



# Genome-Based Characterization of Biological Processes That Differentiate Closely Related Bacteria

Marike Palmer<sup>1</sup>, Emma T. Steenkamp<sup>1</sup>, Martin P. A. Coetzee<sup>2</sup>, Jochen Blom<sup>3</sup> and Stephanus N. Venter<sup>1\*</sup>

<sup>1</sup> Department of Microbiology and Plant Pathology, Forestry and Agricultural Biotechnology Institute, University of Pretoria, Pretoria, South Africa, <sup>2</sup> Department of Genetic, Forestry and Agricultural Biotechnology Institute, University of Pretoria, Pretoria, South Africa, <sup>3</sup> Bioinformatics and Systems Biology, Justus-Liebig-University Giessen, Giessen, Germany

## OPEN ACCESS

### Edited by:

Antonio Ventosa,  
Universidad de Sevilla, Spain

### Reviewed by:

Fabiano Thompson,  
Universidade Federal do Rio de  
Janeiro, Brazil

Alice Rebecca Wattam,  
Virginia Tech, United States

### \*Correspondence:

Stephanus N. Venter  
fanus.venter@up.ac.za

### Specialty section:

This article was submitted to  
Evolutionary and Genomic  
Microbiology,  
a section of the journal  
Frontiers in Microbiology

**Received:** 16 August 2017

**Accepted:** 17 January 2018

**Published:** 06 February 2018

### Citation:

Palmer M, Steenkamp ET,  
Coetzee MPA, Blom J and Venter SN  
(2018) Genome-Based  
Characterization of Biological  
Processes That Differentiate Closely  
Related Bacteria.  
*Front. Microbiol.* 9:113.  
doi: 10.3389/fmicb.2018.00113

Bacteriologists have strived toward attaining a natural classification system based on evolutionary relationships for nearly 100 years. In the early twentieth century it was accepted that a phylogeny-based system would be the most appropriate, but in the absence of molecular data, this approach proved exceedingly difficult. Subsequent technical advances and the increasing availability of genome sequencing have allowed for the generation of robust phylogenies at all taxonomic levels. In this study, we explored the possibility of linking biological characters to higher-level taxonomic groups in bacteria by making use of whole genome sequence information. For this purpose, we specifically targeted the genus *Pantoea* and its four main lineages. The shared gene sets were determined for *Pantoea*, the four lineages within the genus, as well as its sister-genus *Tatumella*. This was followed by functional characterization of the gene sets using the Kyoto Encyclopedia of Genes and Genomes (KEGG) database. In comparison to *Tatumella*, various traits involved in nutrient cycling were identified within *Pantoea*, providing evidence for increased efficacy in recycling of metabolites within the genus. Additionally, a number of traits associated with pathogenicity were identified within species often associated with opportunistic infections, with some support for adaptation toward overcoming host defenses. Some traits were also only conserved within specific lineages, potentially acquired in an ancestor to the lineage and subsequently maintained. It was also observed that the species isolated from the most diverse sources were generally the most versatile in their carbon metabolism. By investigating evolution, based on the more variable genomic regions, it may be possible to detect biologically relevant differences associated with the course of evolution and speciation.

**Keywords:** genome-inferred biology, *Enterobacteriaceae*, phenotype, KEGG, bacterial systematics, *Pantoea*

## INTRODUCTION

Since the early twentieth century scientists have recognized the value of phylogenetic inferences in determining natural relationships between taxa, which is essential for both taxonomy and evolutionary studies (Woese, 1987). However, the move toward a more natural classification system by these early bacteriologists, based on phylogenetics, proved exceedingly difficult as traditionally used morphological traits were not variable enough to group taxa reliably (Stanier and Van Niel, 1941; Woese, 1987; Woese et al., 1990). Although this led to the use of physiological

characters, some researchers already argued early on that such data would not be suitable for developing evolutionary hypotheses. They emphasized that physiological traits would generally not be phylogenetically informative as long as there were no clear understanding of their genetic basis and overall biological importance (Stanier and Van Niel, 1941). The consensus view at the time was thus that phylogenetic inferences were definitely needed for elucidating the natural relationships among bacteria, but that this would only be possible with the use of suitably informative characters (Stanier and Van Niel, 1941; Woese, 1987, 1994; Woese et al., 1990). As a result, scholars mostly abandoned the field of bacterial systematics until more reliable characters became available with the advent of nucleic acid-based molecular phylogenetics in the 1970s (Woese, 1987, 1994, 1998; Woese et al., 1990; McInerney et al., 2011).

For studying bacterial systematics, the ubiquitous 16S ribosomal RNA (16S rRNA) gene was initially the marker of choice (Hillis and Dixon, 1991; Woese, 1994, 1998; Garrity et al., 2005; Gevers et al., 2005; Konstantinidis and Tiedje, 2007). Over time, however, as the diversity of examined samples increased, it became apparent that the 16 rRNA gene sequence alone does not provide sufficient phylogenetic resolution. Therefore, more reliable approaches for phylogenetic inference were sought to obtain better resolved trees. This led to the use of multiple locus sequence analyses (MLSA) (Gevers et al., 2005; Konstantinidis and Tiedje, 2007; Glaeser and Kämpfer, 2015), ribosomal MLSA (Bennett et al., 2012) and more recently core genome phylogenies (Bennett et al., 2012; Chan et al., 2012; Rahman et al., 2015; Schwartz et al., 2015; Palmer et al., 2017). These approaches, especially those based on large numbers of core genes, provide robust evolutionary hypotheses that seems to be resilient to most known phylogenetic errors (Beukes et al., 2017; Palmer et al., 2017) and have recently formed the foundation of taxonomic decisions, particularly in problematic taxa (Zhang et al., 2011; Bennett et al., 2012; Chan et al., 2012; Richards et al., 2014; Ormeno-Orrillo et al., 2015; Rahman et al., 2015).

The next logical step after having used phylogenetics to identify taxa, particularly those above the species level, would be to assign biological characters to them. For example, if bacterial genera or the lineages within them represent natural clusters, it should be possible to identify properties that they share with one another, but that are different from those occurring in other such clusters. Previously, various standardized sets of physiological tests have been used to study phenotypic cohesion of bacterial taxa (Schubert, 1968; Gavini et al., 1989; Mergaert et al., 1993; Brady et al., 2010a, 2013), but these have been mainly developed from a clinical diagnostics perspective (Konstantinidis and Tiedje, 2005a,b; Sutcliffe, 2015). Accordingly, the characters identified by these tests have limited application outside this environment (Sutcliffe et al., 2012; Sutcliffe, 2015). Other than revealing some basic physiological capabilities, these standard phenotypic tests are incapable of capturing the countless traits encoded on bacterial genomes. In addition, the very limited set of traits analyzed rarely differentiates between taxa, as the members of a taxon can show immense physiological variability. Therefore, characters that are biologically meaningful and that potentially

define or distinguish higher-level bacterial groups and taxa would thus have to be sought through other means.

For identifying biological traits that are potentially taxon-defining, whole genome sequences represent invaluable resources. A wealth of traits can be inferred from a bacterium's genome sequence by making use of bioinformatics approaches and databases built from experimental evidence. For example, metabolic and physiological networks or pathways can be inferred from gene sequences by making use of their homology to sequences in the Kyoto Encyclopedia of Genes and Genomes (KEGG) (Kanehisa et al., 2016b). Each sequence in the KEGG database have an associated KEGG Orthology (KO) term, which is in turn coupled to proteins whose functions have been experimentally verified (Kanehisa et al., 2016b). In this workflow for inferring physiological properties, the database of functionally verified protein entries is often regarded as a significant limitation. This is because functional characterization of genes occurs at a much slower pace than gene discovery, thus making it impossible to functionally annotate certain genes (Linghu et al., 2008). As a result, taxa related to the model organisms typically have a higher number genes that can be functionally annotated because of the higher similarity between their genomes (Linghu et al., 2008). Despite this limitation, the current information would still provide valuable biological knowledge, especially as the information in these databases increase.

In this study, we explored the possibility of linking biological characters to higher-level taxonomic groups in bacteria by making use of whole genome sequence information. For this purpose we used the genus *Pantoea* for which genome sequences of 21 species were available (Hong et al., 2012; Kamber et al., 2012; Wan et al., 2015; Palmer et al., 2017). These were chosen to span the diversity of the genus, which includes plant pathogens (*P. agglomerans*, *P. ananatis*, *P. stewartii* amongst others; Coutinho and Venter, 2009) and species that affect humans (*P. brenneri*, *P. conspicua*, *P. eucrina*, and *P. septica*; Walterson and Stavriniades, 2015), as well as species that have been isolated from insects, fungal fruiting bodies and environmental samples (Walterson and Stavriniades, 2015; Ma et al., 2016; Palmer et al., 2016; Rong et al., 2016). Overall the members of this genus appear to be highly adaptable to changing environments and may act opportunistically when in contact with potential eukaryotic hosts (Coutinho and Venter, 2009; De Maayer et al., 2012b, 2014; Walterson and Stavriniades, 2015). From a phylogenetic perspective, *Pantoea* and its sister taxon, *Tatumella*, are nested within the *Enterobacteriaceae* where they are closely related to *Erwinia* (Glaeser and Kämpfer, 2015; Palmer et al., 2017). *Pantoea* further separates into four well-supported lineages, viz. the *P. agglomerans* (containing *P. agglomerans*, *P. anthophila*, *P. brenneri*, *P. conspicua*, *P. deleyi*, *P. eucalypti*, and *P. vagans*), *P. ananatis* (containing *P. allii*, *P. ananatis*, *P. stewartii* ssp. *stewartii* and *P. stewartii* ssp. *indologenes*), *P. rodasii* (containing *P. rodasii*, *P. rwandensis*, and *Pantoea* sp. GM01) and *P. dispersa* (containing *P. dispersa*, *P. eucrina* and *P. wallisii*) lineages (Palmer et al., 2017). Other than a limited set of general biological traits (e.g., colony and cell morphology, respiration

status, growth temperature), characters that potentially define *Pantoea* and its lineages have never been identified.

The overall goal of this study was to link biological properties to the current evolutionary hypothesis of *Pantoea* (Palmer et al., 2017), thus allowing the identification of phenotypic characters that potentially define the genus and its lineages. Our specific aims were three-fold. First, for each of *Pantoea* and *Tatumella*, we functionally compared their shared gene sets (i.e., in terms of the pathways and processes each gene is predicted to be involved in) to evaluate the feasibility of using whole genome sequences for identifying taxon-defining characters at ranks higher than the species level. Secondly, the shared gene sets in each of the four *Pantoea* lineages were functionally compared to identify characters associated with the specific evolutionary path of the lineage and that potentially contributed to its initial emergence or subsequent maintenance. Thirdly, the loci underlying these differential characters were further characterized in order to determine their gene composition and distribution among species and whether their conservation is maintained by purifying selection as suggested before (Fang et al., 2008; Sorrels et al., 2009). Broadly, our strategy (Figure 1) involved the identification of shared gene sets, followed by their functional annotation.

## MATERIALS AND METHODS

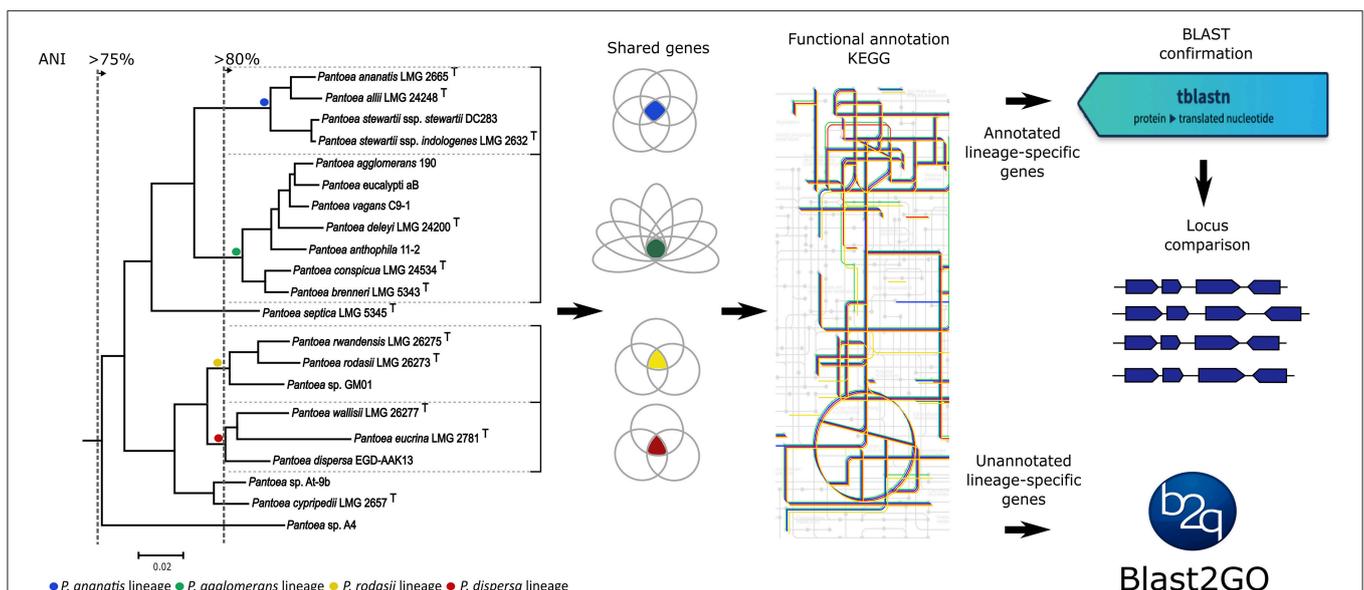
### Genomes Analyzed

All genomes analyzed during this study are publicly available and accessible at the National Centre for Biotechnology Information

(NCBI; <http://www.ncbi.nlm.nih.gov/>). Whole genome sequence data for 21 species of *Pantoea* and three species of *Tatumella* (Tracz et al., 2015) were included in the analyses (Table 1). These *Pantoea* species span the current known phylogenetic and phenotypic diversity of the genus, with most representatives of all of the major lineages (Palmer et al., 2017). For the inter-generic comparisons, all 24 genomes were utilized. For the intra-generic comparisons, we included only 17 *Pantoea* genomes. We excluded three of the known lineages of this genus as one contained only two species (i.e., *Pantoea* sp. At-9b and *P. cypripedii* LMG 2655<sup>T</sup>), while the other two are each represented by single species (i.e., *Pantoea* sp. A4 and *P. septica* LMG 5345<sup>T</sup>).

### Generation of Shared Gene Sets

The shared gene sets of the two genera, as well as the different lineages within *Pantoea* (see Figure 1), were generated with the EDGAR (Efficient Database framework for comparative Genome Analyses using BLAST score Ratios) server (<https://edgar.computational.bio.uni-giessen.de/>; Blom et al., 2016). For each gene set, a representative of the lineage/genus was used for downstream analyses. The representatives used for the different lineages were *P. agglomerans* R190 for the first lineage (encompassing *P. agglomerans*, *P. eucalypti*, *P. vagans*, *P. deleyi*, *P. anthophila*, *P. brenneri*, and *P. conspicua*), *P. ananatis* LMG 2665<sup>T</sup> for the second lineage (comprising of *P. ananatis*, *P. allii*, *P. stewartii* subsp. *stewartii*, and *P. stewartii* subsp. *indologenes*), *P. dispersa* strain number to EGD-AAK13 for the third lineage (encompassing *P. dispersa*, *P. eucrina*, and *P. wallisii*) and *P. rodasii* LMG 26273<sup>T</sup> for the fourth lineage



**FIGURE 1** | Experimental strategy followed for the lineages within *Pantoea* in the study. Lineages were identified from the subtree of *Pantoea* from the protein sequence maximum-likelihood tree of all the shared genes of Palmer et al. (2017). Average nucleotide identity (ANI) values were used as a measure of relatedness between species of a lineage as obtained from Palmer et al. (2017). Shared gene sets were determined from the genome sequences of species within each lineage. Gene sets were then annotated with the Kyoto Encyclopedia of Genes and Genomes (KEGG), followed by BLAST verification and locus comparisons of characterized genes. Uncharacterized genes were subjected to Blast2GO analyses. A similar strategy was followed for the generic comparisons with the exception of the locus comparisons.

**TABLE 1** | Genomes analyzed in the study.

Species	Strain	Origin	Accession number
<i>Pantoea agglomerans</i>	R 190	Apple, Korea	JNGC00000000.1
<i>Pantoea allii</i>	LMG 24248 <sup>T</sup>	Onion seed, South Africa	MLFE00000000.1
<i>Pantoea ananatis</i>	LMG 2665 <sup>T</sup>	Pineapple, Brazil	JFZU00000000.1
<i>Pantoea anthophila</i>	11-2	Hypersaline lake, Hawaii	JXXL00000000.1
<i>Pantoea brenneri</i>	LMG 5343 <sup>T</sup>	Human, USA	MIEI00000000.1
<i>Pantoea calida</i>	LMG 25383 <sup>T</sup>	Infant formula, -	MLFO00000000.1
<i>Pantoea conspicua</i>	LMG 24534 <sup>T</sup>	Human, France	MLFN00000000.1
<i>Pantoea cypripedii</i>	LMG 2657 <sup>T</sup>	Orchid, USA	MLJI00000000.1
<i>Pantoea deleyi</i>	LMG 24200 <sup>T</sup>	Eucalyptus, Uganda	MIPO00000000.1
<i>Pantoea dispersa</i>	EGD-AAK13	Soil, India	AVSS00000000.1
<i>Pantoea eucalypti</i>	aB	Bark beetle, USA	AEDL00000000.1
<i>Pantoea eucria</i>	LMG 2781 <sup>T</sup>	Human, USA	MIPP00000000.1
<i>Pantoea gaviniae</i>	LMG 25382 <sup>T</sup>	Infant formula, -	MLFQ00000000.1
<i>Pantoea rodasii</i>	LMG 26273 <sup>T</sup>	Eucalyptus, Colombia	MLFP00000000.1
<i>Pantoea rwandensis</i>	LMG 26275 <sup>T</sup>	Eucalyptus, Rwanda	MLFR00000000.1
<i>Pantoea septica</i>	LMG 5345 <sup>T</sup>	Human, USA	MLJJ00000000.1
<i>Pantoea stewartii</i> ssp. <i>stewartii</i>	DC 283	Maize, USA	AHIE00000000.1
<i>Pantoea stewartii</i> ssp. <i>indologenes</i>	LMG 2632 <sup>T</sup>	Fox millet, India	JPKO00000000.1
<i>Pantoea vagans</i>	C9-1	Apple, USA	CP001894.1, CP001893.1, CP001894.1
<i>Pantoea wallisii</i>	LMG 26277 <sup>T</sup>	Eucalyptus, South Africa	MLFS00000000.1
<i>Pantoea</i> sp.	At-9b	Leaf cutter ant, USA	CP002433.1, CP002434.1, CP002435.1, CP002436.1, CP002437.1, CP002438.1
<i>Pantoea</i> sp.	A4	Rafflesia flower, Malaysia	ALXE00000000.1
<i>Pantoea</i> sp.	GM01	Poplar, USA	AKUI00000000.1
<i>Tatumella morbiroisei</i>	LMG 23360 <sup>T</sup>	Pineapple, Philippines	CM003276.1
<i>Tatumella ptyseos</i>	ATCC 33301 <sup>T</sup>	Human, USA	ATMJ00000000.1
<i>Tatumella saanichensis</i>	NML 06-3099 <sup>T</sup>	Human, Canada	ATMI00000000.1

(consisting of *P. rodasii*, *P. rwandensis*, and *Pantoea* sp. GM01). For the intergeneric comparisons, *P. agglomerans* R190 was again used as representative of *Pantoea* and *T. ptyseos* ATCC 33301<sup>T</sup> as representative of *Tatumella*.

## Functional Annotation and Identification of Differentially Present Metabolic Pathways

Functional annotation of the different gene sets were first performed by orthology searches against the KEGG database (Kanehisa et al., 2016b) using GhostKOALA (KEGG Orthology and Links Annotation; Kanehisa et al., 2016a) for all gene sets. Genes with KO terms associated with them could be separated based on the functional role of the pathways to which they could be mapped. Specific pathways where differences were detected in the global maps were also considered for comparative purposes (Figure 1).

For the *Pantoea* lineages, genes with no KO associations were then analyzed to assign putative functions using Blast2GO (Conesa et al., 2005; Götz et al., 2008). This was done by subjecting these genes to BLAST analyses for Gene Ontology (GO) associations using Blast2GO implemented in CLC Genomics Workbench (CLC Bio). In these analyses, assignment to more than one GO term per gene was allowed when functional annotation suggested that a gene product is involved

in multiple processes. All Blast2GO analyses were initiated by BLAST searches against the RefSeq non-redundant protein database of NCBI followed by InterproScan (Jones et al., 2014) analyses to identify protein domains as a means of identifying putative functions. Genes remaining without annotation was again subjected to BLAST analyses against the non-redundant database on NCBI to determine the distribution of these genes across taxa.

Individual sets of reconstructed metabolic pathways obtained from the KEGG database were compared to identify differences between lineages and genera. This was done by assigning the set of KEGG pathway maps from each genus/lineage a unique color and then overlaying them onto each other for identifying differences (Figure 1). From these overview pathway maps, specific metabolic pathways were identified for further investigation in eight functional categories used in KEGG. These were carbohydrate, lipid, nucleotide, amino acid and energy metabolism, as well as genes involved in environmental information processing and the metabolism of co-factors, vitamins, and xenobiotics.

Multi-gene pathways that were differentially present or absent were identified from the full set of differences obtained from the KEGG pathway comparisons. This was done to limit the number of genes potentially identified as absent due to sequencing or

assembly errors and also aided in simplifying the overall analysis. For this purpose multi-gene pathways were defined as processes where more than one gene was required to complete a process. From these pathways, the absence of genes from the genomes included in the respective gene sets were verified using local BLAST (Altschul et al., 1990) analyses (tblastn). The genomic coordinates of these genes were then noted to identify clustered genes. The gene clusters were identified and visualized using Geneious 6.1.6 (Biomatters).

Sequences for complete clusters were subsequently extracted from genomes and aligned using the MAFFT 7.309 (Katoh and Standley, 2013) server. When more than three members of a lineage possessed the gene clusters, their sequence alignments were subjected to codon-based selection analyses in MEGA 6.0.6 (Tamura et al., 2013) using HyPhy (Pond and Muse, 2005), to obtain dN (proportion of non-synonymous substitutions) and dS (proportion of synonymous substitutions) values at all codon positions across the alignments. The normalized dN-dS values were then plotted against codon positions in Microsoft Excel 2013.

## RESULTS

### Generation of Shared Gene Sets

In this study, we identified a number of biologically informative characters for *Pantoea* and the four lineages examined. For analyses at both the inter- and intra-generic levels, comparable taxon sets were compiled based on phylogenetic relatedness and Average Nucleotide Identity values (ANI Konstantinidis and Tiedje, 2005b; see Palmer et al., 2017). At the intra-generic level, these sets were also comparable in the sizes of the gene sets (Figure 2), with the exception of the *P. rodasii* lineage. This larger gene set could be attributed to the large size of the genomes of the three current members in the lineage, in comparison to the members of other lineages. However, a large proportion of the genes (>25%) in the respective genomes were present in both *Pantoea* and *Tatumella*. For the *Pantoea* comparisons more than 30% of the genes in respective genomes were present in all taxa, with 55–75% associated with specific lineages and 25–45% apparently species-specific. Overall, the gene sets for the four lineages consisted of 2844 genes for the *P. agglomerans* lineage, 2924 genes for the *P. ananatis* lineage, 3599 genes for the *P. rodasii* lineage and 2872 genes for the *P. dispersa* lineage (Figure 3).

For the inter-generic comparisons, the *Pantoea* gene set (calculated from 21 genomes) consisted of 1862 genes. The *Tatumella* gene set consisted of 2196 genes (calculated from three genomes). This difference in the number of shared genes can most likely be attributed to the number of genomes analyzed in these genera, as the number of genomes available for *Tatumella* is underrepresented.

### Functional Annotation of the *Pantoea* and *Tatumella* Gene Sets with KEGG

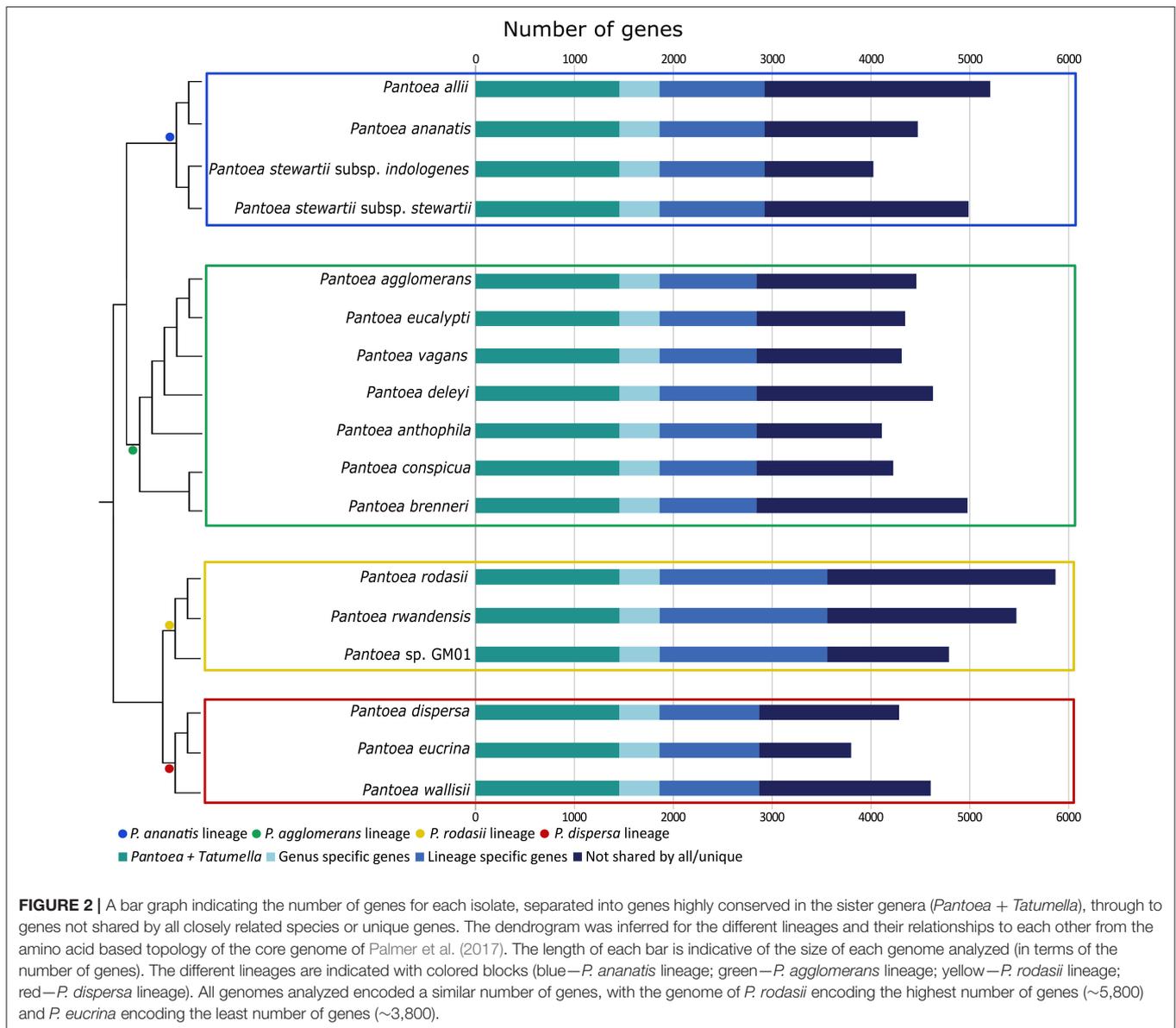
The number of genes with KO associations for *Pantoea* and *Tatumella* were 1,576 (84.6% of the *Pantoea* gene set) and 1,760 (80.1% of the *Tatumella* gene set), respectively

(Supplementary File S1). In both cases, the highest number of genes was involved in “Genetic Information Processing”, followed by “Environmental Information Processing”, with “Unclassified” genes making up the third largest gene group.

The pathways in which we identified differences between *Pantoea* and *Tatumella* were “Metabolic pathways”, “Biosynthesis of secondary metabolites”, “Microbial metabolism in diverse environments”, “Biosynthesis of antibiotics”, “Carbon metabolism”, “Biosynthesis of amino acids”, as well as “2-Oxocarboxylic acid metabolism”, and “Fatty acid metabolism”. Comparison of the relevant global metabolic maps revealed a higher number of reactions predicted for *Tatumella* (as would be expected due to the higher number of shared genes), except for “Fatty acid metabolism”. Closer inspection of the fatty acid metabolism pathways indicated the ability to perform  $\beta$ -oxidation of fatty acids occurred in *Pantoea* but not *Tatumella*.

A total of 124 differences were identified between *Pantoea* and *Tatumella* (Supplementary File S1). These consisted of reactions involved in all functional classes, namely “Carbohydrate metabolism” (citrate cycle, pentose phosphate pathway, fructose and mannose metabolism, ascorbate and aldarate metabolism, starch and sucrose metabolism, glyoxylate and dicarboxylate metabolism and inositol phosphate metabolism), “Energy metabolism” (including methane, sulfur and nitrogen metabolism), “Lipid metabolism” (including fatty acid degradation and sphingolipid metabolism), “Nucleotide metabolism” (purine and pyrimidine metabolism), “Amino acid metabolism” (cysteine and methionine metabolism, lysine degradation, arginine and proline metabolism, histidine metabolism and  $\beta$ -alanine metabolism), “Cofactor metabolism” (nicotinate and nicotinamide metabolism), “Xenobiotics metabolism” (benzoate degradation, chloroalkane and chloroalkene degradation) and “Environmental information processing” [ABC transporters, two-component systems, phosphotransferase systems (PTSs) and chemotaxis]. By limiting the pathways investigated to those where two or more genes are required to complete a pathway, 10 pathways (32 differences) were retained and subsequently absence was confirmed with BLAST analyses (Table 2, Supplementary File S1).

As suggested from the global maps, genes required for  $\beta$ -oxidation of long-chain fatty acids (“Lipid metabolism”—fatty acid degradation) were present in all members of *Pantoea* and absent in all members of *Tatumella*. A number of genes involved in specific pathways in carbohydrate metabolism were detected only in *Pantoea*, namely the “Pentose phosphate pathway” (D-ribose to D-ribose-1-P), “Fructose and mannose metabolism” (D-mannitol to  $\beta$ -D-fructose-6-P), “Starch and sucrose metabolism” (sucrose to ADP-glucose; glycogen to trehalose and amylose, respectively) and “Inositol phosphate metabolism” (myo-inositol to 2-deoxy-5-keto-D-gluconate-6-P). The “Energy metabolism” pathway with differences was sulfur metabolism, where *Pantoea* possessed genes required for the uptake of extracellular taurine and its subsequent conversion to sulfite. *Pantoea* also possessed genes required for the conversion of guanine to (S)-allantoin during “Purine metabolism”. Several pathways involved in “Amino acid metabolism” were also present only in *Pantoea*, specifically those involved in arginine and proline metabolism



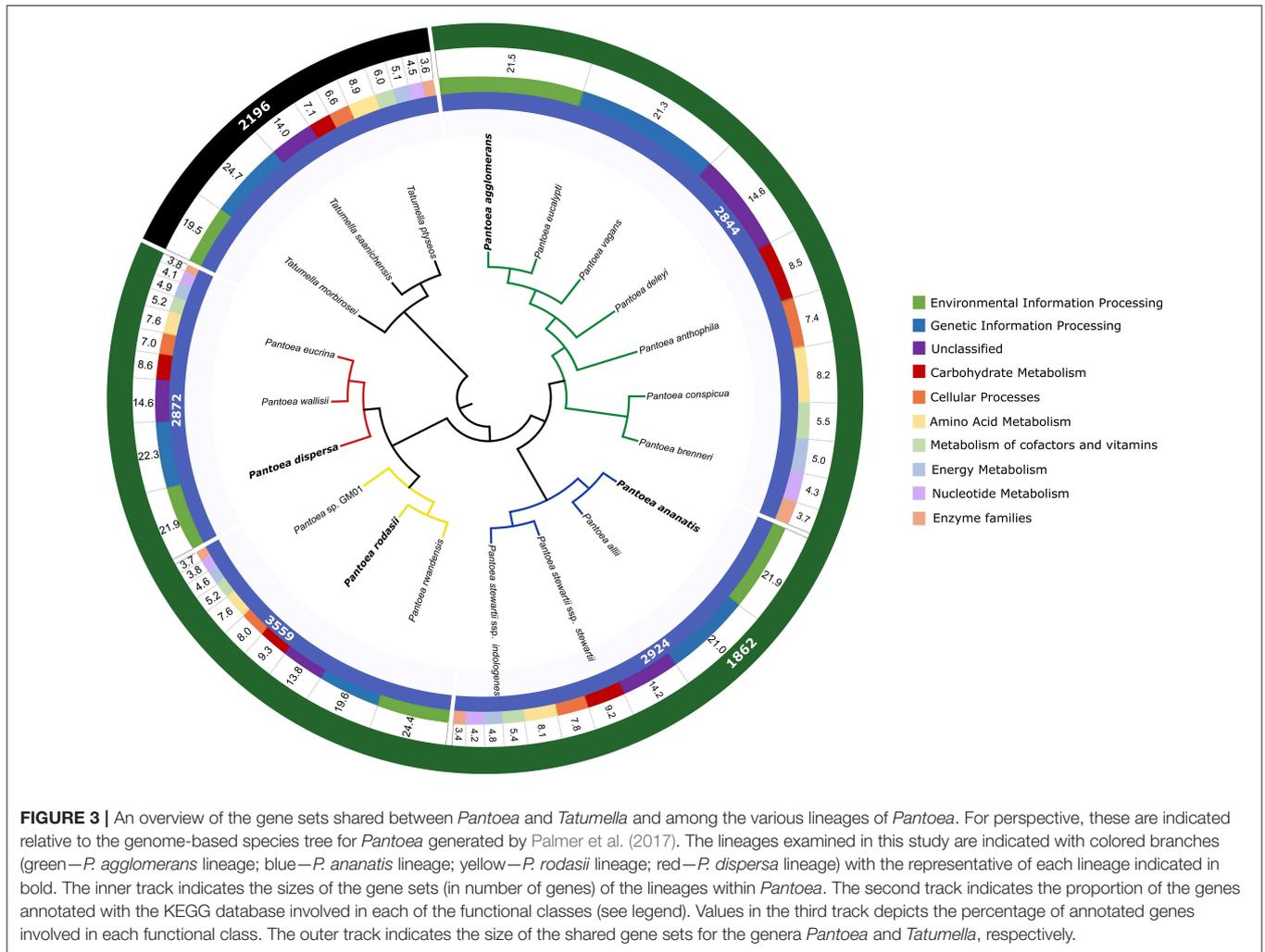
(creatine and N-carbamoylsarcosine to sarcosine) and histidine metabolism (L-histidinol to urocanate). In addition, *Pantoea* also possessed genes for the conversion of L-aspartate to nicotinate-D-ribonucleotide during “Cofactor metabolism” (nicotinate and nicotinamide metabolism).

We also observed differences for genes in the category “Environmental Information Processing”. Amongst the ABC transporters, those for glutamine, glutathione and glycine betaine/proline transporters were found only in *Tatumella*, while those for osmoprotectant, taurine (also seen in sulfur metabolism), L-arabinose, and microcin C were only present in *Pantoea*. In terms of the phosphotransferase systems, *Pantoea* possessed genes necessary to transport and convert N-acetylmuramic acid to N-acetylmuramic acid-6-P, while *Tatumella* possessed genes required for the transport and conversion of N-acetyl-D-glucosamine

to N-acetyl-D-glucosamine-6-P and arbutin/salicin to arbutin-6-P/salicin-6-P.

### Functional Annotation of the Genes Shared by Lineages in *Pantoea* with KEGG

For the *P. agglomerans* lineage, the number of shared genes with KO associations resulted in 2148 genes (75.5% of the gene set). A total of 2197 genes (75.1% of the gene set) in the *P. ananatis* lineage could be annotated using KO terms (Supplementary File S2). The highest percentage of genes with associated KO terms were 75.9% (2181 genes) for the *P. dispersa* lineage, with the lowest being 71.9% for the *P. rodasii* lineage (2559 genes) (Supplementary File S2). In contrast to the inter-generic gene sets, the highest number of genes in all four lineages were involved in “Environmental Information Processing”, followed by “Genetic Information Processing”,



with “Unclassified” genes again being the third most prevalent (Figure 3, Supplementary File S2).

Comparisons of global maps indicated differences between the four lineages in “Biosynthesis of amino acids”, “Biosynthesis of antibiotics”, “Biosynthesis of secondary metabolites”, “Carbon metabolism”, “Overview metabolism” and “Microbial metabolism in diverse environments” (Supplementary File S2). Limiting the reactions investigated to two or more genes acting together to complete a pathway, led to the identification of a number of reactions involved in “Polyketide sugar unit biosynthesis”, “Biosynthesis of siderophore group non-ribosomal peptides”, “Starch and sucrose metabolism”, “Riboflavin metabolism”, “Fructose and mannose metabolism”, “Lysine degradation”, “Chloroalkane and chloroalkene degradation”, “Benzoate degradation”, “Pentose and glucuronate interconversions” and “Cysteine and methionine metabolism” (Supplementary File S3). However, we excluded “Starch and sucrose metabolism” and “Riboflavin biosynthesis” after local BLAST analyses showed that homologs of the respective genes were detected in all taxa (Supplementary File S3). They were likely not recognized previously in our generation of the shared

gene datasets with EDGAR’s strict orthology estimation criteria. The genes involved in the remaining nine processes (two of which were involved in siderophore synthesis) were all found to be clustered and allowed comparison of the gene clusters across all taxa containing these genes (Table 3).

The two genes identified (*rfbC* and *rfbD*) being involved in “Polyketide sugar unit biosynthesis” were present in all the members of the *P. ananatis* lineage, with various members of the other three lineages lacking the genes (*P. eucalypti*, *P. brenneri*, *Pantoea* sp. GM01 and *P. wallisii*; Table 3). Upon examination of the gene cluster containing these genes, two different loci were identified (Figure 4). The first locus was observed in nearly all members of the genus that possessed these genes (including *P. dispersa* and *P. stewartii* subsp. *indologenes*), while the second locus (lacking *rfbB*) was found only in *P. ananatis*, *P. stewartii* subsp. *stewartii*, *P. dispersa* and a partial locus in *P. stewartii* subsp. *indologenes*. The first locus was also slightly different in *P. deleyi* and *P. eucrinea*, as the position of *rfbC* in *P. deleyi* differed (Figure 4—gene indicated in a darker shade) and the locus of *P. eucrinea* contained an additional three genes in comparison to the other taxa (Figure 4). Furthermore,

**TABLE 2** | Intergeneric differences in multi-gene pathways.

Process <sup>a</sup>	Genes	Biological function	Distribution across lineages <sup>b</sup>	References
D-ribose utilization	<i>rhsK, pgm</i>	<ul style="list-style-type: none"> <li>• Converts D-ribose to D-ribose-1-P</li> <li>• Ribose donor during pyrimidine cycling</li> <li>• Involved in co-factor cycling</li> </ul>	<i>Pantoea</i> (21/21) <i>Tatumella</i> (0/3)	Giorgelli et al., 1997; Gossmann et al., 2012; Armenta-Medina et al., 2014
D-mannitol utilization	<i>mtlAD</i>	<ul style="list-style-type: none"> <li>• Converts D-mannitol to <math>\beta</math>-D-fructose-6-P</li> <li>• Used as carbon source</li> </ul>	<i>Pantoea</i> (21/21) <i>Tatumella</i> (0/3)	Berkowitz, 1971; Sprenger, 1993; Boer et al., 1996
Myo-inositol utilization	<i>ioBCDEG</i>	<ul style="list-style-type: none"> <li>• Converts 2-deoxy-5-keto-D-gluconate-6-P to myo-inositol</li> <li>• Used as carbon source</li> </ul>	<i>Pantoea</i> (21/21) <i>Tatumella</i> (0/3)	Anderson and Magasanik, 1971; Yoshida et al., 2004
Sucrose utilization	<i>sacA, MGAM, SI</i>	<ul style="list-style-type: none"> <li>• Degrades sucrose to ADP-glucose and D-fructose</li> <li>• Used as carbon source</li> </ul>	<i>Pantoea</i> (21/21) <i>Tatumella</i> (0/3)	Fouet et al., 1986; Van Beers et al., 1995; Sim et al., 2008
Glycogen degradation	<i>treXYZ, glgB</i>	<ul style="list-style-type: none"> <li>• Converts glycogen to trehalose or amylose</li> <li>• Used as carbon source</li> </ul>	<i>Pantoea</i> (21/21) <i>Tatumella</i> (0/3)	Baecker et al., 1986; Amemura et al., 1988; Chandra et al., 2011
$\beta$ -oxidation	<i>fadABE</i>	<ul style="list-style-type: none"> <li>• Cleaves fatty acids cyclically into acetyl-CoA (even numbered) or succinyl-CoA (odd-numbered)</li> <li>• Used in citric acid cycle</li> </ul>	<i>Pantoea</i> (21/21) <i>Tatumella</i> (0/3)	Schulz, 1991; Fujita et al., 2007; Liu et al., 2010
Purine metabolism	<i>guaD, XDH, uaZ, pucM, PRHOXNB</i>	<ul style="list-style-type: none"> <li>• Converts guanine to (S)-allantoin</li> <li>• Process involved in nitrogen cycling</li> </ul>	<i>Pantoea</i> (21/21) <i>Tatumella</i> (0/3)	Kimiyoshi et al., 1993; Colloc'h et al., 1997; Nygaard et al., 2000; Moffatt and Ashihara, 2002; Cendron et al., 2007
Creatine degradation	E3.5.3.3, E3.5.1.59	<ul style="list-style-type: none"> <li>• Converts creatine and N-carbamoylsarcosine to sarcosine</li> </ul>	<i>Pantoea</i> (21/21) <i>Tatumella</i> (0/3)	Yoshimoto et al., 1976; Deeg et al., 1987; Suzuki, 1991
Histidine degradation	<i>hisD, hutH</i>	<ul style="list-style-type: none"> <li>• Converts L-histidinol to urocanate</li> <li>• Use of histidine as sole carbon source</li> </ul>	<i>Pantoea</i> (21/21) <i>Tatumella</i> (0/3)	Mehler and Tabor, 1953; Schwede et al., 1999; Yang et al., 2011
L-aspartate degradation	<i>nadABCX</i>	<ul style="list-style-type: none"> <li>• Converts L-aspartate to nicotinate-D-ribonucleotide</li> <li>• Process involved in co-factor cycling</li> </ul>	<i>Pantoea</i> (21/21) <i>Tatumella</i> (0/3)	Yang et al., 2003; Ollagnier-de Choudens et al., 2005; Kato et al., 2006; Gossmann et al., 2012
Glutamine ABC Transporter	<i>glnHPQ</i>	<ul style="list-style-type: none"> <li>• Transports glutamine from outside environment into cell through the use of ATP</li> </ul>	<i>Pantoea</i> (0/21) <i>Tatumella</i> (3/3)	Nohno et al., 1986; Walshaw et al., 1997; Hosie and Poole, 2001
Glycine/ betaine/proline ABC Transporter	<i>proVWX</i>	<ul style="list-style-type: none"> <li>• Transports glycine, betaine and proline from outside environment into cell through the use of ATP</li> </ul>	<i>Pantoea</i> (0/21) <i>Tatumella</i> (3/3)	Gowrishankar, 1989; Stirling et al., 1989; Kempf and Bremer, 1995; Ko and Smith, 1999
Glutathione ABC Transporter	<i>gsiABCD</i> ( <i>yliABCD</i> )	<ul style="list-style-type: none"> <li>• Transports glutathione from outside environment into cell through the use of ATP</li> </ul>	<i>Pantoea</i> (0/21) <i>Tatumella</i> (3/3)	Suzuki et al., 2005
Osmoprotectant ABC Transporter	<i>opuABCD</i>	<ul style="list-style-type: none"> <li>• Transports osmoprotectants from outside environment into cell through the use of ATP</li> </ul>	<i>Pantoea</i> (21/21) <i>Tatumella</i> (0/3)	Kempf and Bremer, 1995; Kappes et al., 1999

(Continued)

TABLE 2 | Continued

Process <sup>a</sup>	Genes	Biological function	Distribution across lineages <sup>b</sup>	References
Taurine ABC Transporter	<i>tauABC</i>	<ul style="list-style-type: none"> <li>• Transports taurine from outside environment into cell through the use of ATP</li> <li>• Allows uptake for subsequent degradation of taurine for sulfur cycling</li> </ul>	<i>Pantoea</i> (21/21) <i>Tatumella</i> (0/3)	Van der Ploeg et al., 1996; Javaux et al., 2007
L-arabinose transporter	<i>araFGH</i>	<ul style="list-style-type: none"> <li>• Transports L-arabinose from outside environment into cell through the use of ATP</li> <li>• Required for utilization of L-arabinose as sole carbon source</li> </ul>	<i>Pantoea</i> (21/21) <i>Tatumella</i> (0/3)	Scripture et al., 1987; Kehres and Hogg, 1992; Schneider, 2001
Microcin C ABC Transporter	<i>yejABEF</i>	<ul style="list-style-type: none"> <li>• Transports microcin C from outside environment into cell through the use of ATP</li> <li>• Results in susceptibility to antimicrobial microcin C</li> </ul>	<i>Pantoea</i> (21/21) <i>Tatumella</i> (0/3)	Vanneste et al., 2001; Novikova et al., 2007; Metlitskaya et al., 2009
N-acetylmuramic acid PTS	<i>murP, crr</i>	<ul style="list-style-type: none"> <li>• Transports and phosphorylates N-acetylmuramic acid through the use of ATP</li> <li>• Essential for the use of N-acetylmuramic acid (cell wall component) as carbon source</li> </ul>	<i>Pantoea</i> (21/21) <i>Tatumella</i> (0/3)	Dahl et al., 2004; Uehara et al., 2006,
N-acetyl-D-glucosamine PTS	<i>nagE</i>	<ul style="list-style-type: none"> <li>• Transports and phosphorylates N-acetyl-D-glucosamine through the use of ATP</li> <li>• Essential for the use of N-acetyl-D-glucosamine (cell wall component) as carbon source</li> </ul>	<i>Pantoea</i> (0/21) <i>Tatumella</i> (3/3)	Jaeger and Mayer, 2008; Plumbridge, 2009
Arbutin/salicin PTS	<i>ascF, crr</i>	<ul style="list-style-type: none"> <li>• Transports and phosphorylates arbutin and salicin through the use of ATP</li> <li>• Essential for the use of plant derived glycosides as sole carbon source</li> </ul>	<i>Pantoea</i> (0/21) <i>Tatumella</i> (3/3)	Hall and Xu, 1992; Dahl et al., 2004

<sup>a</sup>PTS refers to phosphotransferase system.

<sup>b</sup>Brackets indicate the number of taxa for which the specific locus is present out of all taxa in the genus.

selection analyses indicated purifying selection for *rfbA* and *rfbB* and diversifying selection for *rfbC* and *rfbD* in the first locus (Figure 4, Supplementary File S4). Contrary to this, both *rfbC* and *rfbD* were under purifying selection in the second locus, with *rfbA* being under mainly purifying selection for the first part and diversifying selection for the second part of the gene (Figure 4, Supplementary File S4).

The genes involved in “Biosynthesis of siderophore group non-ribosomal peptides” and “Lysine degradation” both encoded for different iron acquisition molecules (siderophores) (Table 3). The genes present in most of the members in the genus (“Biosynthesis of siderophore group non-ribosomal peptides”) were identified as being required for the production of enterobactin. The majority of the gene cluster encoding enterobactin appeared to have evolved under purifying selection, with only some regions that evolved mainly under diversifying selection (for example see Figure 5 *entB*; Supplementary File S4). Conversely, the genes involved in “Lysine degradation” in the *P. ananatis* lineage were those required to produce aerobactin from lysine. All members of the *P. ananatis* lineage lacked the genes involved in enterobactin

biosynthesis, but contained the genes required for aerobactin synthesis, while all other members of the genus lacked the genes required for aerobactin biosynthesis (Figure 5). Selection analyses amongst the genes encoding for aerobactin biosynthesis indicated purifying selection in particular for *iucA* and *iucB*. Both these loci were absent from *P. eucalypti*, *P. deleyi* and *P. eucrina*.

The differentially present genes associated with “Fructose and mannose metabolism” consisted of *rhaA*, *rhaB*, and *rhaD* which convert L-rhamnose to glycerone-P and S-lactaldehyde (*rhaD*) (Table 3). This cluster was present in all the members of the *P. agglomerans* and *P. rodasii* lineages, but present only in *P. allii* and *P. ananatis* in the *P. ananatis* lineage, and *P. dispersa* and *P. wallisii* in the *P. dispersa* lineage (Supplementary File S4). From the selection analyses of the *P. agglomerans* and *P. rodasii* lineages it was observed that *rhaB* and *rhaD* evolved under purifying selection, with *rhaA* evolving under diversifying selection (Supplementary File S4).

Our analysis showed that for “Chloroalkane and chloroalkene degradation”, the specific pathway was absent from all the members of the *P. dispersa* lineage (Table 3). This pathway

**TABLE 3** | Distribution multi-gene pathways among the *Pantoea* lineages.

Processes	Pathway reactions	Genes <sup>a</sup>	<i>Pantoea</i> lineages <sup>b</sup>			
			Ag	An	Di	Ro
Polyketide sugar unit biosynthesis	dTDP-4-oxo-6-deoxy-D-glucose <->	<i>rfbC</i>	v	v	v	v
	dTDP-4-dehydro-beta-L-rhamnose					
	dTDP-L-rhamnose <->	<i>rfbD</i>	v	v	v	v
	dTDP-4-dehydro-beta-L-rhamnose					
Biosynthesis of siderophore group non-ribosomal peptides	2,3-Dihydroxybenzoate <-> Enterochelin	<i>entE &amp; F</i>	v	-	v	+
	2,3-Dihydroxybenzoate <->	<i>entA</i>	v	-	v	+
	(2S,3S)-2,3-Dihydro-2,3-dihydroxybenzoate					
	(2S,3S)-2,3-Dihydro-2,3-dihydroxybenzoate <->	<i>entB</i>	v	-	v	+
	Isochorismate					
	Isochorismate <-> Chorismate	<i>entC</i>	v	-	v	+
Fructose and mannose metabolism	L-Rhamnulose <-> L-Rhamnose	<i>rhaA</i>	+	v	v	+
	L-Rhamnulose 1-P <-> L-Rhamnulose	<i>rhaB</i>	+	v	v	+
	Glycerone-P + S-Lactaldehyde <-> L-Rhamnulose 1-P	<i>rhaD</i>	+	v	v	+
Lysine degradation	L-Lysine <-> N6-Hydroxy-L-Lysine	<i>iucD</i>	-	+	-	-
	N6-Hydroxy-L-lysine <->	<i>iucB</i>	-	+	-	-
	N6-Acetyl-N6-hydroxy-L-lysine					
	N6-Acetyl-N6-hydroxy-L-lysine <->	<i>iucA</i>	-	+	-	-
	N2-Citryl-N6-acetyl-N6-hydroxy-L-lysine					
	N2-Citryl-N6-acetyl-N6-hydroxy-L-lysine + N6-Acetyl-N6-hydroxy-L-lysine <-> Aerobactin	<i>iucC</i>	-	+	-	-
Chloroalkane degradation	Chloroacetic acid <-> Glycolate + Hydrochloric acid	2-haloacid DH	+	+	-	+
	Chloroacetaldehyde <-> Chloroacetic acid	aldehyde DH	+	+	-	+
Benzoate degradation	Succinyl-CoA + 3-Oxoadipate <-> Succinate + 3-Oxoadipyl-CoA	<i>pcaI &amp; J</i>	-	+	-	-
	2-Oxo-2,3-dihydrofuran-5-acetate + H2O <->	<i>pcaL</i>	-	+	-	-
	3-Oxoadipate					
	3,4-Dihydroxybenzoate + Oxygen <->	<i>pcaG &amp; H</i>	-	+	-	-
	3-Carboxy-cis,cis-muconate					
Pentose and glucuronate interconversions	D-Glucuronate <-> D-fructuronate	<i>uxaC</i>	+	+	v	+
	D-Galacturonate <-> D-Tagaturonate	<i>uxaC</i>	+	+	v	+
	D-Altronate <-> D-Tagaturonate	<i>uxaB</i>	+	+	v	+
	D-Altronate <-> 2-Dehydro-3-deoxy-D-gluconate	<i>uxaA</i>	+	+	v	+

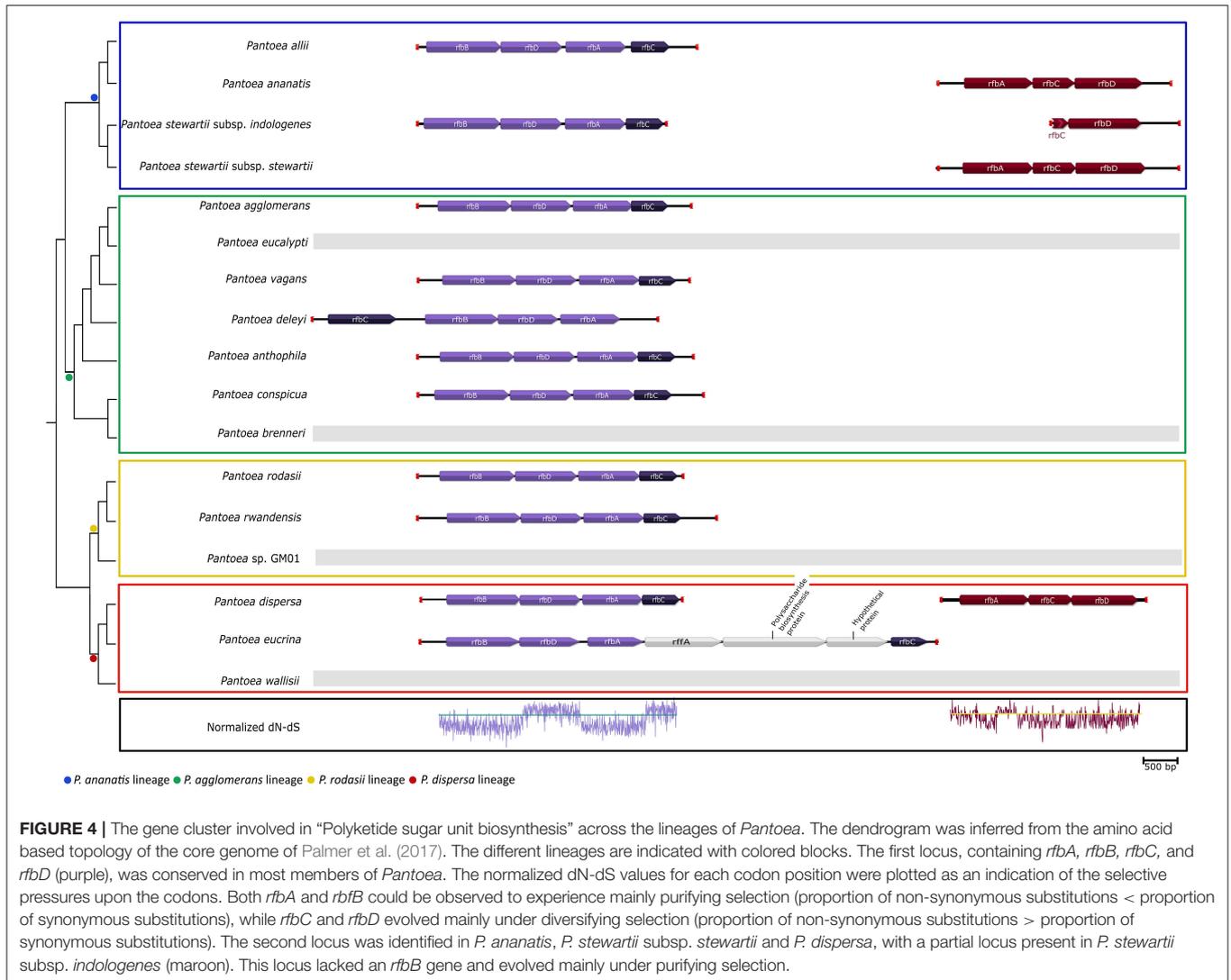
<sup>a</sup>The presence or absence of each pathway was verified using BLAST searches with the relevant sequences against the respective genomes.

<sup>b</sup>The lineages are indicated as follows: Ag = *P. agglomerans* lineage, An = *P. ananatis* lineage, Di = *P. dispersa* lineage and Ro = *P. rodasii* lineage. The presence of genes are indicated with "+", their absence with "-", while "v" is used to indicate the presence in some but not all members of a lineage.

catalyzes the conversion of chloroacetaldehyde to glycolate and hydrochloric acid. All other members of the genus possessed the genes required for this process (**Supplementary File S4**).

The differentially present genes associated with "Benzoate degradation" were involved in the utilization of protocatechuate. They were present only in the *P. ananatis* lineage. However, closer examination revealed that most members of the lineage contained a cluster of 9 genes (*pcaH*, *pcaG*, *pcaQ*, *pcaL*, *pcaB*, KAT, *pcaJ*, *pcaI*, and *pcaR*), but that it contained a deletion in *P. stewartii* subsp. *stewartii* which truncated *pcaL* and removed *pcaB* and KAT from the cluster. Overall, the cluster appeared to be under purifying selection (**Supplementary File S4**).

All members of the genus, except *P. eucrina*, possessed a gene cluster (*uxaA*, *uxaB*, and *uxaC*) involved in "Pentose and glucuronate interconversions" (**Table 3, Supplementary File S4**). The products of *uxaA* and *uxaB* catalyze, respectively, the reversible conversion of 2-dehydro-3-deoxy-D-gluconate to D-altronate and D-altronate to D-tagaturonate, while *uxaC* facilitates interconversions between D-tagaturonate and D-glucuronate and between D-fructuronate and D-galacturonate. These three genes were conserved within the *P. agglomerans*, *P. ananatis*, and *P. rodasii* lineages, with only *uxaA* and *uxaB* being present in *P. dispersa* and *P. wallisii*. Overall, it appeared that these genes were evolving under neutral selection (**Supplementary File S4**).



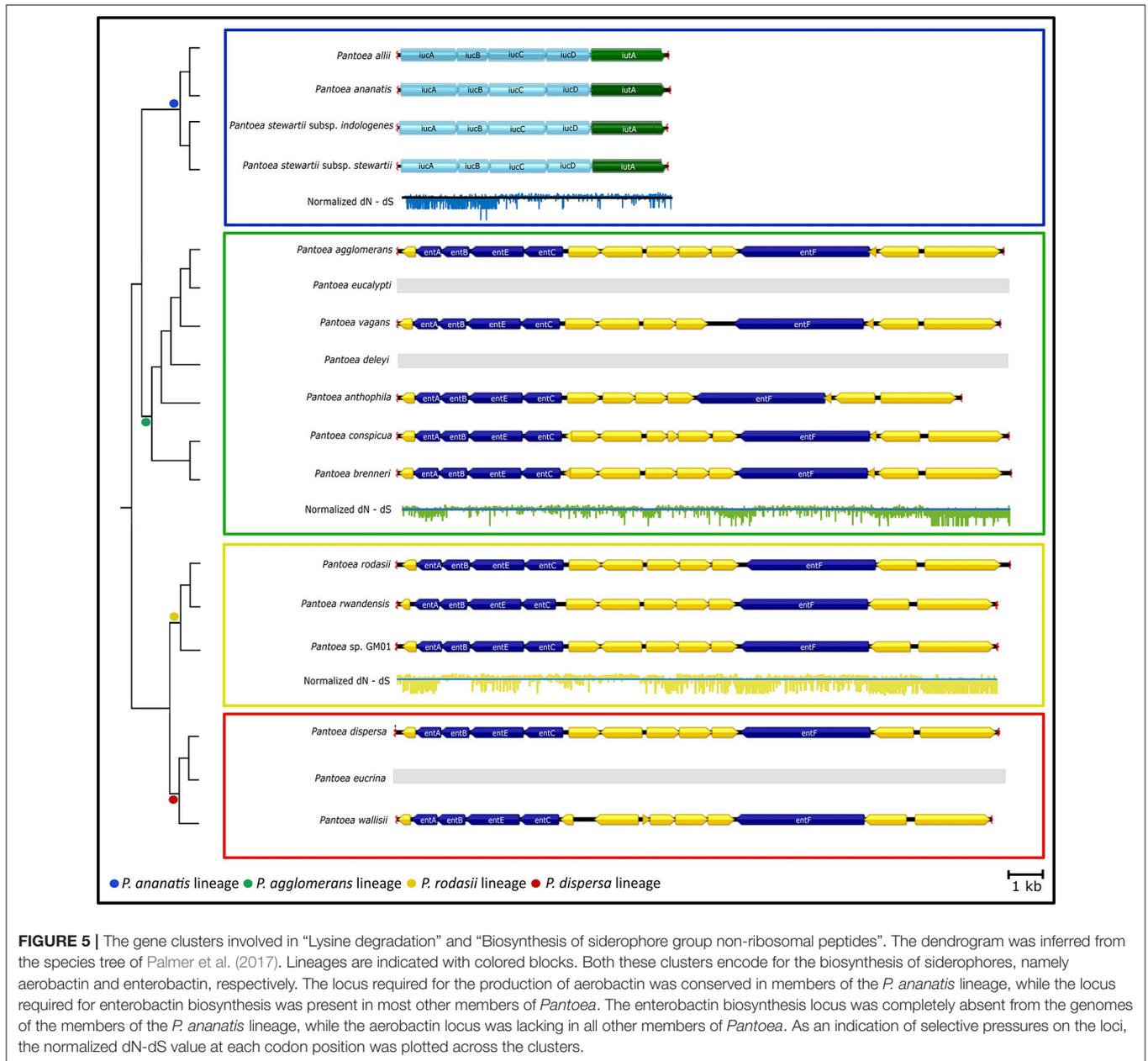
Comparison of the processes involved in “Cysteine and methionine metabolism” showed differences in the synthesis of spermidine (*speD* and *speE*) and the methionine salvage pathway (*mtnA*, *mtnB*, *mtnC*, *mtnD*, and *mtnK*) (Table 3, Supplementary Files S3, S4). The two genes required for the biosynthesis of spermidine allow for the conversion of S-adenosyl-L-methionine and putrescine to 5'-methylthioadenosine and spermidine (*speE*). These two genes were present in all members of the genus except *Pantoea* sp. GM01 (*P. rodasil* lineage), *P. eucrina* and *P. wallisii* (both from the *P. dispersa* lineage; Supplementary File S4). Genes involved in the methionine salvage pathway allow conversion of 5-methylthio-D-ribose to 3-(methylthio)-propanoate through 5 different intermediate reactions (Supplementary File S3). These methionine salvage pathway genes were absent in the *P. dispersa* lineage (*P. dispersa*, *P. eucrina*, and *P. wallisii*) (Supplementary File S4).

We also found several multi-gene systems for two-component systems (2 systems), ABC Transporters (14 systems) and

PTSs (2 systems) that were differentially present within these lineages (Supplementary File S5). Local BLAST analyses allowed identification of taxa where these genes were indeed present, despite not being conserved within the specific gene sets (Figure 6, Supplementary File S5). The two-component systems identified were that for citrate as well as nitrate/nitrite uptake. The ABC transporters identified were the systems for nitrate/nitrite/cyanate, HMP/FAMP, spermidine/putrescine, putrescine, maltose/maltodextrin, D-xylose, myoinositol-1-phosphate, phosphonate, glutamine, arginine, urea, glutathione and iron(II)/manganese. The PTSs detected were those for cellobiose and L-ascorbate.

### Annotation of Lineage-Specific Genes without KEGG Associations

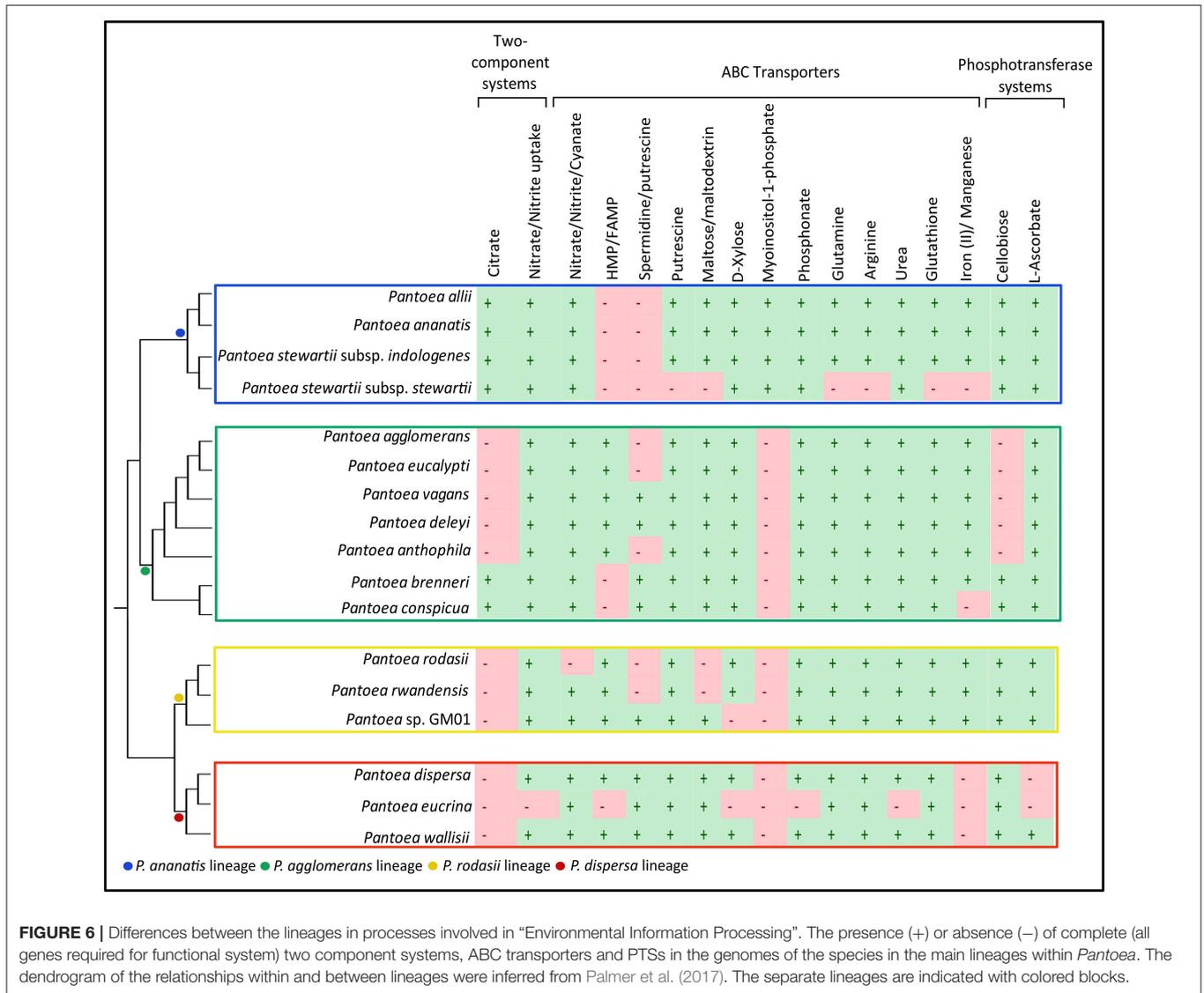
A total of 264 genes were identified as being differentially present in the *Pantoea* lineages, to which no KO term assignment could be made. This set of uncharacterized genes consisted of 62 genes in the *P. ananatis* lineage, 75 genes in the *P. agglomerans*



lineage, 98 genes in the *P. rodasilii* lineage and 29 genes in the *P. dispersa* lineage. Analysis of these genes with Blast2GO allowed annotation of 182 genes. A further six genes could be assigned GO terms, but could not be fully annotated upon merging of annotations due to a lack in InterProScan hits. A total of 76 genes had no functional associations. These unannotated genes could however be used for blastp analyses to identify potential sources of horizontally acquired genes.

The 62 genes of the *P. ananatis* lineage were subjected to Blast2GO analyses, leading to the annotation of 38 genes. In terms of biological processes (GO Level 3), the highest number of genes were involved in “cellular metabolic processes”, “primary metabolic processes” and “organic substance metabolic

processes”, followed by “regulation of cellular processes” and “nitrogen compound metabolic processes” (Figure 7). This lineage thus contained 24 genes present in all members of the lineage, without KEGG or GO functional annotations. Based on blastp hits, 15 of the 24 genes had their closest homologs within other members of the *Enterobacteriaceae*, while two genes had homologs in members of the *Rhizobiaceae* ( $\alpha$ -Proteobacteria). The closest homolog for three genes was respectively from the *Aurantimonadaceae* (*Marteella mediterranea*;  $\alpha$ -Proteobacteria), *Corchorus olitorius* (bush okra; Malvaceae, Eudicots), and *Erwinia* phage ENT90. The remaining four genes had no BLAST hits (blastp) on the non-redundant database (Supplementary File S6).



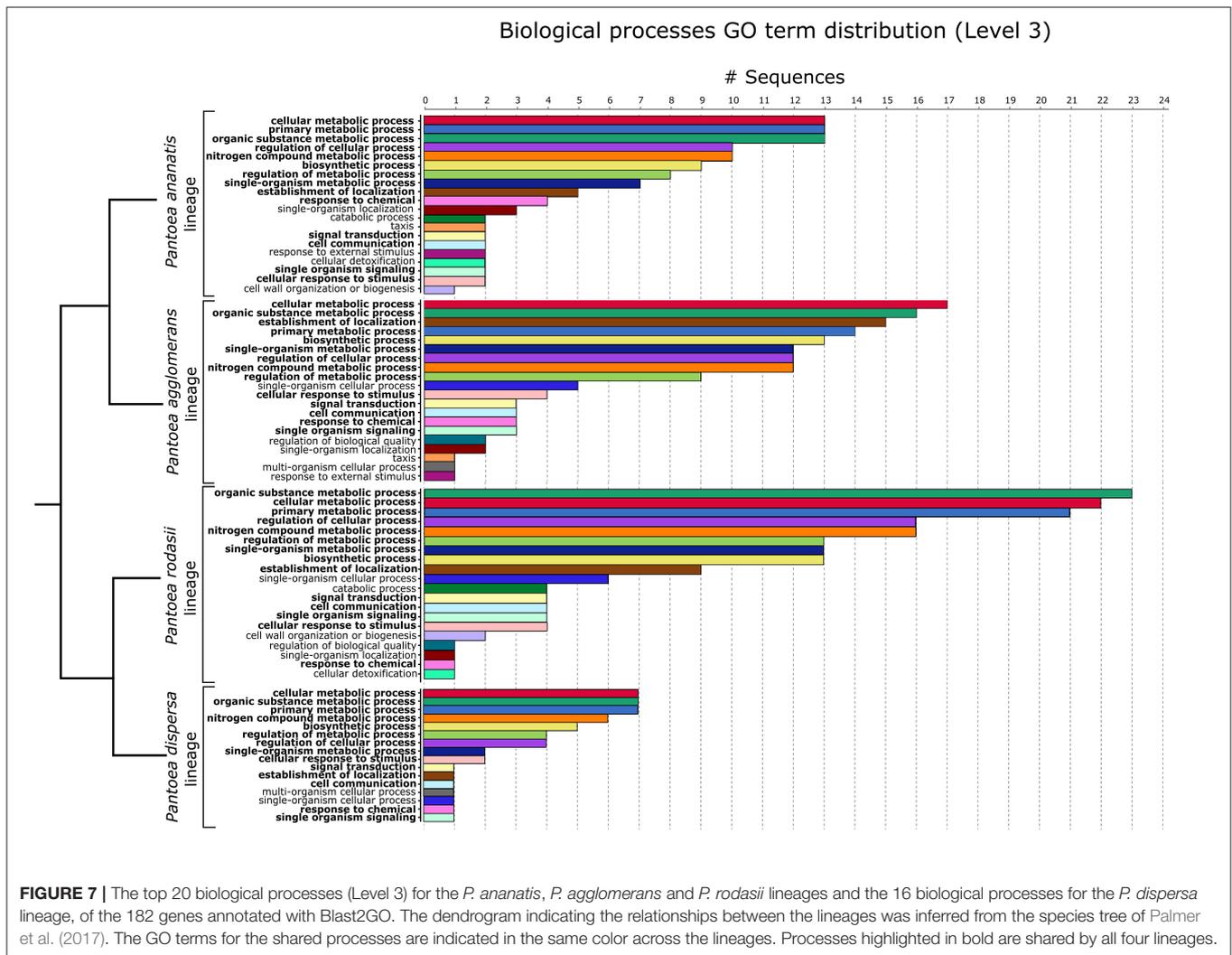
**FIGURE 6** | Differences between the lineages in processes involved in “Environmental Information Processing”. The presence (+) or absence (–) of complete (all genes required for functional system) two component systems, ABC transporters and PTSs in the genomes of the species in the main lineages within *Pantoea*. The dendrogram of the relationships within and between lineages were inferred from Palmer et al. (2017). The separate lineages are indicated with colored blocks.

Of the 75 genes conserved within the *P. agglomerans* lineage not annotated with KEGG, 54 genes could be annotated with Blast2GO. Most of these genes were involved, in descending order, in “cellular metabolic processes”, “organic substance metabolic processes”, “establishment of localization”, “primary metabolic processes” and “biosynthetic processes” (BP GO Level 3; **Figure 7**). The remaining 21 genes with no associated GO terms could not be functionally classified. Homologs of these genes were however, identified in other members of the *Enterobacteriaceae*, often pathogens, with a single gene having its closest homolog in the metagenome of a soil sample from an unknown source (**Supplementary File S6**).

Of the 98 genes conserved within the *P. rodasii* lineage without KEGG annotations, 76 could be annotated with Blast2GO. The five highest biological processes in which these genes were involved were “organic substance metabolic processes”, “cellular metabolic processes”, “primary metabolic processes”,

“regulation of cellular processes” and “nitrogen compound metabolic processes” (GO Level 3; **Figure 7**). This resulted in 22 genes without any functional annotation with either KEGG or GO analyses. Homologs for all 22 genes were identified in other members of the *Enterobacteriaceae*, of which 21 genes were most closely related to genes from human pathogens (**Supplementary File S6**).

Of the 29 unique genes in the *P. dispersa* lineage without any KEGG annotations, 14 genes could be annotated with Blast2GO. These genes were primarily involved in “cellular metabolic processes”, “organic substance metabolic processes”, “primary metabolic processes”, “nitrogen compound metabolic processes” and “biosynthetic processes” (**Figure 7**). Of the 15 unannotated genes, homologs for all genes were identified in other members of the *Enterobacteriaceae*, particularly those associated with the stinkbug, *Plautia stali* (**Supplementary File S6**).



## DISCUSSION

Our findings suggest that *in silico* mining of bacterial genome sequences is a feasible approach for inferring large sets of biological characters for particular taxa. This approach is invaluable for unveiling large repertoires of potential bacterial phenotypes and can thus contribute hugely toward identifying biologically relevant diagnostic characteristics from whole genome sequences. Furthermore, by superimposing such characters onto the phylogeny of a particular bacterial group it appears to be possible to identify those traits that might have contributed toward the initial emergence of a taxon and/or its subsequent stable persistence in nature. Here we identified extensive sets of biological characters specific to *Pantoea* and its main phylogenetic lineages. Our study thus outlines the initial steps toward linking biological functions (based on the variable genomic components) to taxonomy (based on the stable, conserved genomic components).

## Genome-Based Comparisons of Specific Processes between Genera

The methodology employed in this study allowed for the identification of biological characters that potentially define and differentiate bacterial genera from one another. Despite the necessity of these taxonomic ranks, our understanding of what constitutes and distinguishes genera is mostly limited. Previous attempts to obtain natural and logical groupings have always been based on a limited view of the organisms' metabolic potential, often with a focus on what was considered to be clinically relevant data rather than from a biological outlook (Konstantinidis and Tiedje, 2005a,b; Sutcliffe, 2015). Although, the current classification system aims to identify and describe naturally occurring groups by employing an evolution-based approach, it still does not provide any biologically meaningful information for the organisms (Cohan, 2002; Konstantinidis and Tiedje, 2005a; Tindall et al., 2010). However, our study of *Pantoea* and *Tatumella* clearly highlights how diverse sets of biological characters for bacterial genera may be inferred

from genome data. Apart from so-called genus-defining traits that can potentially be used to differentiate these taxa, these characters also provide information on the general biology of the taxa investigated (see **Table 3**). Our findings indicate that such genome-based analyses provide a more informed view of the biology of the organisms, and the information emerging from comparing metabolic differences can be linked to the shared ancestry of groups of organisms.

*Pantoea* appears to be metabolically more versatile than its sister genus *Tatumella*. Different from *Tatumella*, it encodes a range of additional pathways potentially enabling it to use diverse compounds as nutrient sources [e.g., fatty acids (Schulz, 1991; Fujita et al., 2007; Liu et al., 2010) and various carbohydrate derivatives (Mehler and Tabor, 1953; Anderson and Magasanik, 1971; Berkowitz, 1971; Yoshimoto et al., 1976; Baecker et al., 1986; Fouet et al., 1986; Deeg et al., 1987; Amemura et al., 1988; Sprenger, 1993; Van Beers et al., 1995; Boer et al., 1996; Schwede et al., 1999; Yoshida et al., 2004; Sim et al., 2008; Chandra et al., 2011; Yang et al., 2011)]. *Pantoea* also encodes additional nutrient cycling or salvage systems [e.g., purine, pyrimidine and co-factor cycling (Kimiyooshi et al., 1993; Colloch et al., 1997; Giorgelli et al., 1997; Nygaard et al., 2000; Moffatt and Ashihara, 2002; Yang et al., 2003; Ollagnier-de Choudens et al., 2005; Katoh et al., 2006; Cendron et al., 2007; Gossmann et al., 2012; Armenta-Medina et al., 2014)]. These systems have been shown to allow for the recycling of compounds that are no longer utilized in the cell, and might enhance *Pantoea*'s ability to perform basic, yet essential cellular functions under nutrient limiting conditions (Krismer et al., 2014; Shimizu, 2014).

*Pantoea* and *Tatumella* differ markedly in terms of KEGG's "Environmental Information Processing" functional category, which includes all signaling and membrane transport pathways (Kanehisa et al., 2002). Various ABC transporters (Nohno et al., 1986; Scripture et al., 1987; Gowrishankar, 1989; Stirling et al., 1989; Kehres and Hogg, 1992; Kempf and Bremer, 1995; Van der Ploeg et al., 1996; Walshaw et al., 1997; Kappes et al., 1999; Ko and Smith, 1999; Hosie and Poole, 2001; Schneider, 2001; Vanneste et al., 2001; Suzuki et al., 2005; Javaux et al., 2007; Novikova et al., 2007; Metlitskaya et al., 2009) and phosphotransferase systems (PTSs) (Hall and Xu, 1992; Dahl et al., 2004; Uehara et al., 2006; Jaeger and Mayer, 2008; Plumbridge, 2009) were differentially present in the two genera. Among the various ABC transporters identified only in *Pantoea*, one has been associated with susceptibility to microcin C in the absence of a microcin C-specific efflux pump (Metlitskaya et al., 2009), which is part of a group of antibiotic produced by certain *Enterobacteriaceae* (Vanneste et al., 2001; Metlitskaya et al., 2009). The absence of this ABC transporter in *Tatumella* and the concomitant antibiotic resistance may increase ecological competitiveness of species exposed to these compounds (Hacker and Carniel, 2001). Ecological advantages are likely also obtained from some of the predicted PTSs, which have previously been linked to enhanced recycling of cell wall components under nutrient-poor conditions (e.g., PTSs involving N-acetylmuramic acid and N-acetylglucosamine; Jaeger and Mayer, 2008), and the uptake of plant-derived carbon compounds (e.g., PTS involving arbutin and salicin; Zangoui et al., 2015).

Taken together, these findings suggest that evolution has equipped *Pantoea* with extensive repertoires of metabolic processes that make them generally more versatile in their ability to adapt to changing environments. Compared to *Tatumella*, they can utilize a wider range of carbon sources and use available resources more efficiently by recycling metabolic byproducts. This potentially also provides them with a competitive advantage in nutrient-poor environments such as mammalian blood. The various genus-defining traits we identified for *Pantoea* may thus contribute to our understanding of the complex, and often opportunistic, relationships these species have with their plant and animal hosts (De Baere et al., 2004; Cruz et al., 2007; De Maayer et al., 2012a).

## Genome-Based Comparisons of Specific Processes between Lineages of *Pantoea*

Comparisons of the metabolic processes inferred from whole genome sequences allowed for the identification of various sets of traits specific to one or more lineages of *Pantoea*. Based on previous work in diverse bacteria (including *Pantoea*), we attempted to correlate these processes to the lifestyles of the taxa investigated. Although a number of the identified processes were likely related to pathogenicity (see **Table 4**), most probably play roles in niche adaptation and utilization in a non-pathogenic capacity (see **Table 5**).

The processes likely associated with pathogenesis, particularly in the *Enterobacteriaceae*, were those involved in O-antigen (Stevenson et al., 1994; Whitfield, 1995; Wang and Reeves, 1998; Kohchi et al., 2006; Greenfield and Whitfield, 2012), siderophore (Montgomerie et al., 1984; Williams and Carbonetti, 1986; Opal et al., 1990; Fecteau et al., 2001; Fiedler et al., 2001; Torres et al., 2001; Hubertus et al., 2003; Raymond et al., 2003; Garcia et al., 2011; Gao et al., 2012) and polyamine (Khan et al., 1992; Ha et al., 1998; Gugliucci and Menini, 2003; Shah and Swiatlo, 2008; Pegg, 2016) biosynthesis. Differences in the locus involved in O-antigen biosynthesis (of which some *Pantoea* species have two) were previously associated with a pathogen's ability to escape host responses (Whitfield, 1995; Kohchi et al., 2006; Greenfield and Whitfield, 2012). Our results further showed that all species in the *P. ananatis* lineage likely produce aerobactin, while many of those in the other lineages produce enterobactin. Although these siderophores essentially perform the same function, aerobactin is more efficient at scavenging iron during nutrient limitation, and may in some instances even assist in resistance against iron-dependent antimicrobials (Williams and Carbonetti, 1986; Torres et al., 2001; Garcia et al., 2011). Also, all of the examined species in the *P. agglomerans* and *P. ananatis* lineages are predicted to be capable of producing the polyamine spermidine (this process was detected in only some of the species in the other two lineages). Apart from their essential cellular functions (Ha et al., 1998; Shah and Swiatlo, 2008; Pegg, 2016), polyamines have been implicated in biofilm formation, escape from host phagolysosomes and toxin production and activity (Shah and Swiatlo, 2008).

The group of processes likely associated with niche adaptation and utilization were those related to nutrient metabolism and "Environmental Information Processing"

**TABLE 4** | Pathogenicity-associated processes with differences between the lineages.

Process	Gene cluster	Function, biological role and contribution to pathogenicity	Distribution across lineages <sup>a</sup>	References
O-antigen biosynthesis	<i>rfb</i> locus 1	<ul style="list-style-type: none"> <li>• Encodes for the synthesis of polysaccharide unit of O-antigen</li> <li>• Cause variation on outer membrane of cells</li> <li>• Assists in host immune response evasion</li> </ul>	<i>P. ananatis</i> lineage (2/4) <i>P. agglomerans</i> lineage (5/7) <i>P. rodasii</i> lineage (2/3) <i>P. dispersa</i> lineage (2/3)	Stevenson et al., 1994; Whitfield, 1995; Wang and Reeves, 1998; Kohchi et al., 2006; Greenfield and Whitfield, 2012
	<i>rfb</i> locus 2	<ul style="list-style-type: none"> <li>• Encodes for the synthesis of the O-antigen's polysaccharide</li> <li>• Cause variation on outer membrane of cells</li> <li>• Assists in host immune response evasion</li> </ul>	<i>P. ananatis</i> lineage (3/4) <i>P. dispersa</i> lineage (1/4)	Stevenson et al., 1994; Whitfield, 1995; Wang and Reeves, 1998; Kohchi et al., 2006; Greenfield and Whitfield, 2012
Siderophore production	<i>iuc</i> locus	<ul style="list-style-type: none"> <li>• Encodes for the synthesis of aerobactin</li> <li>• Efficient at scavenging iron during nutrient limitation</li> <li>• Assists in resistance against iron-dependent antimicrobials</li> </ul>	<i>P. ananatis</i> lineage (4/4)	Montgomerie et al., 1984; Williams and Carbonetti, 1986; Opal et al., 1990; Fecteau et al., 2001; Torres et al., 2001; Garcia et al., 2011; Gao et al., 2012
	<i>ent</i> locus	<ul style="list-style-type: none"> <li>• Encodes for the synthesis of enterobactin</li> <li>• High affinity for iron, but susceptible to inactivation by lipocalin-2 produced by mammalian hosts</li> <li>• Iron acquisition in hosts</li> </ul>	<i>P. agglomerans</i> lineage (5/7) <i>P. rodasii</i> lineage (3/3) <i>P. dispersa</i> lineage (2/3)	Fiedler et al., 2001; Torres et al., 2001; Hubertus et al., 2003; Raymond et al., 2003; Garcia et al., 2011
Polyamine biosynthesis	<i>spe</i> locus	<ul style="list-style-type: none"> <li>• Synthesis of spermidine from S-adenosyl-L-methionine</li> <li>• Involved in protein and nucleic acid biosynthesis, expression regulation, membrane functioning and ROS scavenging</li> <li>• Contributes toward pathogenesis through involvement in biofilm formation, escape from phagolysosomes and toxin production</li> </ul>	<i>P. ananatis</i> lineage (4/4) <i>P. agglomerans</i> lineage (7/7) <i>P. rodasii</i> lineage (2/3) <i>P. dispersa</i> lineage (1/3)	Khan et al., 1992; Ha et al., 1998; Gugliucci and Menini, 2003; Shah and Swiatlo, 2008; Pegg, 2016

<sup>a</sup>Brackets indicate the number of taxa for which the specific locus is present out of all taxa in a lineage. For complete distribution patterns see **Supplementary File S4**.

(**Supplementary Table S1**). For example, other than those in the *P. dispersa* lineage, all *Pantoea* species were predicted to be capable of converting the environmental mutagen chloroacetaldehyde to glycolate, thus providing the dual means of disposing of the mutagen and accessing glycolate as carbon source (Young Kim et al., 2007; Maciejewska et al., 2010). Similarly, in all species, bar those of the *P. dispersa* lineage, the methionine salvage pathway likely allow increased efficacy under sulfur cycling in nutrient-poor conditions (Sekowska et al., 2000, 2004; Albers, 2009). Species in the *P. ananatis* lineage encode the benzoate degradation pathways, which likely enable their utilization of protocatechuate as carbon source (Song, 2009; Brady et al., 2011; Gueule et al., 2015). Additionally, differences were also observed in rhamnose (Badía et al., 1985; Moralejo et al., 1993; Saxena et al., 2010) and galacturonate (Walton, 1994; Hématy et al., 2009; Richard and Hilditch, 2009) utilization. The four lineages further differed in terms of their ability to transport various nutrients (e.g., the myo-inositol-1-phosphate ABC transporter occurred only in the *P. ananatis* lineage). The same was also true for the predicted two-component signaling systems and other PTSs (e.g., except

for two *P. agglomerans* lineage species, only the *P. ananatis* lineage encoded a two-component signaling system for citrate utilization) (**Supplementary Table S1**).

Overall, we could correlate the versatility in lifestyle and host to increased metabolic potential (in terms of compounds utilized) as well as pathogen-associated traits between the different lineages within *Pantoea*. Despite the association of various *Pantoea* species with only clinical infections (*P. brenneri*, *P. conspicua*, and *P. eucrinea*), it appears that species associated with opportunistic clinical infections (De Baere et al., 2004; Cruz et al., 2007), possessed genes often associated with animal pathogenicity within the *Enterobacteriaceae*. In general, it appears that the lineages with the most diverse niche associated processes also corresponded to those species isolated from the most diverse environments. For example, members of the *P. agglomerans* and *P. ananatis* lineages are routinely isolated from various plant species, as epiphytes, endophytes, or pathogens, as well as from insects, animals and humans (Coutinho and Venter, 2009; Walterson and Stavrinides, 2015), while members of the *P. dispersa* and *P. rodasii* lineages are usually only associated with a single host organism, with the exception of *P. dispersa*.

**TABLE 5** | Niche-associated processes (non-pathogenic) with differences between the lineages.

Process	Gene cluster	Function, biological role and contribution to niche adaptation	Distribution across lineages <sup>a</sup>	References
L-rhamnose utilization	<i>rha</i> locus	<ul style="list-style-type: none"> <li>Convert L-rhamnose to S-lactaldehyde</li> <li>Catabolism of L-rhamnose as carbon source</li> <li>Potentially contributes toward niche expansion</li> </ul>	<ul style="list-style-type: none"> <li><i>P. ananatis</i> lineage (2/4)</li> <li><i>P. agglomerans</i> lineage (7/7)</li> <li><i>P. rodasii</i> lineage (3/3)</li> <li><i>P. dispersa</i> lineage (2/3)</li> </ul>	Badía et al., 1985; Moralejo et al., 1993; Saxena et al., 2010
D-galacturonate utilization	<i>uxa</i> locus	<ul style="list-style-type: none"> <li>Involved in the utilization of glucuronate and galacturonate as carbon sources</li> <li>Integral parts of plant cell walls</li> <li>Can contribute toward colonization of plants</li> </ul>	<ul style="list-style-type: none"> <li><i>P. ananatis</i> lineage (4/4)</li> <li><i>P. agglomerans</i> lineage (7/7)</li> <li><i>P. rodasii</i> lineage (3/3)</li> <li><i>P. dispersa</i> lineage (2/3 partial)</li> </ul>	Walton, 1994; Hématy et al., 2009; Richard and Hilditch, 2009
Benzoate degradation	<i>pca</i> locus	<ul style="list-style-type: none"> <li>Involved in the utilization of protocatechuate</li> <li>Catabolism of protocatechuate as carbon source</li> <li>Potentially contributes toward niche expansion</li> </ul>	<ul style="list-style-type: none"> <li><i>P. ananatis</i> lineage (4/4)</li> <li>(note, <i>P. stewartii</i> ssp. <i>stewartii</i> contains deletion)</li> </ul>	Song, 2009; Brady et al., 2011; Gueule et al., 2015
Chloroalkane and chloroalkene degradation	-	<ul style="list-style-type: none"> <li>Conversion of chloroacetaldehyde to glycolate</li> <li>Chloroacetaldehyde binds to DNA and causes conformational changes leading to mutations</li> <li>Removes mutagen from environment and utilize it as carbon source via glycolate</li> </ul>	<ul style="list-style-type: none"> <li><i>P. ananatis</i> lineage (4/4)</li> <li><i>P. agglomerans</i> lineage (7/7)</li> <li><i>P. rodasii</i> lineage (3/3)</li> </ul>	Young Kim et al., 2007; Maciejewska et al., 2010
Methionine salvage pathway	<i>mtn</i> locus	<ul style="list-style-type: none"> <li>Salvage methionine through the conversion of 5-methylthio-D-ribose to 3-(methylthio)-propanoate</li> <li>Allows sulfur cycling</li> <li>Can survive in low-sulfur environments</li> </ul>	<ul style="list-style-type: none"> <li><i>P. ananatis</i> lineage (4/4)</li> <li><i>P. agglomerans</i> lineage (7/7)</li> <li><i>P. rodasii</i> lineage (3/3)</li> </ul>	Sekowska et al., 2000, 2004; Albers, 2009

<sup>a</sup>Brackets indicate the number of taxa for which the specific locus is present out of all taxa in a lineage. For complete distribution patterns see **Supplementary File S4**.

These characteristics may contribute to the opportunistic nature of lineages containing species like *P. ananatis* and *P. agglomerans* that are proven plant pathogens, but isolated from diverse environments including the clinical setting.

## Evolution of Multi-gene Pathways in *Pantoea* and Its Lineages

Various evolutionary mechanisms likely shaped the presence and distribution of the multi-gene pathways inferred for *Pantoea*. Bacteria propagate asexually and progeny are anticipated to contain the same genetic material as the parent (Daubin et al., 2002). Any changes in an individual's genetic makeup can become fixed in populations if they provide a competitive advantage, or at least have no deleterious effects (Cohan, 2001, 2005; Caro-Quintero and Konstantinidis, 2012). The most common forces facilitating genetic change are random mutations (point mutations as well as insertions and deletions) and horizontal gene transfer (HGT) (Gogarten et al., 2002; Cohen et al., 2011). Accordingly, more closely related species are likely to encode similar pathways, while those subject to HGT would have a more spurious distribution (Gogarten et al., 2002; Cohen et al., 2011).

Evolution of most of the multi-gene pathways in *Pantoea* and its lineages involved complex processes, involving vertical descent with lineage- and species-specific gene losses/gains via

duplication or HGT. For example, lineage- and species-specific gene losses would be characterized by distribution patterns where particular gene clusters are present in all taxa neighboring a species lacking it, because of gene losses at a specific ancestral node (Koskiniemi et al., 2012). In our study, such processes were likely involved in the loci required for the conversion of chloroacetaldehyde to glycolate and in the methionine salvage pathway. In contrast to this, the sudden appearance of genes or loci in a lineage that are lacking in all neighboring taxa suggest they were acquired via HGT (Zaneveld et al., 2008). An example of this is the locus required for protocatechuate utilization. The evolution of the siderophore loci likely involved gene losses together with the horizontal acquisition of genes. This is evident from the absence of the enterobactin locus (despite its presence in all closely related lineages) from the *P. ananatis* lineage, and the presence of the locus encoding aerobactin in only this lineage, thus acquired through an HGT event.

Our results suggested that the gene clusters encoding the various *Pantoea* pathways examined in this study, mainly experience purifying selection. Numerous hypotheses attempt to explain why genes cluster and how clusters are maintained (Carbone et al., 2007; Geddy and Brown, 2007; Fang et al., 2008; Sorrels et al., 2009). However, various studies have showed that purifying selection may contribute to the stability and functionality of gene clusters once they have formed (Carbone

et al., 2007; Geddy and Brown, 2007; Fang et al., 2008; Sorrels et al., 2009). Purifying selection seems to facilitate the maintenance of ancient or ancestral gene clusters by limiting the possibility of non-synonymous mutations becoming fixed, which in turn allows out-competition of individuals undergoing deleterious or lethal mutations inhibiting the functioning of the gene cluster (Fang et al., 2008; Sorrels et al., 2009). In our study, purifying selection appeared to play a role in the maintenance and functioning of the loci for siderophore biosynthesis (aerobactin and enterobactin loci) and protochatechuate utilization, which are evolving mainly under the influence of purifying selection. However, we also detected diversifying selection for some of the genes/gene regions examined [e.g., *rfbC* and *rfbD* (*rfb* locus 1), both associated with the conversion of dTDP-6-deoxy-D-xylo-4-hexulose to dTDP-L-rhamnose (Graninger et al., 1999)]. In these cases, selection is causing non-synonymous changes in the sequences of genes or parts of genes, thus driving the appearance of new alleles.

### ***In Silico* Predictions vs. Experimentally Confirmed Multi-gene Processes**

Although a large number of differential characters were identified from the genome sequences of these organisms, little has been done in terms of experimental verification. Also, a number of characters could not be correlated to current experimental knowledge as false negative and positive results for phenotypic tests are common (Sutcliffe et al., 2012). For instance, despite the presence of genes required for the utilization of arbutin and salicin as carbon source in *Tatumella*, previous phenotypic tests have previously tested negative (Brady et al., 2010b; Tracz et al., 2015). This was also observed for the intra-generic comparisons. Examples within *Pantoea* included the uptake and utilization of cellobiose (Brady et al., 2009) and D-galacturonate, previously identified as a genus-defining attribute (Gavini et al., 1989; Brady et al., 2010a), as well as citrate (Brady et al., 2010a, 2012). In these cases, phenotypic tests previously confirmed the ability to perform these functions, but these genes were lacking in various *Pantoea* species (absence confirmed in available genome sequences of additional isolates of these species). This lack in correlation may either be as a result of gene expression complexities during phenotypic tests, to sequencing, assembly or annotation errors or the presence of as of yet uncharacterized alternate pathways.

There were, however, also a number of *in silico* functional predictions that correlated well with the results of previous phenotypic tests. For example, the lack of genes required for the utilization of histidine in *Tatumella*, correlated entirely with negative results of all previous phenotypic tests (Brady et al., 2010b; Tracz et al., 2015). Moreover, the utilization of protochatechuate has previously tested positive in *P. allii* (Brady et al., 2011), and our findings showed that the locus encoding the necessary gene products is indeed present within this lineage. Other examples of characters supporting previously performed phenotypic tests are the transport for the utilization of D-xylose, maltose, myo-inositol-1-P, and sucrose. In all of the above mentioned examples gene clusters were only observed in

taxa previously testing positive for the associated phenotypic characters (Brady et al., 2010a,b, 2011).

### **Perspectives and Relevance**

The increase in data obtained from bacterial genome sequences has superseded the rate at which gene discovery, characterization, and verification can occur. This means that a number of genes could not be assigned definitive functions as no homologs were detected for these genes in the KEGG database. The majority of these genes, at both the generic and lineage-specific level, could be annotated with gene ontologies with Blast2GO, although the functions of some genes remained unknown. Of the uncharacterized genes, the most abundant GO terms were generally quite similar (data for genera not shown). This indicates that the unknown genes could be performing similar functions within the different lineages or genera, despite not being orthologous amongst the different gene sets. The genes to which no functional annotation could be assigned seemed to have originated mostly within the *Enterobacteriaceae*, either acquired through lateral acquisition of genes from other taxa within this family, or, in the case where genes are conserved in most of the lineages, acquired by an ancestor and subsequently lost in some lineages when the genes were no longer required. A number of unique genes were also identified with no known homologs, but expression of these genes should be confirmed before they are considered for further investigation. Although no functional information is available for the unannotated genes, the distribution across various taxa provides insight into potential HGT events.

The availability of whole genome sequence data has transformed the approaches available for understanding bacterial evolution. This has, however, not yet replaced traditional methods such as DNA-DNA hybridization, monophyly in phylogenetic analyses and physiological and metabolic tests for the delineation of bacterial taxa. Although physiologic capabilities provide an insight into what potential metabolic differentiation may have occurred during speciation, inconsistencies may still occur due to differential expression of genes in different isolates or regulation of expression in specific environments. In recent years, we have been moving toward approaches aiming to identify natural groups at higher taxonomic levels by implementing monophyly as a prerequisite for taxon descriptions, but no light can be shed on the biology of the organisms through these approaches. Instead we need to investigate the more variable genomic compartments reflecting the biology of the organisms to obtain a more natural and robust taxonomic system. By employing this approach, one would be able to supplement or supplant the available diagnostic characters used in bacterial taxonomy. Additionally, obtaining a holistic biological perspective from the genome will provide power to predict the lifestyle and ecology of the organisms and is essentially much more informative than only having discriminative power between taxa. We thus believe that this approach of identifying genome-based characteristics in metabolic networks for the taxonomic levels higher than the species, provide an approach of identifying biologically relevant differences along the course of speciation.

## AUTHOR CONTRIBUTIONS

All authors contributed toward the original concept or data analyses, together with the drafting, revision and approval of the final manuscript. MP, ES, MC, and SV: was involved in the conceptualization and design of the work; MP and JB: was involved in data acquisition and analysis and all authors were involved in the interpretation of data.

## FUNDING

The authors would like to acknowledge the National Research Foundation for student funding through the Centre of Excellence in Tree Health Biotechnology (CTHB) and the University of Pretoria.

## ACKNOWLEDGMENTS

We would like to acknowledge the Centre of Excellence in Tree Health Biotechnology (CTHB) and the Tree Protection Co-operative Programme (TPCP) at the Forestry and Agricultural Biotechnology Institute (FABI), University of Pretoria, for access to computing infrastructure.

## REFERENCES

- Albers, E. (2009). Metabolic characteristics and importance of the universal methionine salvage pathway recycling methionine from 5'-methylthioadenosine. *IUBMB Life* 61, 1132–1142. doi: 10.1002/iub.278
- Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol.* 215, 403–410. doi: 10.1016/S0022-2836(05)80360-2
- Amemura, A., Chakraborty, R., Fujita, M., Noumi, T., and Futai, M. (1988). Cloning and nucleotide sequence of the isoamylase gene from *Pseudomonas amyloclavata* SB-15. *J. Biol. Chem.* 263, 9271–9275.
- Anderson, W. A., and Magasanik, B. (1971). The pathway of myo-inositol degradation in *Aerobacter aerogenes*: conversion of 2-deoxy-5-keto-D-gluconic acid to glycolytic intermediates. *J. Biol. Chem.* 246, 5662–5675.
- Armenta-Medina, D., Segovia, L., and Perez-Rueda, E. (2014). Comparative genomics of nucleotide metabolism: a tour to the past of the three cellular domains of life. *BMC Genomics* 15:800. doi: 10.1186/1471-2164-15-800
- Badía, J., Ros, J., and Aguilar, J. (1985). Fermentation mechanism of fucose and rhamnose in *Salmonella typhimurium* and *Klebsiella pneumoniae*. *J. Bacteriol.* 161, 435–437.
- Baecker, P. A., Greenberg, E., and Preiss, J. (1986). Biosynthesis of bacterial glycogen. Primary structure of *Escherichia coli* 1,4- $\alpha$ -D-glucan:1,4- $\alpha$ -D-glucan 6- $\alpha$ -D-(1, 4- $\alpha$ -D-glucano)-transferase as deduced from the nucleotide sequence of the *glg B* gene. *J. Biol. Chem.* 261, 8738–8743.
- Bennett, J. S., Jolley, K. A., Earle, S. G., Corton, C., Bentley, S. D., Parkhill, J., et al. (2012). A genomic approach to bacterial taxonomy: an examination and proposed reclassification of species within the genus *Neisseria*. *Microbiology* 158, 1570–1580. doi: 10.1099/mic.0.056077-0
- Berkowitz, D. (1971). D-Mannitol utilization in *Salmonella typhimurium*. *J. Bacteriol.* 105, 232–240.
- Beukes, C. W., Palmer, M., Manyaka, P., Chan, W. Y., Avontuur, J. R., van Zyl, E., et al. (2017). Genome Data provides high support for generic boundaries in burkholderia sensu lato. *Front. Microbiol.* 8:1154. doi: 10.3389/fmicb.2017.01154
- Blom, J., Kreis, J., Spänig, S., Juhre, T., Bertelli, C., Ernst, C., et al. (2016). EDGAR 2.0: an enhanced software platform for comparative gene

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmicb.2018.00113/full#supplementary-material>

**Supplementary Table S1** | Environmental information processing traits with differences among the lineages.

**Supplementary File S1** | Output from GhostKOALA for the *Pantoea* and *Tatumella* shared gene sets. This file contains the differences observed for the overview maps and the specific pathways for *Pantoea* and *Tatumella*. The overlay figures of the overview and pathway maps are also indicated.

**Supplementary File S2** | Output from GhostKOALA for the lineages within *Pantoea*. This file contains the differences observed for the overview maps as well as the overlay figures of the overview maps for the different lineages.

**Supplementary File S3** | The summary of differences in pathways requiring 2 or more genes, as well as a summary of the BLAST confirmations of these genes.

**Supplementary File S4** | The results from the selection analyses and the figures for the gene clusters not indicated in text.

**Supplementary File S5** | The differences for pathways involved in “Environmental Information Processing”. A summary of the BLAST confirmation is also included as well as the maps for each lineage for the ABC transporters, two-component systems and the PTSs.

**Supplementary File S6** | Summary of the Blast2GO analyses as well as the BLAST hits for genes not annotated with Blast2GO. A pie chart indicating the distribution of BLAST hits is also indicated.

content analyses. *Nucleic Acids Res.* 44, W22–W28. doi: 10.1093/nar/gkw255

- Boer, H., ten Hoeve-Duurkens, R. H., and Robillard, G. T. (1996). Relation between the oligomerization state and the transport and phosphorylation function of the *Escherichia coli* mannitol transport protein: interaction between mannitol-specific enzyme II monomers studied by complementation of inactive site-directed mutants. *Biochemistry* 35, 12901–12908. doi: 10.1021/bi9611016
- Brady, C. L., Cleenwerck, I., van der Westhuizen, L., Venter, S. N., Coutinho, T. A., and De Vos, P. (2012). *Pantoea rodassii* sp. nov., *Pantoea rwandensis* sp. nov. and *Pantoea wallisii* sp. nov., isolated from *Eucalyptus*. *Int. J. Syst. Evol. Microbiol.* 62, 1457–1464. doi: 10.1099/ijs.0.032615-0
- Brady, C. L., Cleenwerck, I., Venter, S. N., Engelbeen, K., De Vos, P., and Coutinho, T. A. (2010a). Emended description of the genus *Pantoea*, description of four species from human clinical samples, *Pantoea septica* sp. nov., *Pantoea eucrinea* 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.* 60, 2430–2440. doi: 10.1099/ijs.0.017301-0
- Brady, C. L., Goszczynska, T., Venter, S. N., Cleenwerck, I., De Vos, P., Gitaitis, R. D., et al. (2011). *Pantoea allii* sp. nov., isolated from onion plants and seed. *Int. J. Syst. Evol. Microbiol.* 61, 932–937. doi: 10.1099/ijs.0.022921-0
- Brady, C. L., Venter, S. N., Cleenwerck, I., Engelbeen, K., Vancanneyt, M., Swings, J., et al. (2009). *Pantoea vagans* sp. nov., *Pantoea eucalypti* sp. nov., *Pantoea deleyi* sp. nov. and *Pantoea anthophila* sp. nov. *Int. J. Syst. Evol. Microbiol.* 59, 2339–2345. doi: 10.1099/ijs.0.009241-0
- Brady, C. L., Venter, S. N., Cleenwerck, I., Vandemeulebroecke, K., De Vos, P., and Coutinho, T. A. (2010b). 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.* 60, 484–494. doi: 10.1099/ijs.0.012070-0
- Brady, C., Cleenwerck, I., Venter, S., Coutinho, T., and De Vos, P. (2013). Taxonomic evaluation of the genus *Enterobacter* based on multilocus sequence analysis (MLSA): proposal to reclassify *E. nimipressuralis* and *E. amnigenus* into *Lelliottia* gen. nov. as *Lelliottia nimipressuralis* comb. nov. and *Lelliottia amnigena* comb. nov., respectively, *E. gergoviae* and *E. pyrinus* into

- Pluralibacter* gen. nov. as *Pluralibacter gergoviae* comb. nov. and *Pluralibacter pyrinus* comb. nov., respectively, *E. cowanii*, *E. radiccincitans*, *E. oryzae* and *E. arachidis* into *Kosakonia* gen. nov. as *Kosakonia cowanii* comb. nov., *Kosakonia radiccincitans* comb. nov., *Kosakonia oryzae* comb. nov. and *Kosakonia arachidis* comb. nov., respectively, and *E. turicensis*, *E. helveticus* and *E. pulveris* into *Cronobacter* as *Cronobacter zurichensis* nom. nov., *Cronobacter helveticus* comb. nov. and *Cronobacter pulveris* comb. nov., respectively, and emended description of the genera *Enterobacter* and *Cronobacter*. *Syst. Appl. Microbiol.* 36, 309–319. doi: 10.1016/j.syapm.2013.03.005
- Carbone, I., Ramirez-Prado, J. H., Jakobek, J. L., and Horn, B. W. (2007). Gene duplication, modularity and adaptation in the evolution of the aflatoxin gene cluster. *BMC Evol. Biol.* 7:111. doi: 10.1186/1471-2148-7-111
- Caro-Quintero, A., and Konstantinidis, K. T. (2012). Bacterial species may exist, metagenomics reveal. *Environ. Microbiol.* 14, 347–355. doi: 10.1111/j.1462-2920.2011.02668.x
- Cendron, L., Berni, R., Folli, C., Ramazzina, I., Percudani, R., and Zanotti, G. (2007). The structure of 2-Oxo-4-hydroxy-4-carboxy-5-ureidoimidazole decarboxylase provides insights into the mechanism of uric acid degradation. *J. Biol. Chem.* 282, 18182–18189. doi: 10.1074/jbc.M701297200
- Chan, J. Z. M., Halachev, M. R., Loman, N. J., Constantinidou, C., and Pallen, M. J. (2012). Defining bacterial species in the genomic era: insights from the genus *Acinetobacter*. *BMC Microbiol.* 12:302. doi: 10.1186/1471-2180-12-302
- Chandra, G., Chater, K. F., and Bornemann, S. (2011). Unexpected and widespread connections between bacterial glycogen and trehalose metabolism. *Microbiology* 157, 1565–1572. doi: 10.1099/mic.0.044263-0
- Cohan, F. M. (2001). Bacterial species and speciation. *Syst. Biol.* 50, 513–524. doi: 10.1080/10635150118398
- Cohan, F. M. (2002). What are bacterial species? *Annu. Rev. Microbiol.* 56, 457–487. doi: 10.1146/annurev.micro.56.012302.160634
- Cohan, F. M. (2005). “Periodic selection and ecological diversity in bacteria,” in *Selective Sweep*, ed D. Nurminsky (Georgetown, TX: Springer), 78–93.
- Cohen, O., Gophna, U., and Pupko, T. (2011). The complexity hypothesis revisited: connectivity rather than function constitutes a barrier to horizontal gene transfer. *Mol. Biol. Evol.* 28, 1481–1489. doi: 10.1093/molbev/msq333
- Colloch, N., Hajji, M. E., Bachet, B., L’Hermite, G., Schiltz, M., Prange, T., et al. (1997). Crystal structure of the protein drug urate oxidase-inhibitor complex at 2.05 Å resolution. *Nat. Struct. Mol. Biol.* 4, 947–952. doi: 10.1038/nsb1197-947
- Conesa, A., Götz, S., García-Gómez, J. M., Terol, J., Talón, M., and Robles, M. (2005). Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* 21, 3674–3676. doi: 10.1093/bioinformatics/bti610
- Coutinho, T. A., and Venter, S. N. (2009). *Pantoea ananatis*: an unconventional plant pathogen. *Mol. Plant Pathol.* 10, 325–335. doi: 10.1111/j.1364-3703.2009.00542.x
- Cruz, A. T., Cazacu, A. C., and Allen, C. H. (2007). *Pantoea agglomerans*, a plant pathogen causing human disease. *J. Clin. Microbiol.* 45, 1989–1992. doi: 10.1128/JCM.00632-07
- Dahl, U., Jaeger, T., Nguyen, B. T., Sattler, J. M., and Mayer, C. (2004). Identification of a phosphotransferase system of *Escherichia coli* required for growth on N-Acetylmuramic acid. *J. Bacteriol.* 186, 2385–2392. doi: 10.1128/JB.186.8.2385-2392.2004
- Daubin, V., Gouy, M., and Perrière, G. (2002). A phylogenomic approach to bacterial phylogeny: evidence of a core of genes sharing a common history. *Genome Res.* 12, 1080–1090. doi: 10.1101/gr.187002
- De Baere, T., Verhelst, R., Labit, C., Verschraegen, G., Wauters, G., Claeys, G., et al. (2004). Bacteremic infection with *Pantoea ananatis*. *J. Clin. Microbiol.* 42, 4393–4395. doi: 10.1128/JCM.42.9.4393-4395.2004
- De Maayer, P., Chan, W. Y., Rezzonico, F., Bühlmann, A., Venter, S. N., Blom, J., et al. (2012a). Complete genome sequence of clinical isolate *Pantoea ananatis* LMG5342. *J. Bacteriol.* 194, 1615–1616. doi: 10.1128/JB.06715-11.
- De Maayer, P., Chan, W.-Y., Blom, J., Venter, S. N., Duffy, B., Smits, T. H. M., et al. (2012b). The large universal *Pantoea plasmid* LPP-1 plays a major role in biological and ecological diversification. *BMC Genomics* 13:625. doi: 10.1186/1471-2164-13-625
- De Maayer, P., Chan, W.-Y., Rubagotti, E., Venter, S. N., Toth, I. K., Birch, P. R. J., et al. (2014). Analysis of the *Pantoea ananatis* pan-genome reveals factors underlying its ability to colonize and interact with plant, insect and vertebrate hosts. *BMC Genomics* 15:404. doi: 10.1186/1471-2164-15-404
- Deeg, R., Roder, A., Siedel, J., Gauhl, H., Ziegenhorn, J., and Boehringer Mannheim GmbH (1987). *Process and Reagent for the Determination of N-Carbamoylsarcosine with the Use of a New Enzyme*. U.S. Patent 4,645,739.
- Fang, G., Rocha, E. P., and Danchin, A. (2008). Persistence drives gene clustering in bacterial genomes. *BMC Genomics* 9:4. doi: 10.1186/1471-2164-9-4
- Fecteau, G., Fairbrother, J. M., Higgins, R., Van Metre, D. C., Paré, J., Smith, B. P., et al. (2001). Virulence factors in *Escherichia coli* isolated from the blood of bacteremic neonatal calves. *Vet. Microbiol.* 78, 241–249. doi: 10.1016/S0378-1135(00)00299-6
- Fiedler, H.-P., Krastel, P., Müller, J., Gebhardt, K., and Zeeck, A. (2001). Enterobactin: the characteristic catecholate siderophore of Enterobacteriaceae is produced by *Streptomyces* species. *FEMS Microbiol. Lett.* 196, 147–151. doi: 10.1111/j.1574-6968.2001.tb10556.x
- Fouet, A., Klier, A., and Rapoport, G. (1986). Nucleotide sequence of the sucrose gene of *Bacillus subtilis*. *Gene* 45, 221–225. doi: 10.1016/0378-1119(86)90258-1
- Fujita, Y., Matsuoka, H., and Hirooka, K. (2007). Regulation of fatty acid metabolism in bacteria. *Mol. Microbiol.* 66, 829–839. doi: 10.1111/j.1365-2958.2007.05947.x
- Gao, Q., Wang, X., Xu, H., Xu, Y., Ling, J., Zhang, D., et al. (2012). Roles of iron acquisition systems in virulence of extraintestinal pathogenic *Escherichia coli*: salmochelin and aerobactin contribute more to virulence than heme in a chicken infection model. *BMC Microbiol.* 12:143. doi: 10.1186/1471-2180-12-143
- Garcia, E. C., Brumbaugh, A. R., and Mobley, H. L. (2011). Redundancy and specificity of *Escherichia coli* iron acquisition systems during urinary tract infection. *Infect. Immun.* 79, 1225–1235. doi: 10.1128/IAI.01222-10
- Garrity, G. M., Bell, J. A., and Lilburn, T. (2005). “The revised road map to the manual,” in *Bergey’s Manual® of Systematic Bacteriology*, eds D. J. Brenner, N. R. Krieg, J. R. Staley, and G. M. Garrity (Boston, MA: Springer), 159–187.
- Gavini, F., Mergaert, J., Beji, A., Mielcarek, C., Izard, D., Kersters, K., et al. (1989). 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.* 39, 337–345. doi: 10.1099/00207713-39-3-337
- Geddy, R., and Brown, G. G. (2007). Genes encoding pentatricopeptide repeat (PPR) proteins are not conserved in location in plant genomes and may be subject to diversifying selection. *BMC Genomics* 8:30. doi: 10.1186/1471-2164-8-130
- Gevers, D., Cohan, F. M., Lawrence, J. G., Spratt, B. G., Coenye, T., Feil, E. J., et al. (2005). Re-evaluating prokaryotic species. *Nature Rev.* 3, 733–739. doi: 10.1038/nrmicro1236
- Giorgelli, F., Bottai, C., Mascia, L., Scolozzi, C., Camici, M., and Ipata, P. L. (1997). Recycling of  $\alpha$ -D-ribose 1-phosphate for nucleoside interconversion. *Biochim. Biophys. Acta* 1335, 16–22. doi: 10.1016/S0304-4165(96)00117-1
- Glaeser, S. P., and Kämpfer, P. (2015). Multilocus sequence analysis (MLSA) in prokaryotic taxonomy. *Syst. Appl. Microbiol.* 38, 237–245. doi: 10.1016/j.syapm.2015.03.007
- Gogarten, J. P., Doolittle, W. F., and Lawrence, J. G. (2002). Prokaryotic evolution in light of gene transfer. *Mol. Biol. Evol.* 19, 2226–2238. doi: 10.1093/oxfordjournals.molbev.a004046
- Gossmann, T. I., Ziegler, M., Puntervoll, P., de Figueiredo, L. F., Schuster, S., and Heiland, I. (2012). NAD<sup>+</sup> biosynthesis and salvage—a phylogenetic perspective. *FEBS J.* 279, 3355–3363. doi: 10.1111/j.1742-4658.2012.08559.x
- Götz, S., García-Gómez, J. M., Terol, J., Williams, T. D., Nagaraj, S. H., Nueda, M. J., et al. (2008). High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic Acids Res.* 36, 3420–3435. doi: 10.1093/nar/gkn176
- Gowrishankar, J. (1989). Nucleotide sequence of the osmoregulatory proU operon of *Escherichia coli*. *J. Bacteriol.* 171, 1923–1931. doi: 10.1128/jb.171.4.1923-1931.1989
- Granger, M., Nidetzky, B., Heinrichs, D. E., Whitfield, C., and Messner, P. (1999). Characterization of dTDP-4-dehydrorhamnose 3,5-Epimerase and dTDP-4-dehydrorhamnose reductase, required for dTDP-1-rhamnose biosynthesis in *Salmonella enterica* Serovar Typhimurium LT2. *J. Biol. Chem.* 274, 25069–25077. doi: 10.1074/jbc.274.35.25069
- Greenfield, L. K., and Whitfield, C. (2012). Synthesis of lipopolysaccharide O-antigens by ABC transporter-dependent pathways. *Carbohydr. Res.* 356, 12–24. doi: 10.1016/j.carres.2012.02.027

- Gueule, D., Fourny, G., Ageron, E., Le Flèche-Matéos, A., Vandenberg, M., Grimont, P. A. D., et al. (2015). *Pantoea coffeiphila* sp. nov., cause of the 'potato taste' of Arabica coffee from the African Great Lakes region. *Int. J. Syst. Evol. Microbiol.* 65, 23–29. doi: 10.1099/ijms.0.063545-0
- Gugliucci, A., and Menini, T. (2003). The polyamines spermine and spermidine protect proteins from structural and functional damage by AGE precursors: a new role for old molecules? *Life Sci.* 72, 2603–2616. doi: 10.1016/S0024-3205(03)00166-8
- Ha, H. C., Sirisoma, N. S., Kuppusamy, P., Zweier, J. L., Woster, P. M., and Casero, R. A. (1998). The natural polyamine spermine functions directly as a free radical scavenger. *Proc. Natl. Acad. Sci. U.S.A.* 95, 11140–11145. doi: 10.1073/pnas.95.19.11140
- Hacker, J., and Carniel, E. (2001). Ecological fitness, genomic islands and bacterial pathogenicity. *EMBO Rep.* 2, 376–381. doi: 10.1093/embo-reports/kve097
- Hall, B. G., and Xu, L. (1992). Nucleotide sequence, function, activation, and evolution of the cryptic asc operon of *Escherichia coli* K12. *Mol. Biol. Evol.* 9, 688–706.
- Hématy, K., Cherk, C., and Somerville, S. (2009). Host-pathogen warfare at the plant cell wall. *Curr. Opin. Plant Biol.* 12, 406–413. doi: 10.1016/j.pbi.2009.06.007
- Hillis, D. M., and Dixon, M. T. (1991). Ribosomal DNA: molecular evolution and phylogenetic inference. *Q. Rev. Biol.* 66, 411–453. doi: 10.1086/417338
- Hong, K.-W., Gan, H. M., Low, S.-M., Lee, P. K. Y., Chong, Y.-M., Yin, W.-F., et al. (2012). Draft genome sequence of *Pantoea* sp. strain A4, a Rafflesia-associated bacterium that produces N-acylhomoserine lactones as quorum-sensing molecules. *J. Bacteriol.* 194:6610. doi: 10.1128/JB.01619-12
- Hosie, A. H. F., and Poole, P. S. (2001). Bacterial ABC transporters of amino acids. *Res. Microbiol.* 152, 259–270. doi: 10.1016/S0923-2508(01)01197-4
- Hubertus, H., Schoeser, M., Lesuisse, E., Ernst, J. F., Parson, W., Beate, A., et al. (2003). Characterization of the *Aspergillus nidulans* transporters for the siderophores enterobactin and triacetylfulvarinone C. *Biochem. J.* 371, 505–513. doi: 10.1042/bj20021685
- Jaeger, T., and Mayer, C. (2008). N-acetylmuramic acid 6-phosphate lyases (MurNAc etherases): role in cell wall metabolism, distribution, structure, and mechanism. *Cell. Mol. Life Sci.* 65, 928–939. doi: 10.1007/s00018-007-7399-x
- Javaux, C., Joris, B., and De Witte, P. (2007). Functional Characteristics of TauA binding protein from TauABC *Escherichia coli* system. *Protein J.* 26, 231–238. doi: 10.1007/s10930-006-9064-x
- Jones, P., Binns, D., Chang, H.-Y., Fraser, M., Li, W., McAnulla, C., et al. (2014). InterProScan 5: genome-scale protein function classification. *Bioinformatics* 30, 1236–1240. doi: 10.1093/bioinformatics/btu031
- Kamber, T., Smits, T. H. M., Rezzonico, F., and Duffy, B. (2012). Genomics and current genetic understanding of *Erwinia amylovora* and the fire blight antagonist *Pantoea vagans*. *Trees* 26, 227–238. doi: 10.1007/s00468-011-0619-x
- Kanehisa, M., Goto, S., Kawashima, S., and Nakaya, A. (2002). The KEGG databases at GenomeNet. *Nucleic Acids Res.* 30, 42–46. doi: 10.1093/nar/30.1.42
- Kanehisa, M., Sato, Y., and Morishima, K. (2016a). BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. *J. Mol. Biol.* 428, 726–731. doi: 10.1016/j.jmb.2015.11.006
- Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M., and Tanabe, M. (2016b). KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res.* 44, D457–D462. doi: 10.1093/nar/gkv1070
- Kappes, R. M., Kempf, B., Kneip, S., Boch, J., Gade, J., Meier-Wagner, J., et al. (1999). Two evolutionarily closely related ABC transporters mediate the uptake of choline for synthesis of the osmoprotectant glycine betaine in *Bacillus subtilis*. *Mol. Microbiol.* 32, 203–216. doi: 10.1046/j.1365-2958.1999.01354.x
- Katoh, A., Uenohara, K., Akita, M., and Hashimoto, T. (2006). Early steps in the biosynthesis of NAD in arabidopsis start with aspartate and occur in the plastid. *Plant Physiol.* 141, 851–857. doi: 10.1104/pp.106.081091
- Katoh, K., and Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* 30, 772–780. doi: 10.1093/molbev/mst010
- Kehres, D., and Hogg, R. (1992). *Escherichia coli* K12 arabinose-binding protein mutants with altered transport properties. *Protein Sci.* 1, 1652–1660. doi: 10.1002/pro.5560011213
- Kempf, B., and Bremer, E. (1995). OpuA, an osmotically regulated binding protein-dependent transport system for the osmoprotectant glycine betaine in *Bacillus subtilis*. *J. Biol. Chem.* 270, 16701–16713. doi: 10.1074/jbc.270.28.16701
- Khan, N. A., Quemener, V., and Moulinoux, J.-P. (1992). Phorbol esters augment polyamine transport by influencing Na<sup>+</sup>-K<sup>+</sup> pump in murine leukemia cells. *Exp. Cell Res.* 199, 378–382. doi: 10.1016/0014-4827(92)90448-H
- Kimi Yoshi, I., Yoshihiro, A., Kumi, N., Shinsei, M., Tatsuo, H., Osamu, S., et al. (1993). Cloning of the cDNA encoding human xanthine dehydrogenase (oxidase): structural analysis of the protein and chromosomal location of the gene. *Gene* 133, 279–284. doi: 10.1016/0378-1119(93)90652-J
- Ko, R., and Smith, L. T. (1999). Identification of an ATP-Driven, osmoregulated glycine betaine transport system in *Listeria monocytogenes*. *Appl. Environ. Microbiol.* 65, 4040–4048.
- Kochi, C., Inagawa, H., Nishizawa, T., Yamaguchi, T., Nagai, S., and Soma, G.-I. (2006). Applications of lipopolysaccharide derived from *Pantoea agglomerans* (IP-PA1) for health care based on macrophage network theory. *J. Biosci. Bioeng.* 102, 485–496. doi: 10.1263/jbb.102.485
- Konstantinidis, K. T., and Tiedje, J. M. (2005a). Genomic insights that advance the species definition for prokaryotes. *Proc. Natl. Acad. Sci. U.S.A.* 102, 2567–2572. doi: 10.1073/pnas.0409727102
- Konstantinidis, K. T., and Tiedje, J. M. (2005b). Towards a genome-based taxonomy for prokaryotes. *J. Bacteriol.* 187, 6258–6264. doi: 10.1128/JB.187.18.6258-6264.2005
- Konstantinidis, K. T., and Tiedje, J. M. (2007). Prokaryotic taxonomy and phylogeny in the genomic era: advancements and challenges ahead. *Curr. Opin. Microbiol.* 10, 504–509. doi: 10.1016/j.mib.2007.08.006
- Koskiniemi, S., Sun, S., Berg, O. G., and Andersson, D. I. (2012). Selection-driven gene loss in bacteria. *PLoS Genet.* 8:e1002787. doi: 10.1371/journal.pgen.1002787
- Krismer, B., Liebeck, M., Janek, D., Nega, M., Rautenberg, M., Hornig, G., et al. (2014). Nutrient limitation governs *Staphylococcus aureus* metabolism and niche adaptation in the human nose. *PLoS Pathog.* 10:e1003862. doi: 10.1371/journal.ppat.1003862
- Linghu, B., Snitkin, E. S., Holloway, D. T., Gustafson, A. M., Xia, Y., and DeLisi, C. (2008). High-precision high-coverage functional inference from integrated data sources. *BMC Bioinformatics* 9:119. doi: 10.1186/1471-2105-9-119
- Liu, T., Vora, H., and Khosla, C. (2010). Quantitative analysis and engineering of fatty acid biosynthesis in *E. coli*. *Metab. Eng.* 12, 378–386. doi: 10.1016/j.ymben.2010.02.003
- Ma, Y., Yin, Y., Rong, C., Chen, S., Liu, Y., Wang, S., et al. (2016). *Pantoea pleuroti* sp. nov., isolated from the fruiting bodies of *Pleurotus eryngii*. *Curr. Microbiol.* 72, 207–212. doi: 10.1007/s00284-015-0940-5
- Maciejewska, A. M., Ruszel, K. P., Nieminuszczy, J., Lewicka, J., Sokolowska, B., Grzesiuk, E., et al. (2010). Chloroacetaldehyde-induced mutagenesis in *Escherichia coli*: the role of AlkB protein in repair of 3, N 4-ethenocytosine and 3, N 4- $\alpha$ -hydroxyethanocytosine. *Mutat. Res. Fundamental Mol. Mech. Mutagenesis* 684, 24–34. doi: 10.1016/j.mrfmmm.2009.11.005
- McInerney, J. O., Pisani, D., Baptiste, E., and O'Connell, M. J. (2011). The public goods hypothesis for the evolution of life on earth. *Biol. Direct* 6:41. doi: 10.1186/1745-6150-6-41
- Mehler, A. H., and Tabor, H. (1953). Deamination of histidine to form urocanic acid in liver. *J. Biol. Chem.* 201, 775–784.
- Mergaert, J., Verdonck, L., and Kersters, K. (1993). 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.* 43, 162–173. doi: 10.1099/00207713-43-1-162
- Metlitskaya, A., Kazakov, T., Vondenhoff, G. H., Novikova, M., Shashkov, A., Zatsepin, T., et al. (2009). Maturation of the translation inhibitor Microcin C. *J. Bacteriol.* 191, 2380–2387. doi: 10.1128/JB.00999-08
- Moffatt, B. A., and Ashihara, H. (2002). Purine and pyrimidine nucleotide synthesis and metabolism. *Arabidopsis Book* 1:e0018. doi: 10.1199/tab.0018
- Montgomerie, J., Bindereif, A., Neillands, J., Kalmanson, G., and Guze, L. (1984). Association of hydroxamate siderophore (aerobactin) with *Escherichia coli* isolated from patients with bacteremia. *Infect. Immun.* 46, 835–838.

- Moralejo, P., Egan, S., Hidalgo, E., and Aguilar, J. (1993). Sequencing and characterization of a gene cluster encoding the enzymes for L-rhamnose metabolism in *Escherichia coli*. *J. Bacteriol.* 175, 5585–5594. doi: 10.1128/jb.175.17.5585-5594.1993
- Nohno, T., Saito, T., and Hong, J.-S. (1986). Cloning and complete nucleotide sequence of the *Escherichia coli* glutamine permease operon (glnHPQ). *Mol. Gen. Genet.* 205, 260–269. doi: 10.1007/BF00430437
- Novikova, M., Metlitskaya, A., Datsenko, K., Kazakov, T., Kazakov, A., Wanner, B., et al. (2007). The *Escherichia coli* Yej transporter is required for the uptake of translation inhibitor Microcin C. *J. Bacteriol.* 189, 8361–8365. doi: 10.1128/JB.01028-07
- Nygaard, P., Bested, S. M., Andersen, K. A. K., and Saxild, H. H. (2000). *Bacillus subtilis* guanine deaminase is encoded by the yknA gene and is induced during growth with purines as the nitrogen source. *Microbiology* 146, 3061–3069. doi: 10.1099/00221287-146-12-3061
- Ollagnier-de Choudens, S., Loiseau, L., Sanakis, Y., Barras, F., and Fontecave, M. (2005). Quinolinate synthetase, an iron-sulfur enzyme in NAD biosynthesis. *FEBS Lett.* 579, 3737–3743. doi: 10.1016/j.febslet.2005.05.065
- Opal, S. M., Cross, A. S., Gernski, P., and Lyhte, L. W. (1990). Aerobactin and  $\alpha$ -hemolysin as virulence determinants in *Escherichia coli* isolated from human blood, urine, and stool. *J. Infect. Dis.* 161, 794–796. doi: 10.1093/infdis/161.4.794
- Ormeno-Orrillo, E., Servín-Garcidue-as, L. E., Rogel, M. A., González, V., Peralta, H., Mora, J., et al. (2015). Taxonomy of rhizobia and agrobacteria from the Rhizobiaceae family in light of genomics. *Syst. Appl. Microbiol.* 38, 287–291. doi: 10.1016/j.syapm.2014.12.002
- Palmer, M., de Maayer, P., Poulsen, M., Steenkamp, E. T., van Zyl, E., Coutinho, T. A., et al. (2016). Draft genome sequences of *Pantoea agglomerans* and *Pantoea vagans* isolates associated with termites. *Stand. Genomic Sci.* 11, 1–11. doi: 10.1186/s40793-016-0144-z
- Palmer, M., Steenkamp, E. T., Coetzee, M. P. A., Chan, W.-Y., van Zyl, E., De Maayer, P., et al. (2017). Phylogenomic resolution of the bacterial genus *Pantoea* and its relationship with *Erwinia* and *Tatumella*. *Antonie van Leeuwenhoek* 110, 1287–1309. doi: 10.1007/s10482-017-0852-4
- Pegg, A. E. (2016). Functions of polyamines in mammals. *J. Biol. Chem.* 291, 14904–14912. doi: 10.1074/jbc.R116.731661
- Plumbridge, J. (2009). An alternative route for recycling of N-Acetylglucosamine from peptidoglycan involves the N-Acetylglucosamine phosphotransferase system in *Escherichia coli*. *J. Bacteriol.* 191, 5641–5647. doi: 10.1128/JB.00448-09
- Pond, S. L. K., and Muse, S. V. (2005). “HyPhy: hypothesis testing using phylogenies,” in *Statistical Methods in Molecular Evolution*, ed R. Nielsen (New York, NY: Springer New York), 125–181.
- Rahman, N. A., Parks, D. H., Vanwonderghem, I., Morrison, M., Tyson, G. W., and Hugenholtz, P. (2015). A phylogenomic analysis of the bacterial phylum Fibrobacteres. *Front. Microbiol.* 6:1469. doi: 10.3389/fmicb.2015.01469
- Raymond, K. N., Dertz, E. A., and Kim, S. S. (2003). Enterobactin: an archetype for microbial iron transport. *Proc. Natl. Acad. Sci. U.S.A.* 100, 3584–3588. doi: 10.1073/pnas.0630018100
- Richard, P., and Hilditch, S. (2009). D-galacturonic acid catabolism in microorganisms and its biotechnological relevance. *Appl. Microbiol. Biotechnol.* 82, 597–604. doi: 10.1007/s00253-009-1870-6
- Richards, V. P., Palmer, S. R., Pavinski Bitar, P. D., Qin, X., Weinstock, G. M., Highlander, S. K., et al. (2014). Phylogenomics and the dynamic genome evolution of the genus *Streptococcus*. *Genome Biol. Evol.* 6, 741–753. doi: 10.1093/gbe/evu048
- Rong, C., Ma, Y., Wang, S., Liu, Y., Chen, S., Huang, B., et al. (2016). *Pantoea hericii* sp. nov., isolated from the fruiting bodies of *Hericium erinaceus*. *Curr. Microbiol.* 72, 738–743. doi: 10.1007/s00284-016-1011-2
- Saxena, R., Anand, P., Saran, S., Isar, J., and Agarwal, L. (2010). Microbial production and applications of 1, 2-propanediol. *Indian J. Microbiol.* 50, 2–11. doi: 10.1007/s12088-010-0017-x
- Schneider, E. (2001). ABC transporters catalyzing carbohydrate uptake. *Res. Microbiol.* 152, 303–310. doi: 10.1016/S0923-2508(01)01201-3
- Schubert, R. H. (1968). The taxonomy and nomenclature of the genus *Aeromonas* Kluyver and van Niel 1936. *Int. J. Syst. Evol. Microbiol.* 18, 1–7.
- Schulz, H. (1991). Beta oxidation of fatty acids. *Biochimica Biophys. Acta* 1081, 109–120. doi: 10.1016/0005-2760(91)90015-A
- Schwartz, A. R., Potnis, N., Timilsina, S., Wilson, M., Patané, J., Martins, J., et al. (2015). Phylogenomics of *Xanthomonas* field strains infecting pepper and tomato reveals diversity in effector repertoires and identifies determinants of host specificity. *Front. Microbiol.* 6:535. doi: 10.3389/fmicb.2015.00535
- Schwede, T. F., Rétey, J., and Schulz, G. E. (1999). Crystal structure of histidine ammonia-lyase revealing a novel polypeptide modification as the catalytic electrophile. *Biochemistry* 38, 5355–5361. doi: 10.1021/bi982929q
- Scripture, J. B., Voelker, C., Miller, S., O'Donnell, R. T., Polgar, L., Rade, J., et al. (1987). High-affinity l-arabinose transport operon. *J. Mol. Biol.* 197, 37–46. doi: 10.1016/0022-2836(87)90607-3
- Sekowska, A., Dénervaud, V., Ashida, H., Michoud, K., Haas, D., Yokota, A., et al. (2004). Bacterial variations on the methionine salvage pathway. *BMC Microbiol.* 4:9. doi: 10.1186/1471-2180-4-9
- Sekowska, A., Kung, H.-F., and Danchin, A. (2000). Sulfur metabolism in *Escherichia coli* and related bacteria: facts and fiction. *J. Mol. Microbiol. Biotechnol.* 2, 145–177.
- Shah, P., and Swiatlo, E. (2008). A multifaceted role for polyamines in bacterial pathogens. *Mol. Microbiol.* 68, 4–16. doi: 10.1111/j.1365-2958.2008.06126.x
- Shimizu, K. (2014). Regulation systems of bacteria such as *Escherichia coli* in response to nutrient limitation and environmental stresses. *Metabolites* 4:1. doi: 10.3390/metabo4010001
- Sim, L., Quezada-Calvillo, R., Sterchi, E. E., Nichols, B. L., and Rose, D. R. (2008). Human intestinal maltase-glucoamylase: crystal structure of the N-terminal catalytic subunit and basis of inhibition and substrate specificity. *J. Mol. Biol.* 375, 782–792. doi: 10.1016/j.jmb.2007.10.069
- Song, Y.-J. (2009). Characterization of aromatic hydrocarbon degrading bacteria isolated from pine litter. *Kor. J. Microbiol. Biotechnol.* 37, 333–339.
- Sorrels, C. M., Proteau, P. J., and Gerwick, W. H. (2009). Organization, evolution, and expression analysis of the biosynthetic gene cluster for scytonemin, a cyanobacterial UV-absorbing pigment. *Appl. Environ. Microbiol.* 75, 4861–4869. doi: 10.1128/AEM.02508-08
- Sprenger, G. A. (1993). Two open reading frames adjacent to the *Escherichia coli* K-12 transketolase (tkt) gene show high similarity to the mannitol phosphotransferase system enzymes from *Escherichia coli* and various Gram-positive bacteria. *Biochim. Biophys. Acta* 1158, 103–106. doi: 10.1016/0304-4165(93)90103-F
- Stanier, R. Y., and Van Niel, C. (1941). The main outlines of bacterial classification. *J. Bacteriol.* 42:437.
- Stevenson, G., Neal, B., Liu, D., Hobbs, M., Packer, N. H., Batley, M., et al. (1994). Structure of the O antigen of *Escherichia coli* K-12 and the sequence of its rfb gene cluster. *J. Bacteriol.* 176, 4144–4156. doi: 10.1128/jb.176.13.4144-4156.1994
- Stirling, D., Hulton, C., Waddell, L., Park, S., Stewart, G., Booth, I., et al. (1989). Molecular characterization of the proU loci of *Salmonella typhimurium* and *Escherichia coli* encoding osmoregulated glycine betaine transport systems. *Mol. Microbiol.* 3, 1025–1038. doi: 10.1111/j.1365-2958.1989.tb00253.x
- Sutcliffe, I. C. (2015). Challenging the anthropocentric emphasis on phenotypic testing in prokaryotic species descriptions: rip it up and start again. *Front. Genet.* 6:218. doi: 10.3389/fgene.2015.00218
- Sutcliffe, I. C., Trujillo, M. E., and Goodfellow, M. (2012). A call to arms for systematists: revitalising the purpose and practises underpinning the description of novel microbial taxa. *Antonie Van Leeuwenhoek* 101, 13–20. doi: 10.1007/s10482-011-9664-0
- Suzuki, M. (1991). *U.S. Patent No. 5,047,329*. Washington, DC: U.S. Patent and Trademark Office.
- Suzuki, H., Koyanagi, T., Izuka, S., Onishi, A., and Kumagai, H. (2005). The yliA, -B, -C, and -D genes of *Escherichia coli* K-12 encode a novel glutathione importer with an ATP-binding cassette. *J. Bacteriol.* 187, 5861–5867. doi: 10.1128/JB.187.17.5861-5867.2005
- Tamura, K., Stecher, G., Peterson, D., Filipski, A., and Kumar, S. (2013). MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. *Mol. Biol. Evol.* 30, 2725–2729. doi: 10.1093/molbev/mst197
- Tindall, B. J., Rosselló-Mora, R., Busse, H.-J., Ludwig, W., and Kämpfer, P. (2010). Notes on the characterization of prokaryote strains for taxonomic purposes. *Int. J. Syst. Evol. Microbiol.* 60, 249–266. doi: 10.1099/ijs.0.016949-0
- Torres, A. G., Redford, P., Welch, R. A., and Payne, S. M. (2001). TonB-dependent systems of uropathogenic *Escherichia coli*: aerobactin and heme transport and

- TonB are required for virulence in the mouse. *Infect. Immun.* 69, 6179–6185. doi: 10.1128/IAI.69.10.6179-6185.2001
- Tracz, D. M., Gilmour, M. W., Mabon, P., Beniac, D. R., Hoang, L., Kibsey, P., et al. (2015). *Tatumella saanichensis* sp. nov., isolated from a cystic fibrosis patient. *Int. J. Syst. Evol. Microbiol.* 65, 1959–1966. doi: 10.1099/ijs.0.000207
- Uehara, T., Suefuji, K., Jaeger, T., Mayer, C., and Park, J. T. (2006). MurQ esterase is required by *Escherichia coli* in order to metabolize anhydro-N-acetylmuramic acid obtained either from the environment or from its own cell wall. *J. Bacteriol.* 188, 1660–1662. doi: 10.1128/JB.188.4.1660-1662.2006
- Van Beers, E. H., Büller, H. A., Grand, R. J., Einerhand, A. W. C., and Dekker, J. (1995). Intestinal brush border glycohydrolases: structure, function, and development. *Crit. Rev. Biochem. Mol. Biol.* 30, 197–262. doi: 10.3109/10409239509085143
- Van der Ploeg, J., Weiss, M. A., Saller, E., Nashimoto, H., Saito, N., Kertesz, M. A., et al. (1996). Identification of sulfate starvation-regulated genes in *Escherichia coli*: a gene cluster involved in the utilization of taurine as a sulfur source. *J. Bacteriol.* 178, 5438–5446. doi: 10.1128/jb.178.18.5438-5446.1996
- Vanneste, J., Cornish, D., Yu, J., and Voyle, M. (2001). “The peptide antibiotic produced by *Pantoea agglomerans* Eh252 is a microcin,” in *IX International Workshop on Fire Blight* (Napier), 285–290.
- Walshaw, D. L., Lowthorpe, S., East, A., and Poole, P. S. (1997). Distribution of a sub-class of bacterial ABC polar amino acid transporter and identification of an N-terminal region involved in solute specificity. *FEBS Lett.* 414, 397–401. doi: 10.1016/S0014-5793(97)01023-5
- Walterson, A. M., and Stavrinos, J. (2015). *Pantoea*: insights into a highly versatile and diverse genus within the Enterobacteriaceae. *FEMS Microbiol. Rev.* 39, 968–984. doi: 10.1093/femsre/fuv027
- Walton, J. D. (1994). Deconstructing the cell wall. *Plant Physiol.* 104:1113. doi: 10.1104/pp.104.4.1113
- Wan, X., Hou, S., Phan, N., Malone Moss, J. S., Donachie, S. P., and Alam, M. (2015). Draft genome sequence of *pantoea anthophila* strain 11-2 from hypersaline lake laysan, Hawaii. *Genome Announc.* 3:e00321-15. doi: 10.1128/genomeA.00321-15
- Wang, L., and Reeves, P. R. (1998). Organization of *Escherichia coli* O157 O antigen gene cluster and identification of its specific genes. *Infect. Immun.* 66, 3545–3551.
- Whitfield, C. (1995). Biosynthesis of lipopolysaccharide O antigens. *Trends Microbiol.* 3, 178–185. doi: 10.1016/S0966-842X(00)88917-9
- Williams, P. H., and Carbonetti, N. H. (1986). Iron, siderophores, and the pursuit of virulence: independence of the aerobactin and enterochelin iron uptake systems in *Escherichia coli*. *Infect. Immun.* 51, 942–947.
- Woese, C. (1998). The universal ancestor. *Proc. Natl. Acad. Sci. U.S.A.* 95, 6854–6859. doi: 10.1073/pnas.95.12.6854
- Woese, C. R. (1987). Bacterial evolution. *Microbiol. Rev.* 51:221.
- Woese, C. R. (1994). There must be a prokaryote somewhere: microbiology's search for itself. *Microbiol. Rev.* 58, 1–9.
- Woese, C. R., Kandler, O., and Wheelis, M. L. (1990). Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. *Proc. Natl. Acad. Sci. U.S.A.* 87, 4576–4579. doi: 10.1073/pnas.87.12.4576
- Yang, Q., Liu, Y., Huang, F., and He, Z.-G. (2011). Physical and functional interaction between d-ribokinase and topoisomerase I has opposite effects on their respective activity in *Mycobacterium smegmatis* and *Mycobacterium tuberculosis*. *Arch. Biochem. Biophys.* 512, 135–142. doi: 10.1016/j.abb.2011.05.018
- Yang, Z., Savchenko, A., Yakunin, A., Zhang, R., Edwards, A., Arrowsmith, C., et al. (2003). Aspartate dehydrogenase, a novel enzyme identified from structural and functional studies of TM1643. *J. Biol. Chem.* 278, 8804–8808. doi: 10.1074/jbc.M211892200
- Yoshida, K.-I., Yamaguchi, M., Ikeda, H., Omae, K., Tsurusaki, K.-I., and Fujita, Y. (2004). The fifth gene of the iol operon of *Bacillus subtilis*, iolE, encodes 2-keto-myo-inositol dehydratase. *Microbiology* 150, 571–580. doi: 10.1099/mic.0.26768-0
- Yoshimoto, T., Oka, I., and Tsuru, D. (1976). Purification, crystallization, and some properties of creatine amidinohydrolase from *Pseudomonas putida*. *J. Biochem.* 79, 1381–1383. doi: 10.1093/oxfordjournals.jbchem.a131193
- Young Kim, M., Zhou, X., Delaney, J. C., Taghizadeh, K., Dedon, P. C., Essigmann, J. M., et al. (2007). AlkB influences the chloroacetaldehyde-induced mutation spectra and toxicity in the pSP189 supF shuttle vector. *Chem. Res. Toxicol.* 20, 1075–1083. doi: 10.1021/tx700167v
- Zaneveld, J. R., Nemergut, D. R., and Knight, R. (2008). Are all horizontal gene transfers created equal? Prospects for mechanism-based studies of HGT patterns. *Microbiology* 154, 1–15. doi: 10.1099/mic.0.2007/011833-0
- Zangoui, P., Vashishtha, K., and Mahadevan, S. (2015). Evolution of aromatic  $\beta$ -glucoside utilization by successive mutational steps in *Escherichia coli*. *J. Bacteriol.* 197, 710–716. doi: 10.1128/JB.02185-14
- Zhang, Z.-G., Ye, Z.-Q., Yu, L., and Shi, P. (2011). Phylogenomic reconstruction of lactic acid bacteria: an update. *BMC Evol. Biol.* 11:1. doi: 10.1186/1471-2148-11-1

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Palmer, Steenkamp, Coetzee, Blom and Venter. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.