Genomics Data 10 (2016) 63–68

Contents lists available at ScienceDirect

Genomics Data

journal homepage: www.elsevier.com/locate/gdata

Draft genome sequence and detailed analysis of *Pantoea eucrina* strain Russ and implication for opportunistic pathogenesis

Farzaneh Moghadam [a], MB Couger [a], Breeanna Russ [a], Randi Ramsey [a], Radwa A. Hanafy [a], Connie Budd [a], Donald P. French [b], Wouter D. Hoff [a], Noha Youssef [a,*]

[a] Department of Microbiology and Molecular Genetics, Oklahoma State University, Stillwater, OK, United States
[b] Department of Integrative Biology, Oklahoma State University, Stillwater, OK, United States

ARTICLE INFO

Article history:
Received 19 July 2016
Received in revised form 8 September 2016
Accepted 14 September 2016
Available online 20 September 2016

Keywords:
Pantoea eucrina
Draft genome sequence
Detailed annotation
Student Initiated Microbial Discovery (SIMD) project
Metabolic versatility
Secretion systems
Pili

ABSTRACT

The genus *Pantoea* is a predominant member of host-associated microbiome. We here report on the genomic analysis of *Pantoea eucrina* strain Russ that was isolated from a trashcan at Oklahoma State University, Stillwater, OK. The draft genome of *Pantoea eucrina* strain Russ consists of 3,939,877 bp of DNA with 3704 protein-coding genes and 134 RNA genes. This is the first report of a genome sequence of a member of *Pantoea eucrina*. Genomic analysis revealed metabolic versatility with genes involved in the metabolism and transport of all amino acids as well as glucose, fructose, mannose, xylose, arabinose and galactose, suggesting the organism is a versatile heterotroph. The genome also encodes an extensive secretory machinery including types I, II, III, IV, and Vb secretion systems, and several genes for pili production including the new usher/chaperone system (pfam 05,229). The implications of these systems for opportunistic pathogenesis are discussed.

1. Introduction

The strain *Pantoea eucrina Russ* was isolated (by an undergraduate student, BR) from a trashcan surface with frequent human use on Oklahoma State University (OSU) campus in Stillwater, OK. This was part of the Student Initiated Microbial Discovery (SIMD) project at OSU (introduced in [1]). The genus *Pantoea* is a phylogenetically and physiologically diverse genus with members ubiquitously found in host-associated microbiome as plant endophytes, insects symbionts, and members of the human gut microbiomes [2–6]. Endophytic strains range from plant pathogens, plat commensal, to a beneficial strains with growth-promoting effects [7]. *Pantoea* is frequently isolated from the nosocomial environment [8–10] and hence a considerable debate on its role in human infection was recently raised. Genomic analysis of strains belonging to the genus *Pantoea* could potentially contribute majorly to our understanding of opportunistic pathogenesis. Such knowledge can help mitigate the severity of nosocomial infections in immunocompromised patients. Here we report on the first draft genomic sequence, and the detailed annotation and analysis of a *Pantoea eucrina* strain with an emphasis on its pathogenic potential.

2. Genome sequencing information

2.1. Genome project history

The quality draft assembly and annotation were completed in 2015–2016. Table 1 shows the genome project information.

2.2. Growth conditions and genomic DNA preparation

Pantoea eucrina Russ was grown overnight at 30 °C on tryptic soy agar plates. Genomic DNA of high sequencing quality was isolated using the MPBio PowerSoil® DNA extraction kit according to manufacturer's instructions. Negative stain TEM micrographs were obtained using the services of the Oklahoma State University Microscopy Lab. Briefly, the sample was placed on a carbon film TEM grid and allowed to incubate for 2 min, after which the excess liquid was blotted off. Phosphotungstic acid (PTA; 2% w/v) was then added to the grid followed by a 45-sec incubation. Excess PTA was blotted off and the

* Corresponding author at: 1110 S Innovation way, Stillwater, OK 74074, United States.
E-mail address: noha@okstate.edu (N. Youssef).

http://dx.doi.org/10.1016/j.gdata.2016.09.006

Table 1
Project information.

MIGS ID	Property	Term
MIGS 31	Finishing quality	Draft
MIGS-28	Libraries used	Illumina 2X300 paired end chemistry
MIGS 29	Sequencing platforms	Illumina Miseq
MIGS 31.2	Fold coverage	300x
MIGS 30	Assemblers	Velvet 2.0
MIGS 32	Gene calling method	Prodigal, IMG-ER
	Genbank ID	MAYN00000000
	GenBank date of release	July 2016
	GOLD ID	Gp0126758
	BIOPROJECT	PRJNA327384
MIGS 13	Project relevance	Environmental

grid was allowed to dry before it was visualized using JOEL JEM-2100 transmission electron microscope.

2.3. Genome sequencing and assembly

The genome of *Pantoea eucrina* Russ was sequenced using the Illumina MiSeq platform at the University of Georgia Genomics Facility using 2X300 paired end chemistry and an average library insert size of 700 bp. The short read de Brujin graph assembly program Velvet [11] was employed for assembling quality filtered sequence data using the following flags; a kmer value of 101 bp and a minimum contig coverage value of 7×. The genome project is deposited in GOLD (Genomes On-Line Database) and this Whole Genome Shotgun (WGS) project has been deposited in GenBank under the accession MAYN00000000. The version described in this paper is version MAYN01000000.

2.4. Genome annotation

Using the prokaryotic gene calling software package prodigal [12], a total of 3838 gene models were predicted with average gene size of 931.73 bp. Functional annotation involved a combination of NCBI Blast C++ homology search, and HMMER 3.0 [13] hmmscan against the PFAM [14] 26.0 database. Additional gene analysis and functional annotation were carried out through the Integrated Microbial Genomes Expert Review (IMG-ER) platform.

2.5. Comparative genomics

We compared the genome of *Pantoea eucrina* strain Russ to 22 closely related genomes (IMG Ids: 648276708 (*Pantoea* sp. aB), 649633081 (*Pantoea* sp. At-9b), 2511231025 (*Pantoea* sp. YR343), 2511231035 ((*Pantoea* sp. GM01), 2519899784 (*Pantoea* sp. Sc1), 2545824509 (*Pantoea* sp. GL120224-02), 2551306469 (*Pantoea* sp. A4), 2551306543 (*Pantoea* sp. B40), 2582581300 (*Pantoea* sp. 9140), 2602041550 (*Pantoea* sp. AS-PWVM4), 2602042078 (*Pantoea* sp. 9133), 2609460089 (*Pantoea* sp. IMH), 2616644925 (*Pantoea* sp. 3.5.1), 2617271108 (*Pantoea* sp. FF5), 2619619082 (*Pantoea* sp. SL1_M5), 2627853687 (*Pantoea* sp. MBLJ3), 2627853912 (*Pantoea* sp. SM3), 2630968876 (*Pantoea* sp. PSNIH1), 2630968889 (*Pantoea* sp. PSNIH2), 2636415588 (*Pantoea* sp. BL1), 2643221431 (*Pantoea* sp. Isolate 98), 2651869657 (*Pantoea* sp. RIT-PI-b)) using the "Genome clustering" function on the IMG-ER analysis platform based on the KEGG profile. We also used principal component analysis to compare the genomes based on several genomic features including the genome size, the number of genes, the number of transporters identified, the GC content, the number of non-coding bases, the number of genes belonging to COG categories, as well as the number of genes belonging to each COG category. The PCA analysis was conducted using the "princomp" function in the labdsv library of R [15]. The results were visualized using a biplot, where genomes were represented by stars and genomic features or COG categories used for comparison were represented by arrows.

3. Results and discussion

3.1. Classification and features

Cells of *P. eucrina* strain *Russ* are Gram-negative, motile rods that were arranged in singles (Fig. 1). Colonies on TSA agar were yellow.

Within the genus *Pantoea*, 24 species are described with validly published names *P. agglomerans* type strain ATCC 27155^T, *P. allii* type strain BD390^T, *P. ananatis* type strain ATCC 33244^T, *P. anthophila* type strain BD871^T, *P. brenneri* type strain BD873^T, *P. calida* type strain 1400/07^T, *P. citrea* type strain BD875^T, *P. coffeiphila* type strain DSM 28482^T, *P. conspicua* type strain BD805^T, *P. cypripedii* type strain ATCC 29267^T, *P. deleyi* type strain BD767^T, *P. dispersa* type strain ATCC 14589^T, *P. eucalypti* type strain BD769^T, *P. eucrina* type strain BD872^T, *P. gaviniae* type strain A18/07^T, *P. intestinalis* type strain DSM 28113^T, *P. punctata* type strain BD876^T, *P. rodasii* type strain BD943^T, *P. rwandensis* type strain BD944^T, *P. septica* type strain BD874^T, *P. stewartii* type strain ATCC 8199^T, *P. terrea* type strain BD877^T, *P. theicola* type strain DSM 29212^T, *P. vagans* type strain BD765^T, and *P. wallisii* type strain BD946^T. Strain *Russ* shares 96.6% with *P. agglomerans*, 96.3% *P. allii*, 97% *P. ananatis*, 97%*P. anthophila*, 96.6% *P. brenneri*, 96.4% *P. calida*, 94.9% *P. citrea*, 97.9% *P. coffeiphila*, 96.6% *P. conspicua*, 95.7% *P. cypripedii*, 96.8% *P. deleyi*, 98.5% *P. dispersa*, 96.4% *P. eucalypti*, 100% *P. eucrina*, 96.6% *P. gaviniae*, 95.5% *P. intestinalis*, 96.2% *P. punctata*, 97.5% *P. rodasii*, 97.5% *P. rwandensis*, 97.8% *P. septica*, 97.8% *P. stewartii*, 96.6% *P. terrea*, 96.3% *P. theicola*, 96.4% *P. vagans*, and 98.5% *P. wallisii* in the *Pantoea* genus. Phylogenetic analysis based on the 16S rRNA gene placed *Pantoea eucrina* strain Russ in the same node with the *Pantoea eucrina* strains IHB B 10086, C7, and CT194 (Table 2, and Fig. 2).

Compared to other *Pantoea* species with sequenced genomes, strain Russ shares 98% 16S rRNA gene similarity with representatives of *Pantoea dispersa*, 97% similarity with representatives of *Pantoea stewartii*, and 96% similarity with representatives of *Pantoea ananatis*.

3.2. Genome properties

The genome assembly process produced a contig N50 of 2,633,372 bp with a total genome size of 3,939,877 bp. The GC content was 55.98%. One hundred and thirty four RNA genes were identified in the genome including 11 ribosomal RNA and 75 tRNA genes. The ribosomal RNA operon showed a typical bacterial organization with genes for 5S, 16S, and 23S rRNA and tRNAs *tRNA^{Ile} and tRNA^{Ala}*. Of the

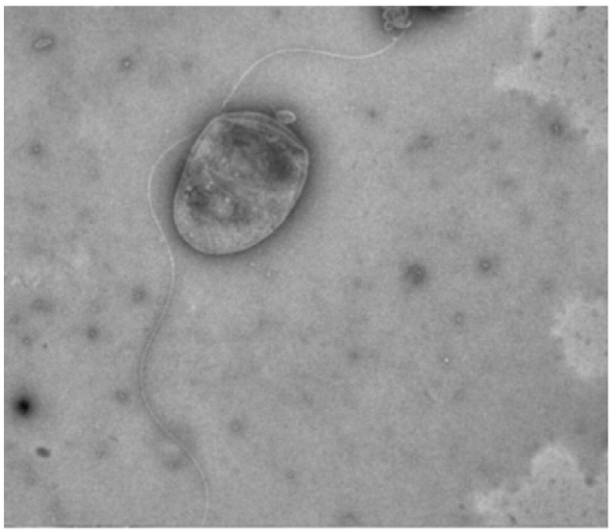

Fig. 1. Negative stain TEM micrograph of *Pantoea eucrina* Russ.

Table 2
Classification and general features of *Pantoea eucrina* Russ [25].

MIGS ID	Property	Term	Evidence code[a]
	Classification	Domain Bacteria	TAS [26]
		Phylum Proteobacteria	TAS [26]
		Class Gammaproteobacteria	TAS [26]
		Order Enterobacteriales	TAS [26]
		Family Enterobacteriaceae	TAS [26]
		Genus *Pantoea*	TAS [26]
		Species *eucrina*	TAS [26]
		Strain: Russ	
	Gram stain	Negative	TAS [26]
	Cell shape	Rod	TAS [26]
	Motility	Motile	TAS [26]
	Sporulation	Non-spore forming	TAS [26]
	Temperature range	Mesophile	TAS [26]
	Optimum temperature	28 °C	TAS [26]
	pH range; optimum	Unknown	
	Carbon source	D-glucose, D-fructose, D-galactose, trehalose, D-mannose, cellobiose, 1-*O*-methylb-D-glucopyranoside, L-arabinose, glycerol, inositol, Dsaccharate, cis-aconitate, D-glucuronate, D-galacturonate, *N*-acetylglucosamine, D-gluconate, DL-lactate, L-histidine, L-aspartate, L-glutamate, L-alanine and L-serine sucrose, maltotriose, maltose, D-arabitol, L-arabitol, xylitol, D-mannitol, adonitol and citrate.	TAS [26]
MIGS-6	Habitat	Trashcan	IDA
MIGS-6.3	Salinity	Growth in TSA (0.5%)	IDA
MIGS-22	Oxygen requirement	Facultative anaerobe	TAS [26]
MIGS-15	Biotic relationship	Free-living	IDA
MIGS-14	Pathogenicity	Unknown	
MIGS-4	Geographic location	Stillwater, Oklahoma	IDA
MIGS-5	Sample collection	March 2015	IDA
MIGS-4.1	Latitude	36.1157	IDA
MIGS-4.2	Longitude	−97.0586	IDA
MIGS-4.4	Altitude	1 M	IDA

[a] Evidence codes - IDA: Inferred from Direct Assay; TAS: Traceable Author Statement (i.e., a direct report exists in the literature); NAS: Non-traceable Author Statement (i.e., not directly observed for the living, isolated sample, but based on a generally accepted property for the species, or anecdotal evidence). These evidence codes are from the Gene Ontology project [27].

3838 detected genes, 3704 genes (96.51%) were protein-coding, of which 80.22% had a function prediction, 74.65% represented a COG functional category, and 8.1% were predicted to have a signal peptide. Using PSORT [16], we classified proteins as 41% cytoplasmic, 0.72% extracellular, and 30.3% associated with the membrane. Based on the presence of 139 single copy genes [17], the genome is predicted to be 81.3% complete. Genome statistics are shown in Table 3. The distribution of genes into COG functional categories is shown in Table 4.

3.3. Insights from the genome sequence

Genome analysis of *Pantoea eucrina* Russ identified a microorganism with a typical Gram-negative cell wall structure. The genome also suggests that the cell envelope contains the polar lipids phosphatidylglycerol, phosphatidylcholine, phosphatidylethanolamine, phosphatidylserine, and cardiolipin since genes for their biosynthesis were identified in the genome. Strain Russ genome also encodes for a complete flagellar assembly, in agreement with the isolate's electron micrograph, as well as a type I pilus system belonging to the newly identified usher/chaperone system [18]. This system was first characterized in *Acinetobacter baumannii* and is linked to the early steps in biofilm formation [18]. The presence of genes encoding this system in *Pantoea eucrina* combined with the observation that it also possesses flagella imply a possible role of these genes to its virulence especially in nosocomial settings by allowing it to establish biofilms. When compared against the virulence factor database [19], the genome of *Pantoea eucrina* Russ showed 1077 virulence factor hits (29% of the protein-coding genes). These included Type I, Type II, Type III, Type IV, and Type Vb secretion systems. Most of these secretion systems have been linked to virulence in Gram-negative organisms [20], and could potentially contribute to opportunistic pathogenesis. Aside from their potential to be opportunistic pathogens, *Pantoea* endophytic strains range in their relationship with plant hosts from pathogenic to beneficial growth-promoters or bioprotectors [7]. Previous research suggested a relationship between the presence of virulence-associated genes on mobile elements and the pathogenicity of the strain towards plants [7]. Even though strain Russ was not isolated from a plant host, we sought to examine the possibility of its potential pathogenicity, or lack thereof, towards plants. Among 11 possible transposases identified in the genome, one putative transposase (IMG gene ID: 2650202385) is present in a cluster (with one other transposase and two insertion elements proteins) downstream from a sucrose utilization cluster and a tellurite resistance cluster. Since sucrose is a predominant disaccharide in higher plants tissues [21], the capability to degrade sucrose would be highly beneficial for plant endophytes. Some plants are also known to accumulate tellurium [22], which would warrant a mechanism of tellurium resistance in the endophytic bacteria affiliated with such plant hosts. The presence of sucrose utilization gene cluster, as well as a tellurium resistance gene cluster upstream from transposases and transposable elements in the genome of strain Russ might suggest its potential for plant pathogenesis. However, this claim requires further investigation.

Further analysis of KEGG pathways identified almost compete to complete catabolic pathways for utilization of glucose, fructose, mannose, xylose, arabinose and galactose, and all amino acids as carbon and energy sources. The genome also suggests the capability of xanthine degradation to glycine as well as uracil degradation to 3-hydroxypropionate, both of which indicate the capability to utilize purines and pyrimidines as energy sources. The genome encodes a complete TCA cycle and electron transport chain with F-type ATPase subunits confirming the aerobic nature of the microorganism. Facultative anaerobiosis is also suggested by the genome based on the presence of genes encoding for enzymes involved in lactate, acetate and formate fermentation were identified. Genomic analysis suggested auxotrophy for VitB12 and Riboflavin. However comparison of the protein-coding genes against the transporter database [23] identified several ABC and secondary transporters that could potentially be used for the import of such molecules.

3.4. Insights from comparative genomics

When the genome of *Pantoea eucrina* Russ was compared to 22 closely related genomes based on their KEGG profiles, the genome clustered with *Pantoea* sp. PSNIH1 previously isolated from patients in a hospital setting and shown to carry several plasmids with antibiotic resistance cassettes [24] (Fig. 3A). This genomic similarity was confirmed when several genomic features (including the genome size, the number of genes, the number of transporters identified, the GC content, the number of non-coding bases, the number of genes belonging to COG categories, as well as the number of genes belonging to each COG category) were used to compare *Pantoea eucrina* Russ genome to the 22 other closely related genomes. The Russ genome was shown to cluster with the genomes of strains PSNIH1, PSNIH2, IMH, and B40 based on the lower number of genes belonging to the COG categories E, K, G, R, and P in these genomes (Fig. 3B).

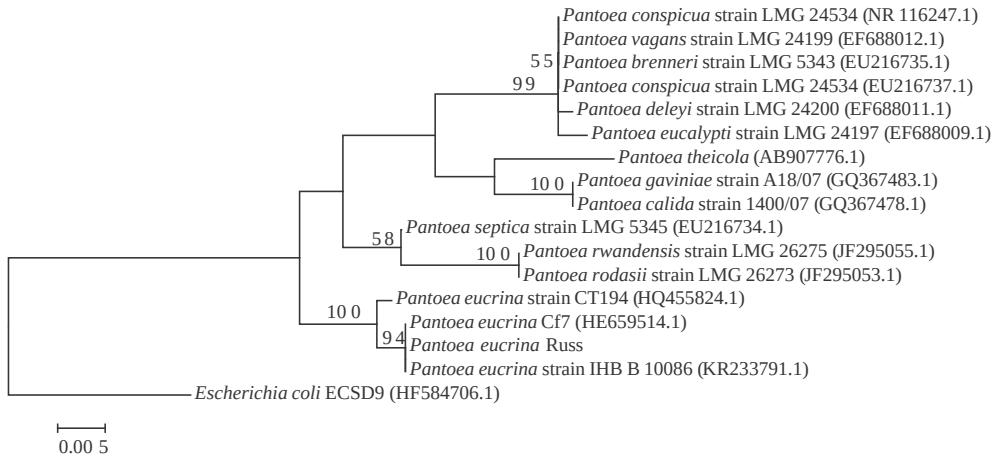

Fig. 2. A maximum likelihood phylogenetic tree constructed using multiple sequence alignments of 16S rRNA genes. "*Pantoea eucrina* Russ" sequence is shown in bold. Reference sequences are also shown and Genbank accession numbers are given in parentheses. The tree was obtained under "K2 + G" model with a variable site γ shape parameter of 0.05. "*Escherichia coli*" was used as the out-group. Bootstrap values, in percent, are based on 100 replicates and are shown for branches with >50% bootstrap support. Multiple sequence alignment, model selection, and maximum likelihood analysis were carried out in MEGA [28].

4. Conclusions

This study presents the first draft genome sequence and functional annotation of a member of the genus *Pantoea eucrina*. The genome of *Pantoea eucrina* strain Russ revealed extensive sugar and amino acid degradation machinery, as well as the potential to use purines and pyrimidines as carbon an energy sources. Type I pili belonging to the new usher/chaperone system (pfam 05,229), and possession of flagella could contribute to the capability to form biofilms. Comparison to the virulence factor database identified 1077 genes in the genome with potential virulence-associated function including type I, II, III, IV, and Vb secretion systems, most of which could potentially contribute to opportunistic pathogenesis that was previously reported for members of the *Pantoea eucrina*. Comparative genomics using general genomic features as well as the KEGG function profile clustered the Russ genome with *Pantoea* strains previously isolated from hospital settings and shown to harbor antibiotic resistance-encoding plasmids.

Authors' contributions

FM, BR, RR, MBC, and NY contributed to the analysis. FM, WDH, DPF, and NY wrote the manuscript. BR, CB, and RAH performed the lab experiments.

Table 3
Genome statistics.

Attribute	Value	% of total
Genome size (bp)	3,939,877	100%
DNA coding (bp)	3,459,667	87.81%
DNA G + C (bp)	2,205,503	55.98%
DNA scaffolds	8	100%
Total genes	3838	100%
Protein coding genes	3704	96.51%
RNA genes	134	3.49%
Pseudo genes	0	
Genes in internal clusters	829	21.60%
Genes with function prediction	3079	80.22%
Genes assigned to COGs	2865	74.65%
Genes with Pfam domains	3267	85.12%
Genes with signal peptides	312	8.10%
Genes with transmembrane helices	853	22.23%
CRISPR repeats	0	

Competing interests

All authors declare no competing interests.

Transparency document

The Transparency document associated with this article can be found, in online version.

Acknowledgements and funding

Pantoea eucrina strain Russ was isolated and selected for sequencing as part of a Howard Hughes Medical Institute funded project at Oklahoma State University. The project aims at improving undergraduate students persistence through authentic laboratory research. During a

Table 4
Number of genes associated with general COG functional categories.

Code	Value	% age	Description
J	242	7.5%	Translation, ribosomal structure and biogenesis
A	1	0.03%	RNA processing and modification
K	258	8%	Transcription
L	124	3.84%	Replication, recombination and repair
B	0	0%	Chromatin structure and dynamics
D	44	1.36%	Cell cycle control, cell division, chromosome partitioning
V	67	2.08%	Defense mechanisms
T	186	5.77%	Signal transduction mechanisms
M	233	7.19%	Cell wall/membrane biogenesis
N	77	2.39%	Cell motility
U	39	1.21%	Intracellular trafficking and secretion
O	120	3.72%	Posttranslational modification, protein turnover, chaperones
C	176	5.46%	Energy production and conversion
G	312	9.67%	Carbohydrate transport and metabolism
E	315	9.76%	Amino acid transport and metabolism
F	95	2.94%	Nucleotide transport and metabolism
H	177	5.49%	Coenzyme transport and metabolism
I	97	3.01%	Lipid transport and metabolism
P	198	6.14%	Inorganic ion transport and metabolism
Q	43	1.33%	Secondary metabolites biosynthesis, transport and catabolism
R	232	7.19%	General function prediction only
S	165	5.11%	Function unknown
–	1640		Not in COGs

The total is based on the total number of protein coding genes in the genome.

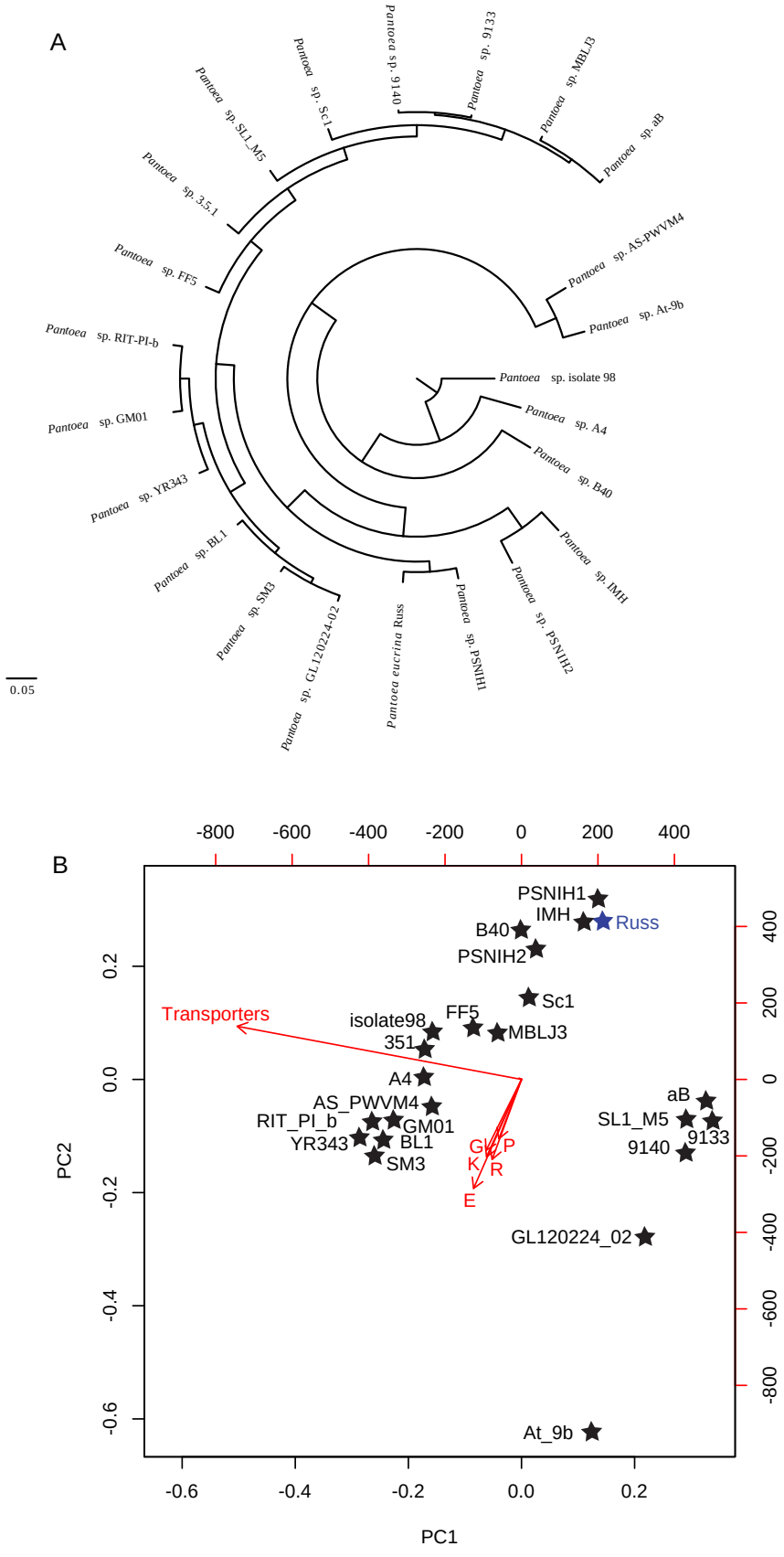

Fig. 3. Comparative genomics of *Pantoea eucrina* strain Russ and 21 closely related genomes. (A) KEGG profile clustering of the genomes compared in this study. (B) PCA biplot of the genomic features and COG category distribution in the genomes compared. Genomes are represented by stars, where the strain name is depicted. Arrows represent genomic features or COG categories used for comparison. The arrow directions follow the maximal abundance, and their lengths are proportional to the maximal rate of change between genomes. The first two components explained 75% of variation.

two-semester long effort, undergraduate students isolate an environmental strain, and extract its genomic DNA. The genome is then sequenced and analyzed by undergraduate students as part of an upper division microbial genomics class. The current genome was analyzed by a team of undergraduate (BR, RR), and graduate (FM) students.

This is Draft Genome #6 in the SIMD project supported in part by a grant from the Howard Hughes Medical Institute (1554854) through the Science Education Program. WDH acknowledges support by NSF grants MCB-1051590, MRI-1338097, and CHE-1412500.

References

[1] M.B. Couger, et al., Draft genome sequence of the environmental isolate *Chryseobacterium* sp. Hurlbut01. Genome Announc. 3 (5) (2015).
[2] D. Coleman-Derr, et al., Plant compartment and biogeography affect microbiome composition in cultivated and native Agave species. New Phytol. 209 (2) (2016) 798–811.
[3] J.P. Gauthier, et al., Bacterial communities associated with host-adapted populations of pea aphids revealed by deep sequencing of 16S ribosomal DNA. PLoS One 10 (3) (2015), e0120664.
[4] P.E. Larsen, Y. Dai, Metabolome of human gut microbiome is predictive of host dysbiosis. Gigascience 4 (2015) 42.
[5] D. Passos da Silva, et al., Bacterial multispecies studies and microbiome analysis of a plant disease. Microbiology 160 (Pt 3) (2014) 556–566.
[6] X. Li, et al., The endophytic bacteria isolated from elephant grass (*Pennisetum purpureum* Schumach) promote plant growth and enhance salt tolerance of hybrid Pennisetum. Biotechnol. Biofuels 9 (1) (2016) 190.
[7] R. Sheibani-Tezerji, et al., The genomes of closely related *Pantoea ananatis* maize seed endophytes having different effects on the host plant differ in secretion system genes and mobile genetic elements. Front. Microbiol. 6 (2015) 440.
[8] E.L. Bicudo, et al., Nosocomial outbreak of *Pantoea agglomerans* in a pediatric urgent care center. Braz. J. Infect. Dis. 11 (2) (2007) 281–284.
[9] I. Boszczowski, et al., Nosocomial outbreak of *Pantoea agglomerans* bacteraemia associated with contaminated anticoagulant citrate dextrose solution: new name, old bug? J. Hosp. Infect. 80 (3) (2012) 255–258.
[10] J. Mardaneh, M.M. Dallal, Isolation, identification and antimicrobial susceptibility of *Pantoea* (Enterobacter) *agglomerans* isolated from consumed powdered infant formula milk (PIF) in NICU ward: first report from Iran. Iran J. Microbiol. 5 (3) (2013) 263–267.
[11] D.R. Zerbino, E. Birney, Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Res. 18 (5) (2008) 821–829.
[12] D. Hyatt, et al., Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinforma. 11 (2010) 119.
[13] J. Mistry, et al., Challenges in homology search: HMMER3 and convergent evolution of coiled-coil regions. Nucleic Acids Res. 41 (12) (2013), e121.
[14] R.D. Finn, et al., The Pfam protein families database: towards a more sustainable future. Nucleic Acids Res. 44 (D1) (2016) D279–D285.
[15] D. Roberts, labdsv: Ordination and Multivariate Analysis for Ecology. 2007.
[16] P. Horton, et al., WoLF PSORT: protein localization predictor. Nucl. Acids Res. 35 (2007) W585–W587 (Web Server issue).
[17] C. Rinke, et al., Insights into the phylogeny and coding potential of microbial dark matter. Nature 499 (7459) (2013) 431–437.
[18] A.P. Tomaras, et al., Attachment to and biofilm formation on abiotic surfaces by *Acinetobacter baumannii*: involvement of a novel chaperone-usher pili assembly system. Microbiology 149 (12) (2003) 3473–3484.
[19] L. Chen, et al., VFDB: a reference database for bacterial virulence factors. Nucleic Acids Res. 33 (Database issue) (2005) D325–D328.
[20] T.R. Costa, et al., Secretion systems in Gram-negative bacteria: structural and mechanistic insights. Nat. Rev. Microbiol. 13 (6) (2015) 343–359.
[21] S.J. Reid, V.R. Abratt, Sucrose utilisation in bacteria: genetic organisation and regulation. Appl. Microbiol. Biotechnol. 67 (3) (2005) 312–321.
[22] U.M. Cowgill, The tellurium content of vegetation. Biol. Trace Elem. Res. 17 (1988) 43–67.
[23] M.H. Saier Jr., et al., The transporter classification database. Nucleic Acids Res. 42 (Database issue) (2014) D251–D258.
[24] S. Conlan, et al., Single molecule sequencing to track plasmid diversity of hospital-associated carbapenemase-producing Enterobacteriaceae. Sci. Trans. Med. 6 (254) (2014) 254ra126.
[25] D. Field, et al., The minimum information about a genome sequence (MIGS) specification. Nat. Biotechnol. 26 (5) (2008) 541–547.
[26] C.L. Brady, et al., Emended description of the genus *Pantoea*, description of four species from human clinical samples, *Pantoea septica* sp. nov., *Pantoea eucrina* sp. nov., *Pantoea brenneri* sp. nov. and *Pantoea conspicua* sp. nov., and transfer of *Pectobacterium cypripedii* (Hori 1911) Brenner et al. 1973 emend. Hauben et al. 1998 to the genus as *Pantoea cypripedii* comb. nov. Int. J. Syst. Evol. Microbiol. 60 (Pt 10) (2010) 2430–2440.
[27] M. Ashburner, et al., Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat. Genet. 25 (1) (2000) 25–29.
[28] S. Kumar, G. Stecher, K. Tamura, MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. Mol. Biol. Evol. 33 (7) (2016) 1870–1874.