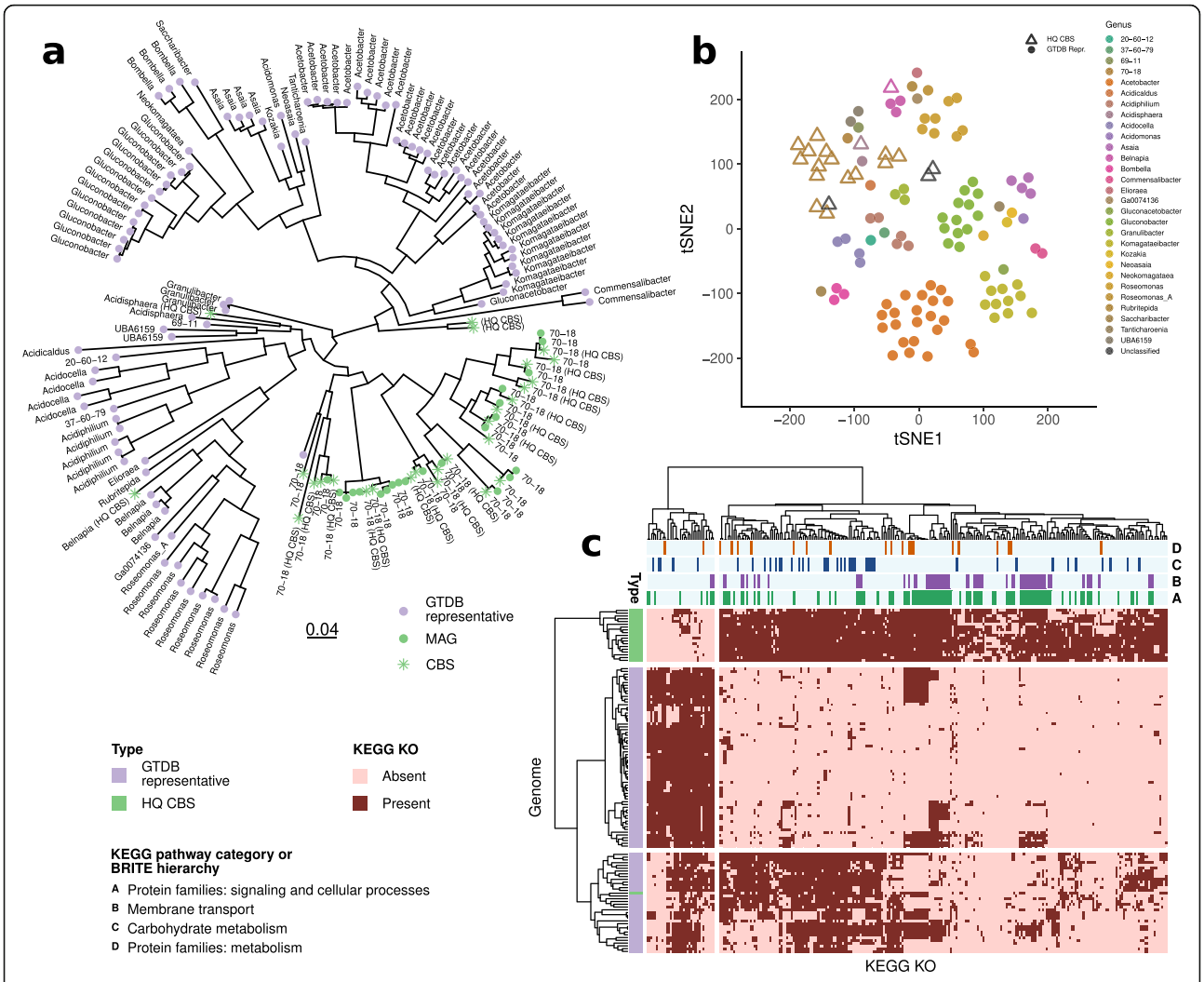

Fig. 4 Antarctic CBS form a distinct clade in the order Acetobacterales, family Acetobacteraceae, with characteristic metabolic potential. **a** Maximum-likelihood phylogenetic tree based on GTDB 120 "core" genes including representative genomes (violet) and Antarctic CBS (green). **b** t-SNE (Jaccard distance) based on the protein cluster profiles (60% identity) of the GTDB representatives (circles) and Antarctic high quality CBS (triangles). The representative genomes of Acetobacter aceti e A. aceti_B are not shown. **c** The Fisher's exact test (Bonferroni corrected $p<0.01$) highlights enriched functional categories. Genomes and KEGG orthologs are clustered according to the Hamming distance between the profiles. The top four KEGG categories significantly more present in the Antarctic CBS are highlighted in the upper bars

Antarctic Frankiales CBS (Fig. S1a) KO profiles form a group clearly distinct from all known genomes of the same order (Fig. S1 and Supplementary Data). Differently for what observed in the Acetobacterales family, the number of genes ($p<0.01$) belonging to membrane transport, carbohydrate, and amino acid metabolisms that are significantly associated with one clade is higher in the reference genomes by factors of 5, 4, and 9, respectively (Fig. S1d, Supplementary Table 8).

To identify genomic features that might explain its widespread presence, we compared the newly identified Candidatus *Jiangella antarctica* to other species from the genus Jiangella, the only represented in the order Jiangellales (Fig. S2a). We found that Antarctic Jiangellales have the smallest genome sizes, with a significant reduction of the number of genes in several functional categories (Fig. S2b,c and Supplementary Table 9). Moreover, several of the KO that were significantly more represented in the Antarctic genomes were involved in the pathway for carotenoid biosynthesis (Fig. S2d).

Additionally, we compared the eight newly identified CBS from the order Thermomicrobiales (class Chloroflexia) to the known genomes from the same order. We found an increase in genome size, with a number of KEGG pathways, in particular those related to transport, more represented in the newly assembled Antarctic genomes (Fig. S3 and Supplementary Table 10).

Discussion

Whole-genome metagenomics has contributed substantially to our understanding of global microbial diversity [53]. Here, we retrieved 497 draft MAGs from environmental DNA extracted from eighteen Antarctic cryptoendolithic communities. These newly assembled genomes were clustered into 269 previously uncharacterized species-level groups. Most of these new candidate bacterial species (CBS) could not be taxonomically classified even at higher taxonomic levels; for instance, out of 269, 81 only were assigned to known genera. These findings demonstrated that a large amount of bacterial diversity remains to be genomically characterized across these environments and that the Antarctic endolithic niches represent a reservoir for unknown bacterial taxa.

These MAGs represent the first example of bacterial genomes recovered from these microbial ecosystems; for instance, to date, only a few Cyanobacteria genomes were constructed with more than 93% estimated completeness from Atacama rocky communities [54]. Overall, Antarctic endolithic microbiomes characterized in this study comprised 12 phyla, 22 classes, and 33 orders. The newly assembled MAGs widen the phylogenetic diversity of bacterial tree of life by more than 50% for Jiangellales, Frankiales, Thermomicrobiales, Isosphaerales, Solirubrobacterales, and UBA5184 lineages.

Actinobacteria, Chloroflexi, and Proteobacteria represent the most abundant phyla and the "core" (i.e., present in almost all samples) members of these communities, as previously reported [12, 13]. On the contrary, Deinococci and Cyanobacteria were generally underrepresented. Actinobacteria are not only the main producers of microbial-derived drugs and play an important role associated with plants [55], but they were found to be widely distributed in ecologically different environments, including extreme terrestrial habitats [56, 57] such as hot and cold deserts [58, 59]. Compared with Actinobacteria from temperate habitats, the adaptation strategies of the members of this phylum colonizing extreme environments are still not well understood. Further work is needed to give insights into how this bacterial group adapts to the severe conditions found in desert habitats. Proteobacteria are the dominant component of Polar habitats including soil biotopes [60], cryoconite holes [61], and rock-inhabiting communities in cold climates [61, 62]. Amongst the most representative species, we found one CBS of Jiangellales [63], an order from the class Actinobacteria that encompasses species isolated from different habitats including indoor environments, cold springs on the Qinghai-Tibet Plateau [64], and caves [65, 66]. This CBS, herein named "Candidatus *Jiangella antarctica*," was present across all samples, suggesting a high adaptation and specialization of this species to the extreme Antarctic environment. The genomes of J. gansuensis, isolated from desert soils in Gansu Province (China) [67], and of Ca. J. antarctica (5.6 and 3.6 Mbp, respectively) showed a smaller size compared with other *Jiangella* species (~ 7 Mbp). Indeed, we found that a significant reduction of the number of genes in several functional categories occurred in the Ca. *J. antarctica*. In particular, genes related to transport, amino, and nucleotide sugar metabolism were underrepresented. These findings support the hypothesis that these microbial species may have undergone a phase of genome reduction to adapt to the hasherst desert conditions to reduce, for instance, the metabolic costs associated with DNA replication and processing. Conversely, genes involved in the carotenoid biosynthesis pathway were enriched in this species, suggesting that the capability to synthesize these pigments is specific to the Antarctic microbes to enhance resistance to UV radiation and freeze-thawing stresses [68, 69].

We found that at least one representative CBS of the class Chloroflexia (i.e., green non-sulfur bacteria), consisting of autotrophic bacteria, was present in all samples. Their capacity of anoxygenic photosynthesis and the presence of bacteriochlorophyll as light-harvesting pigment expand the possibility and the conditions for the community of carbon fixation in highly oligotrophic conditions of the Antarctic desert, a main strategy to conserve energy [70]. Members of Chloroflexi were discovered in Alpine tundra soil, Atacama desert [71] and in microbial mats found in Japanese hot springs [72]. More recently, two novel Chloroflexi, obtained from hot springs in Yellowstone National Park, were identified as putative nitrite-oxidizing bacteria by the presence of nitrite oxidoreductase encoding genes in their genomes [73]. The high abundance of Chloroflexi in such arid environments [74–76] may reflect specific adaptations of this group to survive under arid conditions, but its specific functional role is still to be clarified.

When comparing our newly identified Chloroflexi CBS with the known genomes from the same order, we observed an increase in genome size, with several KEGG pathways, in particular those related to transport, more represented in the newly Antarctic MAGs. This is apparently in contrast to what observed for the genus Jiangella, but it remains rather speculative to generalize considerations at class level.

Our newly assembled MAGs increase by more than 50% the number of representative species in Frankiales (Actinobacteria, G+) that include nitrogen-fixing bacteria in both the free-living and the symbiotic state. Members of *Frankia* genus were found resistant to several stresses such as salinity, heavy metals, extreme pH,

and drought [77]. The high recurrence of this group may suggest a critical role in the Antarctic endolithic ecosystems functioning as contributors for nitrogen fixation.

Since a small number of species are shared among all samples analyzed and the majority of CBS were barely detected, we surmise that dispersal may be not the sole determining factor in shaping the diversity and structure of these communities. Dispersal, in these areas, takes place through transportation of microbial propagules associated to rock fragments blown over long distances by the strong winds. Despite the efficiency of this mechanism, a local diversification apparently occurs; very few adapted species can perpetuate in all locations, while biodiversity remains highly variable regardless of geography. Similar conclusions were reached by Archer and colleagues [78], who recently reported that persistent local airborne inputs were unable to fully explain the composition of Antarctic soil communities. Despite the arguably lower sensitivity of shotgun metagenomics compared with amplicon-based methods for biodiversity description [48, 79], our study confirms earlier findings of high site variability between prokaryotic communities in Antarctica soils [80]. The presence of recurrent species in the Antarctic cryptoendolithic communities has been also observed for the fungal counterpart: for instance, the endemic black fungus Friedmanniomyces endolithicus has been reported in almost all samples collected in the Victoria Land in more than 20 years of Antarctic Campaigns [81, 82], indicating a high degree of adaptation to the prohibitive environmental conditions of this area.

Our molecular clock analysis indicated that most of the Antarctic bacterial clades found here originated during the Tonian glaciations, in a period ranging from 800 to 1000 Ma. before the many glaciations of the Cryogenian [83, 84] when Antarctica was still part of the Supercontinent Rodinia. Even accounting for the uncertainties of the estimated divergence times (see bars in Fig. 2) and the many prior assumptions embedded in the molecular clock of Antarctic organisms [85], our data exclude the hypothesis that the evolution of these bacterial clades was driven in response to the environmental pressure of the more recent Antarctica geological history. In fact, the last cooling events started once Antarctica reached the South Pole in the early Oligocene (~34 Ma), after the separation from Gondwanaland about 200 Ma [86], while the present icy conditions were established round 3 Ma. Our results suggest that these new bacterial clades diversified from a pool of pre-existing frost-evolved species that found the opportunity to spread in Antarctica once the present conditions were established. Based on these data we cannot establish when these organisms reached the continent, but it could be expected that such old clades, or their relatives, may be found searching elsewhere in extreme-cold niches, possibly in continents that were neighboring Antarctica in the era of the Supercontinent Rodinia (i.e., North America, India or Australia). This accomplishes the scenario of "everything is everywhere, but the environment selects" suggested in 1934 by Baas Becking [87]. Similar results were found in a global survey of the hypolithic cyanobacterial genus *Chroococcidiopsis*, where a molecular phylogenetic analysis found that variants from hot and cold deserts were grouped in different lineages, with an estimated time to last common ancestor of the hot and cold clade of ~2400 Ma and regional genetic variability maintained over geological timescales [88].

Further, whole-genome metagenomic sequencing can be employed to investigate not only the composition of the microbial communities, but also the functional roles that these community members may play. In our study, we showed that the set of Antarctic MAGs predicted proteins, typically part of primary metabolism playing a role in normal growth and survival, was significantly consistent with existing representatives in the public domain from the same order. Whereas, at the order level, de novo protein clustering and functional annotation confirmed the results of the phylogenetic analysis indicating that several CBS form separate lineages. The main functional processes which appeared to be potentially enriched in the Acetobacteraceae compared with reference genomes were those related to amino acid and carbohydrate metabolisms, containing proteins with high identity with similar protein sequences in the public domain, while these pathway categories were underrepresented in the order Frankiales. The functional differences observed may be related to a specific adaptation to the Antarctic endolithic niche.

The release of the endolithic MAGs presented here will surely remodel the way we interpret and explore the Antarctic ecosystems data. A more detailed examination of such genomes and additional samples will further increase our understanding of microbial evolution and metabolic diversity and provide important insights into the role of these microorganisms in Antarctic desert functioning.

Conclusions

In conclusion, our data report for the first time the genomes of the dominant bacterial species in Antarctic cryptoendolithic communities; none of the 269 CBS individuated were accounted to already described taxa. Most of the new species found are organized into ancient monophyletic clades that differentiated from known bacterial orders in a time range that predates the estimated origin of modern Antarctica and the establishment of the present glacial climate. Our data point toward a scenario where extant Antarctic bacterial clades

are the remnants of ancient bacterial lineages, dating back up to 1000 Ma, which found in the present frost conditions of the continent a new opportunity to spread and diversify. These findings give also new insights for the possibility of life beyond the Earth (e.g., on Mars) since microbial life, if ever evolved, may have escaped extinction for a timescale of evolutionary significance in proper refugia.

Despite the variability of the bacterial assemblages observed among samples, a "core" of few species was shared among all specimens examined. We did not find a specific set of functions that characterize the Antarctic MAGs; yet, genes for several metabolic pathways were differently represented (both over- or underrepresented, depending on the group considered) compared to reference genomes. A deeper understanding of these mechanisms is likely to contribute substantially to our capacity to predict how these ecosystems respond to the projected climate change which is particularly enhanced at the Poles. Moreover, it would be possible to extrapolate this information in worldwide arid areas deepening our comprehension of the service that these communities provide in an era of rapid desertification.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s40168-021-01021-0.

Additional file 1: Figure S1. Antarctic CBS form two distinct clades in the order Frankiales, with characteristic metabolic potential. **a)** Maximum-likelihood phylogenetic tree based on GTDB 120 core genes including representative genomes (violet) and Antarctic CBS (green). **b)** Principal Coordinate Analysis (Jaccard distance) of the protein cluster profiles (60% identity) **b)** and of the metabolic potential. **d)** The Fisher's exact test (Bonferroni corrected $p<0.01$) highlights enriched functional categories. Genomes and KEGG orthologs are clustered according to the Hamming distance between the profiles. The top four KEGG categories significantly more present in the Antarctic CBS are highlighted in the upper bars. **Figure S2.** Antarctic *Jiangellales* CBS reveal a substantial genome reduction compared to known species, with characteristic differences in metabolic potential. **a)** Maximum-likelihood phylogenetic tree based on GTDB genes, including representative genomes from the GTDB database (violet) and Antarctic CBS (green). **b)** KEGG orthologs that are significantly less frequent in Antarctic Jiagellales compared to reference (uncorrected $p<$ 0.05, Fisher's exact test). Only the first 25 pathways (ranked by the total number of significant orthologs) are shown. **c)** Number of predicted protein coding sequences in Antarctic (green) and reference (violet) Jiangellales **d)** The heatmap shows the presence (dark green) of KEGG orthologs belonging to the carotenoid biosynthesis pathway. The only gene involved in carotenoid biosynthesis detected in both CBS and GTDB reference genomes is the crtD. **e)** The phylogenetic tree inferred on the crtD gene highlights a segregation of Antarctic *Jiangellales*. **Figure S3.** Antarctic *Thermomicrobiales* (class *Chloroflexia*) CBS reveal characteristic metabolic potential. **a)** The Fisher's exact test (uncorrected $p<0.05$) highlights a significant presence, in Antarctic genomes, of orthologs involved in transport, compared to the reference *Thermomicrobiales* genomes. Only the first 30 pathways (ranked by the total number of orthologs called significant) are shown. **b)** The prediction of protein coding sequences shows an increment of the number of genes in Antarctic *Thermomicrobiales* compared to reference genomes. **Figure S4.** Distribution of the number of CBS that are specific to a given number of samples, taxonomically classified at the Class level. We identified a set of 10 CBS (belonging to the classes Actinobacteria and Alphaproteobacteria) that are present in at least 75% (14/18) of the samples. **Figure S5.** Mash Screen was used to validate the presence of CBS in the Antarctic samples. **a)** Distribution of the number of CBS marked as present by the containment score estimated by Mash screen. 1009 out of 1094 (92.2%) CBS have been confirmed by Mash (containment score >0.95, green dashed vertical line). **b)** Distribution of the number of CBS marked as present by the estimated multiplicity. **Figure S6.** Scatter plot of the ANI estimated by mapping versus the containment scores estimated by Mash screen for each sample. Horizontal and vertical dashed lines represent the ideal species-level threshold of 0.95 for the containment score and the estimated ANI, respectively. **Figure S7. a)** Percentage of reads that could be mapped to the CBS representatives, grouped by Class. **b)** Per sample percentage of the reads that could be mapped to the CBS representatives, grouped by Class. **Figure S8.** The " *Candidatus Jiangella antarctica* " was found in each sample. **a)** Scatter plot of the ANI estimated by mapping versus the containment scores estimated by Mash screen ($p < 1.47x10^{-21}$). **b)** Scatter plot of the median depth of coverage estimated by mapping versus the median multeplicity estimated by Mash. The line of equality is represented in black. **Supplementary Figure S9.** Jaccard distance between the KEGG functional profiles for each Order.

Additional file 2: Supplementary Table 1. Results of the CBS detection procedure and the validation using Mash Screen. Each row reports: CBS ID (i.e. the CBS MAG representative), metagenomic sample, estimated depth of coverage (mean, standard deviation, first quartile, median third quartile), number of mapped reads, ANI between the consensus sequences and the CBS representative, coverage breadths at depths from 1 to 5, Mash Screen containment score, number of shared hashes, median multiplicity and containment score p-value. **Supplementary Table 2.** Assembly statistics and taxonomic classification of the MAGs. **Supplementary Table 3.** Abundance of CBS at phylum level, expressed as percentage of reads that could be mapped to the representative CBS. Median: median; Q1 and Q3: first and third quartile; IQR: interquartile range; Mean: mean; SD: standard deviation; #CBS: number of candidate bacterial species belonging to the phylum. **Supplementary Table 4.** Increase in the number of bacterial species for each taxonomic Order provided by the data in the present study, compared to the data available in the GTDB database. **Supplementary Table 5.** Sample metadata. Geographic coordinates of the sampling sites, accession numbers of the raw sequences, accession numbers and N50 of the assembled metagenomes on the JGI IMG/M portal. **Supplementary Table 6.** Prevalence and taxonomic classification for each CBS representative. **Supplementary Table 7.** Summary of Bayesian divergence estimates. For each order we report the mean age of its origin (OO: the split of the order from the closest order) and the 95% CI (OO max and OO min), the origin of the oldest uniquely Antarctic clade (AOO1, the split of the Antarctic clade from a non-Antarctic lineage of the same order), and, where present, the origin of the second oldest antarctic clade (AOO2). See Supplementary Data 1. **Supplementary Table 8.** Number of predicted proteins (NProts) and of proteins that had a match in the EggNOG database (NHitsOG) and that could be associated to a term in the Gene Ontology (NHitsGO) or had a match in the KEGG and COG databases (NHitsKEGG and NHitsCOG, respectively). **Supplementary Table 9.** Number of KEGG orthologs characteristic of the Antarctic or reference *Jiangellales* genomes. The Fisher's exact test (uncorrected $p<0.05$) was performed to identify unevenly distributed orthologs between the two groups. **Supplementary Table 10.** Number of KEGG orthologs characteristic of the Antarctic or reference *Thermomicrobiales* genomes. The Fisher's exact test (uncorrected $p<0.05$) was performed to identify unevenly distributed orthologs between the two groups.

Additional file 3: Supplementary data.

Acknowledgements

L.S. and C.C. wish to thank the Italian National Program for Antarctic Research for funding sampling campaigns and research activities in Italy in the frame of PNRA projects. The Italian Antarctic National Museum (MNA) is kindly acknowledged for financial support to the Mycological Section of the MNA and for providing rock samples used in this study stored in the Culture Collection of Antarctic fungi (MNA-FCC), University of Tuscia, Italy. We wish to thank Davide Pisani for kindly providing mean posterior estimates for

calibration purposes. J.E. Stajich is a CIFAR fellow in the program Fungal Kingdom: Threats and Opportunities.

Authors' contributions

L.S. and J.E.S. are PI and co-PI of the "Metagenomic reconstruction of endolithic communities from Victoria Land, Antarctica" Joint Genome Institute Community Sequencing Project 503708. S.T. is the Leader of the Metagenomic Group. Samples were collected by L.S. during the XXXI Italian Antarctic Expedition (2015–2016); C.C., L.S., J.E.S, D.A., and C.D. designed the research; C.C. performed DNA extraction and quality check control; D.A., C.C., O.R.S., and C.D. analyzed the data; C.C., L.S.,D.A., J.E.S., and C.D. wrote the paper with input from O.R.S., S.O., and S.T. The authors read and approved the final manuscript.

Funding

The work conducted by the U.S. Department of Energy Joint Genome Institute, a DOE Office of Science User Facility, is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.

Availability of data and materials

Raw metagenomes reads and assemblies are deposited under the NCBI accession numbers listed in Supplementary Table 5. Metagenome assemblies, gene predictions, and JGI annotations are available in the IMG/M web site (https://img.jgi.doe.gov) and in the zenodo repository (https://zenodo.org/record/3610489; DOI: https://doi.org/10.5281/zenodo.3610489). MAGs, translated coding sequences and annotations for high-quality MAGs, metadata, and Candidatus *Jiangella antarctica* ribosomal rRNA genes are available at the zenodo repository (DOI: https://doi.org/10.5281/zenodo.3671352).

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

[1]Research and Innovation Centre, Fondazione Edmund Mach, Via E. Mach 1, 38098 San Michele all'Adige, Italy. [2]Department of Ecological and Biological Sciences, University of Tuscia, Largo dell'Università, 01100 Viterbo, Italy. [3]Department of Energy Joint Genome Institute, One Cyclotron Road, Berkeley, CA 94720, USA. [4]Department of Microbiology and Plant Pathology and Institute of Integrative Genome Biology, University of California, Watkins Drive 3401, Riverside, Riverside, CA 92507, USA. [5]Mycological Section, Italian Antarctic National Museum (MNA), Via al Porto Antico, 16128 Genoa, Italy.

Received: 9 December 2020 Accepted: 2 February 2021
Published online: 19 March 2021

References

1. Mergelov N, Mueller CW, Prater I, Shorkunov I, Dolgikh A, Zazovskaya E, et al. Alteration of rocks by endolithic organisms is one of the pathways for the beginning of soils on earth. Sci Rep. 2018;8:3367.
2. Cockell CS. The ultraviolet history of the terrestrial planets—implications for biological evolution. Planet Space Sci. 2000:203–14. https://doi.org/10.1016/s0032-0633(99)00087-2.
3. Friedmann EI, Koriem AM. Life on Mars: how it disappeared (if it was ever there). Adv Space Res. 1989;9:167–72.
4. Friedmann EI. The antarctic cold desert and the search for traces of life on Mars. Adv Space Res. 1986:265–8. https://doi.org/10.1016/0273-1177(86)90095-5.
5. Friedmann EI. Endolithic microorganisms in the antarctic cold desert. Sci sciencesciencemagorg. 1982;215:1045–53.
6. Ju SGIF, Golubic S, Ju FI. The Lithobiontic ecological niche, with special reference to microorganisms. SEPM. J Sediment Res. 1981. https://doi.org/10.1306/212f7cb6-2b24-11d7-8648000102c1865d.
7. Pointing SB, Belnap J. Microbial colonization and controls in dryland systems. Nat Rev Microbiol. 2012:551–62. https://doi.org/10.1038/nrmicro2831.
8. Friedmann EI, Ocampo R. Endolithic blue-green algae in the dry valleys: primary producers in the antarctic desert ecosystem. Science. 1976;193:1247–9.
9. Nienow JA. Extremophiles: dry environments (including Cryptoendoliths). Refer Mod Life Sci. 2017. https://doi.org/10.1016/b978-0-12-809633-8.13046-8.
10. Omelon CR. Endolithic microorganisms and their habitats . Their world: a diversity of microbial. Environments. 2016:171–201. https://doi.org/10.1007/978-3-319-28071-4_4.
11. Archer SDJ, de los Ríos A, Lee KC, Niederberger TS, Craig Cary S, Coyne KJ, et al. Endolithic microbial diversity in sandstone and granite from the McMurdo dry valleys, Antarctica. Polar Biol. 2017:997–1006. https://doi.org/10.1007/s00300-016-2024-9.
12. Coleine C, Stajich JE, Pombubpa N, Zucconi L, Onofri S, Canini F, et al. Altitude and fungal diversity influence the structure of Antarctic cryptoendolithic bacteria communities. Environ Microbiol Rep. 2019;11:718–26.
13. Coleine C, Stajich JE, Zucconi L, Onofri S, Pombubpa N, Egidi E, et al. Antarctic cryptoendolithic fungal communities are highly adapted and dominated by lecanoromycetes and dothideomycetes. Front Microbiol. 2018;9:1392.
14. Walker JJ, Pace NR. Endolithic microbial ecosystems. Annu Rev Microbiol. 2007;61:331–47.
15. Chan Y, Van Nostrand JD, Zhou J, Pointing SB, Farrell RL. Functional ecology of an Antarctic Dry Valley. In: Proceedings of the National Academy of Sciences; 2013. p. 8990–5. https://doi.org/10.1073/pnas.1300643110.
16. Coleine C, Albanese D, Onofri S, Tringe SG, Pennacchio C, Donati C, et al. Metagenomes in the borderline ecosystems of the Antarctic cryptoendolithic communities. Microbiology Res Announc. 2020;9(10).
17. Barrett JE, Virginia RA, Hopkins DW, Aislabie J, Bargagli R, Bockheim JG, et al. Terrestrial ecosystem processes of Victoria Land, Antarctica. Soil Biol Biochem. 2006:3019–34. https://doi.org/10.1016/j.soilbio.2006.04.041.
18. Li H. BFC: correcting Illumina sequencing errors. Bioinformatics. 2015;31:2885–7.
19. Nurk S, Meleshko D, Korobeynikov A, Pevzner PA. metaSPAdes: a new versatile metagenomic assembler. Genome Res. 2017;27:824–34.
20. Kang DD, Li F, Kirton E, Thomas A, Egan R, An H, et al. MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. Peer J. 2019;7:e7359.
21. Langmead B, Salzberg SL. Fast gapped-read alignment with bowtie 2. Nat Methods. 2012;9:357–9.
22. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The sequence alignment/map format and SAMtools. Bioinformatics. 2009;25:2078–9.
23. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res. 2015;25:1043–55.
24. Bowers RM, Kyrpides NC, Stepanauskas R, Harmon-Smith M, Doud D, Reddy TBK, et al. Corrigendum: minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea. Nat Biotechnol. 2018;36:660.
25. Jain C, Rodriguez-R LM, Phillippy AM, Konstantinidis KT, Aluru S. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. Nat Commun. 2018;9:5114.
26. Olm MR, Brown CT, Brooks B, Banfield JF. dRep: a tool for fast and accurate genomic comparisons that enables improved genome recovery from metagenomes through de-replication. ISME J. 2017;11(12):2864–8.
27. Parks DH, Chuvochina M, Waite DW, Rinke C, Skarshewski A, Chaumeil P-A, et al. A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. Nat Biotechnol. 2018;36:996–1004.
28. Chaumeil P-A, Mussig AJ, Hugenholtz P, Parks DH. GTDB-Tk: a toolkit to classify genomes with the genome taxonomy database. Bioinformatics. 2019. https://doi.org/10.1093/bioinformatics/btz848.
29. Oren A, da Costa MS, Garrity GM, Rainey FA, Rosselló-Móra R, Schink B, et al. Proposal to include the rank of phylum in the international code of nomenclature of prokaryotes. Int J Syst Evol Microbiol. 2015;65:4284–7.

30. Whitman WB, Oren A, Chuvochina M, da Costa MS, Garrity GM, Rainey FA, et al. Proposal of the suffix -ota to denote phyla. Addendum to "proposal to include the rank of phylum in the international code of nomenclature of prokaryotes". Int J Syst Evol Microbiol. 2018;68:967–9.
31. Price MN, Dehal PS, Arkin AP. FastTree 2 - approximately maximum-likelihood trees for large alignments. PLoS One. 2010;5(3):e9490
32. Quinlan AR. BEDTools: the Swiss-Army tool for genome feature analysis. Curr Protoc Bioinformatics. 2014;47:11.12.1–34.
33. Li H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics. 2011;27:2987–93.
34. Li H. Tabix: fast retrieval of sequence features from generic TAB-delimited files. Bioinformatics. 2011;27:718–9.
35. Ondov BD, Starrett GJ, Sappington A, Kostic A, Koren S, Buck CB, et al. Mash screen: high-throughput sequence containment estimation for genome discovery. Genome Biol. 2019;20:232.
36. Ondov BD, Treangen TJ, Melsted P, Mallonee AB, Bergman NH, Koren S, et al. Mash: fast genome and metagenome distance estimation using MinHash. Genome Biol. 2016;17:132.
37. Betts HC, Puttick MN, Clark JW, Williams TA, Donoghue PCJ, Pisani D. Integrated genomic and fossil evidence illuminates life's early evolution and eukaryote origin. Nat Ecol Evol. 2018:1556–62. https://doi.org/10.1038/s41559-018-0644-x.
38. Drummond AJ, Rambaut ABEAST. Bayesian evolutionary analysis by sampling trees. BMC Evol Biol. 2007;7:214.
39. Rambaut A, Drummond AJ, Xie D, Baele G, Suchard MA. Posterior summarization in Bayesian phylogenetics using tracer 1.7. Syst Biol. 2018: 901–4. https://doi.org/10.1093/sysbio/syy032.
40. Seemann T. Prokka: rapid prokaryotic genome annotation. Bioinformatics. 2014;30:2068–9.
41. Hyatt D, Chen G-L, Locascio PF, Land ML, Larimer FW, Hauser LJ. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics. 2010;11:119.
42. Huerta-Cepas J, Forslund K, Coelho LP, Szklarczyk D, Jensen LJ, von Mering C, et al. Fast genome-wide functional annotation through orthology assignment by eggNOG-mapper. Mol Biol Evol. 2017;34:2115–22.
43. Huerta-Cepas J, Szklarczyk D, Forslund K, Cook H, Heller D, Walter MC, et al. eggNOG 4.5: a hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences. Nucleic Acids Res. 2016;44:D286–93.
44. Kanehisa M, Goto S, Sato Y, Kawashima M, Furumichi M, Tanabe M. Data, information, knowledge and principle: back to metabolism in KEGG. Nucleic Acids Res. 2014;42:D199–205.
45. Tatusov RL. The COG database: a tool for genome-scale analysis of protein functions and evolution. Nucleic Acids Res. 2000:33–6. https://doi.org/10.1093/nar/28.1.33.
46. Steinegger M, Söding J. Clustering huge protein sequence sets in linear time. Nat Commun. 2018;9:2542.
47. Parks DH, Rinke C, Chuvochina M, Chaumeil P-A, Woodcroft BJ, Evans PN, et al. Recovery of nearly 8,000 metagenome-assembled genomes substantially expands the tree of life. Nat Microbiol. 2017;2:1533–42.
48. Tessler M, Neumann JS, Afshinnekoo E, Pineda M, Hersch R, Velho LFM, et al. Large-scale differences in microbial biodiversity discovery between 16S amplicon and shotgun sequencing. Sci Rep. 2017;7:6589.
49. Cole JR, Wang Q, Fish JA, Chai B, McGarrell DM, Sun Y, et al. Ribosomal database project: data and tools for high throughput rRNA analysis. Nucleic Acids Res. 2014;42:D633–42.
50. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, et al. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res. 2013;41:D590–6.
51. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. 2014;30:1312–3.
52. Kanehisa M. KEGG: Kyoto encyclopedia of genes and genomes. Nucleic Acids Res. 2000:27–30. https://doi.org/10.1093/nar/28.1.27.
53. Quince C, Walker AW, Simpson JT, Loman NJ, Segata N. Shotgun metagenomics, from sampling to analysis. Nat Biotechnol. 2017:833–44. https://doi.org/10.1038/nbt.3935.
54. Crits-Christoph A, Robinson CK, Ma B, Ravel J, Wierzchos J, Ascaso C, et al. Phylogenetic and functional substrate specificity for endolithic microbial communities in hyper-arid environments. Front Microbiol. 2016;7:301.
55. Barka EA, Vatsa P, Sanchez L, Gaveau-Vaillant N, Jacquard C, Klenk H-P, et al. Taxonomy, physiology, and natural products of actinobacteria. Microbiol Mol Biol Rev. 2016:1–43. https://doi.org/10.1128/mmbr.00019-15.
56. Dhakal D, Pokhrel AR, Shrestha B, Sohng JK. Marine rare actinobacteria: isolation, characterization, and strategies for harnessing bioactive compounds. Front Microbiol. 2017;8:1106.
57. Goodfellow M, Nouioui I, Sanderson R, Xie F, Bull AT. Rare taxa and dark microbial matter: novel bioactive actinobacteria abound in Atacama Desert soils. Antonie Van Leeuwenhoek. 2018;111:1315–32.
58. Sun Y, Shi Y-L, Wang H, Zhang T, Yu L-Y, Sun H, et al. Diversity of bacteria and the characteristics of actinobacteria community structure in Badain Jaran Desert and Tengger Desert of China. Front Microbiol. 2018;9:1068.
59. Pointing SB, Chan Y, Lacap DC, Lau MCY, Jurgens JA, Farrell RL. Highly specialized microbial diversity in hyper-arid polar desert. Proc Natl Acad Sci U S A Natl Acad Sci. 2009;106:19964–9.
60. Aislabie JM, Chhour K-L, Saul DJ, Miyauchi S, Ayton J, Paetzold RF, et al. Dominant bacteria in soils of marble point and Wright Valley, Victoria Land, Antarctica. Soil Biol Biochem. 2006:3041–56. https://doi.org/10.1016/j.soilbio.2006.02.018.
61. Christner BC, Kvitko BH 2nd, Reeve JN. Molecular identification of bacteria and Eukarya inhabiting an Antarctic cryoconite hole. Extremophiles. 2003;7: 177–83.
62. Choe Y-H, Kim M, Woo J, Lee MJ, Lee JI, Lee EJ, et al. Comparing rock-inhabiting microbial communities in different rock types from a high arctic polar desert. FEMS Microbiol Ecol. 2018;94. https://doi.org/10.1093/femsec/fiy070.
63. Song L. Jiangella gansuensis gen. Nov., sp. nov., a novel actinomycete from a desert soil in north-West China. Int J Syst Evol Microbiol. 2005:881–4. https://doi.org/10.1099/ijs.0.63353-0.
64. Yang J, Li X, Huang L, Jiang H. Actinobacterial diversity in the sediments of five cold springs on the Qinghai-Tibet plateau. Front Microbiol. 2015;6:1345.
65. Jiao J-Y, Liu L, Park D-J, Kim C-J, Xiao M, Chen J, et al. Draft genome sequence of Jiangella alkaliphila KCTC 19222T, isolated from cave soil in Jeju, Republic of Korea. Genome Announc. 2015;3. https://doi.org/10.1128/genomeA.00721-15.
66. Saygin H, Ay H, Guven K, Sahin N. Genome-based classification of three novel actinobacteria from the Karakum Desert: Jiangella asiatica sp. nov., Jiangella aurantiaca sp. nov. and Jiangella ureilytica sp. nov. Int J Syst Evol Microbiol. 2020:1993–2002. https://doi.org/10.1099/ijsem.0.004011.
67. Jiao J-Y, Carro L, Liu L, Gao X-Y, Zhang X-T, Hozzein WN, et al. Complete genome sequence of Jiangella gansuensis strain YIM 002T (DSM 44835T), the type species of the genus Jiangella and source of new antibiotic compounds. Stand Genomic Sci. 2017. https://doi.org/10.1186/s40793-017-0226-6.
68. Mohana D, Thippeswamy S, Abhishek R. Antioxidant, antibacterial, and ultraviolet-protective properties of carotenoids isolated from micrococcus spp. Radiat Prot Environ. 2013:168. https://doi.org/10.4103/0972-0464.142394.
69. Seel W, Baust D, Sons D, Albers M, Etzbach L, Fuss J, et al. Carotenoids are used as regulators for membrane fluidity by staphylococcus xylosus. Sci Rep. 2020;10:330.
70. Leung PM, Bay SK, Meier DV, Chiri E, Cowan DA, Gillor O, et al. Energetic basis of microbial growth and persistence in desert ecosystems. mSystems. 2020;5. https://doi.org/10.1128/mSystems.00495-19.
71. Costello EK, Schmidt SK. Microbial diversity in alpine tundra wet meadow soil: novel Chloroflexi from a cold, water-saturated environment. Environ Microbiol. 2006;8:1471–86.
72. Ward LM, Hemp J, Shih PM, McGlynn SE, Fischer WW. Evolution of phototrophy in the chloroflexi phylum driven by horizontal gene transfer. Front Microbiol. 2018;9:260.
73. Spieck E, Spohn M, Wendt K, Bock E, Shively J, Frank J, et al. Extremophilic nitrite-oxidizing Chloroflexi from Yellowstone hot springs. ISME J. 2020;14: 364–79.
74. Pointing SB, Warren-Rhodes KA, Lacap DC, Rhodes KL, McKay CP. Hypolithic community shifts occur as a result of liquid water availability along environmental gradients in China's hot and cold hyperarid deserts. Environ Microbiol. 2007:414–24. https://doi.org/10.1111/j.1462-2920.2006.01153.x.
75. Wong FKY, Lacap DC, Lau MCY, Aitchison JC, Cowan DA, Pointing SB. Hypolithic microbial community of quartz pavement in the high-altitude tundra of Central Tibet. Microb Ecol. 2010;60:730–9.
76. Rhind T, Ronholm J, Berg B, Mann P, Applin D, Stromberg J, et al. Gypsum-hosted endolithic communities of the Lake St. Martin impact structure, Manitoba, Canada: spectroscopic detectability and implications for Mars. Int J Astrobiol. 2014:366–77. https://doi.org/10.1017/s1473550414000378.
77. Ngom M, Oshone R, Diagne N, Cissoko M, Svistoonoff S, Tisa LS, et al. Tolerance to environmental stress by the nitrogen-fixing actinobacterium

Frankia and its role in actinorhizal plants adaptation. Symbiosis. 2016:17–29. https://doi.org/10.1007/s13199-016-0396-9.
78. Archer SDJ, Lee KC, Caruso T, Maki T, Lee CK, Cary SC, et al. Airborne microbial transport limitation to isolated Antarctic soil habitats. Nat Microbiol. 2019;4:925–32.
79. Rausch P, Rühlemann M, Hermes BM, Doms S, Dagan T, Dierking K, et al. Comparative analysis of amplicon and metagenomic sequencing methods reveals key features in the evolution of animal metaorganisms. Microbiome. 2019;7:133.
80. Chong C-W, Pearce DA, Convey P. Emerging spatial patterns in Antarctic prokaryotes. Front Microbiol. 2015;6:1058.
81. Schiaparelli S, Selbmann L, Onofri S, Zucconi L, Isola D, Rottigni M, et al. Distributional records of Antarctic fungi based on strains preserved in the culture collection of fungi from extreme environments (CCFEE) mycological section associated with the Italian National Antarctic Museum (MNA). MycoKeys. 2015:57–71. https://doi.org/10.3897/mycokeys.10.5343.
82. Coleine C, Masonjones S, Sterflinger K, Onofri S, Selbmann L, Stajich JE. Peculiar genomic traits in the stress-adapted cryptoendolithic Antarctic fungus Friedmanniomyces endolithicus. Fungal Biol. 2020;124:458–67.
83. Arnaud E, Halverson GP, Shields-Zhou G. The geological record of Neoproterozoic glaciationsGeological Society of London; 2011.
84. Pierrehumbert RT, Abbot DS, Voigt A, Koll D. Climate of the Neoproterozoic. Annu Rev Earth Planet Sci. 2011:417–60. https://doi.org/10.1146/annurev-earth-040809-152447.
85. Guidetti R, McInnes SJ, Cesari M, Rebecchi L, Rota-Stabelli O. Evolutionary scenarios for the origin of an Antarctic tardigrade species based on molecular clock analyses and biogeographic data. Contrib Zool. 2017:97–110. https://doi.org/10.1163/18759866-08602001.
86. Harley SL, Fitzsimons ICW, Zhao Y. Antarctica and supercontinent evolution: historical perspectives, recent advances and unresolved issues: Geological Society, London, Special Publications; 2013. p. 1–34. https://doi.org/10.1144/sp383.9.
87. de Wit R, Bouvier T. "Everything is everywhere, but, the environment selects"; what did baas Becking and Beijerinck really say? Environ Microbiol. 2006;8:755–8.
88. Bahl J, Lau MCY, Smith GJD, Vijaykrishna D, Cary SC, Lacap DC, et al. Ancient origins determine global biogeography of hot and cold desert cyanobacteria. Nat Commun. 2011;2:163.

Publisher's Note