Mixta gen. nov. – A new genus in the *Erwiniaceae*

Marike Palmer [1, 2], Emma T. Steenkamp [1, 2], Martin P. A. Coetzee [2, 3], Juanita R. Avontuur [1,2], Wai-Yin Chan [1,2,4], Elritha van Zyl [1,2], Jochen Blom [5], Stephanus N. Venter *[1, 2]

1. *Department of Microbiology and Plant Pathology, University of Pretoria, Pretoria, South Africa*
2. *DST-NRF Centre of Excellence in Tree Health Biotechnology (CTHB) and Forestry and Agricultural Biotechnology Institute (FABI), University of Pretoria, Pretoria, South Africa*
3. *Department of Genetics, University of Pretoria, Pretoria, South Africa*
4. *Biotechnology Platform (BTP), Agricultural Research Council Onderstepoort Veterinary Institute (ARC-OVI), Onderstepoort, South Africa*
5. *Bioinformatics and Systems Biology, Justus-Liebig-University Giessen, Giessen, Germany*

*Corresponding author: S.N. Venter, fanus.venter@up.ac.za

Keywords: *Pantoea*; genome-inferred phenotype; phylogenomics; *Enterobacteriaceae*; *Enterobacterales*; MLSA

Subject Category: Taxonomic description

Word count: 4,826

Depositories:

The draft genome assembly for *Pantoea theicola* QC88-366 ^{T} is available from the National Centre for Biotechnology Information (NCBI; http://www.ncbi.nlm.nih.gov/) under the accession number NWUO00000000.

Abbreviations:

MLSA, multi-locus sequence analysis; KEGG, Kyoto Encyclopedia of Genes and Genomes; TSA, tryptone soy agar; CTAB, cetyl trimethylammonium bromide; AAI, average amino acid identity;

Abstract

The *Erwiniaceae* contain many species of agricultural and clinical importance. Although relationships among most of the genera in this family are relatively well resolved, the phylogenetic placement of several taxa remain ambiguous. In this study, we aimed to address these uncertainties by using a combination of phylogenetic and genomic approaches. Our MLSA and genome-based maximum likelihood phylogenies revealed that the arsenate reducing strain IMH and plant-associated strain ATCC 700886, both previously presumptively identified as members of *Pantoea*, represent new species of *Erwinia*. Our data also showed that the taxonomy of *E. teleogrylli* requires revision as it is clearly excluded from *Erwinia* and the other genera of the family. Most strikingly, however, five species of *Pantoea* formed a distinct clade within the *Erwiniaceae*, where it had a sister group relationship with the *Pantoea+Tatumella* clade. By making use of gene content comparisons, this new clade is further predicted to encode a range of characters that it shares with or distinguishes it from related genera. We thus propose recognition of this clade as a distinct genus and suggest the name *Mixta* in reference to the diverse habitats from which its species were obtained, including plants, humans and food products. Accordingly, a description for *Mixta* gen. nov. is provided to accommodate the four species *M. calida* comb. nov., *M. gaviniae* comb. nov., *M. intestinalis* comb. nov. and *M. theicola* comb. nov., with *M. calida* as type species for the genus.

Introduction

The recently described bacterial family, *Erwiniaceae*, includes organisms with very diverse lifestyles [1]. It currently contains the genera *Erwinia*, *Pantoea*, *Phaseolibacter*, *Tatumella*, *Buchnera* and *Wigglesworthia* [1]. Of these genera, the best known and most widely sequenced are *Erwinia*, *Pantoea* and to a lesser extent *Tatumella*. *Erwinia* was first described in 1920 [2-4] as a bacterial group containing plant pathogens, with *E. amylovora* as the type species [5]. Taxonomic revision of this group later resulted in the description of a number of additional species, nearly all of which are associated with plants in a pathogenic manner [6-11]. These subsequent studies also lead to the transfer of species to various genera within the *Enterobacterales* [1], particularly *Pantoea* that was described in 1989 with *P. agglomerans* as the type [12]. Since then additional *Pantoea* species have been described, with isolates ranging from fungal and insect associates [13-15], to plant [16, 17] and human pathogens [18, 19], bringing the total number of species with validly published names to 25 with a further three described species without validly published taxon names [13, 14, 20]. *Tatumella* was described in 1981 with *T. ptyseos*, a human pathogen, as type species [21]. This was followed by the transfer and description of five additional species isolated from soil, fruit and clinical samples [22, 23], resulting in a total of six species within this genus.

The relationships between genera in the *Erwiniaceae* have been contentious [1, 7, 16, 18, 22, 24-28], however, monophyletic groups for each genus can generally be recovered, especially *Pantoea, Erwinia* and *Tatumella* [16, 27-30]. There are, however, still a number of uncertainties regarding the taxonomic position of several taxa previously identified as members of these genera. This is evident from the fact that, despite some isolates being denoted as part of certain genera, later studies repeatedly contradicted these taxonomic designations. Examples of these include *Pantoea cedenensis* (taxon name has not yet been validly published) [8, 31, 32] and *Pantoea* sp. IMH [6, 28]. Additionally, some taxa are also phylogenetically independent from these genera, e.g., *P. calida* and *P. gaviniae* [28]. These previous studies thus suggest that one or more of these genera are polyphyletic as currently circumscribed.

A primary criterion for describing taxa at the genus and higher taxonomic levels is monophyly based on molecular-based phylogenetic studies [33-35]. Evaluation of the monophyletic condition is, however, dependent on the availability of a robust phylogenetic hypothesis for the taxa of interest. To generate such hypotheses, researchers typically endeavour to minimize stochastic and systematic error during phylogenetic analyses [36-38]. Stochastic error refers to the use of too little information for obtaining sufficient resolution, while systematic error refers to non-phylogenetic signal that causes tree reconstruction artefacts [36-39]. In general, the use of more characters (i.e., DNA bases or amino acid residues) lessens the effect of stochastic error [37, 38, 40]. Also, the use of appropriate ingroup and outgroup taxa decreases the effect of systematic error, because increased character sampling without appropriate taxon sampling severely impacts phylogeny reconstruction [37, 38, 40, 41].

In this study, we addressed some of the current uncertainties regarding the taxonomic position of various *Pantoea* strains and species. Our first aim was therefore to position these taxa within the phylogenetic hypothesis for *Pantoea*, *Erwinia* and *Tatumella*. To achieve this, we sequenced the genome of *Pantoea theicola* QC88-366 ^{T} to supplement the wide array of

Table 1. Genome sequences used for phylogenomic inference

Genus	Species	Strain *	Accession number	Reference
Erwinia	*Erwinia amylovora*	LA 636	CBVT00000000.1	[76]
	Erwinia billingiae	NCPPB 661 ^{T}	FP236843.1,FP236826.1, FP236830.1	[77]
	Erwinia mallotivora	BT-MARDI	JFHN00000000.1	[78]
	Erwinia pyrifoliae	DSM 12163 ^{T}	FN392235.1, FN392236.1, FN392237.1	[77]
	Erwinia tasmaniensis	Et 1-99 ^{T}	CU468135.1, CU468128.1, CU468130.1, CU468131.1, CU468132.1, CU468133.1	[79]
	Erwinia toletana	DAPP-PG 735	AOCZ00000000.1	[80]
	Erwinia tracheiphila	PSU-1	APJK00000000.1	-
	Pantoea sp.	IMH	JFGT00000000.1	[60]
	Erwinia sp.	9145	JQNE00000000.1	[81]
Mixta	*Pantoea alhagi*	LTYR-11Z ^{T}	CP019706.1, CP019707.1	-
	Mixta calida	LMG 25383 ^{T}	MLFO00000000.1	[28]
	Mixta gaviniae	LMG 25382 ^{T}	MLFQ00000000.1	[28]
	Mixta theicola	QC 88-366 ^{T}		This study
Pantoea	*Pantoea agglomerans*	R 190	JNGC00000000.1	[82]
	Pantoea allii	LMG 24248 ^{T}	MLFE00000000.1	[28]
	Pantoea ananatis	LMG 2665 ^{T}	JFZU00000000.1	[83]
	Pantoea anthophila	11-2	JXXL00000000.1	[84]
	Pantoea brenneri	LMG 5343 ^{T}	MIEI00000000.1	[28]
	Pantoea conspicua	LMG 24534 ^{T}	MLFN00000000.1	[28]
	Pantoea cypripedii	LMG 2657^{T}	MLJI00000000.1	[28]
	Pantoea deleyi	LMG 24200 ^{T}	MIPO00000000.1	[28]
	Pantoea dispersa	EGD-AAK13	AVSS00000000.1	-
	Pantoea eucalypti	aB	AEDL00000000.1	-
	Pantoea eucrina	LMG 2781 ^{T}	MIPP00000000.1	[28]
	Pantoea rodasii	LMG 26273 ^{T}	MLFP00000000.1	[28]
	Pantoea rwandensis	LMG 26275 ^{T}	MLFR00000000.1	[28]
	Pantoea septica	LMG 5345 ^{T}	MLJJ00000000.1	[28]
	Pantoea stewartii ssp. *stewartii*	DC 283	AHIE00000000.1	[85]
	Pantoea stewartii ssp. *indologenes*	LMG 2632 ^{T}	JPKO00000000.1	-

Pantoea	*Pantoea vagans*	C9-1	CP001894.1, CP001893.1, CP001894.1	[86]
	Pantoea wallisii	LMG 26277 ^{T}	MLFS00000000.1	[28]
	Pantoea sp.	At-9b	CP002433.1, CP002434.1, CP002435.1, CP002436.1, CP002437.1, CP002438.1	[87]
	Pantoea sp.	A4	ALXE00000000.1	[88]
	Pantoea sp.	GM01	AKUI00000000.1	[89]
Tatumella	*Tatumella morbirosei*	LMG 23360 ^{T}	CM003276.1	-
	Tatumella ptyseos	ATCC 33301 ^{T}	ATMJ00000000.1	-
	Tatumella saanichensis	NML 06-3099 ^{T}	ATMI00000000.1	[23]
Outgroup taxa	*Brenneria goodwinii*	OBR 1	CGIG00000000.1	-
	Cronobacter sakazakii	ATCC 29544 ^{T}	CP011047.1, CP011048.1, CP011049.1, CP011050.1	-
	Enterobacter cloacae spp. *cloacae*	ATCC 13047 ^{T}	CP001918.1, CP001919.1, CP001920.1	[90]
	Franconibacter helveticus	LMG 23732 ^{T}	AWFX00000000.1	[91]
	Klebsiella pneumoniae ssp. *pneumoniae*	ATCC 13883 ^{T}	JOOW00000000.1	[92]
	Kluyvera ascorbata	ATCC 33433 ^{T}	JMPL00000000.1	-
	Pectobacterium carotovorum ssp. *carotovorum*	NCPPB 312 ^{T}	JQHJ00000000.1	-
	Serratia marcescens ssp. marcescens	ATCC 13880 ^{T}	JMPQ00000000.1	[93]
	Yokenella regensburgei	ATCC 49455 ^{T}	JMPS00000000.1	-

***** Superscript ^{T} indicates type strains for the species

available genome data for these genera (see Table 1 for species and references). These genome data were then used to construct a maximum likelihood phylogeny based on the genes shared by the sequenced members of these genera and the outgroup. For comparative purposes, we also inferred the corresponding phylogenetic tree including almost all of the species in the respective genera by using the multi-locus sequence analysis (MLSA) protein sequences [26], routinely employed within the *Enterobacterales*. Our second aim was to determine the physiological pathways and processes likely characterising the uncertain taxa. This was achieved through core gene set comparisons of the respective taxa and the inference of functional categories by making use of Kyoto Encyclopedia of Genes and Genomes (KEGG) annotations. Finally, we provide descriptions for a proposed new genus (i.e., *Mixta* gen. nov.) and its species.

Materials and Methods

1. Genome sequencing of *P. theicola* QC88-366 ^{T}

The type strain of *P. theicola* [42] was grown on Tryptone Soy Agar (TSA) at 28 °C for 2 days. DNA extraction was performed using the CTAB method previously described by Steenkamp and colleagues [43]. The total genomic DNA was sent for sequencing on the Ion-Torrent™ PGM platform using an Ion PI™ chip, with 200bp chemistry, at the University of Pretoria Sequencing Facility. Raw genome data was trimmed and filtered with FastX Tools v. 0.0.13, followed by assembly with Spades v. 3.9.0 with an optimal k-mer of 127bp. The assembled genome was annotated using the RAST platform (available at http://rast.nmpdr.org/).

2. Average amino acid identity

As a measure of relatedness between isolates, average amino acid identity (AAI) values were determined between all genome sequences analysed in this study. To calculate AAIs the Efficient Database framework for comparative Genome Analyses using BLAST score Ratios (EDGAR) server [44] was used. This was done by determining all pairwise shared gene sets between genomes and calculating the mean amino acid similarity for homologous genes.

3. Phylogenetic analyses

Two rounds of phylogenetic analyses were performed. The first utilized the inferred protein sequences for the MLSA genes *atpD*, *gyrB*, *infB* and *rpoB*. The dataset for this analysis included the sequences for species (taxon names validly published or awaiting valid publication) of *Erwinia*, *Pantoea*, *Phaseolibacter* and *Tatumella*, as well as the outgroup taxa (Table 2). The respective nucleotide sequences were obtained from the database of the National Centre for Biotechnology Information (NCBI; accessed 20/5/2017; http://www.ncbi.nlm.nih.gov/). Where these sequences were not available in NCBI, the corresponding data were extracted from the genome sequences. Individual nucleotide sequence sets were translated to amino acids in BioEdit 7.0.9.0 [45] and aligned with MAFFT v. 7.310

Table 2. Additional species included in the multiple locus sequence analysis

Species *	***atpD***	***gyrB***	***infB***	***rpoB***
Erwinia aphidicola LMG 24877 ^{T}	FN547378.1	FN547377.1	FN547373.1	FN547374.1
Erwinia gerundensis EM 595 ^{T}	JN591389.1	FJ617413.1	JN591392.1	KP070836.1
Erwinia iniecta B 120 ^{T}	NZ_JRXE01000025.1	NZ_JRXE01000025.1	NZ_JRXE01000006.1	NZ_JRXE01000039.1
Erwinia oleae DAPP-PG 531 ^{T}	GU991653.1	GU991654.1	GU991655.1	GU991656.1
Erwinia papaya NCPPB 4294 ^{T}	HQ393588.1	HQ393600.1	HQ393612.1	HQ393624.1
Erwinia persicina NBRC 102418 ^{T}	HQ393598.1	HQ393610.1	HQ393622.1	HQ393634.1
Erwinia piriflorinigrans CFBP 5888 ^{T}	JF311469.1	JF311582.1	JF311695.1	JF311808.1
Erwinia psidii LMG 7039 ^{T}	HQ393594.1	HQ393606.1	HQ393618.1	HQ393630.1
Erwinia rhapontici LMG 2688 ^{T}	EF988751.1	EF988838.1	EF988924.1	EF989010.2
Erwinia teleogrylli SCU-B244 ^{T}	KM108628.1	KM108627.1	KM108629.1	KM108630.1
Erwinia typographi LMG 25347 ^{T}	HQ620539.1	KT073222.1	HQ620540.1	HQ620541.1
Erwinia uzenensis LMG 25843 ^{T}	KT073225.1	AB546201.1	KT073224.1	KT073223.1
Mixta intestinalis 29Y89B ^{T}	KP420529.1	KP420531.1	KP420525.1	KP420527.1
Pantoea beijingensis JZB 2120010 ^{T}	KC969201.1	KC969202.1	KC969203.1	KC969204.1
Pantoea cedenensis ATCC 700886 ^{T}	JF810172.1	-	JF810173.1	JF810174.1
Pantoea coffeiphila Ca04 ^{T}	KM205369.1	KM205365.1	KM205361.1	KJ427825.1
Pantoea hericii JZB 2120024 ^{T}	KU189728.1	KU189731.1	KU189734.1	KU189737.1
Pantoea pleuroti JZB 2120015 ^{T}	KJ654342.1	KJ654343.1	KJ654344.1	KJ654345.1
Phaseolibacter flectens ATCC 12775 ^{T}	JN808190.1	JF745803.1	-	JF745804.1
Tatumella citrea LMG 22049 ^{T}	EF988715.1	EF988802.1	EF988888.1	EF988974.1
Tatumella punctata LMG 22050 ^{T}	EF988716.1	EF988803.1	EF988889.1	EF988975.1
Tatumella terrea LMG 22051 ^{T}	EF988717.1	EF988804.1	EF988890.1	EF988976.1

***** Superscript ^{T} indicates type strains for the species

[46]. The aligned files were concatenated and partitioned using FASconCAT-G v. 1.02 [47], with the appropriate amino acid model of evolution for each partition determined by ProtTest 3.4 [48]. This dataset was then subjected to maximum likelihood analysis with RAxML v. 8.2.1 [49] where branch support was inferred from 1,000 bootstrap replicates.

The second phylogenetic analysis utilized the protein sequences for all of the genes that are shared by representatives of the ingroup and outgroup taxa, included in this study (Table 1). For this purpose the Efficient Database framework for comparative Genome Analyses using BLAST score Ratios (EDGAR) server [44] was used. Individual protein sequence sets were batch-aligned with MUSCLE [50] in CLC Main Workbench 7.6 (CLC Bio). These alignments were then concatenated, partitioned with appropriate substitution model parameters, and subjected to maximum likelihood phylogenetic analysis as described above.

4. Genome-inferred traits

EDGAR was used to determine the shared gene content in each of *Pantoea*, *Erwinia*, *Tatumella* and *Mixta* gen. nov. (see below). Each gene set was then subjected to functional annotation with the KEGG database (5/6/2017; [51]) using BlastKOALA [52]. An overview metabolic map for each genus was investigated to identify reactions conserved in each of the respective genera. From these metabolic reactions, only those conserved in a single genus were further investigated. The genes encoding the enzymes involved in these reactions were then compared to the original genomes of the remaining three genera to confirm the presence or absence of these genes, using BLAST analyses [53].

The full gene set consisting of all genes present in the genomes of taxa investigated in this study were used to identify genes unique to and shared by all members of the novel genus (see below). All genes occurring in the genomes of at least one taxon outside of this monophyletic group were excluded from further analyses. Of these uniquely shared genes, those with no KEGG annotations were subjected to Blast2GO v. 4.1 [54] analyses to identify potential functional associations for these genes.

Results

1. Genome sequence of *P. theicola* QC88-366 ^{T}

Sequencing of the *P. theicola* genomic DNA using the Ion-Torrent platform yielded a total of 16,449,283 single-ended reads. After trimming and quality filtering a total of 15,973,352 reads could be assembled into a draft genome consisting of 4,291,230 bp. The assembly consisted of 72 scaffolds with a minimum coverage of 699x and an N50 of 278,310 nucleotides. The draft genome assembly is available from the National Centre for Biotechnology Information (NCBI; http://www.ncbi.nlm.nih.gov/) under the accession number NWUO00000000.

2. Average amino acid identity

From AAI calculations, it could be observed that members of the respective genera were overall more closely related to each other than to members of different genera (Table S1). Interspecies AAI values for each of the genera ranged from 79.06% (*E. toletana* vs *E. tracheiphila*) to 95.36% (*P. agglomerans* vs *P. vagans*). As was seen from previous average nucleotide identity analyses [28], intergeneric comparisons between *Tatumella* and its closest relatives were noticeably lower (71.9% - 75.19%) than equivalent intergeneric comparisons involving the other three genera. For *Erwinia*, *Mixta* and *Pantoea*, intra- and intergeneric AAI values were comparable.

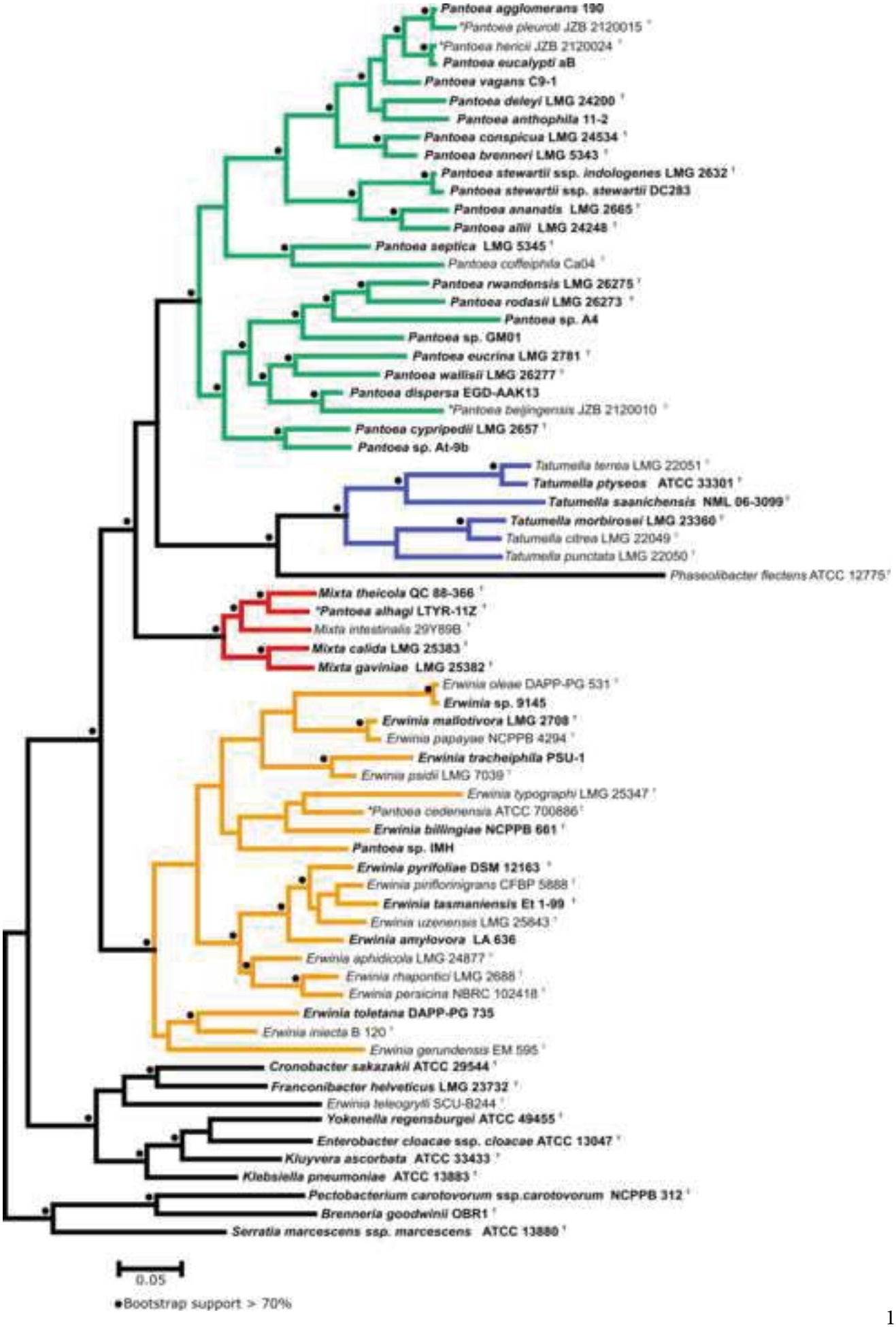
3. Phylogenetic analyses

Taxon selection was based on the availability of gene and whole genome sequences for the ingroup and outgroup taxa. All members of the *Erwiniaceae* with sequences available for at least three of the four genes, *atpD*, *gyrB*, *infB* and *rpoB*, were included for MLSA. Representatives of the phylogenetic diversity of *Erwinia*, *Pantoea* and *Tatumella* with available genome sequences were included in genome-based phylogenetic analyses. Outgroup taxa were selected from the *Enterobacterales*, in the families *Enterobacteriaceae*, *Pectobacteriaceae* and *Yersiniaceae* [1]. From these families, type strains of type species of nine genera with available MLSA and genome sequence data were selected for outgroup purposes. To avoid potential tree building artefacts like long-branch attraction, we did not include isolates of *Buchnera* and *Wigglesworthia*, as well as the recently described Candidatus *Pantoea carbekii*, in any of the analyses, as homologues of only a single gene for some taxa was present. These taxa are rapidly evolving obligate insect endosymbionts that have undergone extensive genome streamlining through gene loss yielding very small genomes [55-59], and are thus not suitable for comparison based on genome content.

The concatenated MLSA dataset consisted of 68 taxa and 3,552 amino acid residues (see Table S2). Maximum likelihood phylogenetic analysis of the dataset separated the ingroup taxa into four clades receiving 70-100% bootstrap support (Fig. 1). The first three clades broadly corresponded to *Pantoea* (with 25 taxa), *Erwinia* (with 20 taxa) and *Tatumella+Phaseolibacter* (6 *Tatumella* species formed a clade with *Phaseolibacter flectens* as its sister taxon). The main exceptions were *Pantoea* strains IMH [60] and *Pantoea cedenensis* ATCC 700886 [8], as well as the recently described *Erwinia teleogrylli* SCU-B244 ^{T}[61]. Strains IMH and ATCC 700886 were nested within the *Erwinia* clade, while *E. teleogrylli* was nested among the outgroup taxa. The fourth clade received 100% bootstrap support and had a sister group relationship with the overall *Pantoea+Tatumella+Phaseolibacter* clade. This fourth clade included five isolates previously described as members of *Pantoea*, i.e., *P. alhagi* [20], *P. calida* [62], *P. gaviniae* [62], *P. intestinalis* [19] and *P. theicola* [42], and corresponds to *Mixta* gen. nov., which we propose in this study.

The phylogenomic dataset was constructed from the genes shared among the genomes of the 46 taxa (37 ingroup and 9 outgroup) included in this study. It consisted of 350,738 amino acids spread across 1,044 proteins (see Table S2). Maximum likelihood phylogenetic analysis of this dataset generated a tree in which most nodes received 100% bootstrap support (Fig. S1).

Figure 1. A maximum-likelihood phylogeny for members of *Mixta* gen. nov., *Erwinia, Pantoea, Tatumella* and *Phaseolibacter*, with appropriate outgroup taxa, constructed from the protein sequences of the genes *atpD, gyrB, infB* and *rpoB*. Full length sequences for isolates with available genomes (indicated in bold), together with partial sequences of the type strains of other species, were used to construct a partitioned, concatenated data matrix, with the appropriate amino acid model of evolution applied to each partition. Branch support was inferred from 1,000 bootstrap replicates, with support values >70% indicated with dots. Asterisks indicate species that have not previously been validated. Colours indicate the genera *Pantoea (*green), *Tatumella* (blue), *Mixta* gen. nov. (red) and *Erwinia* (yellow).



The only exceptions were three branches within the *Pantoea* clade. Also, it differed somewhat from the MLSA tree in terms of certain relationships within the *Pantoea* clade (i.e., the positions of *P. eucalypti*, *P. anthophila* and *Pantoea* sp. A4) and within the *Erwinia* clade (i.e., the positions of *E. mallotivora*, strain IMH and *E. tasmaniensis*). Overall, however, the genome-based tree supported the same phylogenetic hypothesis as the MLSA tree as it also separated the ingroup taxa into four clades corresponding to the three existing genera and one new genus. The latter included only the four taxa (*P. alhagi*, *P. calida*, *P. gaviniae*, and *P. theicola*) for which genome sequences are available. This tree did not include *P. cedenesis* strain ATCC 700886 and *E. teleogrylli*, while *Pantoea* sp. strain IMH was placed in the *Erwinia* clade.

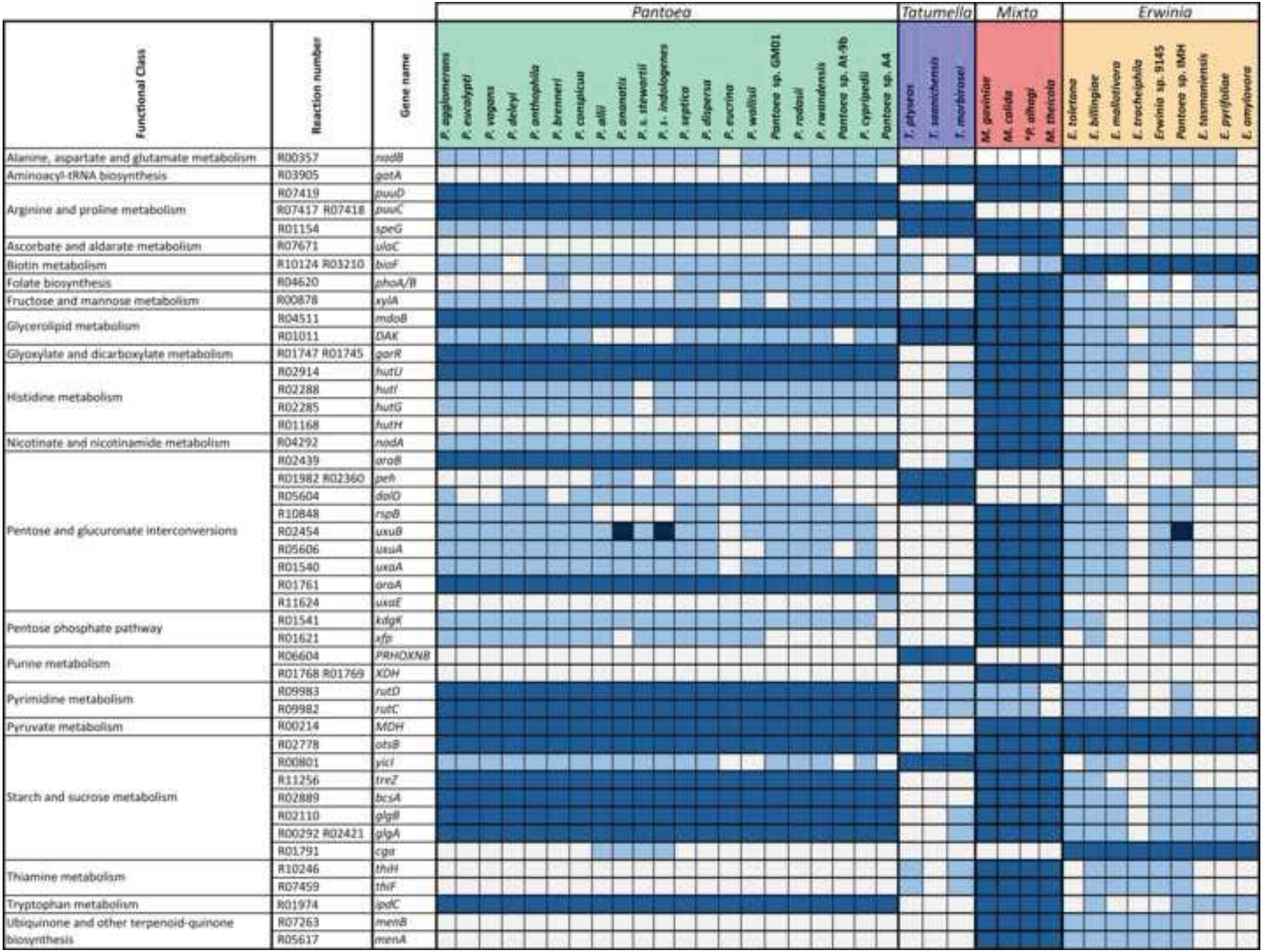
4. Genome-inferred traits

To identify and compare genome-inferred phenotypic traits among the respective genera, the core gene sets (i.e., genes shared by all members of a genus) were identified before annotation. Based on these analyses, the *Pantoea* core contained 1,861 genes, the *Erwinia* core contained 1,644 genes, and the *Tatumella* core had 2,195 genes, while the *Mixta* core contained 2,628 genes (Fig. S2). For these core gene sets, *ca*. 80% (ranging from 78.7% for *Mixta* to 87.5% for *Erwinia*) could be functionally annotated using KEGG (Fig. S2). In all four core gene sets, the highest number of genes were involved in 'Genetic Information Processing', followed by 'Environmental Information Processing', with the third highest number of genes being unclassified (Fig. S2).

Comparison of the functional annotations revealed a number of characteristics common to all members a specific genus (Fig. 2 and Table S3). Three of these characteristics were only present in a single genus. These included the interconversion of L-histidine to urocanate (a potential photoprotectant) and ammonia (encoded for by *hutH*) and hypoxanthine to xanthine (purines; catalysed by XDH, EC 1.17.1.4) present in all members of *Mixta*, and the conversion of 5-hydroxy-2-oxo-4-ureido-2,5-dihydro-1H-imidazole-5-carboxylate to S-allantoin and CO_2 (catalysed by PRHOXNB) solely present in members of *Tatumella*. Two reactions were also present in all members of some genera and absent from others, like the conversion of S-malate to pyruvate (R00214; catalysed by MDH; present in *Pantoea*, *Mixta* and *Erwinia*) and gamma-Glutamyl-gamma-aminobutyraldehyde to 4-(L-gamma-Glutamylamino)-butanoate (R07417 and R07418; encoded by *puuC*; present in *Pantoea* and *Tatumella*).

A number of reactions were present in members of all four genera, although these functions were often present only in some members of the respective genera (Fig. 2). Examples of these reactions are those catalysed by the gene products of *puuD* (present in all members of *Pantoea* and *Mixta* and some members of *Erwinia*), *gatA* (present in all members of *Tatumella* and some members of *Pantoea*), *rspB* (present in *Mixta* and some members of *Pantoea* and *Erwinia*), *speG* (present in *Tatumella* and *Mixta* and members of *Pantoea* and *Erwinia*) and *bioF* (present in *Erwinia* and members of *Mixta*, *Pantoea* and *Tatumella*) (Fig. 2). The majority of the genes present in some members of all genera were involved in 'Pentose and glucuronate

Figure 2. Genes that were differentially present in the genomes of members of the genera *Pantoea, Tatumella, Mixta* gen. nov. and E*rwinia.* The ordering of taxa and coloured sections correspond to the phylogeny constructed from the MLSA (Fig. 1). Asterisks indicate species that have not previously been validated. Genes are ordered based on their functional class. Blue blocks indicate the presence of genes, while white blocks indicate the absence of genes in some species of the genus, darker blue blocks indicate the presence of genes throughout a genus, while navy blue blocks indicate multiple copies in the genome. Reaction numbers correspond to reactions in the KEGG database.



interconversions', followed by 'Starch and sucrose metabolism', with the third highest number of genes involved in 'Histidine metabolism' (Fig. 2).

We also attempted to infer possible functions for the fraction of *Mixta*-specific genes that could not be functionally annotated using KEGG. Among the 21.3% of *Mixta* genes (559 of 2,628) lacking homologs in the KEGG database, only 46 were uniquely present in the members of *Mixta* (Table S4). Based on the Blast2GO analyses conducted on these genes, 26 genes could be annotated, while BLAST hits were obtained for 19 genes without annotation, with one gene having no similarity to genes on the non-redundant database. By examining the Gene Ontologies (GO), the highest number of genes (12) could be annotated as 'integral component of membrane' under cellular components, while three genes each were involved in 'oxidation-reduction processes' and 'transcription regulation' under biological processes (results not shown).

Discussion

This study provided insight into the phylogenetic placement of a number of contentious taxa within the assemblage containing *Erwinia*, *Pantoea*, *Tatumella* and *Phaseolibacter*. Based on the robust evolutionary hypothesis obtained for this group of taxa, both *Erwinia* and *Pantoea* is polyphyletic as it is currently circumscribed. Our findings indicated that *E. teleogrylli* is not related to any of the four genera and that it requires taxonomic revision. More importantly, however, *P. calida*, *P. gaviniae*, *P. intestinalis* and *P. theicola*, together with *P. alhagi* (taxon name awaiting valid publication), form a monophyletic and exclusive cluster within the *Erwiniaceae*. Our phylogenies clearly showed that it represents a unique clade with a sister group relationship to the *Pantoea+Tatumella+Phaseolibacter* clade.

Two approaches are available for achieving the monophyly of *Pantoea*. The first is to amalgamate *Pantoea, Tatumella, Phaseolibacter* and their sister clade (containing the five *Pantoea* species) into a single genus to accommodate all current taxon names with standing in nomenclature of *Pantoea*, *Tatumella* and *Phaseolibacter* species. However, apart from likely being associated with multiple species name changes, such a procedure would be incompatible with the known biological separation of these genera. The second approach is to recognise all of the monophyletic assemblages as distinct. Indeed, *Pantoea* has repeatedly been shown to be definitively distinct from *Tatumella* based on physiological and genomic properties [22, 28].

The phylogenetic hypothesis presented in this study was inferred with a methodology that limited the possible impact of stochastic and systematic errors [36, 38]. In fact, the different, and often incongruent, MLSA phylogenies reported for *Erwinia*, *Pantoea* and *Tatumella* [6, 7, 19, 42, 62], are indicative of a general failure to appropriately account for such errors during the tree building process. For *Erwinia* and *Pantoea* this is especially true regarding (i) the types of characters used, (ii) the tree building methods and (iii) outgroup selection. In terms of the types of characters used, the description of *E. gerundensis* [6] provide an example of where trees inferred from nucleotide characters and amino acid residues differed markedly. This is, however, not uncommon as the higher levels of substitution saturation associated with third codon bases in nucleotide datasets often detracts from the robustness of nucleotide-based

phylogenies [28, 36, 38]. In terms of tree building methods, most of the disparities between published phylogenies for the five taxa included in *Mixta* are due to the use of distance-based methods such as Neighbour-Joining (NJ) [19, 20, 42, 62]. For the descriptions of these species, NJ-based MLSA's were done in which they grouped closely with *P. septica*, nested within the bigger *Pantoea* clade. Although NJ phylogenies may be accurate [63], they are exceptionally sensitive to long-branch attraction, particularly when only a limited number of taxa are included [64], which explains the spurious positions of the *Mixta* species in these previous trees.

The third prominent reason why MLSA trees for *Erwinia*, *Pantoea* and *Tatumella* often differ pertains to the taxa used for outgroup purposes. In phylogenetic analyses, the outgroup is the reference taxa with which a tree is rooted [38, 65], and their selection can have profound effects on the ultimate evolutionary hypothesis [37-40]. A good example of where outgroup selection influenced the tree is in the description of *E. teleogrylli* [61]. In their analysis, the authors [61] rooted their collection of ingroup taxa (consisting of *Erwinia*, *Pantoea* and *Tatumella* species) with *Escherichia coli*, which essentially forced their new taxon to erroneously associate with *Erwinia* in the ingroup. Upon inclusion of a wide selection of outgroup taxa from the *Enterobacterales*, we showed that *E. teleogrylli* is likely not a member of the *Erwinia+Pantoea+Tatumella+Phaseolibacter* assemblage. Future research should thus investigate the taxonomy of *E. teleogrylli* and also investigate the monophyly of the broader *Erwiniaceae* by making use of suitable outgroup taxa.

The generic name *Mixta* is proposed for the novel genus, due to the mixture of different lifestyles of species in the genus. Four species with validated taxon names will be transferred to the novel genus, with the amendment of *P. alhagi* to *P. alhagiae* upon valid publication of the taxon name and subsequent transfer of the species to *Mixta.* Two species, *M. calida* and *M. gaviniae* were originally isolated from infant formula and its production environment [62] and have since been isolated from other environments [28, 66]. Additionally, *P. alhagi* and *M. theicola* was originally isolated from plants, where *P. alhagi* was associated with plant growth promotion and drought tolerance [20] and *M. theicola* was isolated as a potential endophyte of black tea [42]. Furthermore, *M. intestinalis* was isolated from the faecal samples of healthy humans [19]. Based on initial studies, the pathogenic abilities of these species are unclear, although *M. calida* has been associated with post-operative meningitis before [66].

Two additional strains denoted as *Pantoea*, *Pantoea* sp. IMH and *P. cedenensis*, are clearly members of *Erwinia*, as have been noted before [6, 8, 28, 31, 32]. Although *P. cedenensis* has been deposited in a number of culture collections, it has not yet been formally described or validly published and can thus not be transferred to *Erwinia*. Initially, this organism was isolated from olive knots in association with *Pseudomonas savastanoi* from Spain, although various isolates have since been found in association with bark beetles [67, 68]. Similarly, *Pantoea* sp. IMH has not been formally described as a species and can thus not be transferred to *Erwinia*. This strain has been isolated from soil contaminated with arsenic from China and show potential for bioremediation purposes of this compound [60, 69].

As the genomes of *P. alhagi* , *M. calida* and *M. gaviniae* [28] were available, we sequenced genomic DNA for *M. theicola*. Overall, the genome of *M. theicola* is comparable to the other

members of the genus in terms of size and number of genes. The size of the genome for *M. theicola* was 4.29 Mb, while the other genomes in the genus ranged from 4.23 Mb for *M. calida* to 4.44 Mb for *M. gaviniae*. In total 4,217 genes were predicted for *M. theicola*, with the smallest of the additional genomes (*M. calida*) having 4,088 genes predicted and the largest (*M. gaviniae*) having 4,589 genes predicted. Further assembly statistics were also comparable to the genomes of *P. alhagi*, *M. calida* and *M. gaviniae* [28]. Due to regulations, *M. intestinalis* as a potential human pathogen, could unfortunately not be imported for genome sequencing. The availability of these genome sequences allows more robust inferences regarding phylogenetic relationships and provides us with the opportunity to investigate genome evolution in these taxa and infer biological traits from genome data.

Characteristics inferred at genome level provided more comprehensive data on the uniqueness of *Mixta* because the currently available database of phenotypic characters for *Pantoea* and its relatives are incomplete (see Table S5). Based on literature, no unique typically examined physiological traits were observed for *Mixta* [13, 14, 16-20, 22, 23, 27, 42, 61, 62, 70-72]. Similarly, various genome-inferred traits were displayed by all members of the novel genus, while it was displayed only by some species in the closely related genera, like aesculin hydrolysis and the utilisation of D-maltose. There were, however, a number of differences at genome level between *Mixta* and the closely related genera, like the interconversion of L-histidine to urocanate and ammonia and the catalysis of hypoxanthine to xanthine. Overall, the majority of genes annotated were involved in various genetic and environmental information processing, as can be expected for the core genomes of genera [73-75]. Various annotated genes were also present in all members of *Mixta*, with some genes (including hypothetical genes) occurring solely in members of *Mixta*. Although many of these genes are not yet characterised, these gene and protein sequences may be used as delineation tools in conjunction with phylogenetic approaches due to their exclusivity to *Mixta* amongst members of this assemblage.

The increased availability of whole genome sequence data presents the opportunity to perform phylogenomic analyses providing highly supported, robust phylogenies to form the base of taxonomic decisions. In addition, the genomes of novel taxa can be interrogated for unique characteristics, which are often absent following routine phenotypic tests. In this study a highly supported, robust phylogenetic hypothesis was obtained from genome data, and was overall congruent with the MLSA based on protein sequences. The cohesive monophyletic group in question was also supported as separate from the closely related genera *Erwinia*, *Pantoea* and *Tatumella*, based on characteristics inferred from the genome. We propose the generic name *Mixta* gen. nov. for this group of four species with validly published taxon names (with the new combinations: *Mixta calida* comb. nov., *Mixta gaviniae* comb. nov., *Mixta intestinalis* comb. nov and *Mixta theicola* comb. nov.), with the type species *Mixta calida*.

Descriptions

Description of *Mixta* gen. nov.

Mixta [Mix'ta N.L. fem. n. *Mixta* the mixed one, referring to mixed lifestyles of species in genus].

Phenotypic descriptions are inferred from literature (Table S5). For members of this genus, cells are facultative anaerobic, gram-negative rods and all members of the genus are motile. Typically the members of this genus are not producers of the yellow pigment characteristic of *Pantoea* and can grow at 44°C. All members are oxidase-negative and catalase-positive. Most members are Voges-Proskauer positive (production of acetoin) and hydrolyse aesculin but negative for the production of arginine dihydrolase, ornithine and lysine decarboxylases, indole and urease. All isolates were negative for the production of H_2S and positive for the presence of β-galactosidase. All members utilised the following carbon sources after 48h: D-glucose, L-arabinose, D-ribose, D-xylose, D-fructose, D-mannose, L-rhamnose, D-cellobiose, D-galactose, D-maltose, melibiose, trehalose, glycerol, inositol, D-mannitol, N-acetyl-D-glucosamine, gentiobiose, arbutin and aesculin. Variable results were obtained for lactose, sucrose, D-sorbitol, raffinose and salicin. The following carbon sources are not utilised by strains tested previously: L-sorbose, melezitose, erythritol, adonitol, dulcitol, glycogen, xylitol, starch, inulin, L-xylose, turanose, lyxose, tagatose, L-arabitol, L-fucose and D-malate. As inferred from genome comparisons, all members of the genus possessed the following genes: *araA*, *araB*, *bcsA*, DAK, *garR*, *glgA*, *glgB*, *hutG*, *hutH*, *hutI*, *hutU*, *ipdC*, *kdgK*, MDH, *menA*, *menB*, *mdoB*, *nadA*, *nadB*, *otsB*, *phoA/B*, *puuD*, *rspB*, *speG*, *thiF*, *thiH*, *treZ*, *ulaC*, *uxaA*, *uxaE*, *uxuA*, *uxuB*, XDH, *xfp*, *xylA* and *yicI*. Additionally, 46 hypothetical genes were present in all members of the genus. The type species for the genus is *Mixta calida*, isolated from baby formula powder [62].

Description of *Mixta calida* comb. nov.

Mixta calida (ca'li.da. L. fem. adj. *calida* warm, hot, reflecting its ability to grow at 44°C).

Basonym: *Pantoea calida* Popp et al. 2010.
This species displayed the following characteristics in addition to those in the genus description [62]. After 48h, the following compounds are utilised as carbon source: D-sorbitol, raffinose, sucrose, maltotriose, salicin, lactose, lactulose, D-saccharate, mucate, L-malate, cis-aconitate, trans-aconitate, citrate, D-glucuronate, D-galacturonate, 2-keto-D-gluconate, 5-keto-D-gluconate, DL-lactate, succinate, fumarate, DL-glycerate, D-glucosamine, L-aspartate, L-glutamate, L-proline, D-alanine, L-alanine, L-serine and α-ketoglutarate. The following compounds in addition to those in the genus description are not utilised after 48h: palatinose, D-arabitol, maltitol, hydroxyquinoline-β-glucuronide, 1-0-methyl-α-D-glucopyranoside, 3-0-methyl-D-glucopyranoside, L-tartrate, D-tartrate, meso-tartrate, D-malate, tricarballylate, L-tryptophan, phenylacetate, protocatechuate, p-hydroxybenzoate, quinate, gentisate, m-hydroxybenzoate, 3-phenylpropionate, m-coumarate, trigonelline, betaine, putrescine, 4-aminobutyrate, histamine, caprate, caprylate, L-histidine, glutarate, 5-aminovalerate, ethanolamine, tryptamine, itaconate, 3-hydroxybutyrate, malonate, propionate and L-tyrosine.

The type strain for the species is LMG 25383 ^{T} (= 1400/07 ^{T} = DSM 22759 ^{T}).

Description of *Mixta gaviniae* comb. nov.

Mixta gaviniae (ga.vi'ni.ae. N.L. fem. gen. n. *gaviniae* of Gavini, in honour of Françoise Gavini, a French microbiologist who first described the genus *Pantoea*).

Basonym: *Pantoea gaviniae* Popp et al. 2010.
As inferred from literature [62], isolates of this species does not grow at 44°C. In addition to characteristics in the genus description, the following carbon sources are utilised after 4 days: salicin, lactose, sucrose, raffinose, lactulose, maltotriose, D-sorbitol, myo-inositol, D-saccharate, mucate, L-malate, cis-aconitate, trans-aconitate, citrate, D-glucuronate, D-galacturonate, 2-keto-D-gluconate, 5-keto-D-gluconate, D-gluconate, DL-lactate, succinate, fumarate, DL-glycerate, D-glucosamine, L-aspartate, L-glutamate, L-proline, D- and L-alanine, L-serine and α-ketoglutarate. The following carbon sources, in addition to those in the genus description, are not utilised after 4 days: palatinose, D-arabitol, maltitol, hydroxyquinoline-β-glucuronide, 1-0-methyl-α-D-glucopyranoside, 3-0-methyl-D-glucopyranoside, L-tartrate, D-tartrate, meso-tartrate, D-malate, tricarballylate, L-tryptophan, phenylacetate, protocatechuate, p-hydroxybenzoate, quinate, gentisate, m-hydroxybenzoate, 3-phenylpropionate, m-coumarate, trigonelline, betaine, putrescine, 4-aminobutyrate, histamine, caprate, caprylate, L-histidine, glutarate, 5-aminovalerate, ethanolamine, tryptamine, itaconate, 3-hydroxybutyrate, malonate, propionate and L-tyrosine.

The type strain for the species is LMG 25382 ^{T} (= A18/07 ^{T} = DSM 22758 ^{T}).

Description of *Mixta intestinalis* comb. nov.

Mixta intestinalis (in.tes.ti.na'lis. N.L. fem. adj. *intestinalis* pertaining to the intestines, from where the type strain was isolated).

Basonym: *Pantoea intestinalis* Prakash et al. 2015.
In addition to the characteristics in the genus description, the following characteristics were inferred from literature [19]. No growth occurs on trisodium citrate, capric acid, adipic acid and phenylacetic acid.

The type strain for the species is 29Y89B ^{T} (= DSM 28113 ^{T} = MCC 2554 ^{T}).

Description of *Mixta theicola* comb. nov.

Mixta theicola (the.i'co.la. N.L. n. *thea* tea; L. suff. *-cola* (from L. n. *incola*) inhabitant, dweller; N.L. *theicola* inhabitant of tea).

Basonym: *Pantoea theicola* Tanaka et al. 2015.
The follow characteristics were observed for the species in addition to the characteristics in the genus description [42]. Growth does not occur above 40°C. Isolates of this species are Voges-

Proskauer (acetoin production) negative. Amygdalin, citrate, salicin, gentiobiose and D-fucose are utilised as carbon source, but D-sorbitol, lactose, sucrose and raffinose are not utilised.

The type strain for the species is QC88-366 ^{T} (= DSM 29212 ^{T} = NBRC 110557 ^{T}).

Funding information

We want to acknowledge the DST-NRF Centre of Excellence in Tree Health Biotechnology (CTHB) at the University of Pretoria for student funding.

Acknowledgements

The authors would like to acknowledge the DST-NRF Centre of Excellence in Tree Health Biotechnology and the South African Research Chairs Initiative (SARChI) Chair in Fungal Genomics, University of Pretoria, for the use of informatics infrastructure and server access. For genome sequencing, we would like to acknowledge the Ion Torrent Sequencing Facility at the University of Pretoria. We would also like to thank Doctor Stefano Ventura and Professor Aharon Oren for nomenclatural suggestions.

Conflicts of interest

The authors declare no conflict of interest.

Ethical statement

No ethical clearance was required.

References

1. **Adeolu M, Alnajar S, Naushad S, Gupta RS**. Genome-based phylogeny and taxonomy of the '*Enterobacteriales*': proposal for *Enterobacterales* ord. nov. divided into the families *Enterobacteriaceae*, *Erwiniaceae* fam. nov., *Pectobacteriaceae* fam. nov., *Yersiniaceae* fam. nov., *Hafniaceae* fam. nov., *Morganellaceae* fam. nov., and *Budviciaceae* fam. nov. *Int J Syst Evol Microbiol* 2016;66(12):5575-5599.
2. **Winslow CEA, Broadhurst J, Buchanan R, Krumwiede C, Rogers L et al.** The families and genera of the bacteria. Final report of the committee of the Society of American Bacteriologists on characterization and classification of bacterial types. *J Bacteriol* 1920;5:191-229.
3. **Hauben L, Moore ER, Vauterin L, Steenackers M, Mergaert J et al.** Phylogenetic position of phytopathogens within the *Enterobacteriaceae*. *Syst Appl Microbiol* 1998;21(3):384-397.
4. **Skerman VBD, McGowan V, Sneath PHA**. Approved lists of bacterial names. *International Journal of Systematic and Evolutionary Microbiology* 1980;30(1):225-420.
5. **Burrill TJ**. New species of *Micrococcus* (bacteria). *Am Nat* 1883;17:319-320.
6. **Rezzonico F, Smits TH, Born Y, Blom J, Frey JE et al.** *Erwinia gerundensis* sp. nov., a cosmopolitan epiphyte originally isolated from pome fruit trees. *Int J Syst Evol Microbiol* 2016;66(3):1583-1592.
7. **Smits THM, Rezzonico F, López MM, Blom J, Goesmann A et al.** Phylogenetic position and virulence apparatus of the pear flower necrosis pathogen *Erwinia piriflorinigrans* CFBP 5888T as assessed by comparative genomics. *Syst Appl Microbiol* 2013;36(7):449-456.
8. **Gardan L, Christen R, Achouak W, Prior P**. *Erwinia papayae* sp. nov., a pathogen of papaya (*Carica papaya*). *Int J Syst Evol Microbiol* 2004;54(1):107-113.
9. **Mergaert J, Hauben L, Cnockaert MC, Swings J**. Reclassification of non-pigmented *Erwinia herbicola* strains from trees as *Erwinia billingiae* sp. nov. *Int J Syst Evol Microbiol* 1999;49(2):377-383.
10. **Kim W-S, Gardan L, Rhim S-L, Geider K**. *Erwinia pyrifoliae* sp. nov., a novel pathogen that affects Asian pear trees (*Pyrus pyrifolia* Nakai). *Int J Syst Evol Microbiol* 1999;49(2):899-906.
11. **Goto M**. *Erwinia mallotivora* sp. nov., the causal organism of bacterial leaf spot of *Mallotus japonicus* Muell. Arg. *Int J Syst Evol Microbiol* 1976;26(4):467-473.
12. **Gavini F, Mergaert J, Beji A, Mielcarek C, Izard D et al.** Transfer of *Enterobacter agglomerans* (Beijerinck 1888) Ewing and Fife 1972 to *Pantoea* gen. nov. as *Pantoea agglomerans* comb. nov. and description of *Pantoea dispersa* sp. nov. *Int J Syst Bacteriol* 1989;39.
13. **Ma Y, Yin Y, Rong C, Chen S, Liu Y et al.** *Pantoea pleuroti* sp. nov., isolated from the fruiting bodies of *Pleurotus eryngii*. *Curr Microbiol* 2016;72(2):207-212.
14. **Rong C, Ma Y, Wang S, Liu Y, Chen S et al.** *Pantoea hericii* sp. nov., isolated from the fruiting bodies of *Hericium erinaceus*. *Curr Microbiol* 2016;72(6):738-743.
15. **Palmer M, de Maayer P, Poulsen M, Steenkamp ET, van Zyl E et al.** Draft genome sequences of *Pantoea agglomerans* and *Pantoea vagans* isolates associated with termites. *Stand Genomic Sci*, journal article 2016;11(1):1-11.
16. **Brady CL, Cleenwerck I, van der Westhuizen L, Venter SN, Coutinho TA et al.** *Pantoea rodasii* sp. nov., *Pantoea rwandensis* sp. nov. and *Pantoea wallisii* sp. nov., isolated from *Eucalyptus*. *Int J Syst Evol Microbiol* 2012;62(7):1457-1464.
17. **Brady CL, Goszczynska T, Venter SN, Cleenwerck I, De Vos P et al.** *Pantoea allii* sp. nov., isolated from onion plants and seed. *Int J Syst Evol Microbiol* 2011;61(4):932-937.
18. **Brady CL, Cleenwerck I, Venter SN, Engelbeen K, De Vos P et al.** Emended description of the genus *Pantoea*, description of four species from human clinical samples, *Pantoea septica* sp. nov., *Pantoea eucrina* sp. nov., *Pantoea brenneri* sp. nov. and *Pantoea conspicua* sp. nov., and transfer of *Pectobacterium cypripedii* (Hori 1911) Brenner et al. 1973 emend. Hauben et al. 1998 to the genus as *Pantoea cypripedii* comb. nov. *Int J Syst Evol Microbiol* 2010;60(10):2430-2440.
19. **Prakash O, Nimonkar Y, Vaishampayan A, Mishra M, Kumbhare S et al.** *Pantoea intestinalis* sp. nov., isolated from the human gut. *Int J Syst Evol Microbiol* 2015;65(10):3352-3358.

20. **Chen C, Xin K, Liu H, Cheng J, Shen X et al.** *Pantoea alhagi*, a novel endophytic bacterium with ability to improve growth and drought tolerance in wheat. *Sci Rep* 2017;7:41564.
21. **Hollis D, Hickman F, Fanning G, Farmer J, Weaver R et al.** *Tatumella ptyseos* gen. nov., sp. nov., a member of the family *Enterobacteriaceae* found in clinical specimens. *J Clin Microbiol* 1981;14(1):79-88.
22. **Brady CL, Venter SN, Cleenwerck I, Vandemeulebroecke K, De Vos P et al.** Transfer of *Pantoea citrea*, *Pantoea punctata* and *Pantoea terrea* to the genus *Tatumella* emend. as *Tatumella citrea* comb. nov., *Tatumella punctata* comb. nov. and *Tatumella terrea* comb. nov. and description of *Tatumella morbirosei* sp. nov. *Int J Syst Evol Microbiol* 2010;60(3):484-494.
23. **Tracz DM, Gilmour MW, Mabon P, Beniac DR, Hoang L et al.** *Tatumella saanichensis* sp. nov., isolated from a cystic fibrosis patient. *Int J Syst Evol Microbiol* 2015;65(6):1959-1966.
24. **Brady C, Cleenwerck I, Venter S, Vancanneyt M, Swings J et al.** Phylogeny and identification of *Pantoea* species associated with plants, humans and the natural environment based on multilocus sequence analysis (MLSA). *Syst Appl Microbiol* 2008;31(6):447-460.
25. **Kamber T, Smits THM, Rezzonico F, Duffy B**. Genomics and current genetic understanding of *Erwinia amylovora* and the fire blight antagonist *Pantoea vagans*. *Trees*, journal article 2012;26(1):227-238.
26. **Glaeser SP, Kämpfer P**. Multilocus sequence analysis (MLSA) in prokaryotic taxonomy. *Syst Appl Microbiol* 2015;38(4):237-245.
27. **Gueule D, Fourny G, Ageron E, Le Flèche-Matéos A, Vandenbogaert M et al.** *Pantoea coffeiphila* sp. nov., cause of the 'potato taste' of Arabica coffee from the African Great Lakes region. *Int J Syst Evol Microbiol* 2015;65(1):23-29.
28. **Palmer M, Steenkamp ET, Coetzee MPA, Chan W-Y, van Zyl E et al.** Phylogenomic resolution of the bacterial genus *Pantoea* and its relationship with *Erwinia* and *Tatumella*. *Antonie van Leeuwenhoek*, journal article 2017:1-23.
29. **Rezzonico F, Smits THM, Montesinos E, Frey JE, Duffy B**. Genotypic comparison of *Pantoea agglomerans* plant and clinical strains. *BMC Microbiol* 2009;9.
30. **Tambong JT, Xu R, Kaneza C-A, Nshogozabahizi J-C**. An in-depth analysis of a multilocus phylogeny identifies *leus* as a reliable phylogenetic marker for the genus *Pantoea*. *Evol Bioinform Online* 2014;10:115-125.
31. **Moretti C, Hosni T, Vandemeulebroecke K, Brady C, De Vos P et al.** *Erwinia oleae* sp. nov., isolated from olive knots caused by *Pseudomonas savastanoi* pv. *savastanoi*. *Int J Syst Evol Microbiol* 2011;61(11):2745-2752.
32. **Skrodenytė-Arbačiauskienė V, Radžiutė S, Stunžėnas V, Būda V**. *Erwinia typographi* sp. nov., isolated from bark beetle (*Ips typographus*) gut. *Int J Syst Evol Microbiol* 2012;62(4):942-948.
33. **Garrity GM, Bell JA, Lilburn T**. The revised road map to the manual. *Bergey's Manual® of Systematic Bacteriology*: Springer; 2005. pp. 159-187.
34. **Gevers D, Cohan FM, Lawrence JG, Spratt BG, Coenye T et al.** Re-evaluating prokaryotic species. *Nat Rev*, Review 2005;3:733-739.
35. **Konstantinidis KT, Tiedje JM**. Prokaryotic taxonomy and phylogeny in the genomic era: advancements and challenges ahead. *Curr Opin Microbiol* 2007;10(5):504-509.
36. **Jeffroy O, Brinkmann H, Delsuc F, Philippe H**. Phylogenomics: the beginning of incongruence? *Trends Genet* 2006;22(4):225-231.
37. **Heath TA, Hedtke SM, Hillis DM**. Taxon sampling and the accuracy of phylogenetic analyses. *J Syst Evol* 2008;46(3):239-257.
38. **Philippe H, Brinkmann H, Lavrov DV, Littlewood DTJ, Manuel M et al.** Resolving difficult phylogenetic questions: why more sequences are not enough. *PLoS Biol* 2011;9(3):e1000602.
39. **Zwickl DJ, Hillis DM**. Increased taxon sampling greatly reduces phylogenetic error. *Syst Biol* 2002;51(4):588-598.
40. **Hillis DM**. Taxonomic sampling, phylogenetic accuracy, and investigator bias. *Syst Biol* 1998;47(1):3-8.

41. **Mitchell A, Mitter C, Regier JC**. More taxa or more characters revisited: Combining data from nuclear protein-encoding genes for phylogenetic analyses of Noctuoidea (Insecta: Lepidoptera). *Syst Biol* 2000;49(2):202-224.
42. **Tanaka YK, Horie N, Mochida K, Yoshida Y, Okugawa E et al.** *Pantoea theicola* sp. nov., isolated from black tea. *Int J Syst Evol Microbiol* 2015;65(10):3313-3319.
43. **Steenkamp ET, van Zyl E, Beukes CW, Avontuur JR, Chan WY et al.** *Burkholderia kirstenboschensis* sp. nov. nodulates papilionoid legumes indigenous to South Africa. *Syst Appl Microbiol* 2015;38(8):545-554.
44. **Blom J, Kreis J, Spänig S, Juhre T, Bertelli C et al.** EDGAR 2.0: an enhanced software platform for comparative gene content analyses. *Nucleic Acids Res* 2016.
45. **Hall T**. BioEdit: an important software for molecular biology. *GERF Bull Biosci* 2011;2(1):60-61.
46. **Katoh K, Standley DM**. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* 2013;30(4):772-780.
47. **Kück P, Longo GC**. FASconCAT-G: extensive functions for multiple sequence alignment preparations concerning phylogenetic studies. *Frontiers in zoology* 2014;11(1):81.
48. **Abascal F, Zardoya R, Posada D**. ProtTest: selection of best-fit models of protein evolution. *Bioinformatics* 2005;21(9):2104-2105.
49. **Stamatakis A**. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 2014;30(9):1312-1313.
50. **Edgar RC**. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 2004;32(5):1792-1797.
51. **Kanehisa M, Sato Y, Kawashima M, Furumichi M, Tanabe M**. KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res* 2016:D457-D462.
52. **Kanehisa M, Sato Y, Morishima K**. BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. *J Mol Biol* 2016;428(4):726-731.
53. **Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ**. Basic local alignment search tool. *J Mol Biol* 1990;215.
54. **Conesa A, Götz S, García-Gómez JM, Terol J, Talón M et al.** Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* 2005;21(18):3674-3676.
55. **Moran NA, von Dohlen CD, Baumann P**. Faster evolutionary rates in endosymbiotic bacteria than in cospeciating insect hosts. *J Mol Evol*, journal article 1995;41(6):727-731.
56. **Moran NA**. Accelerated evolution and Muller's rachet in endosymbiotic bacteria. *Proc Natl Acad Sci* 1996;93(7):2873-2878.
57. **Moran NA, Wernegreen JJ**. Lifestyle evolution in symbiotic bacteria: insights from genomics. *Trends Ecol Evol* 2000;15(8):321-326.
58. **Moran NA, McCutcheon JP, Nakabachi A**. Genomics and evolution of heritable bacterial symbionts. *Annu Rev Genet* 2008;42(1):165-190.
59. **McCutcheon JP, Moran NA**. Extreme genome reduction in symbiotic bacteria. *Nat Rev Microbiol* 2012;10(1):13.
60. **Tian H, Jing C**. Genome sequence of the aerobic arsenate-reducing bacterium *Pantoea* sp. strain IMH. *Genome Announc* 2014;2(2):e00267-00214.
61. **Liu B, Luo J, Li W, Long X-F, Zhang Y-Q et al.** *Erwinia teleogrylli* sp. nov., a bacterial isolate associated with a chinese cricket. *PLOS ONE* 2016;11(1):e0146596.
62. **Popp A, Cleenwerck I, Iversen C, De Vos P, Stephan R**. *Pantoea gaviniae* sp. nov. and *Pantoea calida* sp. nov., isolated from infant formula and an infant formula production environment. *Int J Syst Evol Microbiol* 2010;60(12):2786-2792.
63. **Huelsenbeck JP**. The robustness of two phylogenetic methods: four-taxon simulations reveal a slight superiority of maximum likelihood over neighbor joining. *Mol Biol Evol* 1995;12(5):843-849.
64. **Bergsten J**. A review of long-branch attraction. *Cladistics* 2005;21(2):163-193.

65. **Hillis DM**. Molecular versus morphological approaches to systematics. *Annu Rev Ecol Syst* 1987;18(1):23-42.
66. **Fritz S, Cassir N, Noudel R, De La Rosa S, Roche P-H et al.** Postsurgical *Pantoea calida* meningitis: a case report. *J Med Case Rep* 2014;8:195-195.
67. **Skrodenyte-Arbaciauskiene V, Sruoga A, Butkauskas D**. Assessment of microbial diversity in the river trout *Salmo trutta* fario L. intestinal tract identified by partial 16S rRNA gene sequence analysis. *Fisheries Science* 2006;72(3):597-602.
68. **Cardoza YJ, Vasanthakumar A, Suazo A, Raffa KF**. Survey and phylogenetic analysis of culturable microbes in the oral secretions of three bark beetle species. *Entomologia Experimentalis et Applicata* 2009;131(2):138-147.
69. **Wu Q, Du J, Zhuang G, Jing C**. *Bacillus* sp. SXB and *Pantoea* sp. IMH, aerobic As (V)-reducing bacteria isolated from arsenic-contaminated soil. *J Appl Microbiol* 2013;114(3):713-721.
70. **Brady CL, Venter SN, Cleenwerck I, Engelbeen K, Vancanneyt M et al.** *Pantoea vagans* sp. nov., *Pantoea eucalypti* sp. nov., *Pantoea deleyi* sp. nov. and *Pantoea anthophila* sp. nov. *Int J Syst Evol Microbiol* 2009;59(9):2339-2345.
71. **Mergaert J, Verdonck L, Kersters K**. Transfer of *Erwinia ananas* (synonym, *Erwinia uredovora*) and *Erwinia stewartii* to the genus *Pantoea* emend. as *Pantoea ananas* (Serrano 1928) comb. nov. and *Pantoea stewartii* (Smith 1898) comb. nov., respectively, and description of *Pantoea stewartii* subsp. *indologenes* subsp. nov. *Int J Syst Bacteriol* 1993;43.
72. **Rezzonico F, Smits THM, Born Y, Blom J, Frey JE et al.** Erwinia gerundensis sp. nov., a cosmopolitan epiphyte originally isolated from pome fruit trees. *International Journal of Systematic and Evolutionary Microbiology* 2016;66(3):1583-1592.
73. **Grote J, Thrash JC, Huggett MJ, Landry ZC, Carini P et al.** Streamlining and core genome conservation among highly divergent members of the SAR11 clade. *mBio* 2012;3(5).
74. **Callister SJ, McCue LA, Turse JE, Monroe ME, Auberry KJ et al.** Comparative bacterial proteomics: Analysis of the core genome concept. *PLOS ONE* 2008;3(2):e1542.
75. **den Bakker HC, Cummings CA, Ferreira V, Vatta P, Orsi RH et al.** Comparative genomics of the bacterial genus *Listeria*: Genome evolution is characterized by limited gene acquisition and limited gene loss. *BMC genomics* 2010;11(1):688.
76. **Smits TH, Guerrero-Prieto VM, Hernández-Escarcega G, Blom J, Goesmann A et al.** Whole-genome sequencing of *Erwinia amylovora* strains from Mexico detects single nucleotide polymorphisms in rpsL conferring streptomycin resistance and in the avrRpt2 effector altering host interactions. *Genome Announc* 2014;2(1):e01229-01213.
77. **Kube M, Migdoll AM, Gehring I, Heitmann K, Mayer Y et al.** Genome comparison of the epiphytic bacteria *Erwinia billingiae* and *E. tasmaniensis* with the pear pathogen *E. pyrifoliae*. *BMC Genomics* 2010;11(1):393.
78. **Redzuan RA, Abu Bakar N, Rozano L, Badrun R, Mat Amin N et al.** Draft genome sequence of *Erwinia mallotivora* BT-MARDI, causative agent of papaya dieback disease. *Genome Announc* 2014;2(3):e00375-00314.
79. **Kube M, Migdoll AM, Müller I, Kuhl H, Beck A et al.** The genome of *Erwinia tasmaniensis* strain Et1/99, a non-pathogenic bacterium in the genus *Erwinia*. *Environ Microbiol* 2008;10(9):2211-2222.
80. **da Silva DP, Devescovi G, Paszkiewicz K, Moretti C, Buonaurio R et al.** Draft genome sequence of *Erwinia toletana*, a bacterium associated with olive knots caused by *Pseudomonas savastanoi* pv. savastanoi. *Genome Announc* 2013;1(3):e00205-00213.
81. **Baltrus DA, Dougherty K, Arendt KR, Huntemann M, Clum A et al.** Absence of genome reduction in diverse, facultative endohyphal bacteria. *Microbial Genomics* 2017;3(2).
82. **Lim J-A, Lee DH, Kim B-Y, Heu S**. Draft genome sequence of *Pantoea agglomerans* R190, a producer of antibiotics against phytopathogens and foodborne pathogens. *J Biotechnol* 2014;188:7-8.

83. **Shyntum DY, Venter SN, Moleleki LN, Toth I, Coutinho TA**. Comparative genomics of type VI secretion systems in strains of *Pantoea ananatis* from different environments. *BMC genomics* 2014;15(1):163.
84. **Wan X, Hou S, Phan N, Malone Moss JS, Donachie SP et al.** Draft genome sequence of *Pantoea anthophila* Strain 11-2 from hypersaline Lake Laysan, Hawaii. *Genome Announc* 2015;3(3).
85. **Stavrinides J, No A, Ochman H**. A single genetic locus in the phytopathogen *Pantoea stewartii* enables gut colonization and pathogenicity in an insect host. *Environ Microbiol* 2010;12(1):147-155.
86. **Smits TH, Rezzonico F, Kamber T, Goesmann A, Ishimaru CA et al.** Genome sequence of the biocontrol agent *Pantoea vagans* strain C9-1. *J Bacteriol* 2010;192(24):6486-6487.
87. **Suen G, Scott JJ, Aylward FO, Adams SM, Tringe SG et al.** An insect herbivore microbiome with high plant biomass-degrading capacity. *PLoS Genet* 2010;6(9):e1001129.
88. **Hong K-W, Gan HM, Low S-M, Lee PKY, Chong Y-M et al.** Draft genome sequence of *Pantoea* sp. strain A4, a *Rafflesia*-associated bacterium that produces N-acylhomoserine lactones as quorum-sensing molecules. *J Bacteriol* 2012;194(23):6610.
89. **Brown SD, Utturkar SM, Klingeman DM, Johnson CM, Martin SL et al.** Twenty-one genome sequences from *pseudomonas* species and 19 genome sequences from diverse bacteria isolated from the rhizosphere and endosphere of *Populus deltoides*. *J Bacteriol* 2012;194(21):5991-5993.
90. **Ren Y, Ren Y, Zhou Z, Guo X, Li Y et al.** Complete genome sequence of *Enterobacter cloacae* subsp. *cloacae* type strain ATCC 13047. *J Bacteriol* 2010;192(9):2463-2464.
91. **Masood N, Moore K, Farbos A, Hariri S, Paszkiewicz K et al.** Draft genome sequences of three newly identified species in the genus *Cronobacter*, *C. helveticus* LMG23732T, *C. pulveris* LMG24059, and *C. zurichensis* LMG23730T. *Genome Announc* 2013;1(5):e00783-00713.
92. **Daligault HE, Davenport KW, Minogue TD, Bishop-Lilly KA, Bruce DC et al.** Draft genome assembly of *Klebsiella pneumoniae* type strain ATCC 13883. *Genome Announc* 2014;2(5):e00939-00914.
93. **Daligault H, Davenport K, Minogue T, Broomall S, Bruce D et al.** Genome assembly of *Serratia marcescens* type strain ATCC 13880. *Genome Announc* 2014;2(5):e00967-00914.