MANUSCRIPT TITLE

Gosha: a database of organisms with defined optimal growth temperatures

AUTHORS

Karla Helena-Bueno[1,†], Charlotte R. Brown[1,†] and Sergey Melnikov[1]

[1] Biosciences Institute, Newcastle University, Newcastle upon Tyne, NE2 4HH, UK

† Joint authors

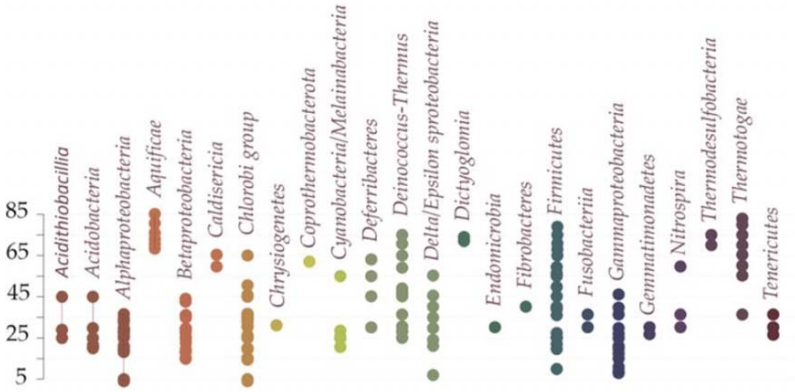
* To whom correspondence should be addressed.

Tel: +44 0191 208 58 10;

Email: sergey.melnikov@ncl.ac.uk

GRAPHICAL ABSTRACT

Gosha (available at www.melnikovlab.com/gshc) is a database that collects information about the optimal growth temperatures of living species. This database aims to facilitate studies of molecular adaptation to specific temperatures.



ABSTRACT

Currently, we are witnessing an explosive accumulation of genomic sequences for organisms across all branches of life. However, typically the genomic data lack the information about optimal growth conditions of corresponding organisms. As a result, it becomes challenging to use the genomic data for studying the adaptations of organisms and biological molecules to diverse environments. To address this problem, we have created a database Gosha, available at http://melnikovlab.com/gshc. This database brings together information about the genomic sequences and optimal growth temperatures for 25,324 species, including ~89% of the bacterial species with known genome sequences. Using this database, one can annotate genomic sequences from thousands of species and correlate variations in genes and genomes with optimal growth temperatures. The database interface allows users to retrieve optimal growth temperatures for bacteria, eukaryotes and archaea, providing a tool to explore how organisms, genomes, and individual proteins and nucleic acids adapt to certain temperatures. We hope that this database will contribute to medicine and biotechnology by helping to create a better understanding of molecular adaptations to heat and cold, leading to new ways to preserve biological samples, engineer useful enzymes, and develop biological materials and organisms with the desired tolerance to heat and cold.

INTRODUCTION

Despite significant research efforts to understand how biological molecules adapt to temperature change (1-36), we are still not able to accurately answer two fundamental questions: What are the most common strategies by which cellular proteins adapt to environmental conditions, such as heat and cold? And can we find a simple and robust approach to alter the thermal tolerance of natural proteins by introducing a minimal number of mutations to the protein sequence?

One challenge in answering these questions stems from the lack of a resource that stores easy-to-use information about the optimal growth conditions of living organisms, together with their genomic data. For example, currently, there are more than 14,400 genome sequences from representative bacterial species publicly available. In principle, we could use these sequences to study thousands of variants of a given protein, observing how its sequence and structure undergo changes upon transition from cold-adapted bacteria (10-16,37-48) to heat-adapted bacteria (26-32,49-69).

In practice, however, it is not immediately possible to organise thousands of organisms (along with their genomic sequences) by their optimal growth temperatures, because the corresponding genomic sequences deposited in public repositories (such as NCBI Genomes) lack information about these organisms' optimal growth conditions. To overcome this challenge, several teams have dedicated their efforts to annotating the existing genomic sequences with experimentally defined growth conditions. Initially, lists of optimal growth conditions were compiled by conducting manual searches through the Bergey's manual of systematic bacteriology (70,71). Later, advancements in automation made it feasible to retrieve data from thousands of species, enabling the creation of a database encompassing 8,645 prokaryotic strains with known genome sequences (72,73). However, despite having access to genome sequences for tens of thousands of representative bacteria, eukaryotes, and archaea, our current capability to efficiently categorize their genomic sequences) by optimal growth conditions remains limited, which hinders our progress in understanding the molecular and organismal adaptation to temperature.

Here, we take one step forward in developing such a resource for scientists and engineers interested in exploring and exploiting molecular adaptations to heat and cold. Using the NCBI database of sequenced genomes as a scaffold, we have created a database Gosha (Genomic sequences and Optimal growth temperatures Show Heat-and-cold Adaptations) in which species with known genome sequences are annotated with information about these species' optimal growth temperature. This database describes the optimal growth temperature of more than 25,000 microorganisms, which significantly increases the volume of the database and includes 89% of the representative bacteria whose genome sequences are deposited in the NCBI database.

With regards to bacterial species, this new resource makes it possible to retrieve up to 12,354 sequences of a given bacterial protein of interest, sort these sequences by the optimal growth temperature of its corresponding species, and explore how each residue in this protein varies in sequence and conservation upon transition from cold-adapted to heat-adapted organisms. Thus, we provide a tool for large-scale studies of the molecular and organismal adaptations to a specific temperature.

DATABASE FEATURES

Downloadable lists of species with known genome sequence and together with their optimal growth temperature

Currently, the Gosha database contains information about the optimal growth temperature for 25,324 species. This information is continuously retrieved by web-scraping 23 public repositories of microorganisms (**Table 1**) and then added to the list of organisms deposited in the NCBI repository of organisms with sequenced genomes. The Gosha site contains the optimal growth temperatures for an organism; however, it does not yet include information about other growth conditions, such as oxygen requirement, pH, pressure, and salt concentration.

Lists of organisms with experimentally defined optimal growth temperatures and a reference to the corresponding genome data can be downloaded from the database. This includes the optimal growth temperature values for 12,265 representative bacteria, 414 representative archaea, and 973 representative eukaryotes. The datasets are updated monthly and are available in .csv format, enabling the species to be sorted by an organism's name, phylogenetic group, genome size, genomic GC-content, number of protein coding genes, and optimal growth temperature. As shown in the example provided for bacterial species (**Figure 1**), organisms contained in the datasets include thermophiles and psychrophiles from all major lineages of species, providing an opportunity to study molecular adaptation to heat and cold at the scale of the entire tree of life.

Optimal growth-temperature checker

In addition to the downloadable data, the database user interface allows searching for the optimal growth temperature for a given species. The user can enter a species name in a search window and retrieve the optimal growth temperature for the species of interest if it is present in the database.

Error and request tracking

To document the rapidly expanding data, most entries in the database are automated, without inspection against published literature. We acknowledge that the optimal growth temperature values for some organisms may contain discrepancies or inaccuracies and therefore encourage users to provide feedback on any unaddressed discrepancies by submitting reports through our feedback tracking system. The submission form allows users to request changes in the database and monitor the progress of each inquiry.

APPLICATIONS

How will this database help researchers to explore molecular mechanisms of thermal adaptation? Below, we propose some example applications:

First, we can now sort genomes or homologous gene sequences by their optimal growth temperature (and not just by phylogenetic origin), making it possible to explore universal strategies of adaptations to heat and cold, as opposed to idiosyncratic adaptations within each lineage of species (**Figure 1**). Such studies would have far-reaching implications in biotechnology as they can simplify the rational design of biological molecules and organisms with a desired thermal stability (32,61,68,74-79).

Second, this database may simplify studies of molecular adaptations to temperature within any given range of temperatures. This is important, because to date most studies have focused on extremophiles (49,50), leaving mostly unexplored how mesophilic organisms adapt to relatively subtle changes in the environment (e.g. temperature increases of a few degrees Celsius as a consequence of climate change (12,20,21,80).

Third, we can monitor how the identity and conservation of each residue in a protein of interest gradually changes across a range of optimal growth temperatures from 2°C to more than 103°C, simplifying studies of structural constraints (21,46,63,79,81-84), and finding new ways to engineer useful proteins with a desired optimal thermal tolerance (24,76-79,84,85) (**Figure 2**).

Fourth, this database can help identify model organisms to observe "extremophiles in the making". Currently, the database contains organisms from the same genus that have almost identical sequences for most of their cellular proteins but exhibit dramatically different optimal growth temperatures. For example, the genus *Clostridium* includes species with an optimal growth temperature ranging from just 5°C (*Clostridium frigoris*) (86), to 55°C (*Clostridium thermobutyricum*) (87). Comparing species in these genera can provide us with a rare opportunity to observe the natural transformation of mesophiles into extremophiles and gain an understanding of how organisms evolve the ability to tolerate heat and cold through minimal changes in their genomes (8,13,32,55,57,69,88). These studies are important, as they may help to simplify the design of economically useful microorganisms with the desired thermal tolerance.

FUTURE DIRECTIONS

We are currently working to expand the database by including additional environmental parameters, such as optimal salt concentration, pH, pressure, and oxygen requirement. We are also testing a new data scraping algorithm to retrieve data not only from repositories of commercially available microorganisms but also from original research papers, to maximise the completeness of our datasets. Finally, we are testing scripts to allow the mapping of temperature-dependent variations in protein sequences to corresponding three-dimensional protein structures that are available in the Protein Data Bank. We hope that in the future our database or similar annotations will be integrated into centralised repositories of genome sequences, such as NCBI, making it possible to explore organismal adaptations to changing

environments using the rapidly expanding collection of genomic sequences for both living and extinct species on our planet.

DATA AVAILABILITY

The data described in this study were deposited in Figshare with the following permanent and citable DOI: https://doi.org/10.6084/m9.figshare.23257940.v1

ACKNOWLEDGEMENTS

We acknowledge Yuri Polikanov for his encouragement that enabled this work to happen. We are also grateful to the members of the Melnikov lab for their contributions, as well as the reviewers of this manuscript, for their critical feedback and valuable suggestions to improve our work and service.

FUNDING

This work was supported by Newcastle University (NUORS 2021 award to K.H-B.), BBSRC UK (4-year PhD studentship BB/T008695/1. to C.R.B.) and the Royal Society UK (RGS\R2\202003 to S.M.). Funding for open access charge: Biotechnology and Biological Sciences Research Council.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

REFERENCES

1. Pinney, M.M., Mokhtari, D.A., Akiva, E., Yabukarski, F., Sanchez, D.M., Liang, R., Doukov, T., Martinez, T.J., Babbitt, P.C. and Herschlag, D. (2021) Parallel molecular mechanisms for enzyme temperature adaptation. *Science*, **371**.
2. Chursov, A., Kopetzky, S.J., Leshchiner, I., Kondofersky, I., Theis, F.J., Frishman, D. and Shneider, A. (2012) Specific temperature-induced perturbations of secondary mRNA structures are associated with the cold-adapted temperature-sensitive phenotype of influenza A virus. *RNA Biology*, **9**, 1266-1274.
3. Sang, P., Liu, S.-Q. and Yang, L.-Q. (2020) New Insight into Mechanisms of Protein Adaptation to High Temperatures: A Comparative Molecular Dynamics Simulation Study of Thermophilic and Mesophilic Subtilisin-Like Serine Proteases. *International Journal of Molecular Sciences*, **21**, 3128.
4. Casanueva, A., Tuffin, M., Cary, C. and Cowan, D.A. (2010) Molecular adaptations to psychrophily: the impact of 'omic' technologies. *Trends Microbiol*, **18**, 374-381.
5. Liao, M.-l., Dong, Y.-w. and Somero, G.N. (2021) Thermal adaptation of mRNA secondary structure: stability versus lability. *Proceedings of the National Academy of Sciences*, **118**, e2113324118.
6. Kashiwagi, A., Sugawara, R., Sano Tsushima, F., Kumagai, T. and Yomo, T. (2014) Contribution of Silent Mutations to Thermal Adaptation of RNA Bacteriophage Qβ. *Journal of Virology*, **88**, 11459-11468.
7. Lorenz, C., Lünse, C. and Mörl, M. (2017) tRNA Modifications: Impact on Structure and Thermal Adaptation. *Biomolecules*, **7**, 35.
8. Goncearenco, A., Ma, B.-G. and Berezovsky, I.N. (2014) Molecular mechanisms of adaptation emerging from the physics and evolution of nucleic acids and proteins. *Nucleic Acids Research*, **42**, 2879-2892.
9. Fields, P.A., Dong, Y., Meng, X. and Somero, G.N. (2015) Adaptations of protein structure and function to temperature: there is more than one way to 'skin a cat'. *Journal of Experimental Biology*, **218**, 1801-1811.
10. Pucci, F. and Rooman, M. (2017) Physical and molecular bases of protein thermal stability and cold adaptation. *Curr Opin Struct Biol*, **42**, 117-128.
11. Feller, G. and Gerday, C. (2003) Psychrophilic enzymes: hot topics in cold adaptation. *Nature Reviews Microbiology*, **1**, 200-208.
12. Russell, R.J., Gerike, U., Danson, M.J., Hough, D.W. and Taylor, G.L. (1998) Structural adaptations of the cold-active citrate synthase from an Antarctic bacterium. *Structure*, **6**, 351-361.
13. Metpally, R. and Reddy, B. (2009) Comparative proteome analysis of psychrophilic versus mesophilic bacterial species: Insights into the molecular basis of cold adaptation of proteins. *BMC Genomics*, **10**, 11.
14. Yang, G., Yao, H., Mozzicafreddo, M., Ballarini, P., Pucciarelli, S. and Miceli, C. (2017) Rational Engineering of a Cold-Adapted α-Amylase from the Antarctic Ciliate Euplotes focardii for Simultaneous Improvement of Thermostability and Catalytic Activity. *Applied and Environmental Microbiology*, **83**, AEM.00449-00417.
15. Siglioccolo, A., Gerace, R. and Pascarella, S. (2010) "Cold spots" in protein cold adaptation: Insights from normalized atomic displacement parameters (B′-factors). *Biophysical Chemistry*, **153**, 104-114.

16. Van Petegem, F., Collins, T., Meuwis, M.A., Gerday, C., Feller, G. and Van Beeumen, J. (2003) The structure of a cold-adapted family 8 xylanase at 1.3 A resolution. Structural adaptations to cold and investgation of the active site. *J Biol Chem*, **278**, 7531-7539.
17. Zeldovich, K.B., Berezovsky, I.N. and Shakhnovich, E.I. (2007) Protein and DNA Sequence Determinants of Thermophilic Adaptation. *PLoS Computational Biology*, **3**, e5.
18. Cipolla, A., Delbrassine, F., Da Lage, J.-L. and Feller, G. (2012) Temperature adaptations in psychrophilic, mesophilic and thermophilic chloride-dependent alpha-amylases. *Biochimie*, **94**, 1943-1950.
19. D'Amico, S., Gerday, C. and Feller, G. (2003) Temperature adaptation of proteins: engineering mesophilic-like activity and stability in a cold-adapted alpha-amylase. *J Mol Biol*, **332**, 981-988.
20. Dong, Y. and Somero, G.N. (2009) Temperature adaptation of cytosolic malate dehydrogenases of limpets (genus Lottia): differences in stability and function due to minor changes in sequence correlate with biogeographic and vertical distributions. *J Exp Biol*, **212**, 169-177.
21. Dong, Y.-W., Liao, M.-L., Meng, X.-L. and Somero, G.N. (2018) Structural flexibility and protein adaptation to temperature: Molecular dynamics analysis of malate dehydrogenases of marine molluscs. *Proceedings of the National Academy of Sciences*, **115**, 1274-1279.
22. De Vendittis, E., Castellano, I., Cotugno, R., Ruocco, M.R., Raimo, G. and Masullo, M. (2008) Adaptation of model proteins from cold to hot environments involves continuous and small adjustments of average parameters related to amino acid composition. *Journal of Theoretical Biology*, **250**, 156-171.
23. Scandurra, R., Condalvi, V., Chiaraluce, R., Politi, L. and Engel, C., Paul. (2000) Protein Stability in Extremophilic Archaea. *Frontiers in Bioscience*, **5**.
24. Cimen, E., Jensen, S.E. and Buckler, E.S. (2020) Building a tRNA thermometer to estimate microbial adaptation to temperature. *Nucleic Acids Res*, **48**, 12004-12015.
25. Liao, M.-L., Somero, G.N. and Dong, Y.-W. (2019) Comparing mutagenesis and simulations as tools for identifying functionally important sequence changes for protein thermal adaptation. *Proceedings of the National Academy of Sciences*, **116**, 679-688.
26. Berezovsky, I.N. and Shakhnovich, E.I. (2005) Physics and evolution of thermophilic adaptation. *Proceedings of the National Academy of Sciences*, **102**, 12742-12747.
27. Brininger, C., Spradlin, S., Cobani, L. and Evilia, C. (2018) The more adaptive to change, the more likely you are to survive: Protein adaptation in extremophiles. *Seminars in Cell & Developmental Biology*, **84**, 158-169.
28. Aguilar, C.F., Sanderson, I., Moracci, M., Ciaramella, M., Nucci, R., Rossi, M. and Pearl, L.H. (1997) Crystal structure of the beta-glycosidase from the hyperthermophilic archeon Sulfolobus solfataricus: resilience as a key factor in thermostability. *Journal of molecular biology*, **271**, 789-802.
29. Smith, C.A., Toogood, H.S., Baker, H.M., Daniel, R.M. and Baker, E.N. (1999) Calcium-mediated thermostability in the subtilisin superfamily: the crystal structure of Bacillus Ak.1 protease at 1.8 A resolution. *J Mol Biol*, **294**, 1027-1040.
30. Suvd, D., Fujimoto, Z., Takase, K., Matsumura, M. and Mizuno, H. (2001) Crystal Structure of Bacillus stearothermophilus a-Amylase: Possible Factors Determining the Thermostability. *Journal of Biochemistry*, **129**, 461-468.

31. Chi, Y.-I., Martinez-Cruz, L.A., Jancarik, J., Swanson, R.V., Robertson, D.E. and Kim, S.-H. (1999) Crystal structure of the β-glycosidase from the hyperthermophile Thermosphaera aggregans : insights into its activity and thermostability. *FEBS Letters*, **445**, 375-383.
32. van Noort, V., Bradatsch, B., Arumugam, M., Amlacher, S., Bange, G., Creevey, C., Falk, S., Mende, D.R., Sinning, I., Hurt, E. *et al.* (2013) Consistent mutational paths predict eukaryotic thermostability. *BMC Evol Biol*, **13**, 7.
33. Elias, M., Wieczorek, G., Rosenne, S. and Tawfik, D.S. (2014) The universality of enzymatic rate–temperature dependency. *Trends in Biochemical Sciences*, **39**, 1-7.
34. Miller, C., Davlieva, M., Wilson, C., White, K.I., Couñago, R., Wu, G., Myers, J.C., Wittung-Stafshede, P. and Shamoo, Y. (2010) Experimental evolution of adenylate kinase reveals contrasting strategies toward protein thermostability. *Biophysical journal*, **99**, 887-896.
35. Kimura, H., Mori, K., Yamanaka, T. and Ishibashi, J. (2013) Growth temperatures of archaeal communities can be estimated from the guanine-plus-cytosine contents of 16S rRNA gene fragments. *Environ Microbiol Rep*, **5**, 468-474.
36. Melnikov, S., Manakongtreecheep, K. and Soll, D. (2018) Revising the Structural Diversity of Ribosomal Proteins Across the Three Domains of Life. *Mol Biol Evol*, **35**, 1588-1598.
37. Feller, G. (2003) Molecular adaptations to cold in psychrophilic enzymes. *Cellular and Molecular Life Sciences (CMLS)*, **60**, 648-662.
38. Mandelman, D., Ballut, L., Wolff, D.A., Feller, G., Gerday, C., Haser, R. and Aghajari, N. (2019) Structural determinants increasing flexibility confer cold adaptation in psychrophilic phosphoglycerate kinase. *Extremophiles*, **23**, 495-506.
39. Russell, N.J. (2000) Toward a molecular understanding of cold activity of enzymes from psychrophiles. *Extremophiles*, **4**, 83-90.
40. Aghajari, N., Haser, R., Feller, G. and Gerday, C. (1998) Crystal structures of the psychrophilic α-amylase from Alteromonas haloplanctis in its native form and complexed with an inhibitor. *Protein Science*, **7**, 564-572.
41. D'Amico, S., Gerday, C. and Feller, G. (2001) Structural Determinants of Cold Adaptation and Stability in a Large Protein. *Journal of Biological Chemistry*, **276**, 25791-25796.
42. D'Amico, S., Claverie, P., Collins, T., Georlette, D., Gratia, E., Hoyoux, A., Meuwis, M.-A., Feller, G. and Gerday, C. (2002) Molecular basis of cold adaptation. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, **357**, 917-925.
43. Cartier, G., Lorieux, F., Allemand, F., Dreyfus, M. and Bizebard, T. (2010) Cold Adaptation in DEAD-Box Proteins. *Biochemistry*, **49**, 2636-2646.
44. Dassarma, S., Capes, M.D., Karan, R. and Dassarma, P. (2013) Amino Acid Substitutions in Cold-Adapted Proteins from Halorubrum lacusprofundi, an Extremely Halophilic Microbe from Antarctica. *PLoS ONE*, **8**, e58587.
45. Georlette, D., Blaise, V., Collins, T., D'Amico, S., Gratia, E., Hoyoux, A., Marx, J.C., Sonan, G., Feller, G. and Gerday, C. (2004) Some like it cold: biocatalysis at low temperatures. *FEMS Microbiology Reviews*, **28**, 25-42.
46. Kwak, K.J., Park, S.J., Han, J.H., Kim, M.K., Oh, S.H., Han, Y.S. and Kang, H. (2011) Structural determinants crucial to the RNA chaperone activity of glycine-rich RNA-binding proteins 4 and 7 in Arabidopsis thaliana during the cold adaptation process. *J Exp Bot*, **62**, 4003-4011.

47. Shokrollahzade, S., Sharifi, F., Vaseghi, A., Faridounnia, M. and Jahandideh, S. (2015) Protein cold adaptation: Role of physico-chemical parameters in adaptation of proteins to low temperatures. *Journal of Theoretical Biology*, **383**, 130-137.
48. Zanphorlin, L.M., De Giuseppe, P.O., Honorato, R.V., Tonoli, C.C.C., Fattori, J., Crespim, E., De Oliveira, P.S.L., Ruller, R. and Murakami, M.T. (2016) Oligomerization as a strategy for cold adaptation: Structure and dynamics of the GH1 β-glucosidase from Exiguobacterium antarcticum B7. *Scientific Reports*, **6**, 23776.
49. Reed, C.J., Lewis, H., Trejo, E., Winston, V. and Evilia, C. (2013) Protein Adaptations in Archaeal Extremophiles. *Archaea*, **2013**, 1-14.
50. Cava, F., Hidalgo, A. and Berenguer, J. (2009) Thermus thermophilus as biological model. *Extremophiles*, **13**, 213-231.
51. Matsuura, Y., Takehira, M., Joti, Y., Ogasahara, K., Tanaka, T., Ono, N., Kunishima, N. and Yutani, K. (2015) Thermodynamics of protein denaturation at temperatures over 100°C: CutA1 mutant proteins substituted with hydrophobic and charged residues. *Scientific Reports*, **5**, 15545.
52. Vetriani, C., Maeder, D.L., Tolliday, N., Yip, K.S., Stillman, T.J., Britton, K.L., Rice, D.W., Klump, H.H. and Robb, F.T. (1998) Protein thermostability above 100 degreesC: a key role for ionic interactions. *Proceedings of the National Academy of Sciences of the United States of America*, **95**, 12300-12305.
53. Fraser, N.J., Liu, J.-W., Mabbitt, P.D., Correy, G.J., Coppin, C.W., Lethier, M., Perugini, M.A., Murphy, J.M., Oakeshott, J.G., Weik, M. *et al.* (2016) Evolution of Protein Quaternary Structure in Response to Selective Pressure for Increased Thermostability. *Journal of Molecular Biology*, **428**, 2359-2371.
54. Leuenberger, P., Ganscha, S., Kahraman, A., Cappelletti, V., Boersema, P.J., von Mering, C., Claassen, M. and Picotti, P. (2017) Cell-wide analysis of protein thermal unfolding reveals determinants of thermostability. *Science*, **355**.
55. Robinson-Rechavi, M. and Godzik, A. (2005) Structural genomics of thermotoga maritima proteins shows that contact order is a major determinant of protein thermostability. *Structure*, **13**, 857-860.
56. Dams, T., Auerbach, G., Bader, G., Jacob, U., Ploom, T., Huber, R. and Jaenicke, R. (2000) The crystal structure of dihydrofolate reductase from Thermotoga maritima: molecular features of thermostability. *J Mol Biol*, **297**, 659-672.
57. Robinson-Rechavi, M., Alibés, A. and Godzik, A. (2006) Contribution of electrostatic interactions, compactness and quaternary structure to protein thermostability: lessons from structural genomics of Thermotoga maritima. *J Mol Biol*, **356**, 547-557.
58. Ma, B.-G., Goncearenco, A. and Berezovsky, I.N. (2010) Thermophilic Adaptation of Protein Complexes Inferred from Proteomic Homology Modeling. *Structure*, **18**, 819-828.
59. Coleman, R.G. and Sharp, K.A. (2009) Thermophilic protein structure adaptation examined with Burial Depth and Travel Depth. *Biophysical Journal*, **96**, 584a.
60. Rosato, V., Pucello, N. and Giuliano, G. (2002) Evidence for cysteine clustering in thermophilic proteomes. *Trends Genet*, **18**, 278-281.
61. Feng, C., Ma, Z., Yang, D., Li, X., Zhang, J. and Li, Y. (2020) A Method for Prediction of Thermophilic Protein Based on Reduced Amino Acids and Mixed Features. *Front Bioeng Biotechnol*, **8**, 285.
62. Feller, G. (2010) Protein stability and enzyme activity at extreme biological temperatures. *Journal of Physics: Condensed Matter*, **22**.

63. Zavodszky, P., Kardos, J., Svingor, A. and Petsko, G.A. (1998) Adjustment of conformational flexibility is a key event in the thermal adaptation of proteins. *Proceedings of the National Academy of Sciences*, **95**, 7406-7411.
64. Kumar, S. and Nussinov, R. (2001) How do thermophilic proteins deal with heat? *Cellular and Molecular Life Sciences*, **58**, 1216-1233.
65. Liao, M.L., Zhang, S., Zhang, G.Y., Chu, Y.M., Somero, G.N. and Dong, Y.W. (2017) Heat-resistant cytosolic malate dehydrogenases (cMDHs) of thermophilic intertidal snails (genus Echinolittorina): protein underpinnings of tolerance to body temperatures reaching 55°C. *J Exp Biol*, **220**, 2066-2075.
66. Fang, X.W., Golden, B.L., Littrell, K., Shelton, V., Thiyagarajan, P., Pan, T. and Sosnick, T.R. (2001) The thermodynamic origin of the stability of a thermophilic ribozyme. *Proceedings of the National Academy of Sciences*, **98**, 4355-4360.
67. Kumar, S., Tsai, C.-J. and Nussinov, R. (2000) Factors enhancing protein thermostability. *Protein Engineering, Design and Selection*, **13**, 179-191.
68. Yang, J., Li, L., Xiao, Y., Li, J., Long, L., Wang, F. and Zhang, S. (2016) Identification and thermoadaption engineering of thermostability conferring residue of deep sea bacterial α-amylase AMY121. *Journal of Molecular Catalysis*, **126**, 56 - 63.
69. Meruelo, A.D., Han, S.K., Kim, S. and Bowie, J.U. (2012) Structural differences between thermophilic and mesophilic membrane proteins. *Protein Sci*, **21**, 1746-1753.
70. Galtier, N. and Lobry, J.R. (1997) Relationships between genomic G+C content, RNA secondary structures, and optimal growth temperature in prokaryotes. *J Mol Evol*, **44**, 632-636.
71. Wang, H.C., Xia, X. and Hickey, D. (2006) Thermal adaptation of the small subunit ribosomal RNA gene: a comparative study. *J Mol Evol*, **63**, 120-126.
72. Sato, Y., Okano, K., Kimura, H. and Honda, K. (2020) TEMPURA: Database of Growth TEMPeratures of Usual and RAre Prokaryotes. *Microbes Environ*, **35**.
73. Engqvist, M.K.M. (2018) Correlating enzyme annotations with a large set of microbial growth temperatures reveals metabolic adaptations to growth at diverse temperatures. *BMC Microbiol*, **18**, 177.
74. Pezeshgi Modarres, H., Mofrad, M.R. and Sanati-Nezhad, A. (2018) ProtDataTherm: A database for thermostability analysis and engineering of proteins. *PLOS ONE*, **13**, e0191222.
75. Bashirova, A., Pramanik, S., Volkov, P., Rozhkova, A., Nemashkalov, V., Zorov, I., Gusakov, A., Sinitsyn, A., Schwaneberg, U. and Davari, M. (2019) Disulfide Bond Engineering of an Endoglucanase from Penicillium verruculosum to Improve Its Thermostability. *International Journal of Molecular Sciences*, **20**, 1602.
76. Furukawa, R., Toma, W., Yamazaki, K. and Akanuma, S. (2020) Ancestral sequence reconstruction produces thermally stable enzymes with mesophilic enzyme-like catalytic properties. *Scientific Reports*, **10**.
77. Xu, Z., Cen, Y.-K., Zou, S.-P., Xue, Y.-P. and Zheng, Y.-G. (2020) Recent advances in the improvement of enzyme thermostability by structure modification. *Critical Reviews in Biotechnology*, **40**, 83-98.
78. Pinto, G.P., Corbella, M., Demkiv, A.O. and Kamerlin, S.C.L. (2021) Exploiting enzyme evolution for computational protein design. *Trends in Biochemical Sciences*.
79. Lee, C.F., Makhatadze, G.I. and Wong, K.B. (2005) Effects of charge-to-alanine substitutions on the stability of ribosomal protein L30e from Thermococcus celer. *Biochemistry*, **44**, 16817-16825.

80. Cavicchioli, R., Ripple, W.J., Timmis, K.N., Azam, F., Bakken, L.R., Baylis, M., Behrenfeld, M.J., Boetius, A., Boyd, P.W., Classen, A.T. *et al.* (2019) Scientists' warning to humanity: microorganisms and climate change. *Nature Reviews Microbiology*, **17**, 569-586.
81. Sočan, J., Purg, M. and Åqvist, J. (2020) Computer simulations explain the anomalous temperature optimum in a cold-adapted enzyme. *Nature Communications*, **11**.
82. Karshikoff, A., Nilsson, L. and Ladenstein, R. (2015) Rigidity versus flexibility: the dilemma of understanding protein thermal stability. *FEBS Journal*, **282**, 3899-3917.
83. Okafor, C.D., Pathak, M.C., Fagan, C.E., Bauer, N.C., Cole, M.F., Gaucher, E.A. and Ortlund, E.A. (2018) Structural and Dynamics Comparison of Thermostability in Ancient, Modern, and Consensus Elongation Factor Tus. *Structure*, **26**, 118-129.e113.
84. Gianese, G., Argos, P. and Pascarella, S. (2001) Structural adaptation of enzymes to low temperatures. *Protein Engineering, Design and Selection*, **14**, 141-148.
85. Santiago, M., Ramírez-Sarmiento, C.A., Zamora, R.A. and Parra, L.P. (2016) Discovery, Molecular Mechanisms, and Industrial Applications of Cold-Active Enzymes. *Frontiers in microbiology*, **7**, 1408-1408.
86. Spring, S., Merkhoffer, B., Weiss, N., Kroppenstedt, R.M., Hippe, H. and Stackebrandt, E. (2003) Characterization of novel psychrophilic clostridia from an Antarctic microbial mat: description of Clostridium frigoris sp. nov., Clostridium lacusfryxellense sp. nov., Clostridium bowmanii sp. nov. and Clostridium psychrophilum sp. nov. and reclassificat. *International Journal of Systematic and Evolutionary Microbiology*, **53**, 1019-1029.
87. Wiegel, J., Kuk, S.U. and Kohring, G.W. (1989) Clostridium thermobutyricum sp. nov., a Moderate Thermophile Isolated from a Cellulolytic Culture, That Produces Butyrate as the Major Product. *International Journal of Systematic Bacteriology*, **39**, 199-204.
88. Szilágyi, A. and Závodszky, P. (2000) Structural differences between mesophilic, moderately thermophilic and extremely thermophilic protein subunits: results of a comprehensive survey. *Structure*, **8**, 493-504.

TABLE AND FIGURES LEGENDS

Table 1. Websites used for web scraping to collect information about the optimal growth temperatures of microbial organisms.

Database	Database Name	URL
ARS (NRRL)	Agricultural Research Service Culture Collection, National Center for Agricultural Utilization Research, USA	https://nrrl.ncaur.usda.gov/
ATCC	American Type Culture Collection, USA	https://www.atcc.org/
BCCM	Belgian Coordinated Collections of Microorganisms, Belgium	https://bccm.belspo.be/
BCRC	Bioresources Collection and Research Center, Taiwan	https://catalog.bcrc.firdi.org.tw
CABRI	Common Access to Biological Resources and Information	http://www.cabri.org/
CCARM	Culture Collection of Antibiotics Resistant Microbes, UK	https://www.phe-culturecollections.org.uk
CCM	Czech Collection of Microorganisms, the Czech Republic	https://www.sci.muni.cz/ccm/
CCUG	Culture Collection, University of Goteborg, Sweden	https://www.ccug.se/
CECT	Spanish Type Culture Collection, Spain	https://www.uv.es/cect/
CICC	China Center of Industrial Culture Collection	http://english.china-cicc.org
CIP	Center for Biological Resources of the Institute Pasteur, France	https://catalogue-crbip.pasteur.fr
DSMZ	German Collection of Microorganisms and Cell Cultures GmbH, Germany	https://www.dsmz.de
GRIN	Agricultural Genetic Resources Information Center, USA	https://www.ars-grin.gov
HAMBI	Culture Collection of Department of Applied Chemistry and Microbiology, University of Helsinki, Finland	https://kotka.luomus.fi/culture/bac
JCM	Japan Collection of Microorganisms (RIKEN Bioresource Center), Japan	https://jcm.brc.riken.jp/en/ordering_e
KEMB	Korea Environmental Microorganisms Bank, South Korea	https://kemb.or.kr
NBIMCC	National Bank for Industrial Microorganisms and Cell Cultures, Bulgaria	https://www.nbimcc.org/en/about.htm
NBRC	NITE Biological Resource Center, Japan	https://www.nite.go.jp
NCIMB	National Collections of Industrial, Food and Marine Bacteria, UK	https://www.ncimb.com/
NCMA	National Center for Marine Algae and Microbiota (NCMA), USA	https://ncma.bigelow.org/
NCTC	National Collection of Type Cultures, UK	https://www.phe-culturecollections.org.uk/
NIES	National Institute for Environmental Studies, Japan	https://mcc.nies.go.jp/
VKM	All-Russian Collection of Microorganisms, Russia	http://www.vkm.ru

Figure 1. Using the database to colour bacterial species by their optimal growth temperatures. The bacterial tree of life shows representative members of bacterial genera, with each organism coloured by the optimal growth temperature ranging from 4°C ("low") to 85°C ("high"). The dot-plot shows lineages of bacterial species in the database and their corresponding optimal growth temperature.

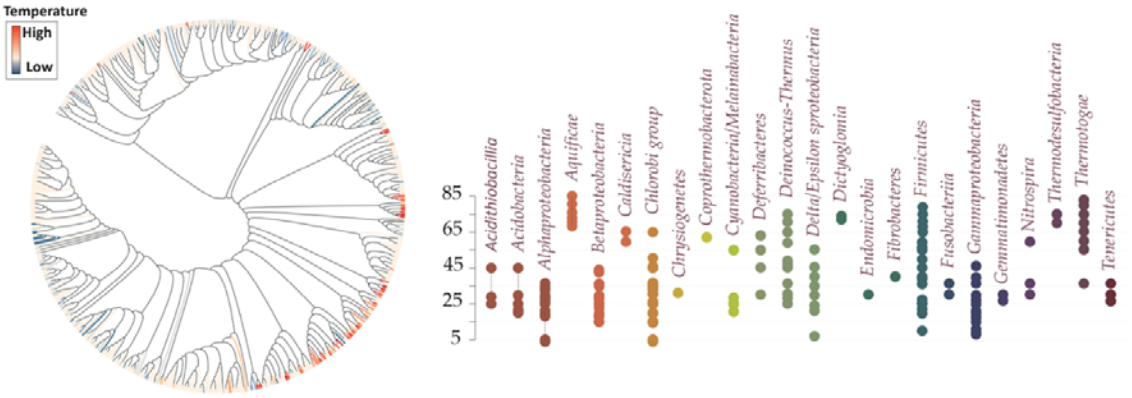
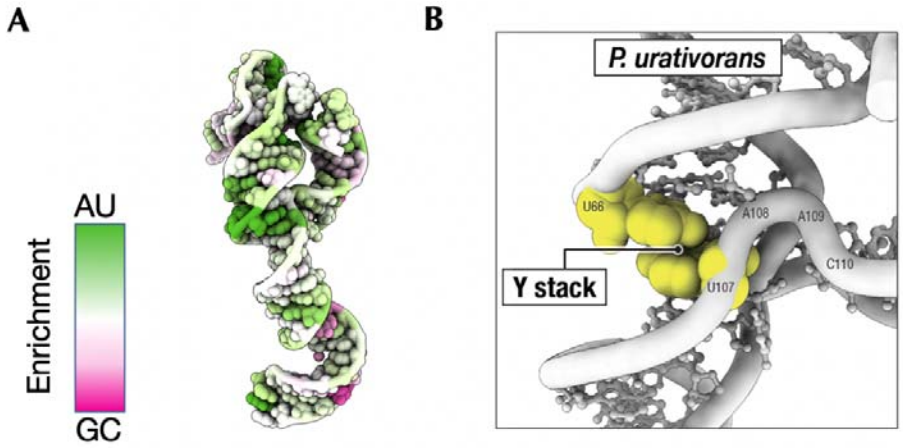



Figure 2. Using the database to colour molecules by thermal adaptation and identify psychrophile-specific variations that may help RNA molecules adapt to cold environments. **a.** The molecular structure of 5S rRNA from the psychrophilic bacterium *Psychrobacter urativorans* is color-coded based on the GC-bias of its residues in psychrophilic species. Residues coloured in green indicate a high preference for mutations from G or C to A or U in psychrophilic bacteria compared to thermophilic bacteria. Residues coloured in magenta highlight residues that exhibit the highest rate of mutation from A or U to G or C in psychrophilic bacteria compared to thermophilic bacteria. **b.** The atomic structure zooming in on the 5S rRNA segment with the highest AU-bias in psychrophilic bacteria reveals the presence of an unusual structural motif in psychrophilic rRNA. This motif is stabilized by the stacking of two uracil bases (U66 and U107), which is a weak interaction that differs from the typical tertiary structures of 5S rRNA in mesophilic and thermophilic bacteria, where this stacking usually involves at least one purine base to form a stronger contact.