC 842 will be further called species group (**A**) The second dominant group will be referred to as species group (**B**) A single strain, currently affiliated to *P. polymyxa*, can be clearly distinguished as an outgroup showing no significant similarity to any other strain within the *P. polymyxa* group nor to any other *Paenibacillus* type strain and will further be called species group C (Fig. 1A & Figure S1). All alignments have a sufficient coverage to be interpreted as reliable results (Table S3). The average whole-genome similarity between species groups A and B is 89,5%, considerably lower than the accepted threshold of 95% in vigor for most species delimitations [31]. Based on nucleotide identity, species group B is closest to *P. polymyxa* species group A (mean ANIm=90.3%). Its next closest species is *Paenibacillus ottowii* (mean ANIm=89,3%). Species group C is closest to *Paenibacillus xylanexedens* (ANIm=94.4%). We will further refer to the strains used in this analysis as the *P. polymyxa* group as they are currently affiliated to this species, even though ANI levels compared to other *Paenibacillus* species suggest that this
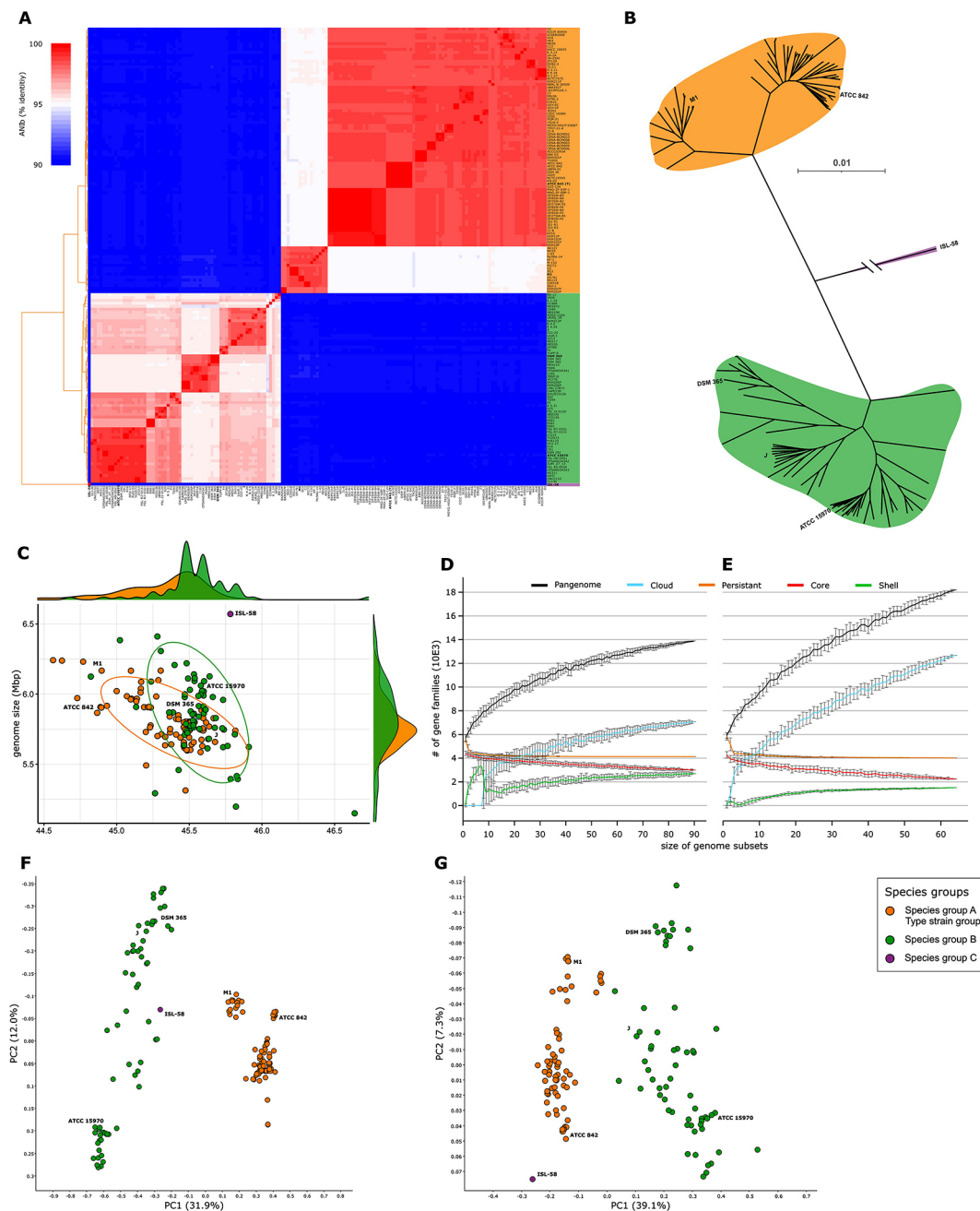
Wallner *et al. BMC Genomics*   (2024) 25:720

Page 4 of 11

Fig. 1 Segregation of *Paenibacillus polymyxa* strains. (**A**) Average nucleotide identity analysis based on pairwise blastn comparison of whole genome sequences between *P. polymyxa* strains (*n* = 157). (**B**) Phylogenetic reconstruction of *P. polymyxa* affiliated strains. Single copy orthologous protein sequences were detected and aligned using Orthofinder resulting in an alignment of 1,397 sequences for a total of 406,116 amino acids. The tree was build using FastTree with default parameters. A rooted tree with local support values and complete leaf labels is available as supplemental Figure S3. Branch length for ISL-58 is 0.43 (**C**) Genome size and GC content comparison between *P. polymyxa* strains (*n* = 157). The samples were colored according to the species groups determined through ANIb. (**D**) Rarefaction plots of pangenome calculations for species group A (**E**) and species group B. The different gene categories vary based on their frequency in the total observed population with cloud genes present in at least one organism, shell genes present in more than 8%, persistent genes present in more than 90% and core genes present in 100%. (**F**) Bray-Curtis PCoA representation of the genetic profile of *P. polymyxa* affiliated strains based on their pgpt-db affiliation or (**G**) their SEED affiliation

group is not monophyletic (Figure S1). A phylogenomic reconstruction including five *P. ottowii* strains confirms that *P. polymyxa* species group A shares a more recent common ancestor with *P. ottowii* than with *P. polymyxa*

species group B (Figure S2). Genomes identified as *P. polymyxa* only on GTDB are distributed in both major species groups (21 in species group A and 26 in species group B).

Clustering into additional subgroups can be observed in both species group A and B (Fig. 1). *P. polymyxa* species group A is clustered in two subgroups (hereafter subgroup A1 and A2) while species group B has 3 clusters (hereafter subgroup B1, B2 and B3). These subgroup repartition partially matches the automatic species separation for *P. polymyxa* found on GTDB. GTDB'S *P. polymyxa*_A matches with subgroup A2, *P. polymyxa*_C with subgroup A1, *P. polymyxa*_E with group C while *P. polymyxa*_B and *P. polymyxa*_D are not homogenously distributed across subspecies B1, B2 and B3 (Table S1).

To corroborate the ANI results, additional Overall Genome relatedness indices (OGRI) were used. Average Amino acid Identity (AAI) functions in a similar manner to ANI but uses protein sequences as input [32]. Digital DNA-DNA hybridization (dDDH) infers genome distances from whole genome alignments [34]. Finally, we compared the genomes considering the number of bases impacted by substitutions or insertion/deletion events. All OGRI confirm the results obtained through ANI. The species group described before are supported by a 95% AAI similarity (Table S4). The standard 70% dDDH threshold for species delineation is suitable to segregate Species A and Species B genomes (Table S6). All additional OGRI also indicate a strong clustering into species subgroups for species subgroup A1 and A2. For dDDH, the threshold for subspecies clustering was suggested to be 79% [49] which is reached for every species subgroup except B1.

A phylogenetic reconstruction was modelized using 1,397 single copy orthologous genes, conserved among the 157 strains in the *P. polymyxa* group (Fig. 1B & S4). The different species and subgroups appear clearly and are robustly supported by Shimodaira-Hasegawa values [50]. The strain identified as species group C serves as outgrouping sample.

Detailed information about the isolation sources of the different *P. polymyxa* strains is only partially available (Table S1). In both dominant species groups, the most abundant isolation sources are plant environments and soil with 35.0% and 17.8% of *P. polymyxa* isolates respectively. Although only 9 isolates originate from animal environments, mostly the digestive tract, all belong to *P. polymyxa* species group A.

Genome completeness levels are predicted to be similarly distributed across strains from the two major species groups (Table S1). Also, all strains included in the study were analyzed using GTDB-Tk to validate their current taxonomic standing and affiliation to *P. polymyxa* [19]. The strains have a genome size ranging from 5.15 Mbp for strain YUPP-8 to 6.57 Mbp for strain ISL-58. The lowest GC content is found in strain SC2 with 44.6% and the highest in strain YUPP-8 with 46.6% (Fig. 1C & Table 1). Strains of species group A have a tendentially smaller genome with lower GC% content than species group B strains.

Gene detection was conducted *de novo* on all genomic sequences to homogenize their annotation using Bakta [27]. Then, pangenome analyses were performed using Ppanggolin [37]. For every pangenome analysis, the rarefaction of the total number of genes indicates that there is still strain diversity to be found within the different species as no saturation of gene diversity is reached (Fig. 1D and E). The core genomes of species group A and B represent a high fraction of their total genome with 3,004 and 2,247 CDSs respectively (Table S7). Expectedly, species group A shows more conservation as suggested by the ANI levels (Fig. 1A). The larger strain diversity in species group B is also displayed through the size of its pangenome with 18,203 CDSs compared to 13,889 for species group A (Table S7). Another metric depicting the higher genome dissimilarity in species group B is the genomic fluidity of 0.167 compared to 0.12 in species group A [51]. Finally, unique genes are more abundant on average in species group B with 104 copies per strain compared to 31 in species group A. The persistent genomes of species group A and B comprise 4,129 and 4,014 CDSs respectively or 79.93% and 76.63% of their mean gene number (Table S7).

The detected coding sequences were further compared to the PlaBA-db and Annotree databases [38, 40, 41, 52]. In both cases, gene function does segregate the different

Table 1 Genomic features of *P. polymyxa* strains (*n* = 157). The average genome size in base pairs, number of coding sequences, coding density, GC percentage, number of ribosomal RNAs and number of non-coding RNAs were calculated for each species group and subspecies group

Species group	Genome size (bp)	CDSs	Coding density	GC (%)	rRNAs	ncRNAs
Group A (*P. polymyxa*)	**5,813,992**	**5,166**	**86.9**	**45.35**	**20.7**	**6.7**
Subgroup A1	5,858,337	5,237	86.2	45.13	24.4	7.0
Subgroup A2	5,804,532	5,150	87.1	45.40	20.0	6.6
Group B	**5,861,318**	**5,238**	**86.2**	**45.55**	**18.4**	**7.4**
Subgroup B1	5,964,228	5,283	86.1	45.57	18.9	7.5
Subgroup B2	5,804,772	5,221	86.4	45.44	13.4	6.7
Subgroup B3	5,758,371	5,199	86.3	45.66	18.4	7.4
Group C	**6,570,682**	**5,795**	**86.1**	**45.80**	**5.0**	**9.0**

Wallner *et al. BMC Genomics*   (2024) 25:720

Page 6 of 11

strains into the previously determined species groups, suggesting different metabolic capacities and different ecologies (Fig. 1F and G).

The *P. polymyxa* species groups display several distinct metabolic functions

While *P. polymyxa* is generally described as a nitrogen-fixing species, the irregular presence of this trait has been repeatedly observed (Xie et al., 2016; L. Zhou et al., 2020). Indeed, the presence of the *nifHDKENB* gene set is strongly correlated with the genomic background of species group B. The genes responsible for nitrogen fixation can be found in 87.7% of species group B, while they are only present in 5.5% of species group A and strictly restricted to the species subgroup A1 (Fig. 2A and F). Several nutrient transport systems were found to be segregating between species group. An operonic structure containing genes putatively involved in polyamine import was found exclusively in strains with species group B background (Fig. 2B and F). In species group A, strains are enriched with the *frl* and *lic* operons, respectively responsible for fructoselysine and β-glucoside utilization (Fig. 2C, D and F). While the *frl* operon is rare in species group B, a modified *lic* operon is common in the species subgroup B1. It contains an additional 6-phospho-β-glucosidase. A 42.9 kb-long genomic region responsible for secondary metabolism was detected to be preferentially present in species group A strains (Fig. 2E and F). The pathway is predicted to function as a trans-acyltransferase PKS/NRPS hybrid system and to generate a metabolite unknown to the antiSMASH database.

P. polymyxa has been repeatedly reported to possess multiple antimicrobial resistances [2, 16–18, 53]. Indeed, using the Comprehensive Antibiotic Resistance Database (CARD) complemented with a similarity search for resistance genes reported in *P. polymyxa* A18 [2], the *P. polymyxa* group displays genes for resistance against penicillin, tellurium, daunorubicin, fosmidomycin, erythromycin, kanamycin, teicoplanin, chloramphenicol, clindamycin, fosfomycin, tetracycline, vancomycin as well as benzalkonium chloride. However, these resistance genes are not all conserved across the different species groups. Species subgroups B1 and B3 are fully devoid of erythromycin resistance genes and only 47% of species group A2 possess them. Resistance against daunorubicin was lost in 73% of species subgroup A2 whereas kanamycin resistance was lost in 69% of species subgroup B3. All other resistance genes were detected in more than 70% of strains from each group (Fig. 3).

Genomes were screened for the presence of macromolecular systems using MacSyFinder [46]. The presence of a Tad-pilus and a flagellum is ubiquitous, except for species group C lacking the flagellum coding genes. In accordance to what has been reported in a previous comparative genomics approach, the Gram positive twin-arginine translocation pathway (*tatAC*) was found in all *P. polymyxa* strains [17]. All the components for the general secretory pathway *secAYEG* were also detected.

P. polymyxa A18, a gut isolate, was screened for CRISPR/Cas systems putatively conferring resistance against phage invasion which are present in high numbers in the gut environment [2]. Out of the 157 strains screened here, 77 across all species groups except species group C present complete CRISPR/Cas systems. Most of these strains display a Type I system against double stranded DNA (20.8% species group A, 46.2% species group B). However, the Type III system, targeting single strand RNA, is enriched in species subgroup A2 with 15 positive strains compared to only three in species group B (Fig. 3).

A screening of the secondary metabolism of the 157 strains using antiSMASH revealed interesting patterns for the production of anabaenopeptin NZ857/nostamide A, paenibacillin, paenilan and paenilipoheptin [54] (Fig. 3). The complete metabolic pathways to synthetize fusaricidine were detected in all but five strains. The capacity to synthetize polymyxin is conserved in all but twelve strains, distributed in the two major species groups with no apparent segregating pattern (Table S11).

From the pgpt-db we can infer what functional categories contribute most to the segregation between *P. polymyxa* species groups (Fig. 1F). Gene functions linked to plant growth promotion were essentially found to be enriched in species group B compared to species group A (Fig. 3). Species group B could benefit from an enhanced capacity to utilize plant derived compounds, improving its competitivity in the rhizosphere environment. Especially the species subgroup B2 subgroup is enriched in genes functioning in peptide, opine and beta-glucan utilization. Species subgroup B3 displays an enrichment in genes for fructan glycolysis. The direct benefits of species group B for plants can be derived from genes functioning in vitamin production, xenobiotic degradation, nickel detoxication, iron acquisition and sulfur assimilation (Fig. 3). Additional functions, directly or indirectly linked to plant growth promotion potential, such as nitrogenase activity, antimicrobial resistance or CRISPR-Cas systems were developed in more detail above.

Discussion

Comparative genomics analyses were performed on 157 *P. polymyxa* strains, resulting in a composite description of the species. Whole-genome alignments and phylogenomic reconstructions strongly suggest the presence of two dominant species groups that do not share a monophyletic position on the *Paenibacillus* genus tree. Ninety-one strains share genomic and functional similarities with the type strain ATCC 842 of *P. polymyxa*, while 65
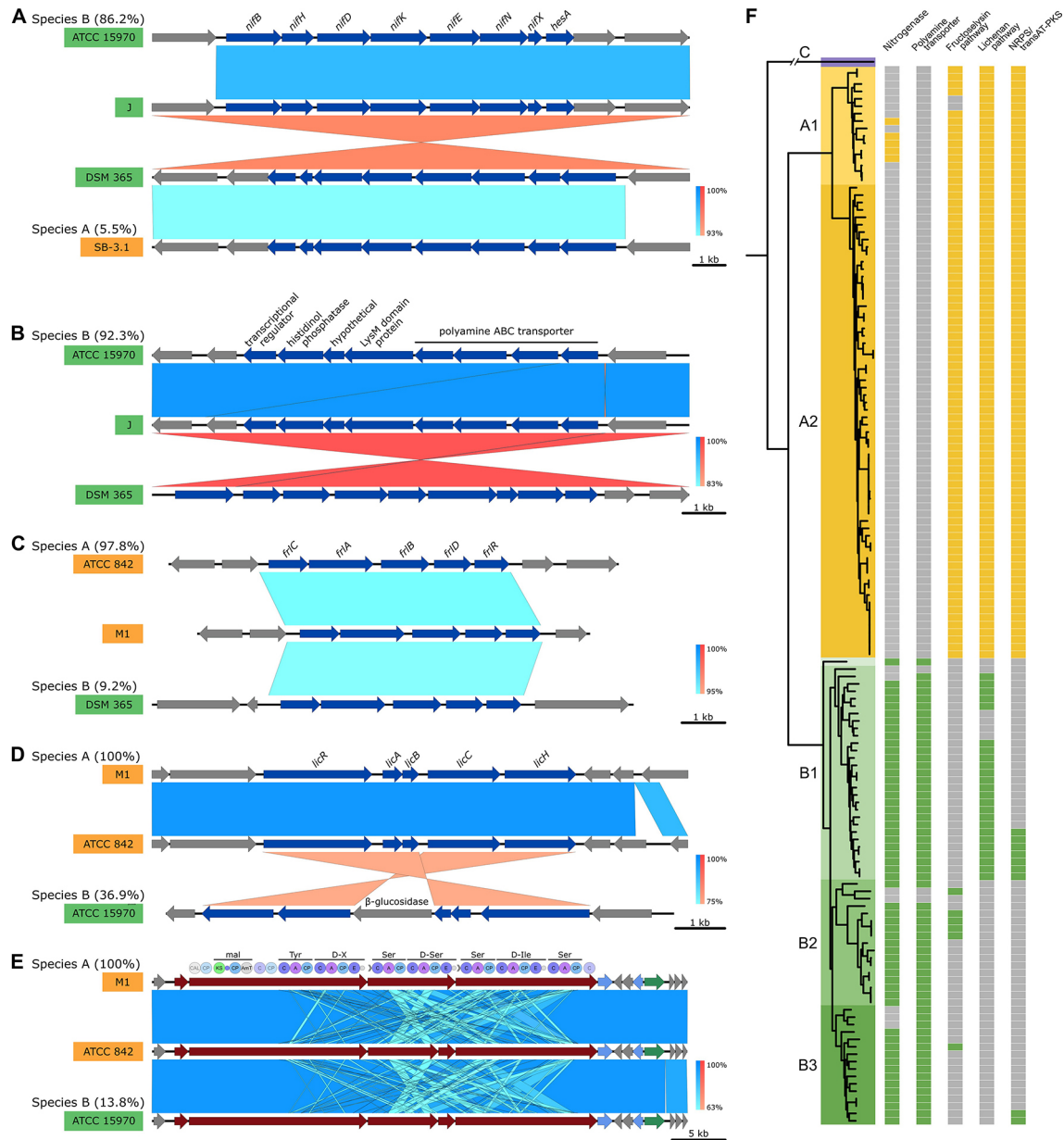
Fig. 2 Conserved enrichments of genetic pathways in *P. polymyxa* species groups. The synteny of a non-exhaustive selection of genetic pathways with differential abundances depending on the species group as inferred through Lefse (LDA > 1.9, pval < 1E-5) has been generated using EasyFig. In each case, the synteny consists of one representative of each species subgroup where the genomic pathway was detected. The total ratio of strains possessing the pathway is given for each species group. Similarity levels were inferred through blastn and are displayed using a blue color scale or a red color scale for inverted sequence identities. For the nitrogenase pathway (**A**) the polyamine transporter pathway (**B**) the fructoselysine degradation pathway (**C**) and the lichenan degradation operon (**D**) genes with a significant Lefse output (blue) are distinguished from the remaining ones (grey). For the NRPS/transAT-PKS hybrid system (**E**) genes were colored according to the AntiSmash color code distinguishing biosynthetic (red), transport (blue) and regulatory (green) genes. In addition, the module domain-composition is provided. Detailed presence absence profiles of the different pathways in the *P. polymyxa* strains (**F**)

strains show strong evidence for the existence of a separate species. Further based on genomic and functional criteria, several subgroups can be delimited within each *P. polymyxa* species group. While the species A subgroups are supported by all OGRI, those of species group B are less clearly defined and display less divergence.

The genomic distance between species A subgroups satisfies the requirements for subspecies delineation as they display an ANI of 94–95% and a dDDH of 69–79% [49]. Additional and dedicated analyses are required to reveal the true nature of putative species subgroups of *P. polymyxa*.
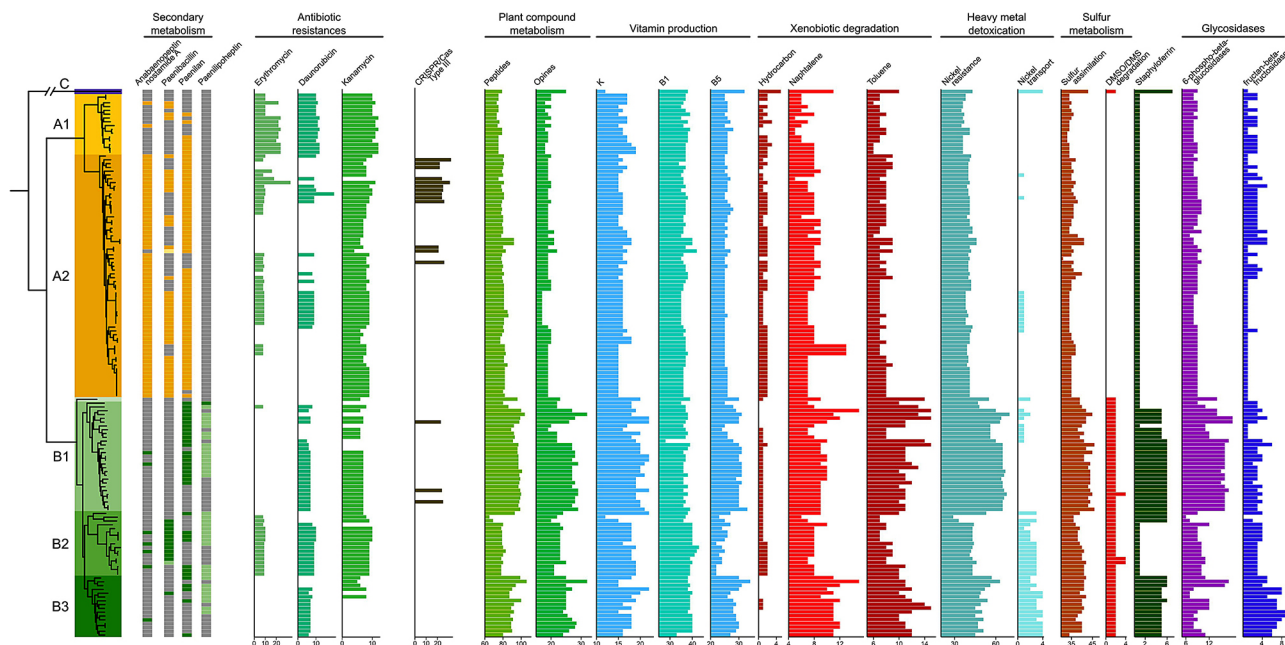
Wallner *et al. BMC Genomics*   (2024) 25:720

Page 8 of 11

Fig. 3 Gene of interest distribution across *P. polymyxa* strains. Genes involved in secondary metabolism were inferred through antiSMASH. Dark-color squares indicate the presence of pathways with > 90% similarity. Light-color squares represent incomplete pathways with > 40% similarity. Antibiotic resistance genes were detected using the CARD database. The CRISPR/Cas systems were detected using MacSyFinder. All remaining functions were detected with the pgpt-db. The list is not exhaustive, and only pathways with differential distributions between species groups or subgroups are represented

The genomic diversity of *P. polymyxa* strains has been a source of confusion in the past. In 2001, an olive-mill wastewater isolate, strain KACC 10,925, was compared to *P. polymyxa* CECT 153 and resulted in significantly low DNA-DNA relatedness, prompting the authors to propose *Paenibacillus jamilae* as a new species [55]. Almost 20 years later, it was argued that when using the type strain *P. polymyxa* ATCC 842, the comparison with KACC 10,925 yields sufficiently high similarity to be classified as the same species, relegating *P. jamilae* as a heterotypic synonym of *P. polymyxa* [56]. This confusion arose from the heterogeneity of strains classified as *P. polymya*, which we described here. Strain CECT 153 is part of the species B group, which is indeed genomically distant from the type strain group, including KACC 10,925. Divergence from the *P. polymyxa* type strain, belonging to species group A, was also found in the MALDI-TOF MS profiles of *P. polymyxa* strains DSM 365 and DSM 292, both part of species group B [57].

We have put the focus on the plant interactive traits which are often reported for *P. polymyxa* [4, 18]. As it is a promising plant growth promotion agent, it is intriguing to verify how the different species groups differ in their genetic baggage dedicated for plant interaction. A major recognition, is that the inconsistent distribution of nitrogen fixation genes, repeatedly described in the literature, is heavily correlated with the genetic background of the strain [9, 18]. This feature is only rarely found in strains with the genetic background of the *P. polymyxa*

type strain, and largely enriched in the newly separated species. This species group shows further enrichment in genes directly or indirectly involved in plant growth promotion, as well as plant substrate usage.

The pangenomic analysis revealed presence/absence patterns for further complete metabolic pathways, correlating with the different genetic backgrounds of the *P. polymyxa* species groups. The putative polyamine transporter detected exclusively in the species group B background could also have an impact on plant interaction. *Pseudomonas* sp. WCS365 uses the polyamine putrescine as a signaling molecule with its host during colonization of *Arabidopsis thaliana* [58]. The implication of putrescine-related pathways in plant colonization was also underlined for *Burkholderia* species interacting with rice [59, 60]. Polyamines further play a role in biofilm formation in different bacterial models [61].

The *P. polymyxa* type strain species group and the species group B group display a strong conservation in secondary metabolism. Both species groups retain the capacity to synthetize polymyxin, often described as a last resort antimicrobial against multidrug-resistant pathogens [62]. They also have the capacity to produce fusaricidin a potent antimicrobial against Gram-positive bacteria and phytopathogenic fungi [15, 63]. However, a predicted trans-acyltransferase PKS/NRPS system is conserved in every strain of species group A, while it is only present in few species group B strains. It is predicted to be composed of seven amino acids but displays no

similarity to any synthesis pathway in secondary metabolite databases. Its strong conservation level within the *P. polymyxa* type strain group make it an interesting target for future research.

The *lic* operon was studied in *Bacillus subtilis* and allows the bacterium to use oligomeric β-glucosides as sole carbon source [64]. In *Clostridiodes difficile* it enables the utilization of cellobiose [65]. Cellobiose is a product of cellulose degradation, but it can also be found in maize grains and is present in maize root where it participates in immune responses [66]. The presence of the *lic* operon could benefit organisms associating with plants producing this sugar or with cellulose-degrading microorganisms, scavenging the breakdown compounds. Interestingly, the species subgroup B1 subgroup possesses a *lic* operon containing an additional gene identifies as a 6-phospho-β-glucosidase. This function is also found in the cellobiose-utilization operon of *Escherichia coli* [67]. While in *E. coli* this function is essential for the utilization of cellobiose, a homologous gene is present outside of the operon in strains of species group A. The *frl* operon is almost completely absent in species group B but highly conserved in species group A. It enables the use of fructoselysine, a stable glycation product of glucose and lysine formed nonenzymatically [68].

Both cellobiose and fructoselysine are degradation products and found in animal digestive systems where they are used by intestinal bacteria. All *P. polymyxa* strains originating from animal associated environments belong to the genomic background where *frl* and *lic* operons are enriched. This species group is also enriched in antimicrobial resistance genes as well as in CRISPR-Cas systems which could benefit strains in environment with increased viral pressure such as the gut environment.

Conclusion
Our approach of combining multiple bioinformatics analyses to compare 157 genomes of *P. polymyxa* at the genomic and functional levels revealed the existence of two species groups, as 65 strains show significant genomic divergences and specific functional traits compared to the 91 strains showing similarities to the type strain *P. polymyxa* ATCC 842. We identified different functions with significant enrichment levels suggesting a specialization towards different lifestyles for the two species groups. Strains sharing similarity with the *P. polymyxa* type strain are predictively less apt to colonize the plant environment and do not possess the genomic capacities for nitrogen fixation.

Our analysis reveals two potential new species of *Paenibacillus* (here species groups B and C) which deserve further investigation and biochemical characterization to be formally classified as separate species with new type strains. Also, the strict genomic separation between species A subgroups deserves further scrutiny for their classification as *P. polymyxa* subspecies.

Abbreviations
AAI	Average Amino-acid Identity
ANI	Average Nucleotide Identity
CARD	Comprehensive Antibiotic Resistance Database
dDDH	digital DNA-DNA Hybridization
OGRI	Overall Genome Relatedness Indices

Supplementary Information
The online version contains supplementary material available at https://doi.org/10.1186/s12864-024-10610-w.

Supplementary Material 1

Supplementary Material 2

Data availability
All genomes used in this study are available on NCBI's GenBank database (https://www.ncbi.nlm.nih.gov/genbank/). Accession numbers for all genomes can be found in Table S1.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

References
1. Padda KP, Puri A, Chanway CP. *Paenibacillus polymyxa*: a prominent biofertilizer and biocontrol agent for sustainable agriculture. Agriculturally important microbes for sustainable agriculture. Singapore: Springer Singapore; 2017. p. 165–91. https://doi.org/10.1007/978-981-10-5343-6_6.
2. Pasari N, Gupta M, Eqbal D, Yazdani SS. Genome analysis of *Paenibacillus polymyxa* A18 gives insights into the features associated with its adaptation

to the termite gut environment. Sci Rep. 2019;9(1):6091. https://www.nature.com/articles/s41598-019-42572-5.

3. Huang E, Yousef AE. Draft genome sequence of *Paenibacillus polymyxa* OSY-DF, which coproduces a lantibiotic, paenibacillin, and polymyxin E1. J Bacteriol. 2012;194(17):4739–40. https://doi.org/10.1128/JB.00846-12.

4. Langendries S, Goormachtig S. *Paenibacillus polymyxa*, a jack of all trades. Environ Microbiol. 2021;23(10):5659–69. https://doi.org/10.1111/1462-2920.15450.

5. Daudu R, Singh NK, Wood JM, Debieu M, O'Hara NB, Mason CE et al. Draft genome sequences of *Bacillaceae* strains isolated from the international space station. Gill SR, editor. Microbiol Resour Announc. 2020;9(44). https://doi.org/10.1128/MRA.00701-20.

6. Ash C, Priest FG, Collins MD. Molecular identification of rRNA group 3 bacilli (Ash, Farrow, Wallbanks and Collins) using a PCR probe test. Antonie Van Leeuwenhoek. 1994;64(3–4):253–60. https://doi.org/10.1007/BF00873085.

7. Judicial Commission of the International Committee on Systematics of Prokaryotes. The type species of the genus *Paenibacillus* Ash et al. 1994 is *Paenibacillus polymyxa*. Opinion 77. Int J Syst Evol Microbiol et al. 2005;55(1):513–513. https://www.microbiologyresearch.org/content/journal/ijsem/https://doi.org/10.1099/ijs.0.63546-0.

8. Pandey AK, Barbetti MJ, Lamichhane JR. *Paenibacillus polymyxa*. Trends Microbiol. 2023;31(6):657–9. https://linkinghub.elsevier.com/retrieve/pii/S0966842X22003183.

9. Xie J, Shi H, Du Z, Wang T, Liu X, Chen S. Comparative genomic and functional analysis reveal conservation of plant growth promoting traits in *Paenibacillus polymyxa* and its closely related species. Sci Rep. 2016;6:21329. https://www.nature.com/articles/srep21329.

10. Puri A, Padda KP, Chanway CP. Evidence of nitrogen fixation and growth promotion in canola (*Brassica napus* L.) by an endophytic diazotroph *Paenibacillus polymyxa* P2b-2R. Biol Fertil Soils. 2016;52(1):119–25. https://doi.org/10.1007/s00374-015-1051-y.

11. Raza W, Shen Q, Growth. Fe3 + reductase activity, and siderophore production by *Paenibacillus polymyxa* SQR-21 under differential iron conditions. Curr Microbiol. 2010;61(5):390–5. https://doi.org/10.1007/s00284-010-9624-3.

12. Zhou C, Guo J, Zhu L, Xiao X, Xie Y, Zhu J, et al. *Paenibacillus polymyxa* BFKC01 enhances plant iron absorption via improved root systems and activated iron acquisition mechanisms. Plant Physiol Biochem. 2016;105:162–73. https://linkinghub.elsevier.com/retrieve/pii/S0981942816301413.

13. Rybakova D, Rack-Wetzlinger U, Cernava T, Schaefer A, Schmuck M, Berg G. Aerial warfare: a volatile dialogue between the plant pathogen *Verticillium longisporum* and its antagonist *Paenibacillus polymyxa*. Front Plant Sci. 2017;8. https://doi.org/10.3389/fpls.2017.01294/full.

14. Lee B, Farag MA, Park HB, Kloepper JW, Lee SH, Ryu CM. Induced resistance by a long-chain bacterial volatile: elicitation of plant systemic defense by a C13 volatile produced by *Paenibacillus polymyxa*. Vinatzer BA, editor. PLoS One. 2012;7(11):e48744. https://doi.org/10.1371/journal.pone.0048744.

15. Li Y, Chen S. Fusaricidin produced by *Paenibacillus polymyxa* WLY78 induces systemic resistance against *Fusarium* wilt of cucumber. Int J Mol Sci. 2019;20(20):5240. https://www.mdpi.com/1422-0067/20/20/5240.

16. Jeong H, Choi SK, Ryu CM, Park SH. Chronicle of a soil bacterium: *Paenibacillus polymyxa* E681 as a tiny guardian of plant and human health. Front Microbiol. 2019;10. https://doi.org/10.3389/fmicb.2019.00467/full.

17. Eastman AW, Heinrichs DE, Yuan ZC. Comparative and genetic analysis of the four sequenced *Paenibacillus polymyxa* genomes reveals a diverse metabolism and conservation of genes relevant to plant-growth promotion and competitiveness. BMC Genomics. 2014;15:851. https://doi.org/10.1186/1471-2164-15-851.

18. Zhou L, Zhang T, Tang S, Fu X, Yu S. Pan-genome analysis of *Paenibacillus polymyxa* strains reveals the mechanism of plant growth promotion and biocontrol. Antonie Van Leeuwenhoek. 2020;113(11):1539–58. https://doi.org/10.1007/s10482-020-01461-y.

19. Chaumeil PA, Mussig AJ, Hugenholtz P, Parks DH. GTDB-Tk v2: memory friendly classification with the genome taxonomy database. Borgwardt K, editor. Bioinformatics. 2022;38(23):5315–6. https://academic.oup.com/bioinformatics/article/38/23/5315/6758240.

20. Matsen FA, Kodner RB, Armbrust EV. Pplacer: linear time maximum-likelihood and bayesian phylogenetic placement of sequences onto a fixed reference tree. BMC Bioinformatics. 2010;11(1):538. https://doi.org/10.1186/1471-2105-11-538.

21. Jain C, Rodriguez-R LM, Phillippy AM, Konstantinidis KT, Aluru S. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species

boundaries. Nat Commun. 2018;9(1):5114. https://www.nature.com/articles/s41467-018-07641-9.

22. Hyatt D, Chen GL, LoCascio PF, Land ML, Larimer FW, Hauser LJ. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics. 2010;11(1):119. https://doi.org/10.1186/1471-2105-11-119.

23. Price MN, Dehal PS, Arkin AP. FastTree 2 – approximately maximum-likelihood trees for large alignments. PLoS ONE. 2010;5(3):e9490. https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0009490.

24. Eddy SR, Accelerated Profile HMM, Searches. Pearson WR, editors. PLoS Comput Biol. 2011;7(10):e1002195. https://dx.plos.org/10.1371/journal.pcbi.1002195.

25. Ondov BD, Treangen TJ, Melsted P, Mallonee AB, Bergman NH, Koren S, et al. Mash: fast genome and metagenome distance estimation using MinHash. Genome Biol. 2016;17(1):132. https://doi.org/10.1186/s13059-016-0997-x.

26. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res. 2015;25(7):1043–55. https://doi.org/10.1101/gr.186072.114.

27. Schwengers O, Jelonek L, Dieckmann MA, Beyvers S, Blom J, Goesmann A. Bakta: rapid and standardized annotation of bacterial genomes via alignment-free sequence identification. Microb Genomics. 2021;7(11). https://doi.org/10.1099/mgen.0.000685.

28. Pritchard L, Glover RH, Humphris S, Elphinstone JG, Toth IK. Genomics and taxonomy in diagnostics for food security: soft-rotting enterobacterial plant pathogens. Anal Methods. 2016;8(1):12–24. http://xlink.rsc.org/?DOI=C5AY02550H.

29. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol. 1990;215(3):403–10.

30. Kurtz S, Phillippy AM, Delcher AL, Smoot M, Shumway M, Antonescu C, et al. Versatile and open software for comparing large genomes. Genome Biol. 2004;5:R12.

31. Richter M, Rosselló-Móra R. Shifting the genomic gold standard for the prokaryotic species definition. Proc Natl Acad Sci. 2009;106(45):19126–31. https://doi.org/10.1073/pnas.0906412106.

32. Kim D, Park S, Chun J. Introducing EzAAI: a pipeline for high throughput calculations of prokaryotic average amino acid identity. J Microbiol. 2021;59(5):476–80. https://doi.org/10.1007/s12275-021-1154-0.

33. Steinegger M, Söding J. MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. Nat Biotechnol. 2017;35(11):1026–8. https://www.nature.com/articles/nbt.3988.

34. Meier-Kolthoff JP, Carbasse JS, Peinado-Olarte RL, Göker M. TYGS and LPSN: a database tandem for fast and reliable genome-based classification and nomenclature of prokaryotes. Nucleic Acids Res. 2022;50(D1):D801–7. https://academic.oup.com/nar/article/50/D1/D801/6389592.

35. Emms DM, Kelly S. OrthoFinder: phylogenetic orthology inference for comparative genomics. Genome Biol. 2019;20(1):238. https://genomebiology.biomedcentral.com/articles/10.1186/s13059-019-1832-y.

36. Segata N, Izard J, Waldron L, Gevers D, Miropolsky L, Garrett WS, et al. Metagenomic biomarker discovery and explanation. Genome Biol. 2011;12(6):1–18. https://doi.org/10.1186/gb-2011-12-6-r60.

37. Gautreau G, Bazin A, Gachet M, Planel R, Burlot L, Dubois M et al. PPanG-GOLiN: Depicting microbial diversity via a partitioned pangenome graph. Ouzounis CA, editor. PLOS Comput Biol. 2020;16(3):e1007732. https://doi.org/10.1371/journal.pcbi.1007732.

38. Patz S, Rauh M, Gautam A, Huson DH, mgPGPT. Metagenomic analysis of plant growth-promoting traits. bioRxiv. 2024.

39. Bağcı C, Patz S, Huson DH, DIAMOND + MEGAN. Fast and easy taxonomic and functional analysis of short and long microbiome sequences. Curr Protoc. 2021;1(3):e59. https://doi.org/10.1002/cpz1.59.

40. Gautam A, Felderhoff H, Bağci C, Huson DH. Using AnnoTree to get more assignments, faster, in DIAMOND + MEGAN microbiome analysis. mSystems. 2022;7(1). https://journals.asm.org/journal/msystems.

41. Mendler K, Chen H, Parks DH, Lobb B, Hug LA, Doxey AC. AnnoTree: visualization and exploration of a functionally annotated microbial tree of life. Nucleic Acids Res. 2019;47(9):4442–8. https://academic.oup.com/nar/article/47/9/4442/5432638.

42. Blin K, Shaw S, Kloosterman AM, Charlop-Powers Z, van Wezel GP, Medema MH, et al. antiSMASH 6.0: improving cluster detection and comparison capabilities. Nucleic Acids Res. 2021;49(W1):W29–35. https://academic.oup.com/nar/article/49/W1/W29/6274535.

43. Alcock BP, Huynh W, Chalil R, Smith KW, Raphenya AR, Wlodarski MA, et al. CARD 2023: expanded curation, support for machine learning, and resistome

Wallner *et al. BMC Genomics*   (2024) 25:720

Page 11 of 11

prediction at the Comprehensive Antibiotic Resistance Database. Nucleic Acids Res. 2023;51(D1):D690–9. https://academic.oup.com/nar/article/51/D1/D690/6764414.

44. Néron B, Denise R, Coluzzi C, Touchon M, Rocha EPC, Abby SS. MacSyFinder v2: improved modelling and search engine to identify molecular systems in genomes. Peer Community J. 2023;3:e28. https://doi.org/10.24072/pcjournal.250/.

45. Abby SS, Cury J, Guglielmini J, Néron B, Touchon M, Rocha EPC. Identification of protein secretion systems in bacterial genomes. Sci Rep. 2016;6(1):23080. https://www.nature.com/articles/srep23080.

46. Abby SS, Néron B, Ménager H, Touchon M, Rocha EPC, MacSyFinder. A program to mine genomes for molecular systems with an application to CRISPR-Cas systems. PLoS ONE. 2014;9(10):e110726. https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0110726.

47. Sullivan MJ, Petty NK, Beatson SA. Easyfig: a genome comparison visualizer. Bioinformatics. 2011;27(7):1009–10. https://academic.oup.com/bioinformatics/article/27/7/1009/230508.

48. Goris J, Konstantinidis KT, Klappenbach JA, Coenye T, Vandamme P, Tiedje JM. DNA–DNA hybridization values and their relationship to whole-genome sequence similarities. Int J Syst Evol Microbiol. 2007;57(1):81–91. https://doi.org/10.1099/ijs.0.64483-0.

49. Meier-Kolthoff JP, Hahnke RL, Petersen J, Scheuner C, Michael V, Fiebig A, et al. Complete genome sequence of DSM 30083T, the type strain (U5/41T) of *Escherichia coli*, and a proposal for delineating subspecies in microbial taxonomy. Stand Genomic Sci. 2014;9(1):2. https://doi.org/10.1186/1944-3277-9-2.

50. Shimodaira H, Hasegawa M. Multiple comparisons of log-likelihoods with applications to phylogenetic inference. Mol Biol Evol. 1999;16(8):1114–6. https://academic.oup.com/mbe/article-lookup/doi/10.1093/oxfordjournals.molbev.a026201.

51. Kislyuk AO, Haegeman B, Bergman NH, Weitz JS. Genomic fluidity: an integrative view of gene diversity within microbial populations. BMC Genomics. 2011;12(1):32. https://doi.org/10.1186/1471-2164-12-32.

52. Patz S, Gautam A, Becker M, Ruppel S, Rodríguez-Palenzuela P, Huson D. PLaBAse: a comprehensive web resource for analyzing the plant growth-promoting potential of plant-associated bacteria. bioRxiv. 2021. https://doi.org/10.1101/2021.12.13.472471.

53. Storm DR, Rosenthal KS, Swanson PE. Polymyxin and related peptide antibiotics. Annu Rev Biochem. 1977;46(1):723–63. https://doi.org/10.1146/annurev.bi.46.070177.003451.

54. Blin K, Shaw S, Augustijn HE, Reitz ZL, Biermann F, Alanjary M, et al. antiSMASH 7.0: new and improved predictions for detection, regulation, chemical structures and visualisation. Nucleic Acids Res. 2023;51(W1):W46–50. https://academic.oup.com/nar/article/51/W1/W46/7151336.

55. Aguilera M, Monteoliva-Sánchez M, Suárez A, Guerra V, Lizama C, Bennasar A, et al. *Paenibacillus jamilae* sp. nov., an exopolysaccharide-producing bacterium able to grow in olive-mill wastewater. Int J Syst Evol Microbiol. 2001;51(5):1687–92. https://doi.org/10.1099/00207713-51-5-1687.

56. Kwak MJ, Choi SB, Ha S, min, Kim EH, Kim BY, et al. Genome-based reclassification of *Paenibacillus jamilae* Aguilera 2001 as a later heterotypic synonym of *Paenibacillus polymyxa* (Prazmowski 1880) Ash 1994. Int J Syst Evol Microbiol. 2020;70(5):3134–8. https://www.microbiologyresearch.org/content/journal/ijsem/https://doi.org/10.1099/ijsem.0.004140.

57. Lebano I, Fracchetti F, Vigni ML, Mejia JF, Felis G, Lampis S. MALDI-TOF as a powerful tool for identifying and differentiating closely related microorganisms: the strange case of three reference strains of *Paenibacillus polymyxa*. Sci Rep. 2024;14(1):2585. https://www.nature.com/articles/s41598-023-50010-w.

58. Liu Z, Beskrovnaya P, Melnyk RA, Hossain SS, Khorasani S, O'Sullivan LR, et al. A genome-wide screen identifies genes in rhizosphere-associated *Pseudomonas* required to evade plant defenses. MBio. 2018;9(6). https://doi.org/10.1128/mbio.00433-18.

59. Wallner A, Busset N, Lachat J, Guigard L, King E, Rimbault I, et al. Differential genetic strategies of *Burkholderia vietnamiensis* and *Paraburkholderia kururiensis* for root colonization of *Oryza sativa* subsp. Japonica and *O. sativa* subsp. indica, as revealed by transposon mutagenesis sequencing. Appl Environ Microbiol. 2022;88(14). https://doi.org/10.1128/aem.00642-22.

60. Wallner A, Klonowska A, Guigard L, King E, Rimbault I, Ngonkeu E, et al. Comparative genomics and transcriptomic response to root exudates of six rice root-associated *Burkholderia sensu lato* species. Peer Community J. 2023;3:e25. https://doi.org/10.24072/pcjournal.252/.

61. Karatan E, Michael AJ. A wider role for polyamines in biofilm formation. Biotechnol Lett. 2013;35(11):1715–7. http://link.springer.com/10.1007/s10529-013-1286-3.

62. Zavascki AP, Goldani LZ, Li J, Nation RL. Polymyxin B for the treatment of multidrug-resistant pathogens: a critical review. J Antimicrob Chemother. 2007;60(6):1206–15. https://doi.org/10.1093/jac/dkm357.

63. Yu WB, Yin CY, Zhou Y, Ye BC. Prediction of the mechanism of action of fusaricidin on *Bacillus subtilis*. Herrera-Estrella A, editor. PLoS One. 2012;7(11):e50003. https://doi.org/10.1371/journal.pone.0050003.

64. Tobisch S, Stülke J, Hecker M. Regulation of the lic operon of *Bacillus subtilis* and characterization of potential phosphorylation sites of the LicR regulator protein by site-directed mutagenesis. J Bacteriol. 1999;181(16):4995–5003. https://doi.org/10.1128/JB.181.16.4995-5003.1999.

65. Hasan MK, Dhungel BA, Govind R. Characterization of an operon required for growth on cellobiose in *Clostridioides difficile*. Microbiology. 2021;167(8). https://doi.org/10.1099/mic.0.001079.

66. Saravanakumar K, Fan L, Fu K, Yu C, Wang M, Xia H, et al. Cellulase from *Trichoderma harzianum* interacts with roots and triggers induced systemic resistance to foliar disease in maize. Sci Rep. 2016;6(1):35543. https://www.nature.com/articles/srep35543.

67. Verma SC, Mahadevan S. The chbG gene of the chitobiose (chb) operon of *Escherichia coli* encodes a chitooligosaccharide deacetylase. J Bacteriol. 2012;194(18):4959–71. https://doi.org/10.1128/JB.00533-12.

68. Wiame E, Delpierre G, Collard F, Van Schaftingen E. Identification of a pathway for the utilization of the amadori product fructoselysine in *Escherichia coli*. J Biol Chem. 2002;277(45):42523–9. https://linkinghub.elsevier.com/retrieve/pii/S0021925819718353.

Publisher's Note