Whole-genome sequencing of five proteobacteria isolated from Lechuguilla Cave, New Mexico, USA: Insights into taxonomy and quorum-sensing

Han Ming Gan[1,2*], Peter C. Wengert[4], Hazel A. Barton[3], André O. Hudson[4], Michael A. Savka[4]

[1] Centre for Integrative Ecology, School of Life and Environmental Sciences, Deakin University, Geelong, Victoria, Australia

[2] Deakin Genomics Centre, Deakin University, Geelong, Victoria, Australia

[3] Department of Biology, University of Akron, Akron, Ohio, USA

[4] The Thomas H. Gosnell School of Life Sciences, Rochester Institute of Technology, Rochester, New York, USA

*Corresponding author

Name: Han Ming Gan

Email: han.gan@deakin.edu.au

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

Main Text:

Adequate genomic resources are crucial for the understanding of adaptation for culturable microbes in nutrient-limited cave environments (1). To date, bacterial genomes isolated from caves are poorly represented in the public databases (2-5). Here, we sequenced the genomes of five non-clonal Lechuguilla cave isolates and used genome-based phylogeny to refine their taxonomic assignment that was previously based on the variable regions 3 and 4 of the 16S rRNA gene. We also identified genomic potential for the biosynthesis of cell-cell communication signal in two *Rhizobiaceae* isolates and confirmed the predicted phenotype using an *Agrobacterium tumefaciens* reporter assay (6).

Bacteria were grown in half strength tryptic soy broth with shaking at 30°C for 2 days. DNA extraction was performed using the GenElute bacterial genomic DNA kit (MilliporeSigma, St Louis, MO). The gDNA was processed with the Illumina Nextera XT DNA sample prep kit (Illumina, San Diego, CA) according to the manufacturer's instructions and sequenced on an Illumina MiSeq (run configuration of 2 × 150 bp). The paired-end reads were adapter-trimmed with Trimmomatic v0.36 (7) and assembled

using Unicycler v0.4.7 (8). The assembled genomes range from 4.5 to 6.5 Mb with N_{50} length of more than 100 kb (Table 1).

The genome of strain LC387 was assembled into a single contig (Table 1). Its full-length 16S rRNA sequence is 100% identical to that of *Afipia massiliensis* CCUG 45153 ^{T} (RefSeq Accession: NR_025646.1), making LC387 the first member of *A. massiliensis* with genome availability. Using phylogenomic approach, GToTree v1.2.1 (9), confident taxonomic assignment to the genus level was obtained for three of the four remaining strains (Figure 1A). The basal placement of LC148 in the *Neorhizobium* clade suggests that it is a divergent member within the genus or a member of an undescribed genus closely related to *Neorhizobium*. Therefore, strain LC148 was classified as *Rhizobacteriaceae* sp. pending future taxonomic investigation.

We used a previously described Hidden Markov Model (HMM) approach (10) to identify the genomic potential for biosynthesis of acyl-homoserine lactone (AHL) molecules involved in the regulation of gene expression in response to cell density (11). AHL synthase homologs (LuxI) were identified in strains LC34 (GenBank Protein ID: TKT57503.1) and LC148 (GenBank Protein IDs: TKT46115.1 and TKT66962.1). The genes coding for these proteins were also classified as *luxI* homologs with AntiSMASH 4 (12). To confirm the predicted phenotype, LC34 and LC148 strains were grown in YM media and ethyl acetate extracts from the cultures were tested for the presence of AHL signals using the AHL-dependent biosensor *Agrobacterium tumefaciens* NTL4(pZLR4) as previously described(13). Under the described culturing and detection setting, LC34 produced one type of AHL signal with retardation factor (RF) that is to similar 3-oxo-C8. LC148 produced two distinct AHLs, one with RF similar to C6 and the other with RF smaller than the included AHL standards (Figure 1B). Future work investigating the role of quorum sensing in these cave isolates through transposon mutagenesis (14), targeted gene deletion (15) or transcriptome sequencing (16) will be informative in understanding bacteria from cave environments.

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

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

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

68 Table 1. Genome assembly metrics and data availability

BioProject	Assembly Accession	Organism	Strain	Genome Size (bp)	GC content (%)	No. of contigs (>500 bp)	N_{50} length (bp)
PRJNA281560	LBCQ02	*Aquamicrobium* sp.	LC103	6,044,022	64.06	34	330,476
PRJNA532460	SZVV01	*Rhizobiaceae* sp.	LC148	5,822,494	61.45	86	124,816
PRJNA281572	LBHX02	*Agrobacterium* sp.	LC34	5,303,297	59.35	36	377,718
PRJNA281679	LBIA02	*Afipia massiliensis*	LC387	4,495,198	61.35	1	4,495,198
PRJNA281681	LBHW02	*Achromobacter* sp.	LC458	6,547,009	64.55	102	143,492

69

70

71 **Figure Legends**

72 Figure 1. (A) Maximum likelihood tree constructed based on the concatenated alignment of 119
73 conserved proteobacteria single-copy gene sets. Branch lengths indicate the number of substitutions per
74 site while node labels are Shimodaira-Hasegawa (SH) local support values computed in FastTree2 (17).
75 (B) Separation of AHL molecules from the LC34 and LC148 extracts by C_{18} reversed-phase thin-layer plate
76 developed with methanol/water (60:40, vol/vol). Standards include: C6, N-hexanoyl-L-homoserine
77 lactone; C6-O, N-(3-oxohexanoyl)-L-homoserine lactone; C8-O, N-(3-oxo-octanoyl)-L-homoserine
78 lactone.

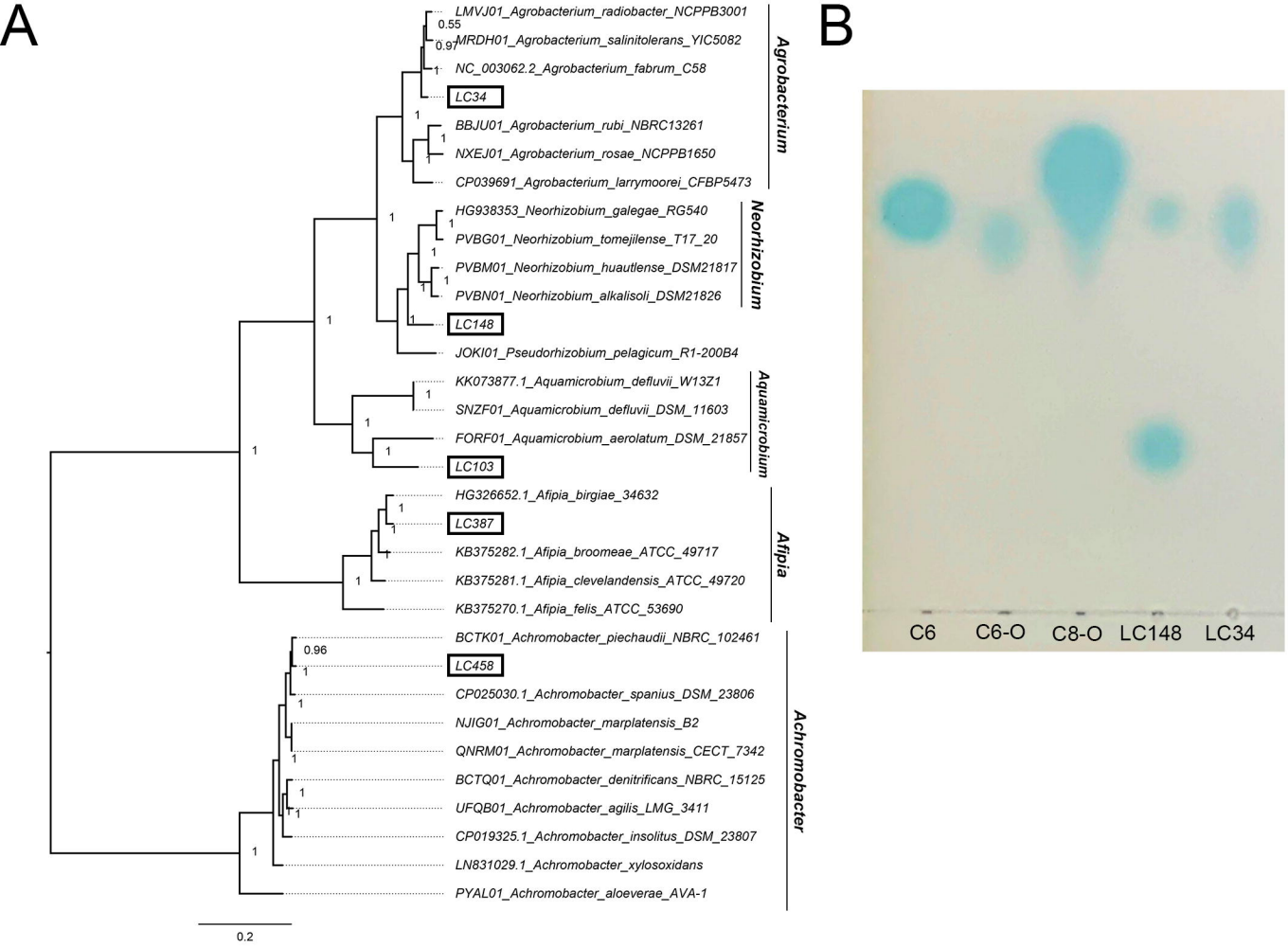
79

80 **References**

81 1. Barton MD, Petronio M, Giarrizzo JG, Bowling BV, Barton HA. 2013. The genome of
82 *Pseudomonas fluorescens* strain R124 demonstrates phenotypic adaptation to the mineral
83 environment. J Bacteriol 195:4793-803.
84 2. Gosse JT, Ghosh S, Sproule A, Overy D, Cheeptham N, Boddy CN. 2019. Whole Genome
85 Sequencing and Metabolomic Study of Cave Streptomyces Isolates ICC1 and ICC4. Frontiers in
86 Microbiology 10.
87 3. Pawlowski AC, Westman EL, Koteva K, Waglechner N, Wright GD. 2018. The complex resistomes
88 of Paenibacillaceae reflect diverse antibiotic chemical ecologies. The ISME Journal 12:885-897.
89 4. Pawlowski AC, Wang W, Koteva K, Barton HA, McArthur AG, Wright GD. 2016. A diverse intrinsic
90 antibiotic resistome from a cave bacterium. Nature Communications 7:13803.
91 5. Gan HY, Gan HM, Tarasco AM, Busairi NI, Barton HA, Hudson AO, Savka MA. 2014. Whole-
92 genome sequences of five oligotrophic bacteria isolated from deep within Lechuguilla Cave, New
93 Mexico. Genome Announc 2:e01133-14.
94 6. Scott RA, Weil J, Le PT, Williams P, Fray RG, von Bodman SB, Savka MA. 2006. Long- and Short-
95 Chain Plant-Produced Bacterial N-Acyl-Homoserine Lactones Become Components of
96 Phyllosphere, Rhizosphere, and Soil. Molecular Plant-Microbe Interactions 19:227-239.
97 7. Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence
98 data. Bioinformatics 30:2114-2120.
99 8. Wick RR, Judd LM, Gorrie CL, Holt KE. 2017. Unicycler: resolving bacterial genome assemblies
100 from short and long sequencing reads. PLoS computational biology 13:e1005595.

101 9. Lee MD. 2019. GToTree: a user-friendly workflow for phylogenomics. Bioinformatics
102 doi:10.1093/bioinformatics/btz188.
103 10. Gan HM, Gan HY, Ahmad NH, Aziz NA, Hudson AO, Savka MA. 2015. Whole genome sequencing
104 and analysis reveal insights into the genetic structure, diversity and evolutionary relatedness of
105 luxI and luxR homologs in bacteria belonging to the Sphingomonadaceae family. Frontiers in
106 Cellular and Infection Microbiology 4.
107 11. Miller MB, Bassler BL. 2001. Quorum sensing in bacteria. Annual Reviews in Microbiology
108 55:165-199.
109 12. Blin K, Wolf T, Chevrette MG, Lu X, Schwalen CJ, Kautsar SA, Suarez Duran HG, de Los Santos EL,
110 Kim HU, Nave M. 2017. antiSMASH 4.0—improvements in chemistry prediction and gene cluster
111 boundary identification. Nucleic acids research 45:W36-W41.
112 13. Lowe N, Gan HM, Chakravartty V, Scott R, Szegedi E, Burr TJ, Savka MA. 2009. Quorum-sensing
113 signal production by *Agrobacterium vitis* strains and their tumor-inducing and tartrate-catabolic
114 plasmids. FEMS microbiology letters 296:102-109.
115 14. Gan HM, Ibrahim Z, Shahir S, Yahya A. 2011. Identification of genes involved in the 4-
116 aminobenzenesulfonate degradation pathway of *Hydrogenophaga* sp. PBC via transposon
117 mutagenesis. FEMS microbiology letters 318:108-114.
118 15. Marx CJ. 2008. Development of a broad-host-range *sacB*-based vector for unmarked allelic
119 exchange. BMC research notes 1:1-1.
120 16. Wagner VE, Gillis RJ, Iglewski BH. 2004. Transcriptome analysis of quorum-sensing regulation
121 and virulence factor expression in *Pseudomonas aeruginosa*. Vaccine 22:S15-S20.
122 17. Price MN, Dehal PS, Arkin AP. 2010. FastTree 2–approximately maximum-likelihood trees for
123 large alignments. PloS one 5:e9490.

124


A
LMVJ01_Agrobacterium_radiobacter_NCPPB3001
MRDH01_Agrobacterium_salinitolerans_YIC5082
NC_003062.2_Agrobacterium_fabrum_C58
LC34
BBJU01_Agrobacterium_rubi_NBRC13261
NXEJ01_Agrobacterium_rosae_NCPPB1650
CP039691_Agrobacterium_larrymoorei_CFBP5473
Agrobacterium
HG938353_Neorhizobium_galegae_RG540
PVBG01_Neorhizobium_tomejilense_T17_20
PVBM01_Neorhizobium_huautlense_DSM21817
PVBN01_Neorhizobium_alkalisoli_DSM21826
Neorhizobium
LC148
JOKI01_Pseudorhizobium_pelagicum_R1-200B4
KK073877.1_Aquamicrobium_defluvii_W13Z1
SNZF01_Aquamicrobium_defluvii_DSM_11603
FORF01_Aquamicrobium_aerolatum_DSM_21857
LC103
Aquamicrobium
HG326652.1_Afipia_birgiae_34632
LC387
KB375282.1_Afipia_broomeae_ATCC_49717
KB375281.1_Afipia_clevelandensis_ATCC_49720
KB375270.1_Afipia_felis_ATCC_53690
Afipia
BCTK01_Achromobacter_piechaudii_NBRC_102461
LC458
CP025030.1_Achromobacter_spanius_DSM_23806
NJIG01_Achromobacter_marplatensis_B2
QNRM01_Achromobacter_marplatensis_CECT_7342
BCTQ01_Achromobacter_denitrificans_NBRC_15125
UFQB01_Achromobacter_agilis_LMG_3411
CP019325.1_Achromobacter_insolitus_DSM_23807
LN831029.1_Achromobacter_xylosoxidans
PYAL01_Achromobacter_aloeverae_AVA-1
Achromobacter
0.55
0.97
0.96
0.2
B
C6
C6-O
C8-O
LC148
LC34