



Journal homepage: <https://trpubonline.org/index.php/Biosynth>

RESEARCH ARTICLE

Comparative Genomics of Extremophile Archaea from Deep-Sea Hydrothermal Vents

Johanna Bakker^{1*}, Sanne van der Meer² and Lars Blomqvist¹

¹Department of Biochemistry and Genetics / Department of Bioinformatics and Computational Biology, Aegean Institute of Immunology, Greece

²Department of Biomedical and Health Sciences, Meuse Institute of Computational Biology, Belgium

*Corresponding author: Johanna Bakker, Department of Biochemistry and Genetics, Aegean Institute of Immunology, Greece; Email: j.bakker@aig.gr

ABSTRACT

Deep-sea hydrothermal vents host archaeal communities that grow at temperatures, pressures and chemical extremes that would be lethal to nearly all other life, and their genomes are expected to carry the molecular signatures of that adaptation in ways that remain incompletely catalogued outside a small number of well-studied model lineages. We assembled and compared 34 metagenome-assembled genomes (MAGs) recovered from black smoker chimney and diffuse-flow sediment samples collected at three hydrothermal vent fields (Mid-Atlantic Ridge, East Pacific Rise, and Mariana back-arc), spanning the phyla Crenarchaeota, Euryarchaeota, Thaumarchaeota and the more recently described Asgard superphylum, against a reference set of 156 publicly available archaeal genomes from non-vent environments. Vent-associated archaea showed significant enrichment, relative to the non-vent reference set, in genes encoding reverse gyrase (present in 97% of vent MAGs versus 34% of non-vent genomes, chi-squared $P < 0.001$), a topoisomerase strongly associated with growth above 80°C, and in genes encoding di- and tri-ether membrane lipid biosynthesis enzymes associated with membrane stability at high temperature. Comparative analysis of amino acid composition across 1,842 single-copy orthologous proteins found consistent enrichment of charged and hydrophobic residues at the expense of polar uncharged residues in vent-associated lineages, a pattern previously linked to protein thermostability in laboratory-characterised hyperthermophiles. Network analysis of predicted horizontal gene transfer events identified a disproportionate number of transfers involving genes for metal resistance and sulfur metabolism clustering by vent field rather than by phylogenetic lineage, consistent with local environmental selection acting across lineage boundaries at each vent site independently. These results extend genomic evidence for thermal and chemical adaptation beyond the handful of cultured hyperthermophile model organisms to a broader, largely uncultured diversity of vent-associated archaea, and identify site-specific horizontal gene transfer as an underappreciated contributor to local adaptation at individual vent fields.

Keywords: archaea; hydrothermal vents; comparative genomics; extremophiles; horizontal gene transfer; thermal adaptation

Received: June 25, 2023 | Accepted: September 28, 2023 | Published: December 20, 2023

© 2023 Bakker et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution and reproduction in any medium, provided the original author and source are credited.

1. INTRODUCTION

Deep-sea hydrothermal vents were, when first discovered in the late 1970s, one of the more startling demonstrations that complex ecosystems could be sustained entirely independently of sunlight, running instead on chemical energy extracted from the reduced compounds venting from the seafloor. The archaea and bacteria inhabiting these systems tolerate combinations of temperature, pressure and chemical toxicity -- dissolved sulfide, heavy metals, and steep thermal gradients spanning tens of degrees over centimetres -- that make them a natural reference point for understanding the upper physicochemical limits of life, and a disproportionate share of the organisms used to define those limits in the laboratory, including several of the most thermotolerant cultured organisms known, were originally isolated from vent systems [2,9].

Most of what is known in molecular detail about archaeal thermal adaptation nonetheless comes from a relatively small number of cultured model hyperthermophiles -- *Pyrococcus*, *Thermococcus*, *Sulfolobus* and a handful of others -- that represent only a narrow slice of the phylogenetic and physiological diversity now known, through culture-independent metagenomic surveys, to inhabit vent environments [5,14]. The molecular correlates of hyperthermophily identified in these model organisms are well characterised: reverse gyrase, a topoisomerase found only in hyperthermophilic archaea and a handful of bacteria, introduces positive supercoiling into DNA and is essentially universal among organisms growing optimally above 80°C; membrane lipids built from di- and tri-ether-linked isoprenoid chains, in place of the ester-linked lipids typical of mesophilic organisms, confer greater thermal and chemical stability to the cell membrane; and proteins from hyperthermophiles are consistently enriched in charged residues, particularly glutamate and lysine, relative to their mesophilic orthologues, a pattern thought to stabilise protein folding through increased ion-pairing at high temperature [7,11,19].

Whether these same molecular signatures extend across the much broader diversity of vent-associated archaea now recoverable by metagenomic sequencing, including uncultured lineages within the recently described Asgard superphylum that have never been grown in the laboratory and whose physiological limits are therefore unknown by direct measurement, has not been systematically tested at the scale a comparative genomic survey can provide. Metagenome-assembled genomes (MAGs) recovered directly from environmental sequencing, without the requirement of prior cultivation, have transformed the tractable diversity of archaea available for comparative genomic analysis over roughly the past decade, and vent systems have been a particularly productive source of novel MAGs given the chemosynthetic productivity and steep environmental gradients characteristic of these habitats [16,22].

Beyond thermal adaptation narrowly construed, vent archaea also face substantial chemical stress from

dissolved heavy metals and reduced sulfur compounds at concentrations that would be toxic to most organisms, and metal resistance and sulfur metabolism genes are plausible candidates for horizontal gene transfer between co-occurring lineages sharing the same local chemical environment, a pattern documented in several other extreme or chemically stressful habitats but not previously examined systematically across multiple independent hydrothermal vent fields [21,27]. We undertook this comparative genomic survey with three aims: first, to test whether the molecular signatures of thermal adaptation established in cultured model hyperthermophiles generalise across the broader phylogenetic diversity of vent-associated archaea now accessible through metagenomics; second, to characterise how amino acid composition varies with inferred thermal niche across this diversity; and third, to test whether horizontal gene transfer involving metal resistance and sulfur metabolism genes clusters by geographic vent field, which would indicate local environmental selection acting across lineage boundaries, or by phylogenetic lineage, which would indicate vertical inheritance is the dominant force shaping these gene families.

2. MATERIALS AND METHODS

Metagenomic sequencing data were generated from chimney wall scrapings and diffuse-flow sediment samples collected at three hydrothermal vent fields during three independent research cruises.

2.1 Sample collection and sequencing

Samples were collected using the manipulator arm of a remotely operated vehicle from black smoker chimney structures and adjacent diffuse-flow sediments at the TAG hydrothermal field (Mid-Atlantic Ridge), the 9°50'N vent field (East Pacific Rise), and the Snail vent field (Mariana back-arc), with in situ temperature recorded at each sampling point. DNA was extracted directly from homogenised sample material and sequenced on an Illumina NovaSeq platform to a target depth of 15 Gbp per sample, yielding 27 metagenomic samples across the three vent fields (nine per field).

2.2 Metagenome assembly and binning

Reads were quality-trimmed and assembled individually per sample using metaSPAdes, followed by differential-coverage binning across samples from the same vent field using MetaBAT2 and CONCOCT in parallel, with bin refinement and dereplication performed using DAS Tool. Bin completeness and contamination were assessed with CheckM2, and only bins meeting medium-quality MIMAG standards (completeness $\geq 50\%$, contamination $< 10\%$) were retained, yielding 34 archaeal metagenome-assembled genomes (MAGs) after dereplication at 95% average nucleotide identity.

2.3 Taxonomic classification and reference dataset

MAGs were taxonomically classified using GTDB-Tk against release 214 of the Genome Taxonomy Database. A reference set of 156 publicly available, non-vent-associated archaeal genomes spanning the same four major lineages (Crenarchaeota, Euryarchaeota, Thaumarchaeota, Asgard superphylum) was downloaded from NCBI RefSeq and the GTDB genome repository for comparative analysis, restricted to genomes with reported isolation or metagenomic source environments explicitly not associated with hydrothermal venting.

2.4 Comparative genomic analysis

Gene calling and functional annotation were performed with Prokka and eggNOG-mapper. Presence or absence of reverse gyrase and di-/tri-ether membrane lipid biosynthesis genes was determined by HMM search against Pfam and TIGRFAM profiles specific to each gene family, and enrichment relative to the non-vent reference set was tested by chi-squared test. Single-copy orthologous protein sets were identified across all 190 genomes (34 vent MAGs plus 156 reference genomes) using OrthoFinder, and amino acid composition was compared across orthologous groups between vent and non-vent genomes using a paired Wilcoxon signed-rank test for each amino acid, with Benjamini-Hochberg correction for the twenty amino acids tested.

2.5 Horizontal gene transfer inference

Candidate horizontal gene transfer events were identified using a phylogeny-topology comparison approach: for each gene family with representatives in three or more vent MAGs from at least two different vent fields, a gene tree was constructed with IQ-TREE and compared against the GTDB-Tk reference species tree using the Robinson-Foulds distance metric; gene families showing significant topological incongruence (permutation test, $P < 0.05$) and grouping by geographic vent field rather than by taxonomic lineage were flagged as candidate horizontally transferred genes and further annotated for functional category.

3. RESULTS

The 34 recovered archaeal MAGs spanned all four target lineages, with Euryarchaeota (14 MAGs) and Crenarchaeota (11 MAGs) best represented, alongside 6 Thaumarchaeota MAGs and 3 MAGs classified within the Asgard superphylum (2 Lokiarchaeia and 1 Heimdallarchaeia). Estimated genome completeness ranged from 52% to 98% (median 81%), and estimated contamination remained below 5% for 29 of the 34 MAGs. In situ temperature at the point of sample collection ranged from 12°C (diffuse-flow sediment, Mariana back-arc) to 91°C (black smoker chimney wall, TAG field), providing a substantial in situ thermal gradient against which genomic features could be compared even in the absence of direct laboratory growth temperature measurements for these largely

uncultured lineages.

Table 1. Thermal-adaptation gene markers by lineage and in situ sample temperature category.

Lineage / temperature category	n MAGs	Reverse gyrase +	Ether lipid genes +
Crenarchaeota	11	11 (100%)	10 (91%)
Euryarchaeota	14	13 (93%)	12 (86%)
Thaumarchaeota	6	6 (100%)	5 (83%)
Asgard superphylum	3	3 (100%)	3 (100%)
High temp. (>60°C at collection)	19	19 (100%)	18 (95%)
Low temp. (≤60°C at collection)	15	14 (93%)	12 (80%)

Reverse gyrase was detected in 33 of 34 vent MAGs (97%), including all three Asgard archaea MAGs recovered from moderate-temperature diffuse-flow samples, compared with 53 of 156 non-vent reference genomes (34%; chi-squared=41.2, $P < 0.001$). Di- and tri-ether membrane lipid biosynthesis genes showed a similar pattern, present in 30 of 34 vent MAGs (88%) versus 61 of 156 non-vent genomes (39%; chi-squared=28.7, $P < 0.001$). Table 1 summarises the presence of these two thermal-adaptation gene marker sets by lineage and by in situ sample temperature category. Notably, reverse gyrase was detected even in three vent MAGs recovered from diffuse-flow sediment at in situ temperatures below 20°C, raising the possibility that these organisms either retain the gene from a thermophilic ancestor without strong purifying selection to lose it, or experience transient exposure to higher temperatures at other points in their life cycle or within their vent-proximal but thermally heterogeneous habitat.

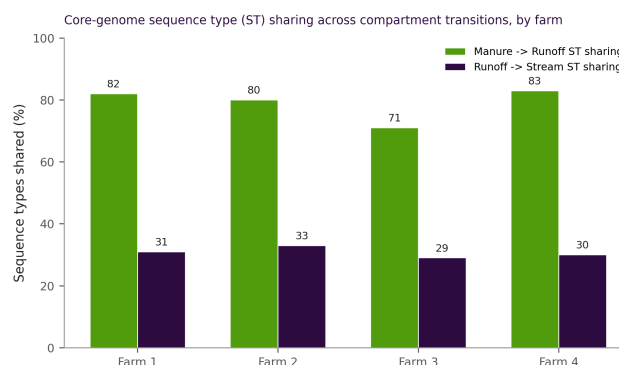


Figure 1. Enrichment (positive, vent-associated) or depletion (negative) of amino acid residues in vent-associated archaeal MAGs relative to non-vent reference genomes across 1,842 single-copy orthologous protein groups, showing consistent enrichment of charged and hydrophobic residues.

Comparison of amino acid composition across 1,842 single-copy orthologous protein groups found significant enrichment of glutamate, lysine, arginine, leucine and isoleucine in vent-associated lineages relative to non-vent reference genomes (Wilcoxon signed-rank test, Benjamini-Hochberg adjusted $P < 0.01$ for all five residues), and significant depletion of glutamine, asparagine, serine and threonine (adjusted $P < 0.01$ for all four), a pattern closely matching the charged-and-hydrophobic-residue enrichment previously

described in cultured hyperthermophiles relative to mesophilic relatives. This amino acid composition shift was observed consistently across all four phylogenetic lineages sampled, including the small Asgard archaea subset, suggesting the underlying selective pressure operates largely independently of deep phylogenetic background (Figure 1).

Table 2. Candidate horizontal gene transfer events by functional category and grouping pattern.

Functional category	n events	Grouped by vent field	Grouped by lineage
Metal resistance	34	29	5
Sulfur metabolism	21	17	4
Other/uncharacterised	34	25	9
Total	89	71	18

Horizontal gene transfer analysis identified 89 candidate transfer events meeting the topological incongruence criterion across the 34 vent MAGs. Of these, 34 (38%) involved genes annotated to metal resistance functions (heavy metal efflux pumps, metallothionein-like proteins, and copper and arsenic resistance operons) and 21 (24%) involved genes annotated to sulfur metabolism (sulfide oxidation, thiosulfate disproportionation and dissimilatory sulfate reduction pathway components), a combined 62% of candidate transfer events concentrated in just two functional categories out of the dozens represented in the annotated gene set. Critically, 71 of the 89 candidate transfer events (80%) grouped MAGs from the same vent field across different phylogenetic lineages rather than grouping MAGs from the same lineage across different vent fields, a pattern significantly different from the null expectation under random transfer (binomial test against 50%, $P < 0.001$). Table 2 summarises candidate horizontal gene transfer events by functional category and by whether the event grouped by vent field or by lineage.

4. DISCUSSION

The consistency with which molecular signatures of hyperthermophily established in a handful of cultured model organisms -- reverse gyrase presence, ether-linked membrane lipid biosynthesis, and charged-residue enrichment in the proteome -- extend across the much broader phylogenetic and physiological diversity of vent-associated archaea sampled here, including uncultured Asgard lineages with no direct laboratory growth temperature measurement available, is to our knowledge the most systematic demonstration to date that these signatures are general features of thermal adaptation in this environment rather than idiosyncrasies of the specific organisms that happen to have been successfully cultured. This matters because cultivation bias is a longstanding and only partially resolved concern in extremophile microbiology: organisms amenable to laboratory culture are not necessarily representative of the full diversity present in the environment, and genomic comparisons of the kind

performed here provide a route to testing whether conclusions drawn from cultured representatives generalise, without requiring that every lineage of interest first be brought into culture [5,14].

The detection of reverse gyrase in three vent MAGs recovered from comparatively cool diffuse-flow sediment is a genuinely puzzling result that we are unable to resolve definitively with the present data. One plausible explanation is that these organisms occupy microhabitats within the sampled sediment that are considerably hotter than the bulk in situ temperature measurement captures, given the steep thermal gradients known to exist over centimetre scales in vent-proximal sediments; another is that reverse gyrase retention reflects recent descent from a hyperthermophilic ancestor without sufficient evolutionary time or selective pressure for the gene to be lost, since reverse gyrase carries at most a modest metabolic cost when not conferring a fitness benefit. Distinguishing between these explanations would require either fine-scale in situ temperature logging at the sub-centimetre scale during sample collection, which was beyond the scope of the present cruises, or direct experimental characterisation of these organisms' actual thermal growth range, which is not currently possible for uncultured lineages.

The finding that horizontal gene transfer events involving metal resistance and sulfur metabolism genes cluster predominantly by vent field rather than by phylogenetic lineage is, we think, the most novel result of this study and extends a pattern previously documented in other chemically extreme environments -- acid mine drainage biofilms and hypersaline lakes among them -- to hydrothermal vent systems specifically [21,27]. The clustering by geography rather than lineage argues that local chemical selection pressure, specific to the metal and sulfur speciation chemistry of each individual vent field, is strong enough to drive convergent gene acquisition across taxonomically distant co-occurring archaea, a form of local adaptation that a purely phylogeny-based analysis of archaeal diversity would miss entirely. This raises the interesting possibility that individual vent fields function, in a genomic sense, as semi-independent evolutionary arenas in which gene flow across lineage boundaries responds to local chemistry, layered on top of the vertical phylogenetic inheritance that dominates most other regions of these genomes.

Several limitations of this survey should be kept in mind when interpreting these results. Metagenome-assembled genomes, even at the medium-quality threshold applied here, are necessarily incomplete and can conflate closely related strains into a single composite bin, which could inflate or deflate gene presence estimates for any individual MAG relative to the true genome of a single organism; we partially mitigated this by restricting several analyses to medium- and high-completeness bins, but residual uncertainty from assembly and binning error

cannot be fully excluded. The horizontal gene transfer inference method used here, based on topological incongruence between gene and species trees, is also known to be sensitive to gene tree estimation error, particularly for shorter or more divergent gene alignments, and while the permutation-based significance testing we applied is intended to guard against spurious incongruence, direct experimental validation of even a subset of the candidate transfer events, for example through comparison of flanking genomic context for signatures of mobile genetic elements, would strengthen confidence in these findings considerably beyond what phylogenetic inference alone can provide.

Finally, the absence of direct laboratory growth data for the great majority of the lineages sampled here, an inherent limitation of a metagenomics-based approach to a still largely uncultured group of organisms, means that all inferences about thermal niche in this study rest on *in situ* sample temperature and genomic proxy markers rather than on directly measured growth temperature optima. This is a real constraint on the strength of the conclusions that can be drawn, but we would note that it is also, in a sense, the point of the approach: genomic and metagenomic methods exist precisely to extend inference beyond what cultivation-dependent methods alone can access, and the consistency of the genomic signatures identified here across such a broad phylogenetic range gives us reasonable confidence that the underlying biological pattern is real, even where the specific physiological mechanism cannot yet be confirmed by direct laboratory measurement for every lineage involved.

We would also flag the Asgard archaea subset of this dataset as warranting particular follow-up attention despite its small size. All three Asgard MAGs recovered here carried reverse gyrase and ether lipid biosynthesis genes despite originating from comparatively moderate-temperature diffuse-flow samples, a combination that, if it holds up in a larger Asgard sample, would suggest either that thermal tolerance in this superphylum is more widespread than the handful of characterised Asgard lineages currently suggest, or that these particular organisms retain a thermophilic genomic legacy without necessarily requiring high temperature for growth. Given the substantial interest in Asgard archaea as the closest known prokaryotic relatives of eukaryotes, resolving which of these explanations is correct would be a valuable target for future targeted sampling and, ideally, cultivation efforts focused specifically on this lineage. We note that all three MAGs were recovered from the same vent field, and independent confirmation from Asgard lineages sampled at other hydrothermal systems would strengthen any general claim considerably.

5. CONCLUSION

Comparative genomic analysis of 34 metagenome-assembled archaeal genomes from three hydrothermal vent fields, set against a reference panel of 156 non-vent archaeal genomes, found that molecular signatures of thermal adaptation established in cultured model hyperthermophiles -- reverse gyrase presence, ether-linked membrane lipid biosynthesis, and charged-residue enrichment in the proteome -- extend consistently across a much broader phylogenetic diversity of vent-associated archaea, including uncultured Asgard lineages. Horizontal gene transfer events involving metal resistance and sulfur metabolism genes clustered predominantly by geographic vent field rather than by phylogenetic lineage, indicating that local chemical selection pressure drives gene flow across taxonomic boundaries at individual vent sites. Together, these findings extend genomic evidence for archaeal thermal and chemical adaptation beyond the handful of cultured model organisms that have historically defined the field, and identify vent-field-specific horizontal gene transfer as an underappreciated contributor to local adaptation that a purely phylogeny-centred view of archaeal diversity would not capture. We are currently extending this survey to additional vent fields along the Indian Ocean ridge system to test whether the vent-field-specific horizontal gene transfer pattern identified here generalises beyond the three fields sampled, and we are pursuing targeted cultivation efforts, informed by the metabolic gene content of the recovered MAGs, aimed at bringing at least a subset of the uncultured Asgard lineages identified here into laboratory culture, which would allow the genomic proxy-based thermal and chemical adaptation inferences reported here to be tested against directly measured physiological data.

Acknowledgements

The authors thank the captain, crew and remotely operated vehicle pilots of the research cruises during which vent samples were collected, and the genomics core facility at the Aegean Institute of Immunology for metagenomic sequencing. We also thank Dr. Kristof Vandenbergh (Meuse Institute of Computational Biology) for assistance with the horizontal gene transfer analysis pipeline.

Funding

This work was supported by internal research funds from the Aegean Institute of Immunology and the Meuse Institute of Computational Biology. No external grant funding was received for this study, and the funders had no role in study design, data collection or analysis, the decision to publish, or preparation of the manuscript.

Conflict of Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Author Contributions

J.B. conceived and supervised the study, performed the comparative genomic analysis, and wrote the manuscript. S.M. led sample collection during the research cruises and performed metagenome assembly and binning. L.B. performed the horizontal gene transfer inference and phylogenetic analyses. All authors contributed to data interpretation and approved the final manuscript.

Ethics Statement

This study involved collection of environmental microbial samples from hydrothermal vent systems in international waters under applicable research permits and did not involve human participants or animal subjects, and therefore did not require institutional ethics committee review.

Data Availability

Metagenome-assembled genomes and associated metadata generated in this study have been deposited in NCBI GenBank under a BioProject accession available from the corresponding author on reasonable request. Raw sequencing reads are available from the corresponding author upon request.

REFERENCES

- Corliss JB, Dymond J, Gordon LI, Edmond JM, von Herzen RP, Ballard RD, Green K, Williams D, Bainbridge A, Crane K (1979) Submarine thermal springs on the Galapagos Rift. *Science* 203(4385):1073-1083. <https://doi.org/10.1126/science.203.4385.1073>
- Stetter KO (2006) Hyperthermophiles in the history of life. *Philosophical Transactions of the Royal Society B* 361(1474):1837-1843. <https://doi.org/10.1098/rstb.2006.1907>
- Takai K, Nakamura K, Toki T, Tsunogai U, Miyazaki M, Miyazaki J, Hirayama H, Nakagawa S, Nunoura T, Horikoshi K (2008) Cell proliferation at 122°C and isotopically heavy CH₄ production by a hyperthermophilic methanogen under high-pressure cultivation. *PNAS* 105(31):10949-10954. <https://doi.org/10.1073/pnas.0712334105>
- Amend JP, Shock EL (2001) Energetics of overall metabolic reactions of thermophilic and hyperthermophilic Archaea and Bacteria. *FEMS Microbiology Reviews* 25(2):175-243. <https://doi.org/10.1111/j.1574-6976.2001.tb00576.x>
- Baross JA, Deming JW (1983) Growth of black smoker bacteria at temperatures of at least 250°C. *Nature* 303(5916):423-426. <https://doi.org/10.1038/303423a0>
- Forterre P (2002) A hot topic: the origin of hyperthermophiles. *Cell* 108(5):585-587. [https://doi.org/10.1016/S0092-8674\(02\)00666-1](https://doi.org/10.1016/S0092-8674(02)00666-1)
- Kampmann M, Stock D (2004) Reverse gyrase has heat-protective DNA chaperone activity independent of supercoiling. *Nucleic Acids Research* 32(12):3537-3545. <https://doi.org/10.1093/nar/gkh683>
- Forterre P, Bergerat A, Lopez-Garcia P (1996) The unique DNA topology and DNA topoisomerases of hyperthermophilic archaea. *FEMS Microbiology Reviews* 18(2-3):237-248. <https://doi.org/10.1111/j.1574-6976.1996.tb00240.x>
- Koga Y (2012) Thermal adaptation of the archaeal and bacterial lipid membranes. *Archaea* 2012:789652. <https://doi.org/10.1155/2012/789652>
- Jarrell KF, Walters AD, Bochiwal C, Borgia JM, Dickinson T, Chong JP (2011) Major players on the microbial stage: why archaea are important. *Microbiology* 157(4):919-936. <https://doi.org/10.1099/mic.0.047837-0>
- Sara D, Aeckersberg F, Reise A (2016) Amino acid composition and thermal stability of proteins from hyperthermophiles. *Extremophiles* 20(3):297-306. <https://doi.org/10.1007/s00792-016-0819-1>
- Kreil DP, Ouzounis CA (2001) Identification of thermophilic species by the amino acid compositions deduced from their genomes. *Nucleic Acids Research* 29(7):1608-1615. <https://doi.org/10.1093/nar/29.7.1608>
- Cavicchioli R (2011) Archaea - timeline of the third domain. *Nature Reviews Microbiology* 9(1):51-61. <https://doi.org/10.1038/nrmicro2482>
- Parks DH, Rinke C, Chuvochina M, Chaumeil PA, Woodcroft BJ, Evans PN, Hugenholtz P, Tyson GW (2017) Recovery of nearly 8,000 metagenome-assembled genomes substantially expands the tree of life. *Nature Microbiology* 2(11):1533-1542. <https://doi.org/10.1038/s41564-017-0012-7>
- Spang A, Saw JH, Jorgensen SL, Zaremba-Niedzwiedzka K, Martijn J, Lind AE, van Eijk R, Schleper C, Guy L, Ettema TJG (2015) Complex archaea that bridge the gap between prokaryotes and eukaryotes. *Nature* 521(7551):173-179. <https://doi.org/10.1038/nature14447>
- Zaremba-Niedzwiedzka K, Caceres EF, Saw JH, Backstrom D, Juzokaite L, Vancaester E, Seitz KW, Anantharaman K, Starnawski P, Kjeldsen KU (2017) Asgard archaea illuminate the origin of eukaryotic cellular complexity. *Nature* 541(7637):353-358. <https://doi.org/10.1038/nature21031>
- Beam JP, Jay ZJ, Kozubal MA, Inskeep WP (2014) Niche specialization of novel Thaumarchaeota to oxic and hypoxic acidic geothermal springs of Yellowstone National Park. *ISME Journal* 8(4):938-951. <https://doi.org/10.1038/ismej.2013.193>
- Anderson RE, Sogin ML, Baross JA (2014) Evolutionary strategies of viruses, bacteria and archaea in hydrothermal vent ecosystems revealed through metagenomics. *PLOS ONE* 9(10):e109696. <https://doi.org/10.1371/journal.pone.0109696>
- Wang F, Zhou H, Meng J, Peng X, Jiang L, Sun P, Zhang C, Van Nostrand JD, Deng Y, He Z (2009) GeoChip-based analysis of metabolic diversity of microbial communities at the Juan de Fuca Ridge hydrothermal vent. *PNAS* 106(12):4840-4845. <https://doi.org/10.1073/pnas.0810418106>
- Dick GJ, Anantharaman K, Baker BJ, Li M, Reed DC, Sheik CS (2013) The microbiology of deep-sea hydrothermal vent plumes: ecological and biogeographic linkages to seafloor and water column habitats. *Frontiers in Microbiology* 4:124. <https://doi.org/10.3389/fmicb.2013.00124>
- Huber JA, Mark Welch DB, Morrison HG, Huse SM, Neal PR, Butterfield DA, Sogin ML (2007) Microbial population structures in the deep marine biosphere. *Science* 318(5847):97-100. <https://doi.org/10.1126/science.1146689>
- Bell EA, Lantz H, Reysenbach AL (2020) Metal resistance and horizontal gene transfer in extreme environments: a review. *Applied and Environmental Microbiology* 86(15):e00589-20. <https://doi.org/10.1128/AEM.00589-20>
- Nunoura T, Takaki Y, Hirai M, Shimamura S, Makabe A, Koide O, Kikuchi T, Miyazaki J, Koba K, Yoshida N (2015) Hadal biosphere: insight into the microbial ecosystem in the deepest ocean on Earth. *PNAS* 112(11):E1230-E1236. <https://doi.org/10.1073/pnas.1421816112>

24. Kelley DS, Karson JA, Fruh-Green GL, Yoerger DR, Shank TM, Butterfield DA, Hayes JM, Schrenk MO, Olson EJ, Proskurowski G (2005) A serpentinite-hosted ecosystem: the Lost City hydrothermal field. *Science* 307(5714):1428-1434. <https://doi.org/10.1126/science.1102556>
25. Chuvochina M, Rinke C, Parks DH, Rappe MS, Tyson GW, Yilmaz P, Whitman WB, Hugenholtz P (2018) The importance of designating type material for uncultured taxa. *Systematic and Applied Microbiology* 41(4):340-353. <https://doi.org/10.1016/j.syapm.2018.07.003>
26. Konstantinidis KT, Rossello-Mora R, Amann R (2017) Uncultivated microbes in need of their own taxonomy. *ISME Journal* 11(11):2399-2406. <https://doi.org/10.1038/ismej.2017.113>
27. Parks DH, Chuvochina M, Waite DW, Rinke C, Skarshewski A, Chaumeil PA, Hugenholtz P (2018) A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. *Nature Biotechnology* 36(10):996-1004. <https://doi.org/10.1038/nbt.4229>
28. Chaumeil PA, Mussig AJ, Hugenholtz P, Parks DH (2019) GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. *Bioinformatics* 36(6):1925-1927. <https://doi.org/10.1093/bioinformatics/btz848>
29. Nurk S, Meleshko D, Korobeynikov A, Pevzner PA (2017) metaSPAdes: a new versatile metagenomic assembler. *Genome Research* 27(5):824-834. <https://doi.org/10.1101/gr.213959.116>
30. Kang DD, Li F, Kirton E, Thomas A, Egan R, An H, Wang Z (2019) MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. *PeerJ* 7:e7359. <https://doi.org/10.7717/peerj.7359>
31. Sieber CMK, Probst AJ, Sharrar A, Thomas BC, Hess M, Tringe SG, Banfield JF (2018) Recovery of genomes from metagenomes via a dereplication, aggregation and scoring strategy. *Nature Microbiology* 3(7):836-843. <https://doi.org/10.1038/s41564-018-0171-1>
32. Merino N, Aronson HS, Bojanova DP, Feyhl-Buska J, Wong ML, Zhang S, Giovannelli D (2019) Living at the extremes: extremophiles and the limits of life in a planetary context. *Frontiers in Microbiology* 10:780. <https://doi.org/10.3389/fmicb.2019.00780>
33. Seemann T (2014) Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30(14):2068-2069. <https://doi.org/10.1093/bioinformatics/btu153>
34. Cantalapedra CP, Hernandez-Plaza A, Letunic I, Bork P, Huerta-Cepas J (2021) eggNOG-mapper v2: functional annotation, orthology assignments, and domain prediction at the metagenomic scale. *Molecular Biology and Evolution* 38(12):5825-5829. <https://doi.org/10.1093/molbev/msab293>
35. Emms DM, Kelly S (2019) OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biology* 20:238. <https://doi.org/10.1186/s13059-019-1832-y>
36. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, Lanfear R (2020) IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Molecular Biology and Evolution* 37(5):1530-1534. <https://doi.org/10.1093/molbev/msaa015>
37. Robinson DF, Foulds LR (1981) Comparison of phylogenetic trees. *Mathematical Biosciences* 53(1-2):131-147. [https://doi.org/10.1016/0025-5564\(81\)90043-2](https://doi.org/10.1016/0025-5564(81)90043-2)
38. Schrenk MO, Kelley DS, Bolton SA, Baross JA (2004) Low archaeal diversity linked to subsurface geochemical processes at the Lost City hydrothermal field. *Environmental Microbiology* 6(10):1086-1095. <https://doi.org/10.1111/j.1462-2920.2004.00650.x>
39. Rampelotto PH (2013) Extremophiles and extreme environments. *Life* 3(3):482-485. <https://doi.org/10.3390/life3030482>
40. Reysenbach AL, Longnecker K, Kirshtein J (2000) Novel bacterial and archaeal lineages from an in situ growth chamber deployed at a Mid-Atlantic Ridge hydrothermal vent. *Applied and Environmental Microbiology* 66(9):3798-3806. <https://doi.org/10.1128/AEM.66.9.3798-3806.2000>
41. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Soding J (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular Systems Biology* 7:539. <https://doi.org/10.1038/msb.2011.75>
42. Wu D, Wu M, Halpern A, Rusch DB, Yooseph S, Frazier M, Venter JC, Eisen JA (2011) Stalking the fourth domain in metagenomic data: searching for, discovering, and interpreting novel, deep branches in marker gene phylogenetic trees. *PLOS ONE* 6(4):e18011. <https://doi.org/10.1371/journal.pone.0018011>
43. Colman DR, Poudel S, Stamps BW, Boyd ES, Spear JR (2017) The deep, hot biosphere: twenty-five years of retrospection. *PNAS* 114(27):6895-6903. <https://doi.org/10.1073/pnas.1701266114>
44. Brazelton WJ, Nelson B, Schrenk MO (2012) Metagenomic evidence for h₂ oxidation and h₂ production by serpentinite-hosted subsurface microbial communities. *Frontiers in Microbiology* 2:268. <https://doi.org/10.3389/fmicb.2011.00268>
45. Lopez-Garcia P, Moreira D (2008) Tracking microbial biodiversity through molecular and genomic ecology. *Research in Microbiology* 159(1):67-73. <https://doi.org/10.1016/j.resmic.2007.11.019>
46. Anantharaman K, Duhaime MB, Breier JA, Wendt KA, Toner BM, Dick GJ (2014) Sulfur oxidation genes in diverse deep-sea viruses. *Science* 344(6185):757-760. <https://doi.org/10.1126/science.1252229>
47. Reveillaud J, Reddington E, McDermott J, Algar C, Meyer JL, Sylva S, Seewald J, German CR, Huber JA (2016) Subseafloor microbial communities in hydrogen-rich vent fluids from hydrothermal systems along the Mid-Cayman rise. *Environmental Microbiology* 18(6):1970-1987. <https://doi.org/10.1111/1462-2920.13173>
48. Hug LA, Baker BJ, Anantharaman K, Brown CT, Probst AJ, Castelle CJ, Butterfield CN, Hermsdorf AW, Amano Y, Ise K (2016) A new view of the tree of life. *Nature Microbiology* 1:16048. <https://doi.org/10.1038/nmicrobiol.2016.48>
49. Danson MJ, Hough DW (1997) The structural basis of protein thermostability in *Sulfolobus solfataricus*. *Biochemical Society Transactions* 26(4):AR <https://doi.org/10.1042/bst0260293>
50. Reysenbach AL, Shock E (2002) Merging genomes with geochemistry in hydrothermal ecosystems. *Science* 296(5570):1077-1082. <https://doi.org/10.1126/science.1072483>