Supporting Information

Whole-Genome Shotgun Sequencing of Two β-Proteobacterial Species in Search of the Bulgecin Biosynthetic Cluster

Mark E. Horsman,†# Daniel R. Marous,†# Rongfeng Li,§ Ryan A. Oliver,§ Byungjin Byun,† Scott J. Emrich,‡ Bill Boggess,¶ Craig A. Townsend,* and Shahriar Mobashery†*

†Departments of Chemistry and Biochemistry, and ‡Computer Science and Engineering, University of Notre Dame, Notre Dame, Indiana 46556, USA
§Department of Chemistry, Johns Hopkins University, Baltimore, Maryland 21218, USA

# These authors contributed equally
Methods:

Classification of draft genome taxonomy. 16S rRNA sequence comparisons were conducted on RDP\textsuperscript{1} using the “RDP Naive Bayesian rRNA Classifier (v2.11)” classifier and the “RDP 16S rRNA training set 16” Taxonomical Hierarchy. The detected 16S sequences in the ATCC31363 and ATCC31433 assemblies were both categorized as genus \textit{Burkholderia} in origin with 100% confidence. (Note: this system also files \textit{Paraburkholderia graminis},\textsuperscript{2} the type species of \textit{Paraburkholderia},\textsuperscript{3} as \textit{Burkholderia} at 100% confidence) Broad genomic typing of ATCC 31363 and ATCC31433 were performed using RAST’s nearest neighbor genome tool\textsuperscript{4} and MASH genome distance estimates.\textsuperscript{5} RAST compares the SPAdes assembly annotations of candidate genomes against unduplicated and universal proteins within published genomes.

Table S1. RAST nearest neighbor bacterial genomes found on the RAST database

| ATCC 31433 |   |   |
|-------------|---|---|
| Genome name  | RAST Score | RAST Genome ID |
| Burkholderia vietnamiensis G4 | 509 | 269482.11 |
| Burkholderia vietnamiensis strain G4 | 449 | 269482.4 |
| Burkholderia cenocepacia J2315 | 437 | 216591.1 |

| ATCC 31363 |   |   |
|-------------|---|---|
| Genome name  | RAST Score | RAST Genome ID |
| Burkholderia sp. Ch1-1 | 523 | 243261.3 |
| Burkholderia xenovorans LB400 | 515 | 266265.5 |
| Burkholderia phymatum STM815 | 437 | 391038.7 |

\textit{Burkholderia vietnamiensis} G4, \textit{Burkholderia vietnamiensis} strain G4, and \textit{Burkholderia cenocepacia} J2315 are \textit{Burkholderia} species found in the \textit{Burkholderia cenocepacia} complex. \textit{Burkholderia sp. Ch1-1}, \textit{Burkholderia xenovorans} LB400, and \textit{Burkholderia phymatum} STM815 are listed as members of the recently proposed \textit{Paraburkholderia} genus.

In order to compare the draft genomes against custom subsets of the RefSeq genome set, the draft genome sequences were reduced to MASH sketches and compared against Ondov \textit{et al.}’s NCBI reference set of 54,118 NCBI RefSeq genomes.\textsuperscript{5}

For ATCC 31433, the comparison identified similarities to \textit{Burkholderia ubonensis} strain MSHR3999 (Bunnell \textit{et al.}, GenBank: JRFO00000000.1), \textit{Burkholderia ubonensis} MSMB22\textsuperscript{6} (Los Alamos Lab, GenBank: PRJNA239258), and \textit{Burkholderia ubonensis} Bu. (NZ_ABBE00000000).\textsuperscript{7} After creating individual sketches of each of these reference genomes and generating a table of pairwise MASH distances from ATCC 31433, MSHR3999 was selected as a close reference genome (MASH distance of 0.05), and Loveridge \textit{et al.}’s taxonomical assignment of ATCC 31433 as \textit{Burkholderia}\textsuperscript{8} was cross-validated. We propose the novel subspecies \textit{Burkholderia ubonensis} subsp. \textit{mesacidiphila}.

When ATCC 31363 was compared with the RefSeq genome sketch, it displayed larger MASH distances from all included reference sequences. ATCC 31363 showed moderate similarity to \textit{Burkholderia} sp. 9120 (Baltrus \textit{et al.}, GenBank: JQNA00100001.1), \textit{Paraburkholderia} CCGE1004, \textit{Paraburkholderia graminis} C4D1M, and \textit{Paraburkholderia} Ch1-1. Pairwise MASH distances demonstrated \textit{Burkholderia} sp. 9120 as a moderately similar genome reference for ATCC 31363 (MASH distance 0.164, two-way ANI 81.98\%). Due to the composition of the \textit{Burkholderia} sp. 9120 genome, we suggest that both sp. 9120 and ATCC 31363 are also members of the newly formed Paraburkholderia genus. As such, we propose the designation of ATCC 31363 as \textit{Paraburkholderia acidiphila}. 

\textsuperscript{1}Baltrus, S., \textit{et al.} (2017) GenBank: JQNA00100001.1
\textsuperscript{2}Los Alamos Lab, GenBank: PRJNA239258
\textsuperscript{3}Los Alamos Lab, GenBank: PRJNA239258
\textsuperscript{4}Ondov, B., \textit{et al.} (2014) GenBank: JRFO00000000.1
\textsuperscript{5}Loveridge, C., \textit{et al.} (2017) GenBank: NZ_ABBE00000000
\textsuperscript{6}Burkholderia ubonensis MSMB22: Los Alamos Lab, GenBank: PRJNA239258
\textsuperscript{7}Burkholderia ubonensis Bu: (NZ_ABBE00000000)
\textsuperscript{8}Burkholderia ubonensis subsp. mesacidiphila: loveridge, C., \textit{et al.} (2017) GenBank: NZ_ABBE00000000
\textsuperscript{9}Burkholderia ubonensis subsp. mesacidiphila: loveridge, C., \textit{et al.} (2017) GenBank: NZ_ABBE00000000
**Primary and Secondary metabolism gene annotation.** The final contig sets of both genomes were annotated by PATRIC RASTtk, antiSMASH,\(^\text{10}\) and NCBI PGAP\(^\text{11}\) to characterize genes involved in primary and secondary metabolism. FASTA sequences of MTZU00000000 and MTZV00000000 were submitted for secondary metabolism annotation by antiSMASH v3.0.5.\(^\text{10}\) For genes specific to the putative bulgecin cluster, predicted protein functions were assigned based on antiSMASH homology predictions for bulC, D, F, G and sat1. Annotations for bulA, B, H, and sat2 were determined from close matches from NCBI BLAST searches.

The GenBank files submitted to NCBI contain the *ab initio* PGAP annotations and supports supplemented with manual references to the PATRIC and antiSMASH matches for the sulfazecin and bulgecin-type pathways.\(^\text{8}\)

**Search for coding sequences uniquely conserved in genomes containing sulfazecin-type clusters.** Two close reference genomes for ATCC 31433 (*Burkholderia* MSHR3999, NCBI accession PRJNA260165, and *Burkholderia ubonensis* MSMB22,\(^\text{6}\) determined from MASH distance) were searched for analogs of *sbnB*, which encodes an L-ornithine cyclodeaminase-type protein. A homolog of *sbnB*, *sulG* has been shown to be essential for sulfazecin biosynthesis in ATCC31433.\(^\text{12}\) With no similar hits detected, these two reference genomes were designated as close genome-scale matches of ATCC 31433 that are devoid of a complete sulfazecin cluster.

The inventory of genes unique to the reported bulgecin-producers, ATCC 31363 and ATCC 31433, was generated using the output from PATRIC’s proteome comparison tool; by comparing the ATCC 31363 genome against the two reference genomes (*B. MSHR3999* and *B. ubonensis* MSMB22) as well as two monobactam cluster-containing genomes (ATCC 31433 and *B. gladiola* ATCC 10248), a list of genes found uniquely in monobactam-positive genomes was compiled (Table S5). We inventoried genes that had a significantly higher identity percentage (+30%) with genes in ATCC 31433 and *B. gladiola* ATCC 10248 compared to the two *B. ubonensis* strains. Relatively few proteins met this criterion (70 of 6535), 58 of which did not cluster with other distinct genes based on the order of genes in ATCC 31363. One large block (7 singles and a contiguous block of 3 within 28 genes of the ATCC 31363 genome) of high identity proteins featured prominently, having genes adjacent to each other in the ATCC 31363 genome, and corresponded to many of the previously identified gene in the sulfazecin cluster as well as gene directly downstream.

**Identification of possible promoter sites.** The contig containing the monobactam cluster in ATCC 31433 was input to the genome2D webserver (PePPER: Prokaryote promoters, [http://pepper.molgenrug.nl/index.php/prokaryote-promoters]\(^\text{13}\)) to identify possible promoter sites within the cluster. The possible promoter sites are shown in Figure S1.

![Figure S1](http://pepper.molgenrug.nl/index.php/prokaryote-promoters)

**Figure S1.** Potential promoter sites within *sulD-sat1* of ATCC 31433.

The locations of potential promoters, indicated by black lines, were determined using the genome 2D software.\(^\text{13}\) Promoters reading in the reverse direction are elevated above those reading left to right. Regions containing high densities of possible promoter sites are indicated by asterisks.
**Identifying variants of gene-regions from sulI to sat2 in previously reported genomes.** On the PATRIC database, all genomes containing at least one copy of “Gll3516 protein” (PATRIC global family ID PGF_00007735, sulN) and “Glycosyl transferase, group 2 family” (PATRIC global family PGF_00009528, bulH) were combined to form a genome group. This genome group was ranked by blast similarity score to the sulG to BulH region of ATCC 31363.

The ordered set of genomes was inspected at the genome region containing a PATRIC annotated SerC-type phosphoserine aminotransferase (PATRIC local family plf_32008_00003145, bulG), and the genome region was categorizes for the presence of sulG-sat1 gene orthologues and common alternate structures.

**Phyre2 and I-Tasser Methods.** Review of the predicted BulE protein sequence using protein-threading programs Phyre2 and I-Tasser gave indications that BulE may adopt the fold of an NADH-dependent dehydrogenase similar in structure to PDB 4PLC, 4ND4, and 3QVO.

![Figure S2](image)

**Figure S2.** Structural predictions for BulE (B) compared against documented apicomplexan dehydrogenase 4PLC (A).

**Ordering of the contigs.** After reference genomes were identified for ATCC 31433 and ATCC 31363, *B. ubonensis* MSHR3999 and *B. sp.* 9120, respectively, the sequencing reads for both bulgecin-producing species were reassembled in PATRIC BRC using the corresponding similar genome for each as a reference assembly. The contigs of both new assemblies were then ordered to the sequences of their reference genomes using Mauve. Genome annotation was performed in PATRIC MultiGenomeComparison tool. Figures S3 and S4 show the similarity between ATCC 31433 and ATCC 31363 and their corresponding reference genomes. The more prominent diagonal line in the ATCC 31433/*B. ubonensis* MSHR3999 comparison (Figure S3) indicates that these two genomes have a higher degree of similarity than ATCC 31363 and *B. sp.* 9120 (Figure S4).
Figure S3. Organization of ATCC 31433 contigs using the published contig order of B. ubonensis MSHR3999

The contigs of ATCC 31433 were reordered using Mauve and a comparison was made to the reference genome of B. ubonensis MSHR3999 using the RAST sequence-based comparison BlastDotPlot feature. Visual inspection confirmed that contig 10 (as numbered in NCBI submission) is sorted two contigs out of order and should be repositioned between contig 007 and 009. Additionally, contigs 24 and 45 should be inverted to complete the manual reorder process. Parts A and B show before and after reordering, respectively. The prominent diagonal line demonstrates the similarity of the overall architecture of the two genomes. The purple box highlights a unique monobactam pathway on ATCC 31433 contig 11.

Figure S4. Comparison of ATCC 31363 with Burkholderia sp. 9120

The contigs of ATCC 31363 were reordered using Mauve’s local collinear sequence DNA sequence block method; a comparison between the resulting assembly and the reference genome of B. sp. 9120 using the RAST MultiGenomeComparison DotBlot feature. The diagonal line highlights which sections of the genomes align well. The black oval highlights a unique monobactam pathway on ATCC 31363 contig 6, which is discussed below in section 3.5.

The draft genomes can also be compared to their reference genomes in a complementary coding sequence-based ranking system. The draft genome contigs can be scored by the number of bidirectional best matches of genes found in each portion of the reference genome.
Table S2. ATCC 31433 Genome contig association table.

The groups of ATCC 31433 contigs are suspected to co-localize based on gene identity with MSHR3999 RAST bidirectional gene matches. Assigned contigs must have at least 60% of bidirectional matches in cognate reference contig. The recent ATCC 31433 assembly by Loveridge et al. is used as an anchor.\(^8\)

| ATCC 31433 Chromosome\(^8\) | ATCC 31433 contig (this study) | MSHR3999 contig |
|-----------------------------|--------------------------------|-----------------|
| 1                           | 30-54, 79-132                  | 2, 7            |
| 2                           | 1-29, 69-75                    | 1, 5, 6         |
| 3                           | 55-65                          | 3               |
| Plasmid                     | 67                             | 4               |

This analysis also suggests the possible configuration of the contigs in MSHR3999.

Table S3. Association table between the chromosomes of Burkholderia sp. 9120 and the contigs of ATCC 31363

| Burkholderia sp. 9120 Chromosome | ATCC 31363 contig |
|----------------------------------|-------------------|
| 1                                | 3-8               |
| 2                                | -                 |
| Conflicted contigs                | 1, 2              |
| Unassociated                      | 9-11              |

Table S2 also outlines the possible configuration of MSHR3999. As shown in Table S3 and Figure S4, the gene order between 31363 and sp. 9120 varies to a larger extent than the relative gene order of MSHR3999 and ATCC 31433 (Table S2).

Table S4. Putative secondary metabolite pathways detected by antiSMASH. Sulfazecin/bulgecin cluster in bold.

|                          | ATCC 31433 | ATCC 31363 |
|--------------------------|------------|------------|
| Arylpolyene              | 2          | 2          |
| Bacteriocin              | 3          | 2          |
| **Blactam-Arylpolyene-Ectoine-Nrps** | | **1** |
| Butyrolactone            | 1          |            |
| Ectoine-Arylpolyene-Nrps | 1          |            |
| Hserlactone-Terpene      | 1          |            |
| Lasso peptide            | 1          |            |
| Lantipeptide             | 1          |            |
| NRPS                     | 9          | 5          |
| Nrps-T1pks               | 2          |            |
| Nrps-T1pks-Transatpks    | 1          |            |
| Other                    | 5          |            |
| Otherks-Nrps-T1pks-Hserlactone | 1    |            |
| Phosphonate              | 1          | 1          |
| Siderophore              | 1          |            |
| Terpene                  | 3          | 3          |
| TransAT PKS              | 2          |            |
| **Total**                | **31**     | **17**     |
Table S5. List of genes found uniquely in monobactam-positive genomes. Sulfazecin/bulgecin region in bold.

| ATCC 31363 gene number | Annotated gene function (for ATCC 31363) |
|------------------------|-----------------------------------------|
| 78                     | hypothetical protein                     |
| 137                    | Putative transmembrane protein           |
| 190                    | 2-Oxobutyrate oxidase, putative          |
| 203                    | probable ring-cleaving dioxygenase PA0880|
| 221                    | FIG00454099: hypothetical protein        |
| 740                    | Methionine sulfoxide reductase-associated methionine-rich protein |
| 890                    | Predicted membrane protein               |
| 1047                   | FIG00462729: hypothetical protein        |
| 1226                   | Glycine betaine/L-proline transport system permease protein ProW (TC 3.A.1.12.1) |
| 1294                   | Lipase 1 (EC 3.1.1.3)                    |
| 1428                   | Uncharacterized protein conserved in bacteria |
| 1445                   | hypothetical protein                     |
| 1719                   | FIG00455658: hypothetical protein        |
| 1753                   | cAMP-binding proteins - catabolite gene activator and regulatory subunit of cAMP-dependent protein kinases |
| 1775                   | FIG00456194: hypothetical protein        |
| 1793                   | Superfamily I DNA and RNA helicases      |
| 1905                   | probable competence protein              |
| 1993                   | Lipase 1 (EC 3.1.1.3)                    |
| 1996                   | hypothetical protein                     |
| 2014                   | hypothetical protein                     |
| 2112                   | Phosphoglycerate mutase family protein   |
| 2207                   | FIG00453291: hypothetical protein        |
| 2688                   | hypothetical protein                     |
| 2852                   | Putative permease often clustered with de novo purine synthesis |
| 3259                   | Arsenic resistance protein ArsH          |
| 3266                   | hypothetical protein                     |
| 3719                   | FIG00454007: hypothetical protein        |
| 3996                   | Isoprenylcysteine carboxyl methyltransferase |
| 4008                   | 2-Oxobutyrate oxidase, putative          |
| 4124                   | hypothetical protein                     |
| 4332                   | Hydroxyacylglutathione hydrolase (EC 3.1.2.6) |
| 4643                   | Proline-rich region                      |
| 4895                   | hypothetical protein                     |
| 4905                   | L-ectoine synthase (EC 4.2.1.108)        |
| 4912                   | Glycosyl transferase, group 2 family     |
| 4915                   | FIG00453866: hypothetical protein        |
| 4918                   | FIG00453797: hypothetical protein        |
| 4919                   | FIG00454727: hypothetical protein        |
| 4920                   | FIG00454024: hypothetical protein        |
Gll3516 protein

FIG00454264: hypothetical protein

Peptide synthetase

putative membrane protein

Putative transporter

hypothetical protein

putative transmembrane anti-sigma factor

4-hydroxy-2-oxovalerate aldolase (EC 4.1.3.39)

Transcriptional regulator, AcrR family

FIG00454545: hypothetical protein

putative esterase

NAD synthetase (EC 6.3.1.5) / Glutamine amidotransferase chain of NAD synthetase

Superfamily I DNA and RNA helicases

PROBABLE LIPOPROTEIN

hypothetical protein

FIG00457602: hypothetical protein

Similar to Agmatine deiminase

Agmatine deiminase (EC 3.5.3.12)

Outer membrane protein

Xaa-Pro aminopeptidase (EC 3.4.11.9)

Lipase 1 (EC 3.1.1.3)

hypothetical protein

hypothetical protein

Citrate/H+ symporter of CitMHS family

Protein of unknown function DUF1446

Small uncharacterized protein Bpro_4170

hypothetical protein

Cyanate hydratase (EC 4.2.1.104)

Transcriptional regulator, AcrR family

Purine-cytosine permease
Table S6 Genomes indexed on PATRIC BRC to contain an equivalent to SulN, BulG, BulH, and SulG (SbnB)

| Genome                                      | Genome ID | Accession | PATRIC ID position of SbnB | Start |
|---------------------------------------------|-----------|-----------|-----------------------------|-------|
| *Burkholderia gladioli* NBRC 13700          | 1218079   | BBJG01000047 | fig|1218079.3.peg.4590 | 92490 |
| *Chromobacterium* sp. F49                   | 1777131   | LQNP01000019 | fig|1777131.3.peg.2509 | 38349 |
| *Chromobacterium subtsugae* strain MWU2920  | 251747.1  | LCWP01000016 | fig|251747.14.peg.3426 | 97315 |
| *Chromobacterium subtsugae* strain MWU2387  | 251747.2  | LCWR01000011 | fig|251747.15.peg.630  | 3820  |
| *Chromobacterium subtsugae* strain MWU2576  | 251747.2  | LCWQ01000021 | fig|251747.16.peg.3875 | 51101 |
| *Chromobacterium subtsugae* strain MWU3525  | 251747.2  | LCWO01000019 | fig|251747.17.peg.3501 | 38321 |
| *Chromobacterium* subtsugae strain PRAA4-1 | 251747.2  | JYKA01000020 | fig|251747.20.peg.1992 | 38320 |
| *Chromobacterium subtsugae* F49             | 251747.7  | JWJN01000024 | fig|251747.7.peg.3314  | 38340 |
| *Burkholderia gladioli* strain ATCC 25417   | 28095.11  | KN150849   | fig|28095.11.peg.4179  | 4253427 |
| *Burkholderia gladioli* strain ATCC 10248   | 28095.12  | CP009323   | fig|28095.12.peg.2978  | 3045230 |
| *Burkholderia gladioli* strain SN82F6       | 28095.13  | LAVN01000006 | fig|28095.13.peg.3590 | 85952  |
| *Chromobacterium violaceum* strain Cv017    | 536.13    | LKIW01000108 | fig|536.13.peg.2705  | 39312  |
Table S8: Extended comparison of bulA–H and sat1–2 in Table 1.

| Gene Name | Similar function | Length | Identity/Similarity | Locus Tag |
|-----------|------------------|--------|---------------------|-----------|
|           |                  | ATCC 31433 | ATCC 31363 | ATCC 10248 | ATCC 31363 | ATCC 10248 | ATCC 31433 | ATCC 31363 | ATCC 10248 |
| bulA      | ATP-grasp family | 259    | 265    | 267    | 85/90 | 84/88 | BZL54_04205 | BWP39_23675 | BM43_2804 |
| bulB      | Sulfotransferase ThDP-dependent Transketolase, N-terminal thiamine diphosphate binding domain | 341    | 340    | 313    | 77/85 | 67/76 | BZL54_04210 | BWP39_23670 | BM43_2805 |
| bulC      | Transketolase, C-terminal pyrimidine binding domain | 272    | 272    | 272    | 85/89 | 81/86 | BZL54_04215 | BWP39_23665 | BM43_2806 |
| bulD      | Dehydrogenase (see SI) | 308    | 302    | 317    | 85/88 | 83/88 | BZL54_04220 | BWP39_23660 | BM43_2807 |
| bulE      | Acyltransferase 3 Phosphoserine transaminase Glycosyl transferase group 2 family | 355    | 363    | 369    | 68/77 | 66/73 | BZL54_04225 | BWP39_23655 | BM43_2808 |
| bulF      | Sulfate adenyllytransferase small subunit | 375    | 375    | 376    | 76/86 | 80/88 | BZL54_04230 | BWP39_23650 | BM43_2809 |
| bulG      | Sulfate adenyllytransferase large subunit | 364    | 364    | 364    | 78/87 | 82/89 | BZL54_04235 | BWP39_23645 | BM43_2810 |
| bulH      | Sulfate adenyllytransferase small subunit | 270    | 299    | 270    | 84/91 | 80/88 | BZL54_04240 | BWP39_23640 | BM43_2811 |
| sat2      | Sulfate adenyllytransferase large subunit | 320    | 320    | 320    | 91/93 | 94/95 | BZL54_13320 | BWP39_23635 | BM43_2812 |
| sat1      | Sulfate adenyllytransferase large subunit | 438    | 437    | 438    | 85/90 | 88/92 | BZL54_13325 | BWP39_23630 | BM43_2813 |
Supplemental References

1. Wang, Q., Garrity, G. M., Tiedje, J. M., and Cole, J. R. (2007) Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy, *Applied and Environmental Microbiology* 73, 5261-5267.

2. Viallard, V., Poirier, I., Cournoyer, B., Haurat, J., Wiebkin, S., Ophel-Keller, K., and Balandreau, J. (1998) Burkholderia graminis sp. nov., a rhizospheric Burkholderia species, and reassessment of Pseudomonas phenazinium, Pseudomonas pyrocinia and Pseudomonas glathei as Burkholderia, *International Journal of Systematic Bacteriology* 48, 549-563.

3. Sawana, A., Adeolu, M., and Gupta, R. S. (2014) Molecular signatures and phylogenomic analysis of the genus Burkholderia: proposal for division of this genus into the emended genus Burkholderia containing pathogenic organisms and a new genus Paraburkholderia gen. nov harboring environmental species, *Frontiers in Genetics* 5.

4. Overbeek, R., Olson, R., Pusch, G. D., Olsen, G. J., Davis, J. J., Disz, T., Edwards, R. A., Gerdes, S., Parrello, B., Shukla, M., Vonstein, V., Wattam, A. R., Xia, F., and Stevens, R. (2014) The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST), *Nucleic Acids Research* 42, D206-D214.

5. Ondov, B. D., Treangen, T. J., Melsted, P., Mallonee, A. B., Bergman, N. H., Koren, S., and Phillippy, A. M. (2016) Mash: fast genome and metagenome distance estimation using MinHash, *Genome Biology* 17, 132.

6. Johnson, S. L., Bishop-Lilly, K. A., Ladner, J. T., Daligault, H. E., Davenport, K. W., Jaissle, J., Frey, K. G., Koroleva, G. I., Bruce, D. C., Coyne, S. R., Broomall, S. M., Li, P.-E., Teshima, H., Gibbons, H. S., Palacios, G. F., Rosenzweig, C. N., Redden, C. L., Xu, Y., Minogue, T. D., and Chain, P. S. (2015) Complete Genome Sequences for 59 Burkholderia Strains, Both Pathogenic and Near Neighbor, *Genome Announcements* 3, e00159-00115.

7. Mukhopadhyay, S., Thomason, M. K., Lentz, S., Nolan, N., Willner, K., Gee, J. E., Glass, M. B., Inglis, T. J. J., Merritt, A., Levy, A., Sozhamannan, S., Mateczun, A., and Read, T. D. (2010) High-Redundancy Draft Sequencing of 15 Clinical and Environmental Burkholderia Strains, *Journal of Bacteriology* 192, 6313-6314.

8. Loveridge, E. J., and Jones, C. (2017) Reclassification of the specialized metabolite producer Pseudomonas mesoacidiphila ATCC 31433 as a member of the Burkholderia cepacia complex, *Journal of Bacteriology*.

9. Goris, J., Konstantinidis, K. T., Klappenbach, J. A., Coenye, T., Vandamme, P., and Tiedje, J. M. (2007) DNA-DNA hybridization values and their relationship to whole-genome sequence similarities, *International Journal of Systematic and Evolutionary Microbiology* 57, 81-91.

10. Weber, T., Blin, K., Duddela, S., Krug, D., Kim, H. U., Bruccoleri, R., Lee, S. Y., Fischbach, M. A., Muller, R., Wohlleben, W., Breitling, R., Takano, E., and Medema, M. H. (2015) antiSMASH 3.0-a comprehensive resource for the genome mining of biosynthetic gene clusters, *Nucleic Acids Research* 43, W237-W243.

11. Tatusova, T., DiCuccio, M., Badretdin, A., Chetverin, V., Nawrocki, E. P., Zaslavsky, L., Lomsadze, A., Pruitt, K., Borodovsky, M., and Ostell, J. (2016) NCBI prokaryotic genome annotation pipeline, *Nucleic Acids Research* 44, 6614-6624.
12. Li, R. F., Oliver, R. A., and Townsend, C. A. (2017) Identification and Characterization of the Sulfazecin Monobactam Biosynthetic Gene Cluster, *Cell Chemical Biology* 24, 24-34.

13. Baerends, R. J. S., Smits, W. K., de Jong, A., Hameon, L. W., Kok, J., and Kuipers, O. P. (2004) Genome2D: a visualization tool for the rapid analysis of bacterial transcriptome data, *Genome Biology* 5.

14. Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N., and Sternberg, M. J. E. (2015) The Phyre2 web portal for protein modeling, prediction and analysis, *Nature Protocols* 10, 845-858.

15. Zhang, Y. (2008) I-TASSER server for protein 3D structure prediction, *BMC Bioinformatics* 9.

16. Yang, J. Y., and Zhang, Y. (2015) I-TASSER server: new development for protein structure and function predictions, *Nucleic Acids Research* 43, W174-W181.

17. Boucher, J. I., Jacobowitz, J. R., Beckett, B. C., Classen, S., and Theobald, D. L. (2014) An atomic-resolution view of neofunctionalization in the evolution of apicomplexan lactate dehydrogenases, *Elife* 3.

18. Darling, A. E., Mau, B., and Perna, N. T. (2010) progressiveMauve: Multiple Genome Alignment with Gene Gain, Loss and Rearrangement, *Plos One* 5.

19. Snauwaert, I., Stragier, P., De Vuyst, L., and Vandamme, P. (2015) Comparative genome analysis of Pediococcus damnosus LMG 28219, a strain well-adapted to the beer environment, *BMC Genomics* 16.