

1 Draft genome sequences of five Proteobacteria isolated from Lechuguilla Cave, New Mexico, USA:  
2 Insights into taxonomy and quorum-sensing

3

4 Han Ming Gan<sup>1,2\*</sup>, Peter C. Wengert<sup>4</sup>, Hazel A. Barton<sup>3</sup>, André O. Hudson<sup>4</sup>, Michael A. Savka<sup>4</sup>

5 <sup>1</sup> Centre for Integrative Ecology, School of Life and Environmental Sciences, Deakin University, Geelong,  
6 Victoria, Australia

7 <sup>2</sup> Deakin Genomics Centre, Deakin University, Geelong, Victoria, Australia

8 <sup>3</sup> Department of Biology, University of Akron, Akron, Ohio, USA

9 <sup>4</sup> The Thomas H. Gosnell School of Life Sciences, Rochester Institute of Technology, Rochester, New York,  
10 USA

11

12 \*Corresponding author

13 Name: Han Ming Gan

14 Email: han.gan@deakin.edu.au

15

16 **Abstract.** Genomic resources remain scarce for bacteria isolated from oligotrophic caves. We  
17 sequenced the genomes of five Proteobacteria isolated from Lechuguilla Cave in New Mexico, USA.  
18 Genome-based phylogeny indicates that each strain belongs to a distinct genus. Two *Rhizobiaceae*  
19 isolates possess the genomic potential for the biosynthesis of acyl-homoserine lactone.

20 **Main Text:**

21 Adequate genomic resources are crucial for the understanding of adaptation for culturable  
22 microbes in nutrient-limited cave environments (1). To date, bacterial genomes isolated from caves are  
23 poorly represented in the public databases (2-5). Here, we report the genomes of five Lechuguilla cave  
24 isolates and used genome-based phylogeny to refine their taxonomic assignment. We also identified  
25 genomic potential for the biosynthesis of cell-cell communication signal in two *Rhizobiaeae* isolates and  
26 confirmed the predicted phenotype using an *Agrobacterium tumefaciens* reporter assay (6).

27 Initial isolation of the bacterial strains from remote sample sites in Lechuguilla Cave was  
28 previously described by Bhullar *et al* (7). Strains were grown in half strength tryptic soy broth with  
29 shaking at 30°C for 2 days. DNA extraction used the GenElute bacterial genomic DNA kit (MilliporeSigma,  
30 St Louis, MO). Sequencing library was generated using the tagmentation-based Illumina Nextera XT DNA  
31 sample prep kit (Illumina, San Diego, CA) and sequenced on an Illumina MiSeq (run configuration of 2 ×

32 150 bp). The paired-end reads were adapter-trimmed and assembled with Trimmomatic v0.36 (8) and  
33 Unicycler v0.4.7 (9), respectively, using the default setting.

34 The genome of LC387 was assembled into a single contig (Table 1). Based on BLASTN alignment  
35 (10), its full-length 16S rRNA sequence is 100% identical to that of *Afipia massiliensis* CCUG 45153<sup>T</sup>  
36 (RefSeq Accession: NR\_025646.1). For the generation of genome-based phylogeny using GToTree v1.2.1  
37 (11), genome assemblies of bacterial type strains exhibiting high 16S rRNA gene sequence similarity to  
38 the cave isolates were downloaded and included in the pipeline. Based on phylogenomic clustering,  
39 confident taxonomic assignment to the genus level was obtained for LC34, LC103 and LC458 (Figure 1A).  
40 The basal placement of LC148 in the *Neorhizobium* clade suggests that it is a divergent member within  
41 the genus or a member of an undescribed genus. Therefore, strain LC148 was classified as  
42 *Rhizobacteriaceae* sp. pending future taxonomic investigation.

43 We used a previously described Hidden Markov Model (HMM) approach (12) to identify the  
44 genomic potential for biosynthesis of acyl-homoserine lactone (AHL) molecules involved in the  
45 regulation of gene expression in response to cell density (13). AHL synthase homologs (LuxI) were  
46 identified in strains LC34 (GenBank Protein ID: TKT57503.1) and LC148 (GenBank Protein IDs:  
47 TKT46115.1 and TKT66962.1). The genes coding for these proteins were similarly classified as *luxI*  
48 homologs by AntiSMASH 4 and the NCBI prokaryotic genome annotation pipeline (14, 15). To confirm  
49 the predicted phenotype, LC34 and LC148 were grown in YM media and ethyl acetate extracts from the  
50 cultures were tested for the presence of AHL signals using the AHL-dependent biosensor *Agrobacterium*  
51 *tumefaciens* NTL4(pZLR4) (16). LC34 produced one type of AHL signal with retardation factor (RF) that is  
52 similar to 3-oxo-C8. LC148 produced two distinct AHLs, one with RF similar to C6 and the other with RF  
53 smaller than the included AHL standards (Figure 1B). Future work investigating the role of quorum  
54 sensing in these cave isolates through transposon mutagenesis (17), targeted gene deletion (18) or  
55 transcriptome sequencing (19) will be informative in understanding bacteria from cave environments.

56 **Data availability.** The raw Illumina paired-end reads and genome assemblies have been deposited in  
57 GenBank under the BioProject numbers as listed in Table 1. Bacterial strains can be requested from  
58 Professor Michael A. Savka, Rochester Institute of Technology, New York, USA.

59 **Acknowledgements.** The authors acknowledge the Thomas H. Gosnell School of Life Sciences (GSOLS)  
60 and the College of Science (COS) at the Rochester Institute of Technology (RIT) for ongoing support.  
61 PCW was supported by a 2019 COS Summer Undergraduate Research Fellowship from RIT.

62 **Competing Interests.** The authors have declared that no competing interests exist.

63

64

65

66

67 Table 1. Genome assembly metrics and data availability

BioProject	Assembly Accession	Organism	Strain	Number of reads	Number of bases (Mb)	Estimated genome coverage (x)	Genome Size (bp)	GC content (%)	No. of contigs (>500 bp)	N <sub>50</sub> length (bp)
<a href="#">PRJNA281560</a>	<a href="#">LBCQ02</a>	<i>Aquamicrobium</i> sp.	LC103	1,246,649	417	69.5	6,044,022	64.06	34	330,476
<a href="#">PRJNA532460</a>	<a href="#">SZVV01</a>	<i>Rhizobiaceae</i> sp.	LC148	399,832	110	18.8	5,822,494	61.45	86	124,816
<a href="#">PRJNA281572</a>	<a href="#">LBHX02</a>	<i>Agrobacterium</i> sp.	LC34	843,163	247	46.6	5,303,297	59.35	36	377,718
<a href="#">PRJNA281679</a>	<a href="#">LBIA02</a>	<i>Afipia massiliensis</i>	LC387	777,637	216	48.0	4,495,198	61.35	1	4,495,198
<a href="#">PRJNA281681</a>	<a href="#">LBHW02</a>	<i>Achromobacter</i> sp.	LC458	1,286,382	448	68.4	6,547,009	64.55	102	143,492

68

## 69 Figure Legends

70 Figure 1. (A) Maximum likelihood tree based on the concatenated alignment of 119 conserved  
71 proteobacteria single-copy gene sets generated using the default setting of GToTree v1.2.1. Branch  
72 lengths indicate the number of substitutions per site while node labels are Shimodaira-Hasegawa (SH)  
73 local support values computed in FastTree2 (20). (B) Separation of AHL molecules from the LC34 and  
74 LC148 extracts by  $C_{18}$  reversed-phase thin-layer plate developed with methanol/water (60:40, vol/vol).  
75 Standards include: C6, N-hexanoyl-L-homoserine lactone; C6-O, N-(3-oxohexanoyl)-L-homoserine  
76 lactone; C8-O, N-(3-oxo-octanoyl)-L-homoserine lactone.

77

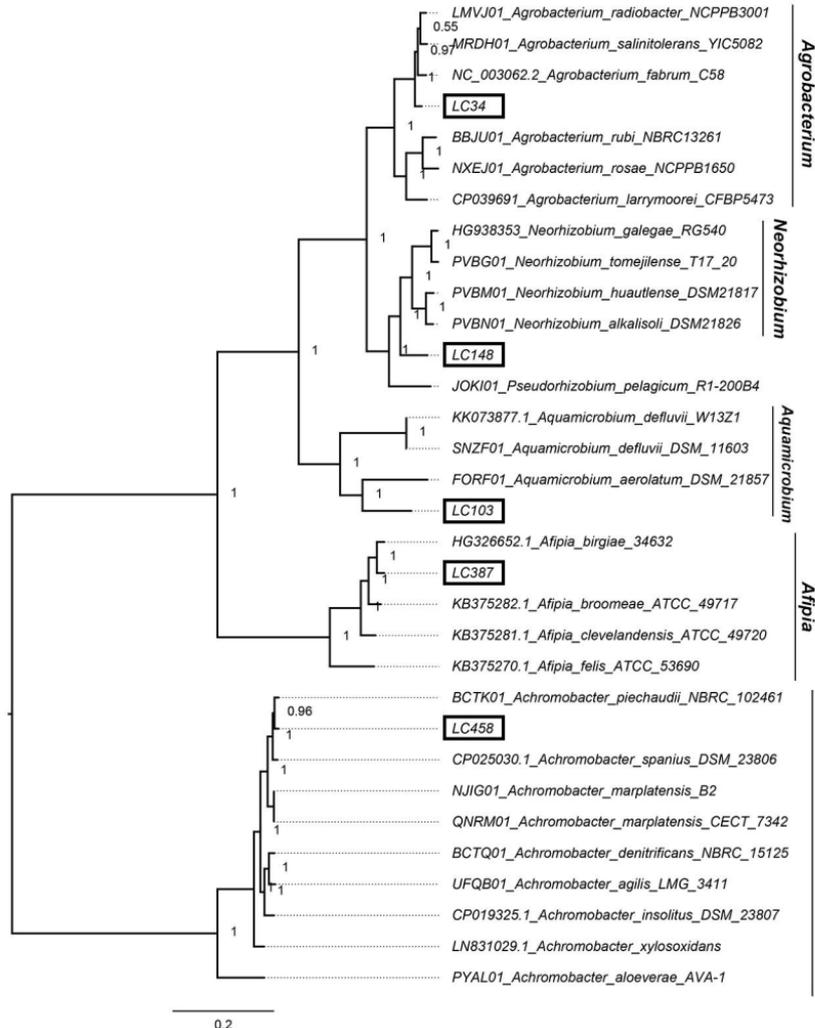
## 78 References

- 79 1. Barton MD, Petronio M, Giarrizzo JG, Bowling BV, Barton HA. 2013. The genome of  
80 *Pseudomonas fluorescens* strain R124 demonstrates phenotypic adaptation to the mineral  
81 environment. *J Bacteriol* 195:4793-803.
- 82 2. Gosse JT, Ghosh S, Sproule A, Overy D, Cheeptham N, Boddy CN. 2019. Whole Genome  
83 Sequencing and Metabolomic Study of Cave *Streptomyces* Isolates ICC1 and ICC4. *Frontiers in*  
84 *Microbiology* 10.
- 85 3. Pawlowski AC, Westman EL, Koteva K, Waglechner N, Wright GD. 2018. The complex resistomes  
86 of Paenibacillaceae reflect diverse antibiotic chemical ecologies. *The ISME Journal* 12:885-897.
- 87 4. Pawlowski AC, Wang W, Koteva K, Barton HA, McArthur AG, Wright GD. 2016. A diverse intrinsic  
88 antibiotic resistome from a cave bacterium. *Nature Communications* 7:13803.
- 89 5. Gan HY, Gan HM, Tarasco AM, Busairi NI, Barton HA, Hudson AO, Savka MA. 2014. Whole-  
90 genome sequences of five oligotrophic bacteria isolated from deep within Lechuguilla Cave, New  
91 Mexico. *Genome Announc* 2:e01133-14.
- 92 6. Scott RA, Weil J, Le PT, Williams P, Fray RG, von Bodman SB, Savka MA. 2006. Long- and Short-  
93 Chain Plant-Produced Bacterial N-Acyl-Homoserine Lactones Become Components of  
94 Phyllosphere, Rhizosphere, and Soil. *Molecular Plant-Microbe Interactions* 19:227-239.
- 95 7. Bhullar K, Waglechner N, Pawlowski A, Koteva K, Banks ED, Johnston MD, Barton HA, Wright GD.  
96 2012. Antibiotic resistance is prevalent in an isolated cave microbiome. *PloS one* 7:e34953-  
97 e34953.
- 98 8. Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence  
99 data. *Bioinformatics* 30:2114-2120.
- 100 9. Wick RR, Judd LM, Gorrie CL, Holt KE. 2017. Unicycler: resolving bacterial genome assemblies  
101 from short and long sequencing reads. *PLoS computational biology* 13:e1005595.
- 102 10. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *J*  
103 *Mol Biol* 215:403-10.
- 104 11. Lee MD. 2019. GToTree: a user-friendly workflow for phylogenomics. *Bioinformatics*  
105 doi:10.1093/bioinformatics/btz188.
- 106 12. Gan HM, Gan HY, Ahmad NH, Aziz NA, Hudson AO, Savka MA. 2015. Whole genome sequencing  
107 and analysis reveal insights into the genetic structure, diversity and evolutionary relatedness of  
108 luxI and luxR homologs in bacteria belonging to the Sphingomonadaceae family. *Frontiers in*  
109 *Cellular and Infection Microbiology* 4.
- 110 13. Miller MB, Bassler BL. 2001. Quorum sensing in bacteria. *Annual Reviews in Microbiology*  
111 55:165-199.

- 112 14. Blin K, Wolf T, Chevrette MG, Lu X, Schwalen CJ, Kautsar SA, Suarez Duran HG, de Los Santos EL,  
113 Kim HU, Nave M. 2017. antiSMASH 4.0—improvements in chemistry prediction and gene cluster  
114 boundary identification. *Nucleic acids research* 45:W36-W41.
- 115 15. Tatusova T, DiCuccio M, Badretdin A, Chetvernin V, Nawrocki EP, Zaslavsky L, Lomsadze A, Pruitt  
116 KD, Borodovsky M, Ostell J. 2016. NCBI prokaryotic genome annotation pipeline. *Nucleic Acids  
117 Res* 44:6614-24.
- 118 16. Lowe N, Gan HM, Chakravartty V, Scott R, Szegedi E, Burr TJ, Savka MA. 2009. Quorum-sensing  
119 signal production by *Agrobacterium vitis* strains and their tumor-inducing and tartrate-catabolic  
120 plasmids. *FEMS microbiology letters* 296:102-109.
- 121 17. Gan HM, Ibrahim Z, Shahir S, Yahya A. 2011. Identification of genes involved in the 4-  
122 aminobenzenesulfonate degradation pathway of *Hydrogenophaga* sp. PBC via transposon  
123 mutagenesis. *FEMS microbiology letters* 318:108-114.
- 124 18. Marx CJ. 2008. Development of a broad-host-range *sacB*-based vector for unmarked allelic  
125 exchange. *BMC research notes* 1:1-1.
- 126 19. Wagner VE, Gillis RJ, Iglewski BH. 2004. Transcriptome analysis of quorum-sensing regulation  
127 and virulence factor expression in *Pseudomonas aeruginosa*. *Vaccine* 22:S15-S20.
- 128 20. Price MN, Dehal PS, Arkin AP. 2010. FastTree 2—approximately maximum-likelihood trees for  
129 large alignments. *PLoS one* 5:e9490.

130

A



B

