



# Genome-Annotated Bacterial Collection of the Barley Rhizosphere Microbiota

Senga Robertson-Albertyn,<sup>a</sup> James C. Abbott,<sup>b</sup> Federico Concas,<sup>a\*</sup> Lynn H. Brown,<sup>a</sup> Jamie N. Orr,<sup>c</sup> Timothy S. George,<sup>d</sup>  
Davide Bulgarelli<sup>a</sup>

<sup>a</sup>Plant Sciences, School of Life Sciences, University of Dundee, Dundee, United Kingdom

<sup>b</sup>Computational Biology, School of Life Sciences, University of Dundee, Dundee, United Kingdom

<sup>c</sup>Cell and Molecular Sciences, The James Hutton Institute, Dundee, United Kingdom

<sup>d</sup>Ecological Sciences, The James Hutton Institute, Dundee, United Kingdom

**ABSTRACT** A culture collection of 41 bacteria isolated from the rhizosphere of cultivated barley (*Hordeum vulgare* subsp. *vulgare*) is available at the Division of Plant Sciences, University of Dundee (UK). The data include information on genes putatively implicated in nitrogen fixation, HCN channels, phosphate solubilization, and linked whole-genome sequences.

The microbial communities thriving at the root-soil interface, that is, the rhizosphere microbiota, represent an untapped resource of plant probiotic functions (1, 2). Bacterial members of the microbiota capable of enhancing a plant's mineral uptake from soil and pathogen protection, namely, plant growth-promoting rhizobacteria (PGPRs), have gained prominence in both basic scientific and translational applications (3–5). As a resource for comparative investigations of the plant microbiota across host species, we present a collection of 41 bacterial strains encompassing 15 genera with the presence of at least 5 putative plant growth-promoting (PGP)-associated gene orthologs, including, *dinG*, *hcn*, *nif*, *pho*, and *pqq* (Table 1) (6).

Strains were isolated from the rhizosphere of cultivated barley (*Hordeum vulgare* L. subsp. *vulgare*), the fourth most cultivated cereal worldwide (7), which was grown in an agricultural soil used for previous barley-microbiota investigations (8, 9). Bacterial rhizosphere fractions were obtained by detaching the soil adhering the uppermost 6 cm of barley roots by vortexing in phosphate-buffered saline (PBS) buffer. Serial dilutions were plated onto R2A and nutrient agar media and incubated at 20°C for 48 to 72 h (10, 11). Individual CFUs were selected for isolation based on morphological variation; clean isolate liquid cultures were stored at –80°C in 70% glycerol following 24 to 48 h of shaking incubation at 27°C.

DNA was extracted as per the manufacturer's instructions using the FastDNA spin kit for soil (MP Biomedicals, USA). Individual bacterial isolates were subjected to whole-genome sequencing using the “standard service” of MicrobesNG (Birmingham, UK). Briefly, bacterial genomic DNA libraries were prepared using the Nextera XT library prep kit (Illumina, USA) following the manufacturer's protocol with the following modifications: 2 ng of DNA were input, and PCR elongation time was increased to 1 min. DNA quantification and library preparation were conducted 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 quantitative PCR (qPCR) machine. Libraries were sequenced by using an Illumina HiSeq instrument with a 250-bp paired-end protocol. Reads were adapter trimmed using Trimmomatic (v0.30) with a sliding window quality cutoff of Q15 (12). *De novo* assembly was performed using SPAdes (v3.7), and contigs were annotated using Prokka

**Editor** Irene L. G. Newton, Indiana University, Bloomington

**Copyright** © 2022 Robertson-Albertyn et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](#).

Address correspondence to Davide Bulgarelli, [d.bulgarelli@dundee.ac.uk](mailto:d.bulgarelli@dundee.ac.uk).

\*Present address: Federico Concas, GlaxoSmithKline Research and Development, Stevenage, United Kingdom.

The authors declare no conflict of interest.

**Received** 1 November 2021

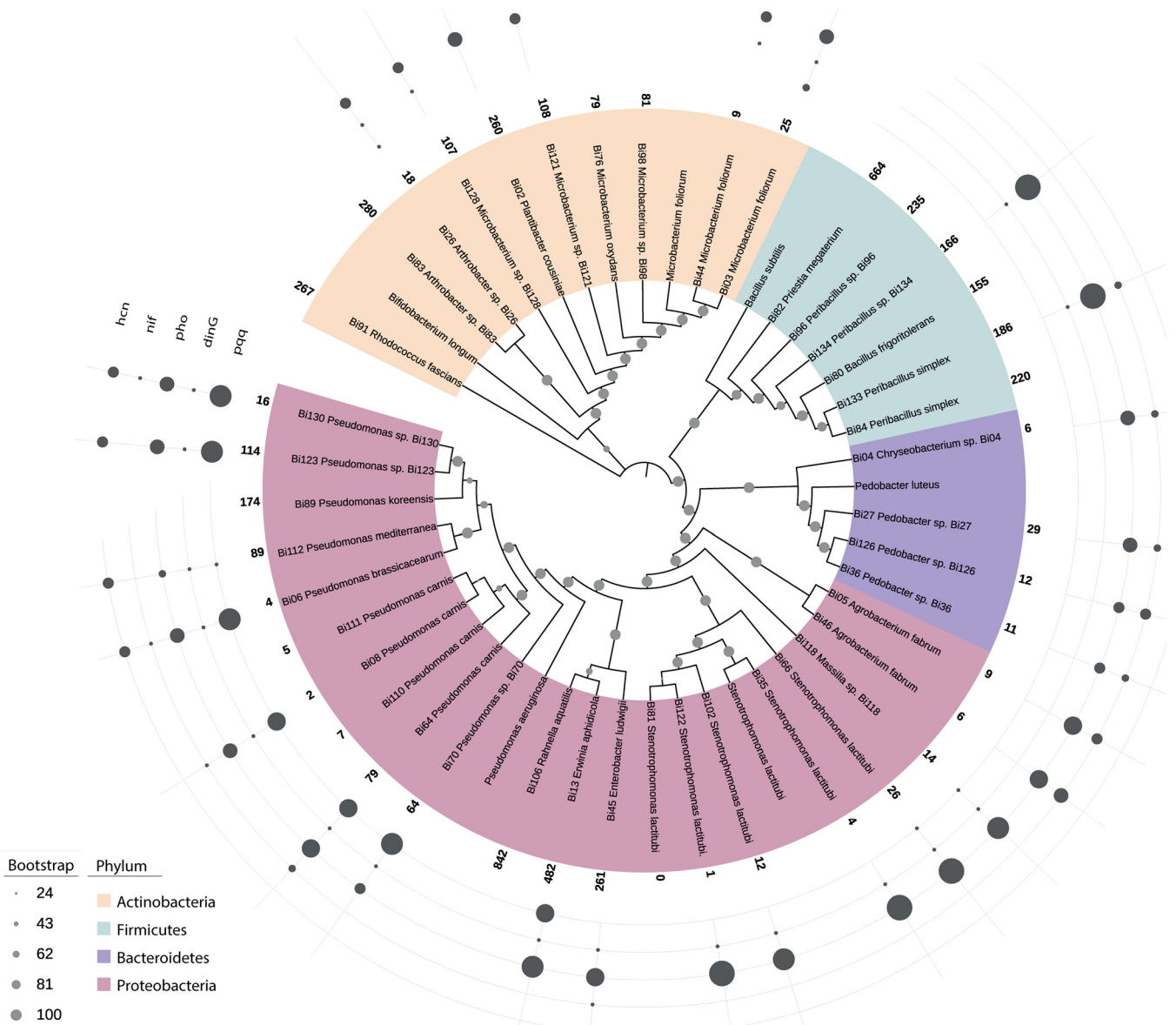
**Accepted** 28 January 2022

**Published** 17 February 2022

**TABLE 1** Taxonomic affiliation, genomic characteristics, and accession numbers of genomes of 41 individual bacterial isolates<sup>a</sup> described in this study

Isolate	Bacterial species <sup>b</sup>	Genome size (bp)	No. of contigs	N <sub>50</sub> (bp)	PGPR operon gene(s) identified				GC content (%)	ENA accession no.
					pqq	dinG <sup>c</sup>	pho	nif		
Bi02	<i>Plantibacter cousiniae</i>	3,994,224	78	112,363			CH,R,U		69.52	ERS5639569
Bi03	<i>Microbacterium foliorum</i>	3,548,807	95	63,748		+	D,H,R,U		67.79	ERS5639570
Bi04	<i>Chryseobacterium</i> sp.	5,066,124	34	280,276	C,D		A,D,H,R	A,U	36.22	ERS5639571
Bi05	<i>Agrobacterium fabrum</i>	5,161,912	21	417,533			E,H,T,R,U	A,S,U	59.17	ERS5639572
Bi06	<i>Pseudomonas brassicacearum</i>	6,570,464	168	68,008	B,C,D,E,F,H	+	D,H,R,U	A	60.84	ERS5639573
Bi08	<i>Pseudomonas carnis</i>	6,697,465	117	94,982	B,C,D,E,F	+	D,H,R,U	A	59.44	ERS5639574
Bi13	<i>Erwinia aphidicola</i>	4,939,014	42	383,165	B,C,D,E,F	+	A,C,E,H,R,U		56.33	ERS5639577
Bi26	<i>Arthrobacter</i> sp. Bi26	5,112,477	217	43,594	F	+	H,R,U		66.4	ERS5639578
Bi27	<i>Pedobacter</i> sp. Bi27	6,285,703	23	908,772			A,H,R,U	A,U	39.01	ERS5639579
Bi35	<i>Stenotrophomonas lactitubi</i>	4,595,590	106	67,604		+	A,C,D,H,Q,R,U		66.03	ERS5639581
Bi36	<i>Pedobacter</i> sp. Bi36	6,291,372	23	901,067			A,H,R	A,U	38.98	ERS5639582
Bi44	<i>Microbacterium foliorum</i>	3,535,885	57	117,718		+	H,R,U		68.79	ERS5639584
Bi45	<i>Enterobacter ludwigii</i>	4,753,187	20	697,319		+	A,E,H,R,U	J	54.61	ERS5639585
Bi46	<i>Agrobacterium fabrum</i>	5,161,469	20	379,239			E,H,R,T,U	A, L, S, U	59.17	ERS5639586
Bi64	<i>Pseudomonas carnis</i>	6,761,930	127	163,868	B,C,D,E,F	+	A,D,H,R,U	A,U	58.02	ERS5639588
Bi66	<i>Stenotrophomonas lactitubi</i>	4,755,071	138	114,998	F	+	A,C,D,H,Q,R,U		65.19	ERS5639589
Bi70	<i>Pseudomonas</i> sp. Bi70	5,263,541	236	33,755	B,C,D,E,F,H	+	H,R,U		62.36	ERS5639590
Bi76	<i>Microbacterium oxydans</i>	3,812,380	110	78,635					68.34	ERS5639593
Bi80	<i>Bacillus frigoritolerans</i>	5,107,215	1,016	7,008					40.48	ERS5639594
Bi81	<i>Stenotrophomonas lactitubi</i>	4,657,418	91	99,706					65.95	ERS5639595
Bi82	<i>Priestia megaterium</i>	5,805,678	65	369,445					37.9	ERS5639596
Bi83	<i>Arthrobacter</i> sp. Bi83	4,893,187	121	76,950					65.19	ERS5639597
Bi84	<i>Peribacillus simplex</i>	5,576,719	465	27,035					39.86	ERS5639598
Bi89	<i>Pseudomonas korensis</i>	6,136,482	73	165,400					60.15	ERS5639600
Bi91	<i>Rhodococcus fascians</i>	5,679,420	54	276,010					64.41	ERS5639601
Bi96	<i>Peribacillus</i> sp. Bi96	5,513,925	75	186,917					39.53	ERS5639602
Bi98	<i>Microbacterium</i> sp. Bi98	3,651,233	25	247,909					67.94	ERS5639603
Bi102	<i>Stenotrophomonas lactitubi</i>	4,657,928	103	75,776		+	A,C,D,H,Q,R		65.92	ERS5639604
Bi106	<i>Rahnella aquatilis</i>	4,557,699	1,614	3,019					51.67	ERS5639653
Bi110	<i>Pseudomonas carnis</i>	6,696,265	62	160,812					59.44	ERS5639654
Bi111	<i>Pseudomonas carnis</i>	6,695,877	169	75,930					59.44	ERS5639655
Bi112	<i>Pseudomonas mediterranea</i>	6,109,875	377	28,693	B	+	R,U	A,B,C	61.4	ERS5639656
Bi118	<i>Massilia</i> sp. Bi118	5,786,026	48	291,100	L	+	A,D,H,R,U,X		65.3	ERS5639658
Bi121	<i>Microbacterium</i> sp. Bi121	3,088,809	20	392,560		+	H,R,U		67.75	ERS5639659
Bi122	<i>Stenotrophomonas lactitubi</i>	4,652,813	125	66,023		+	A,C,D,H,Q,R,U		65.96	ERS5639660
Bi123	<i>Pseudomonas</i> sp. Bi123	6,405,172	52	250,449	B,C,D,E,F,H	+	D,H,R,U	A,B,C	59.37	ERS5639661
Bi126	<i>Pedobacter</i> sp. Bi126	6,291,746	22	1,097,689			A,H,R	A, L, U	38.98	ERS5639662
Bi128	<i>Microbacterium</i> sp. Bi128	6,566,320	1,694	4,998		+	H,R,U		68.6	ERS6138326
Bi130	<i>Pseudomonas</i> sp. Bi130	6,574,124	54	239,349	B,C,D,E,F,H	+	D,H,R,U	A	59.41	ERS5640634
Bi133	<i>Peribacillus simplex</i>	5,310,979	214	42,292		+	A,B,D,H,Q,R,U	S, U	40.03	ERS5640636
Bi134	<i>Peribacillus</i> sp. Bi134	5,473,506	31	522,968		+	A,B,D,H,Q,R,U		40.2	ERS5640637

<sup>a</sup>ANI cutoff, 96%. Capital letters depict actual genes identified within the inspected metabolic processes.<sup>b</sup>Strain taxonomy reflects the lowest and unique rank as defined by GTDBTK (v1.6.0) with data version r202.<sup>c</sup>The identification in each bacterial genome is depicted by the plus sign.



**FIG 1** Whole-genome phylogenetic tree of individual genomes (ANI cutoff, 96%) constructed incorporating additional sequences for *Bifidobacterium longum* NCC2705 (GCA\_000007525), *Microbacterium foliorum* DSM 12966 (GCA\_000956415), *Bacillus subtilis* NCIB 3610 (GCA\_006088795), *Pedobacter luteus* DSM 22385 (GCA\_900168015), *Stenotrophomonas lactitubi* M15 (GCA\_0002803515), and *Pseudomonas aeruginosa* PAO1 (GCA\_000006765). Protein predictions were obtained using Prokka (v1.14.6), and the tree was constructed with 100 bootstrap iterations and annotated with iTOL (24). The size of circular shapes on the periphery of the tree reflects the number of the indicated PGPR genes ranging from 1 to 7 present in each individual sample.

(v1.12) (13, 14). On the basis of GC content, unambiguous taxonomic annotations generated using amphora classification (15) and whole-genome average nucleotide identity (ANI) to identify individual “founder” members (ANI cutoff, 96%) yielded 41 genomes retained for downstream analyses. To compare only components of characterized metabolic pathways, predicted genes were concatenated and annotated with eggNOG-Mapper (v1.0.3) (16, 17). The resultant annotation file was parsed in Python to generate a table of taxonomic identities (IDs) of Kyoto Encyclopedia of Genes and Genomes (KEGG) ortholog (KO) identifiers. From this table, a presence-absence matrix of all KOs predicted at least once in each isolate was generated in R (<https://www.r-project.org>). Predicted proteomes were clustered using OrthoFinder (v2.2.1) and functionally annotated using InterProScan (v5.29-68.0) (18, 19). Clusters and annotations were aggregated using KinFin (v1.0) (20). Cluster and KO intersections were defined using UpSetR (v1.3.3) (21). The phylogenetic tree (Fig. 1) was constructed using

bcgTree (v1.1.0) and RAXML (v8.2.12), using RAXML's GTRGAMMA model and 100 bootstrap iterations (22, 23); default parameters were used for all analyses unless otherwise noted.

The collection is available as frozen isolates preserved in 300 to 500  $\mu$ L of nutrient or R2A medium containing 50% glycerol and maintained at  $-70^{\circ}\text{C}$ . To revive the frozen cultures, we recommend using a sterile inoculating loop to transfer a small amount (e.g., 50  $\mu$ L equivalent) of the frozen culture onto a nutrient or R2A agar medium base following standard microbiological procedures. The plates should be incubated at  $27^{\circ}\text{C}$  for 24 to 48 h.

**Data availability.** The genomic sequences reported in this study are deposited in the European Nucleotide Archive (ENA). Accession numbers for the individual genomes are provided in Table 1. To acquire isolates, or for questions or suggestions, please contact Davide Bulgarelli at d.bulgarelli@dundee.ac.uk.

## ACKNOWLEDGMENTS

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 D.B. (BB/M016811/1) and partnered by the James Hutton Limited (Invergowrie, UK). F.C. was supported by an Erasmus+ Traineeship program (European Commission). L.H.B. was supported by a James Black Prize Studentship (University of Dundee). J.N.O. was supported by an ERC advanced grant "Shuffle" (project ID 669182) awarded to Robbie Waugh/The James Hutton Institute. S.R.-A., J.C.A., and D.B. are currently supported by the Horizon 2020 Framework Program 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

- Compant S, Samad A, Faist H, Sessitsch A. 2019. A review on the plant microbiome: ecology, functions, and emerging trends in microbial application. *J Adv Res* 19:29–37. <https://doi.org/10.1016/j.jare.2019.03.004>.
- Schlaeppi K, Bulgarelli D. 2015. The plant microbiome at work. *Mol Plant Microbe Interact* 28:212–217. <https://doi.org/10.1094/MPMI-10-14-0334-FI>.
- Escudero-Martinez C, Bulgarelli D. 2019. Tracing the evolutionary routes of plant-microbiota interactions. *Curr Opin Microbiol* 49:34–40. <https://doi.org/10.1016/j.mib.2019.09.013>.
- Lugtenberg B, Kamilova F. 2009. Plant-growth-promoting rhizobacteria. *Annu Rev Microbiol* 63:541–556. <https://doi.org/10.1146/annurev.micro.62.081307.162918>.
- Vorholt JA, Vogel C, Carlström CI, Müller DB. 2017. Establishing causality: opportunities of synthetic communities for plant microbiome research. *Cell Host Microbe* 22:142–155. <https://doi.org/10.1016/j.chom.2017.07.004>.
- Kanehisa M, Sato Y, Kawashima M, Furumichi M, Tanabe M. 2016. KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res* 44:D457–D462. <https://doi.org/10.1093/nar/gkv1070>.
- Newton AC, Flavell AJ, George TS, Leat P, Mullholland B, Ramsay L, Revoredo-Giha C, Russell J, Steffenson BJ, Swanston JS, Thomas WTB, Waugh R, White PJ, Bingham IJ. 2011. Crops that feed the world 4. Barley: a resilient crop? Strengths and weaknesses in the context of food security. *Food Sec* 3:141. <https://doi.org/10.1007/s12571-011-0126-3>.
- Rodrigo AT, Katharin B-C, Morris J, Hedley PE, Russell J, Paterson E, Baggs EM, Eyal F, Davide B. 2020. A footprint of plant eco-geographic adaptation on the composition of the barley rhizosphere bacterial microbiota. *Sci Rep* 10:12916. <https://doi.org/10.1038/s41598-020-69672-x>.
- Robertson-Albertyn S, Alegria Terrazas R, Balbirnie K, Blank M, Janiak A, Szarejko I, Chmielewska B, Karcz J, Morris J, Hedley PE, George TS, Bulgarelli D. 2017. Root hair mutations displace the barley rhizosphere microbiota. *Front Plant Sci* 8:1094. <https://doi.org/10.3389/fpls.2017.01094>.
- Atlas RM. 2005. Handbook of media for environmental microbiology. CRC Press, Boca Raton, FL.
- Reasoner DJ, Geldreich EE. 1985. A new medium for the enumeration and subculture of bacteria from potable water. *Appl Environ Microbiol* 49:1–7. <https://doi.org/10.1128/aem.49.1.1-7.1985>.
- Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30:2114–2120. <https://doi.org/10.1093/bioinformatics/btu170>.
- Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* 19:455–477. <https://doi.org/10.1089/cmb.2012.0021>.
- Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30:2068–2069. <https://doi.org/10.1093/bioinformatics/btu153>.
- Kerepesi C, Szalkai B, Grolmusz V. 2015. Visual analysis of the quantitative composition of metagenomic communities: the AmphoraVizu webserver. *Microb Ecol* 69:695–697. <https://doi.org/10.1007/s00248-014-0502-6>.
- Huerta-Cepas J, Forslund K, Coelho LP, Szklarczyk D, Jensen LJ, von Mering C, Bork P. 2017. Fast genome-wide functional annotation through orthology assignment by eggNOG-Mapper. *Mol Biol Evol* 34:2115–2122. <https://doi.org/10.1093/molbev/msx148>.
- Huerta-Cepas J, Szklarczyk D, Heller D, Hernandez-Plaza A, Forslund SK, Cook H, Mende DR, Letunic I, Rattei T, Jensen LJ, von Mering C, Bork P. 2019. eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Res* 47:D309–D314. <https://doi.org/10.1093/nar/gky1085>.
- Emms DM, Kelly S. 2019. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol* 20:238. <https://doi.org/10.1186/s13059-019-1832-y>.
- Jones P, Binns D, Chang HY, Fraser M, Li W, McAnulla C, McWilliam H, Maslen J, Mitchell A, Nuka G, Pesseat S, Quinn AF, Sangrador-Vegas A, Scheremetjev

- M, Yong SY, Lopez R, Hunter S. 2014. InterProScan 5: genome-scale protein function classification. *Bioinformatics* 30:1236–1240. <https://doi.org/10.1093/bioinformatics/btu031>.
20. Laetsch DR, Blaxter ML. 2017. KinFin: software for taxon-aware analysis of clustered protein sequences. *G3 (Bethesda)* 7:3349–3357. <https://doi.org/10.1534/g3.117.300233>.
  21. Conway JR, Lex A, Gehlenborg N. 2017. UpSetR: an R package for the visualization of intersecting sets and their properties. *Bioinformatics* 33:2938–2940. <https://doi.org/10.1093/bioinformatics/btx364>.
  22. Ankenbrand MJ, Keller A. 2016. bcgTree: automatized phylogenetic tree building from bacterial core genomes. *Genome* 59:783–791. <https://doi.org/10.1139/gen-2015-0175>.
  23. Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30:1312–1313. <https://doi.org/10.1093/bioinformatics/btu033>.
  24. Letunic I, Bork P. 2021. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res* 49:W293–W296. <https://doi.org/10.1093/nar/gkab301>.