A genome-annotated bacterial collection of the barley rhizosphere

microbiota

Senga Robertson-Albertyn[a], Federico Concas[a±], Lynn H Brown[a], Jamie N Orr[b],

James C Abbott[c], Timothy S George[d], Davide Bulgarelli[a]#

[a]Plant Sciences, School of Life Sciences, University of Dundee, Dundee, United

Kingdom

[b]Cell and Molecular Sciences, The James Hutton Institute, Dundee, United Kingdom

[c]Computational Biology, School of Life Sciences, University of Dundee, United

Kingdom

[d]Ecological Sciences, The James Hutton Institute, Dundee, United Kingdom

[±]Present address: GlaxoSmithKline Research and Development, Stevenage, United

Kingdom

[#]Address correspondence to Davide Bulgarelli, d.bulgarelli@dundee.ac.uk

ABSTRACT

We generated a bacterial collection from the rhizosphere of cultivated barley (*Hordeum vulgare* L. ssp. *vulgare*) to assess taxonomic distribution of culturable members of the barley microbiota and their plant growth-promoting potential. From this we retrieved strains belonging to the dominant phyla of the plant microbiota— Actinobacteria, Bacteroidetes, Firmicutes and Proteobacteria—and gathered evidence they code for functional genes implicated in nitrogen fixation, hydrogen cyanide channels and phosphate solubilisation. Here we present an initial comparative genomic analysis of the collection revealing that plant growth-promoting potential of the culturable barley bacterial microbiota appears to have a relatively broad phylogenetic base while retaining some strain-specificity.

MAIN TEXT

The microbial communities thriving at the root-soil interface, collectively referred to as the rhizosphere microbiota, represent an untapped resource of plant probiotic functions [1, 2]. Bacterial members of the microbiota capable of enhancing plant's mineral uptake from soil and pathogen protection, collectively referred to as plant growth-promoting rhizobacteria (PGPRs) have gained prominence in both basic science and translational applications [3-5].

As a resource for comparative investigations of the plant microbiota across host species, we provide a collection of 53 bacterial strains, isolated from the rhizosphere of barley (*Hordeum vulgare* L. ssp. *vulgare*), the fourth most cultivated cereal worldwide [6]. Bacteria were isolated from the rhizosphere of plants grown in an agricultural soil previously used for investigations in barley-microbiota interactions [7, 8]. To facilitate further experimentations, we subjected individual isolates to whole-genome sequencing and we performed a preliminary comparative analysis which identified varying presence of genes associated with PGPRs.

Using the table of taxonomic ID to Kyoto Encyclopaedia of Genes and Genomes (KEGG) ortholog (KO) identifier, we predict 6,429 unique KOs across all isolates. When examining the KOs that overlapped between isolates it was observed that there was a group of 514 KOs that were conserved between all isolates (Figure 1). In the case of Microbacterium and Stenotrophomonas groups, all isolates assigned to each group shared the same KOs based on this analysis. However, this same observation did not extend to isolates assigned to the Pseudomonas and Bacillus groups: Bacillus samples had 240 gene intersections shared only between themselves. In addition, among the analysed *Pseudomonas* sp, where 4 isolates

share 24 KOs that did not feature in any other sample (within or outwith

Pseudomonas sp.), which were assigned to 3 different *P. brassicacearum* isolates

according to 16S rRNA gene identity (bi75, bi06, bi11, bi55). A further *P.*

brassicacearum isolate, bi70, assigned to yet another OTU, had its own set of 24

unique KEGG orthologs versus the other isolates.

In order to identify bacterial isolates that may have downstream applications e.g. in

inoculant trials, we compared whole-genome average nucleotide identity (ANI)

across isolates which allowed us to identify individual "founder" members (ANI cut-off

96%, Figure 2). Next, we further inspected these genomes, representing 11 distinct

genera, for the presence of known PGPR genes focussing on, *dinG, hcn, nif, pho*

and *pqq* (Table 1). These genes have been previously implicated in the acquisition of

the essential elements phosphorus and nitrogen and enhancing the efficacy of

biological fertilisers to enhance plant nutrient uptake [9-13]. Congruent with the

overall KO analysis, we detected within genera variation for the plant-growth

promoting potential of the identified isolates. For instance, our results showed that a

variety of *pqq* genes belonging to the PQQ operon were present in a number of

bacterial isolates. Notably, *ppq*B, C, D, E, F and (in 7 isolates) H were almost

exclusively found together in *Pseudomonas* derived isolates, apart from the bi13,

Enterobacterium genus, isolate in which *ppq*B, C, D, E and F were also detected.

Although the majority of isolates did not carry any *ppq* gene orthologs, in addition to

these there were 4 isolates that carried a single *pqq* gene ortholog (Arthrobacter

FB24 (bi26); *pqq*F, *Stenotrophomonas maltophilia* (bi66); *pqq*F, *Pseudomonas*

aeruginosa (bi112); *pqq*B and *Janthinobacterium* (bi118); *pqq*L and one

(*Microbacterium testaceum* (bi03)) that carried a pair of *pqq* gene orthologs, *pqq*C

and *pqq*D. Interestingly, *pqq*E was not detected in this isolate when it has been

demonstrated that *pqq* genes are present in a conserved order and that,

occasionally, *pqq*CDE can be fused with *pqq*D being fused to the C-terminus of

*pqq*C and/ or the N-terminus of *pqq*E [14]. Likewise, we observed between genera

variation for genes putatively implicated in nitrogen metabolism. For instance, isolate

bi46 *Agrobacterium* H13-3 was identified as carrying the most *nif* gene orthologs,

*nif*ALSU. However, 20 isolates across a variety of genera including *Janthinobacter,*

Pedobacter and *Pseudomonas* were observed to have *nif*A orthologs which have

previously been identified in and identified in nitrogen-fixing *Pseudomonas* [15] as

well as in bacteria implicated in denitrification [16].

Taken together, our data indicate that the growth promoting potential of the barley

bacterial microbiota appears to have a relatively broad phylogenetic base while

retaining some strain-specificity, a similar observation has been recently reported for

the root microbiota of the model plant *Arabidopsis thaliana* [17]. Our collection

complements Synthetic Communities (SynComs) already available for the root

endospheric microbiota of important cereal crops, such as maize [18] and rice [19],

and it is ideally placed to foster further collaborative studies in molecular plant-

microbe interactions.

METHODS
102

Rhizosphere specimens were prepared as in previous studies [7, 8] and serial

dilutions of rhizosphere suspension were plated on Nutrient [20] or R2A [21] agar

plates and incubated for 48 - 72 hours at either 27 ˚C or at room temperature

(approximately 20 ˚C). We initially selected 194 distinct CFUs on the basis of

different morphology and colour. Of these, we retained 84 strains for full genome

sequencing. Individual bacterial genomes were sequenced using the 'standard

service' of MicrobesNG (Birmingham, UK; https://microbesng.com/). Briefly, bacterial

genomic DNA libraries were prepared using Nextera XT Library Prep Kit (Illumina,

San Diego, USA) following the manufacturer's protocol with the following

modifications: 2 ng of DNA instead of 1 ng were used as input, and PCR elongation

time was increased to 1 min from 30 seconds. DNA quantification and library

preparation were carried out on a Hamilton Microlab STAR automated liquid

handling system. Pooled libraries were quantified using the Kapa Biosystems Library

Quantification Kit for Illumina on a Roche light cycler 96 qPCR machine. Libraries

were sequenced on the Illumina HiSeq using a 250bp paired end protocol. Reads

were adapter trimmed using Trimmomatic 0.30 with a sliding window quality cut off of

Q15 [22]. De novo assembly was performed on samples using SPAdes version 3.7

[23], and contigs were annotated using Prokka 1.11 [24]. On the basis of GC content

and unambiguous taxonomic annotation generated using amphora [25] classification,

we retained 53 genomes for downstream analyses. To compare only components of

characterised metabolic pathways, predicted genes from all isolates were

concatenated and annotated with eggNOG-mapper (v1.0.3) [26]. The resultant

annotation file was parsed in Python to generate a table of taxonomic ID to Kyoto

Encyclopaedia of Genes and Genomes (KEGG) ortholog (KO) identifier. From this

table a presence-absence matrix of all KOs predicted at least once in each isolate

was generated in R. To compare all predicted proteins between isolates, predicted

proteomes were clustered using OrthoFinder (v2.2.1) [27], and functionally

annotated using InterProScan (v5.29-68.0) [28]. Clusters and annotations were

aggregated using KinFin (v1.0) [29]. Cluster and KO intersections were plotted using

UpSetR (v1.3.3) [30]. The phylogenetic tree was constructed using bcgTree 1.1.0

[31] and RAxML 8.2.12 [32] , using RAxML's GTRGAMMA model and 100 bootstrap

iterations.

FIGURE LEGEND

Figure 1. KEGG ortholog intersection map of all high-quality genome isolates.

Colours depict the assigned genus of each isolate based on Amphora taxonomic

assignment from full genome sequence reads. Intersection size denotes the number

of KEGG orthologs shared by the isolates indicated by the corresponding circle

below the intersection. The number in brackets next to the isolate description

denotes the isolate ID and the length of the coloured bar next to each isolate

represents the length of the sequence assembly of the individual sample.

Figure 2. Whole genome phylogenetic tree of individual bacterial genomes (ANI cut-

off 96%) constructed incorporating additional sequences for *Bacillus subtilis* NCIB

3610 (GCA_006088795), *Bifidobacterium longum* NCC2705 (GCA_000007525),

Escherichia coli K12 MG1655 (GCA_000005845) and *Staphylococcus aureus* NCTC

8325 (GCA_000013425). Protein predictions were obtained using Prokka 1.14.6 and

the tree was constructed with 100 bootstrap iterations.

TABLE LEGEND

Table 1. Taxonomic affiliation, genomic characteristics and accession numbers of

genomes of 41 individual bacterial isolates (ANI cut-off 96%) described in this study.

Capital letters depict actual genes identified within the inspected metabolic

processes. For the *dinG* gene, the identification in each bacterial genome is depicted

by the plus sign.

^aStrain taxonomy reflects the lowest and unique rank representing over 90% of the

amphora classification.

DATA AVAILABILITY

The genomic sequences reported in this study are deposited in the European

Nucleotide Archive (ENA) under the study number PRJEB42773. Accession

numbers for the individual genomes of the "founder" members are provided in Table

1.

AUTHORS CONRTIBUTION

SRA, TSG, DB: study design. SRA, FC, LHB: bacterial collection construction. SRA,

JNO, JCA: genome data analysis. SRA and DB wrote the manuscript with inputs

from all co-authors.

ACKNOWLEDGEMENTS
174

We are thankful to Federica Caradonia (University of Modena and Reggio Emilia,

Italy) and Carmen Escudero-Martinez (University of Dundee, UK) for their technical

assistance during the development of the collection. This work was supported by a

BBSRC iCASE studentship awarded to DB (BB/M016811/1) and partnered by the

James Hutton Limited (Invergowrie, UK). FC was supported by an Erasmus+

Traineeship programme (European Commission). LB was supported by a James

Black Prize Studentships (University of Dundee). JNO was supported by an ERC

advanced grant "Shuffle" (Project ID: 669182) awarded to Robbie Waugh/The James

Hutton Institute. SR-A, JA and DB are currently supported by the Horizon 2020

Framework Programme Innovation Action 'CIRCLES' (European Commission, Grant

agreement 818290) awarded to the University of Dundee. James Hutton researchers

receive financial support from the Rural and Environment Science and Analytical

Service Division of the Scottish Government.

REFERENCES

1. Compant, S., et al., *A review on the plant microbiome: ecology, functions, and emerging trends in microbial application.* Journal of advanced research, 2019. **19**: p. 29-37.
2. Schlaeppi, K. and D. Bulgarelli, *The plant microbiome at work.* Molecular Plant-Microbe Interactions, 2015. **28**(3): p. 212-217.
3. Escudero-Martinez, C. and D. Bulgarelli, *Tracing the evolutionary routes of plant-microbiota interactions.* Curr Opin Microbiol, 2019. **49**: p. 34-40.
4. Lugtenberg, B. and F. Kamilova, *Plant-Growth-Promoting Rhizobacteria.* Annual Review of Microbiology, 2009. **63**(1): p. 541-556.
5. Vorholt, J.A., et al., *Establishing Causality: Opportunities of Synthetic Communities for Plant Microbiome Research.* Cell Host Microbe, 2017. **22**(2): p. 142-155.
6. Newton, A.C., et al., *Crops that feed the world 4. Barley: a resilient crop? Strengths and weaknesses in the context of food security.* Food Security, 2011. **3**(2): p. 141.
7. Rodrigo, A.T., et al., *A footprint of plant eco-geographic adaptation on the composition of the barley rhizosphere bacterial microbiota.* Scientific Reports (Nature Publisher Group), 2020. **10**(1).
8. Robertson-Albertyn, S., et al., *Root hair mutations displace the barley rhizosphere microbiota.* Frontiers in plant science, 2017. **8**: p. 1094.
9. Berna, A., et al., *DING proteins; novel members of a prokaryotic phosphate-binding protein superfamily which extends into the eukaryotic kingdom.* The International Journal of Biochemistry & Cell Biology, 2008. **40**(2): p. 170-175.
10. Santos-Beneit, F., *The Pho regulon: a huge regulatory network in bacteria.* Frontiers in Microbiology, 2015. **6**(402).
11. Otieno, N., et al., *Plant growth promotion induced by phosphate solubilizing endophytic Pseudomonas isolates.* Frontiers in Microbiology, 2015. **6**(745).
12. Young, J., *Phylogenetic classification of nitrogen-fixing organisms.* Biological nitrogen fixation, 1992. **1544**: p. 43-86.
13. Ahemad, M. and M. Kibret, *Mechanisms and applications of plant growth promoting rhizobacteria: Current perspective.* Journal of King Saud University - Science, 2014. **26**(1): p. 1-20.
14. Shen, Y.-Q., et al., *Distribution and properties of the genes encoding the biosynthesis of the bacterial cofactor, pyrroloquinoline quinone.* Biochemistry, 2012. **51**(11): p. 2265-2275.
15. Xie, Z., et al., *Interaction between NifL and NifA in the nitrogen-fixing Pseudomonas stutzeri A1501.* Microbiology, 2006. **152**(12): p. 3535-3542.
16. Bueno, E., et al., *NifA is required for maximal expression of denitrification genes in Bradyrhizobium japonicum.* Environmental microbiology, 2010. **12**(2): p. 393-400.
17. Harbort, C.J., et al., *Root-secreted coumarins and the microbiota interact to improve iron nutrition in Arabidopsis.* Cell host & microbe, 2020. **28**(6): p. 825-837. e6.
18. Niu, B., et al., *Simplified and representative bacterial community of maize roots.* Proceedings of the National Academy of Sciences, 2017. **114**(12): p. E2450-E2459.
19. Zhang, J., et al., *NRT1. 1B is associated with root microbiota composition and nitrogen use in field-grown rice.* Nature biotechnology, 2019. **37**(6): p. 676-684.
20. Atlas, R.M., *Handbook of Media for Environmental Microbiology.* 2005: CRC Press.
21. Reasoner, D.J. and E.E. Geldreich, *A new medium for the enumeration and subculture of bacteria from potable water.* Applied and environmental microbiology, 1985. **49**(1): p. 1-7.
22. Bolger, A.M., M. Lohse, and B. Usadel, *Trimmomatic: a flexible trimmer for Illumina sequence data.* Bioinformatics, 2014. **30**(15): p. 2114-20.
23. Bankevich, A., et al., *SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing.* Journal of computational biology : a journal of computational molecular cell biology, 2012. **19**(5): p. 455-477.

24. Seemann, T., *Prokka: rapid prokaryotic genome annotation.* Bioinformatics, 2014. **30**(14): p.
2068-9.
25. Kerepesi, C., B. Szalkai, and V. Grolmusz, *Visual analysis of the quantitative composition of
metagenomic communities: the AmphoraVizu webserver.* Microb Ecol, 2015. **69**(3): p. 695-7.
26. Huerta-Cepas, J., et al., *Fast genome-wide functional annotation through orthology
assignment by eggNOG-mapper.* Molecular biology and evolution, 2017. **34**(8): p. 2115-
2122.
27. Emms, D.M. and S. Kelly, *OrthoFinder: solving fundamental biases in whole genome
comparisons dramatically improves orthogroup inference accuracy.* Genome biology, 2015.
16(1): p. 1-14.
28. Jones, P., et al., *InterProScan 5: genome-scale protein function classification.* Bioinformatics,
2014. **30**(9): p. 1236-1240.
29. Laetsch, D.R. and M.L. Blaxter, *KinFin: software for Taxon-Aware analysis of clustered protein
sequences.* G3: Genes, Genomes, Genetics, 2017. **7**(10): p. 3349-3357.
30. Conway, J.R., A. Lex, and N. Gehlenborg, *UpSetR: an R package for the visualization of
intersecting sets and their properties.* Bioinformatics, 2017. **33**(18): p. 2938-2940.
31. Ankenbrand, M.J. and A. Keller, *bcgTree: automatized phylogenetic tree building from
bacterial core genomes.* Genome, 2016. **59**(10): p. 783-791.
32. Stamatakis, A., *RAxML version 8: a tool for phylogenetic analysis and post-analysis of large
phylogenies.* Bioinformatics, 2014. **30**(9): p. 1312-1313.

Isolate	Taxonomy[a]	Genome size (bp)	No. of contigs	N50	PGPR-Genes identified					G_C content (%)	ENA accession no.
					Pqq	dinG	pho	nif	hcn		
Bi02	Microbacteriaceae family	3994224	78	112363			C, H, R, U,			69.52	ERS5639569
Bi03	Microbacterium testaceum	3548807	95	63748	C, D	+	D, H, R, U,			67.79	ERS5639570
Bi04	Flavobacteriaceae bacterium 3519-10	5066124	34	280276			A, D, H, R,	A, U,		36.22	ERS5639571
Bi05	Agrobacterium H13-3	5161912	21	417533			E, H, T, R, U,	A, S, U,		59.17	ERS5639572
Bi06	Pseudomonas brassicacearum	6570464	168	68008	B, C, D, E, F, H,	+	D, H, R, U,	A	A, B, C,	60.84	ERS5639573
Bi08	Pseudomonas fluorescens PfO-1	6697465	117	94982	B, C, D, E, F,	+	D, H, R, U,	A		59.44	ERS5639574
Bi13	Enterobacter genus	4939014	42	383165	B, C, D, E, F,	+	A, C, E, H, R, U,			56.33	ERS5639577
Bi26	Arthrobacter sp. FB24	5112477	217	43594	F	+	H, R, U,			66.4	ERS5639578
Bi27	Pedobacter heparinus	6285703	23	908772			A, H, R, U,	A, U,		39.01	ERS5639579
Bi35	Stenotrophomonas maltophilia	4595590	106	67604		+	A, C, D, H, Q, R, U			66.03	ERS5639581
Bi36	Pedobacter heparinus	6291372	23	901067			A, H, R	A, U,		38.98	ERS5639582
Bi44	Microbacterium testaceum	3535885	57	117718		+	H, R, U,			68.79	ERS5639584
Bi45	Enterobacter genus	4753187	20	697319		+	A, E, H, R, U,	J		54.61	ERS5639585
Bi46	Agrobacterium H13-3	5161469	20	379239			E, H, R, T, U,	A, L, S, U		59.17	ERS5639586
Bi64	Pseudomonas fluorescens PfO-1 C	6761930	127	163868	B, C, D, E, F,	+	A, D, H, R, U,	A, U,		58.02	ERS5639588
Bi66	Stenotrophomonas maltophilia	4755071	138	114998	F	+	A, C, D, H, Q, R, U,			65.19	ERS5639589
Bi70	Pseudomonas brassicacearum	5263541	236	33755	B, C, D, E, F, H,	+	H, R, U,			62.36	ERS5639590
Bi76	Microbacterium testaceum	3812380	110	78635						68.34	ERS5639593
Bi80	Bacillus sp.	5107215	1016	7008						40.48	ERS5639594
Bi81	Stenotrophomonas maltophilia	4657418	91	99706						65.95	ERS5639595
Bi82	Bacillus megaterium	5805678	65	369445						37.9	ERS5639596
Bi83	Arthrobacter Genus	4893187	121	76950						65.19	ERS5639597
Bi84	Bacillus genus	5576719	465	27035						39.86	ERS5639598
Bi89	Pseudomonas	6136482	73	165400						60.15	ERS5639600

	florescens PfO-1										
Bi91	*Rhodococcus jostii* RHA1	5679420	54	276010						64.41	ERS5639601
Bi96	*Bacillus* megaterium	5513925	75	186917						39.53	ERS5639602
Bi98	*Microbacterium testaceum*	3651233	25	247909						67.94	ERS5639603
Bi102	*Stenotrophomonas maltophilia*	4657928	103	75776		+	A, C, D, H, Q, R,			65.92	ERS5639604
Bi110	*Pseudomonas aeruginosa*	6696265	62	160812						59.44	ERS5639654
Bi111	*Pseudomonas aeruginosa*	6695877	169	75930						59.44	ERS5639655
Bi112	*Pseudomonas aeruginosa*	6109875	377	28693	B	+	R, U,		A, B, C,	61.4	ERS5639656
Bi115	*Microbacterium testaceum*	6476559	979	18468		+	H, R, U,	A		68.45	ERS5639657
Bi118	*Janthinobacterium* sp	5786026	48	291100	L	+	A, D, H, R, U, X,			65.3	ERS5639658
Bi121	*Microbacterium* Genus	3088809	20	392560			H, R, U,			67.75	ERS5639659
Bi122	*Stenotrophomonas maltophilia*	4652813	125	66023		+	A, C, D, H, Q, R, U,			65.96	ERS5639660
Bi123	*Pseudomonas brassicacearum*	6405172	52	250449	B, C, D, E, F, H,	+	D, H, R, U,		A, B, C,	59.37	ERS5639661
Bi126	*Pedobacter heparinus*	6291746	22	1097689			A, H, R,	A, L, U,		38.98	ERS5639662
Bi128	Marine actinobacterium PHSC2011	6566320	1694	4998		+	H, R, U,			68.6	ERS5639663
Bi130	*Pseudomonas brassicacearum*	6574124	54	239349	B, C, D, E, F, H,	+	D, H, R, U,	A	A, B, C,	59.41	ERS5639664
Bi133	*Bacillus* genus	5310979	214	42292		+	A, B, D, H, Q, R, U,	S, U,		40.03	ERS5640636
Bi134	*Bacillus* genus	5473506	31	522968		+	A, B, D, H, Q, R, U,			40.2	ERS5640637

bi96
bi84
bi80
Bacillus_subtilis
bi04
bi27
bi126
bi36
bi05
bi46
bi118
bi66
bi35
bi81
bi102
bi122
bi45
Escherichia_coli
bi13
bi70
bi08
bi64
bi111
bi110
bi112
bi06
bi89
bi130
bi123
bi91
Bifidobacterium_longum
bi03
bi85
bi44
bi115
bi98
bi76
bi121
bi02
bi128
bi83
bi26
Staphylococcus_aureus
bi82

0.2