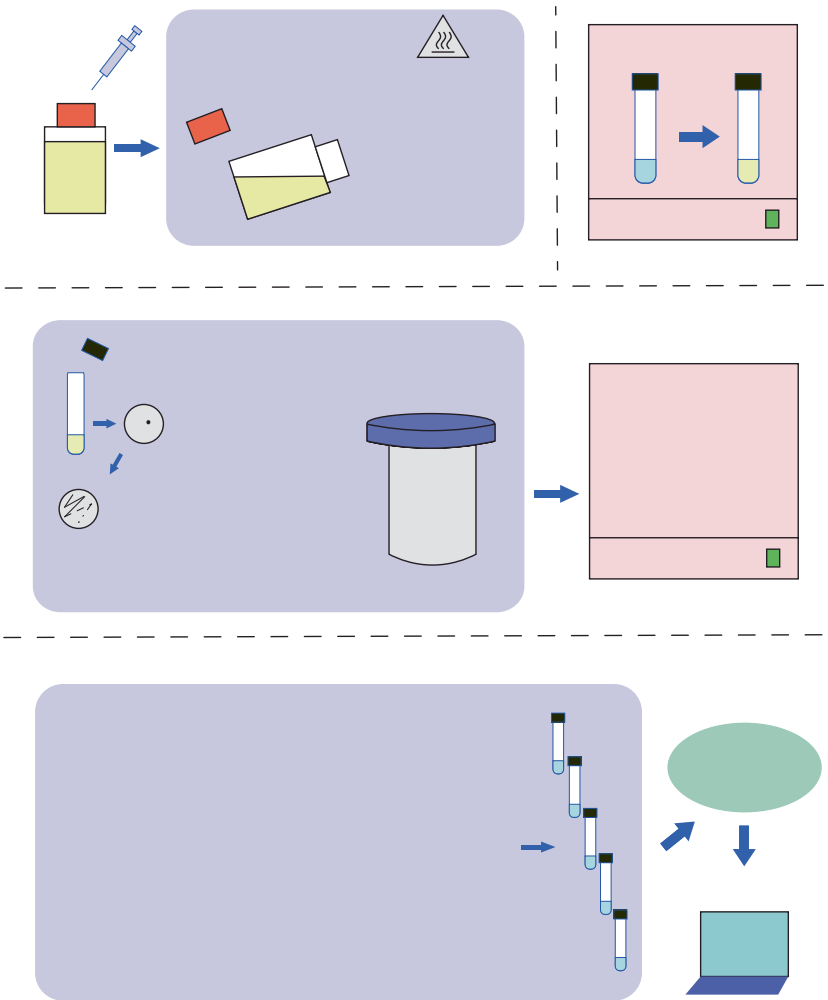
The ISME Journal (2021) 15:1862–1869
https://doi.org/10.1038/s41396-020-00888-5

BRIEF COMMUNICATION

mesophiles [4]. This explanation is also supported by experimental assays showing nearly neutral mutations in temperate conditions become strongly deleterious at high temperature [5]. Furthermore, fluctuation tests on a hyperthermophilic archeaon *Sulfolobus acidocaldarius* [6] and a hyperthermophilic bacterium *Thermus thermophilus* [7] consistently showed that hyperthermophiles have much lower mutation rate compared to mesophiles. This appears to support the hypothesis that selection favors high replication fidelity at high temperature [4



sites (288) relative to those (104) at synonymous sites (χ^2 test; $p = 0.014$) and the accumulated mutations at intergenic regions (65) relative to protein-coding regions (392) (χ^2 test; $p = 0.013$) showed marginally significant differences.

To date, over 20 phylogenetically diverse free-living bacterial species and two archaeal species isolated from various environments have been assayed with MA/WGS, and their mutation rates vary from 0.79×10^{-10} to $97.80 \times$

10^{-10} per cell division per site [20]. The only prokaryote that displays a mutation rate (97.80×10^{-10} per cell division per site) comparable to A501 is *Mesoplasma florum* L1 [21], a host-dependent wall-less bacterium with highly reduced genome (~700 genes). Our PCR validation of randomly chosen 20 base-substitution mutations from two MA lines displaying highest mutation rates and of all nine INDEL mutations involving >10 bp changes across all lines (Table S2) indicates that the calculated high mutation rate did not result from false bioinformatics predictions.

The extremely high mutation rate of *T. eurythermalis* is unexpected. One potential explanation in line with the "mutator theory" [22–24] is that high mutation rate may allow the organisms to gain beneficial mutations more rapidly and thus is selectively favored in deep-sea hydrothermal vents where physicochemical parameters are highly fluctuating. Alternatively, high mutation rate is the result of random genetic drift according to the "drift-barrier model" [21]. In this model, increased mutation rates are associated with increased load of deleterious mutations, so natural selection favors lower mutation rates. On the other hand, increased improvements of replication fidelity come at an increased cost of investments in DNA repair activities. Therefore, natural selection pushes the replication fidelity to a level that is set by genetic drift, and further improvements are expected to reduce the fitness advantages [11, 21]. These two explanations for the high mutation rate of *T. eurythermalis* are mutually exclusive, and resolving them requires the calculation of the power of genetic drift, which is inversely proportional to N_e.

A common way to calculate N_e for a prokaryotic population is derived from the equation $\pi_S = 2 \times N_e \times \mu$, where π_S represents the nucleotide diversity at synonymous (silent) sites among randomly sampled members of a panmictic population [25]. We therefore sequenced genomes of another eight *T. eurythermalis* isolates available in our culture collections. Like *T. eurythermalis* A501, these

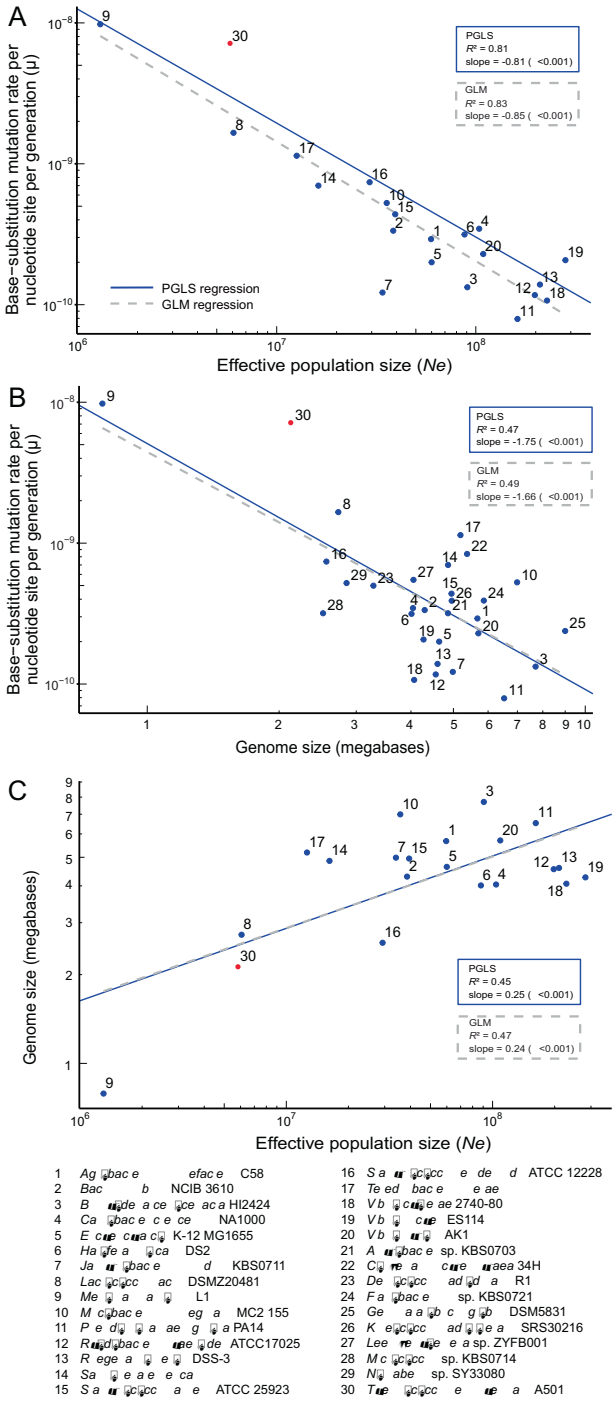


Fig. 2 T a a a - a a a ( ), ff a (N_e), a a 28 a a a a a . All three traits' values were logarithmically transformed. The mutation rates of these species are all determined with the mutation accumulation experiment followed by whole-genome sequencing of the mutant lines. The mutation rate of species numbered 1–29 (blue) is collected from literature and that of the species 30 (red) is determined in the present study. Among the numbered species shown in the figure, the species #6 *Haloferax volcanii* is facultative anaerobic halophilic archaeon, and the species #30 is an obligate anaerobic hyperthermophilic archaeon. **A** The scaling relationship between μ and N_e. **B** The scaling relationship between μ and genome size. **C** The scaling relationship between genome size and N_e. Numbered data points 21–29 are not shown in **A** and **C** because of the lack of population dataset for estimation of N_e. The dashed gray lines and blue lines represent the generalized linear model (GLM) regression and the phylogenetic generalized least square (PGLS) regression, respectively. The Bonferroni adjusted outlier test for the GLM regression show that #7 *Janthinobacterium lividum* is an outlier in the scaling relationship between μ and N_e, and #9 *Mesoplasma florum* is an outlier in the scaling relationship between genome size and N_e. No outlier was identified in the PGLS regression results.

(dashed gray line in Fig. 2B [$r^2 = 0.49$, slope $= -1.66$, s.e.m. $= 0.32$, $p < 0.001$]) between genome size and base-substitution mutation rate, which is consistent with the hypothesis that increased mutation rate drives microbial genome reduction. We also showed a positive linear relationship (dashed gray line in Fig. 2C [$r^2 = 0.47$, slope $=$ 0.24, s.e.m. $= 0.06$, $p < 0.001$]) between genome size and N_e, which suggests that random genetic drift drives genome reduction across prokaryotes. These correlations remain robust when the data were analyzed as phylogenetically independent contrasts (blue solid lines in Fig. 2B [$r^2 = 0.47$, slope $= -1.75$, s.e.m. $= 0.34$, $p < 0.001$] and in Fig. 2C [$r^2 = 0.45$, slope $= 0.25$, s.e.m. $= 0.06$, $p < 0.001$]). Our results are consistent with recent studies which employed mathematical modeling and/or comparative sequence analyses to show random genetic drift [36] and increased mutation rate [37] driving genome reduction across diverse bacterial lineages including both free-living and host-dependent bacteria. One benefit of the present study is that it directly measures μ and N_e, as compared to those recent advances which relied on proxies for these metrics (e.g., using the ratio of nonsynonymous substitution rate to synonymous substitution rate to represent N_e) to infer mechanisms of genome reduction.

Despite this advantage, there are important caveats to our conclusions related to the mechanisms of genome reduction. The correlation analyses performed here are inspired by Lynch and colleagues' work, who had great success explaining eukaryotic genome expansion with genetic drift [11, 38]. However, there are a few key differences of genomic features between prokaryotes and eukaryotes, which makes it more difficult to explain the correlation observed in prokaryotes. Importantly, genome sizes of eukaryotes can vary over several orders of magnitude, whereas those of free-living prokaryotes differ by only an order of magnitude [11], so there is much less variability to explain in prokaryotes. Moreover, eukaryotic genomes experience dramatic expansions of transposable elements which are often considered as genomic parasites, whereas prokaryotic genomes including those large ones are usually depleted with transposable elements and their genome size variations are largely driven by gene content [39]. Aside

from these conceptual difficulties, the plots (Fig. 2B, C) are poorly populated with typical free-living species carrying small genomes such as the *Prochlorococcus* (mostly 1.6–8 Mb) and *Pelagibacterales* (1.3–1.5 Mb), which dominate the photosynthetic and heterotrophic microbial communities, respectively, in the ocean [40]. It has been generally postulated that bacterial species in these lineages have very large N_e [39–41], though there has been little direct evidence for it [42, 43]. If confirmed through the measurement of the unbiased mutation rate (μ) followed by the calculation of N_e based on μ, it might compromise the linear relationship between genome size and N_e observed here (Fig. 2C). It is also not necessarily appropriate to translate correlations to causal relationships. For example, the correlation between increased mutation rates and decreased genome sizes (Fig. 2B) does not necessarily mean that increased mutation rate drives genome reduction. This is because high mutation rates are observed in species with small N_e. Given that deletion bias is commonly found in prokaryotes [44, 45], genome reduction can be easily explained by increased fixation of deletional mutations in species with smaller N_e. High mutation rates in these species are simply the result of random genetic drift as explained by the drift-barrier theory, and they may have a limited role in driving genome reduction.

Whereas our analysis based on the available data did not support natural selection as a universal mechanism driving genome reduction across prokaryotes (Fig. 2B, C), it does not mean that selection has no role in genome reduction of a particular taxon. In the case of thermophiles, proponents for selection acting to reduce genomes explained that genome size, due to its positive correlation with cell volume, may be an indirect target of selection which strongly favors smaller cell volume [35]. The underlying principle is that high temperature requires cells to increase the lipid content and change the lipid composition of the cell membranes, which consumes a large part of the cellular energy, and thus lower cell volume is selectively favored at high temperature [35]. Our calculations of a relatively small N_e in *T. eurythermalis* does not necessarily contradict with this selective argument, given that the fitness gained by decreasing cell volume and thus reducing genome size is large enough to overcome the power of random genetic drift. On the other hand, our data strongly indicate that neutral forces dictate the genome evolution of *T. eurythermalis*, and they are not negligible with regard to its genome reduction process. The significantly more deletion over insertion events (*t* test; 95 versus 37 events with $p < 0.001$ and 48 versus 20 events with $p < 0.05$ before and after removing the 14 genes enriched in mutations, respectively) and the significantly more nucleotides involved in deletions over insertions (*t* test; 433 versus 138 bp with $p < 0.05$ and 386 versus 121 bp with $p < 0.001$ before and after removing the 14 genes enriched in mutations, respectively) suggest that the deletion bias, combined with increased chance fixation of deletion mutants due to low N_e, is a potentially important neutral mechanism giving rise to the small genomes of *T. eurythermalis* (2.12 Mbp).

The globally distributed deep-sea hydrothermal vents are microbe-driven ecosystems, with no known macroorganisms surviving at the vent fluids. Sample collections, microbial isolations, and laboratory propagations of mutation lines at hot and anoxic conditions are challenging. In the present study, we determined that *T. eurythermalis*, and perhaps *Thermococcaceae* in general, has a highly increased mutation rate and a highly decreased effective population size compared to all other known free-living prokaryotic lineages. While it remains to be tested whether this is a common feature among the vents' microbes, the present study nevertheless opens a new avenue for investigating the hyperthemophile ecology and evolution in the deep sea.

Data availability

All the datasets generated, analyzed, and presented in the current study are available in the Supplementary Information. Genomic sequences of the eight *Thermococcus eurythermalis* strains are available at the JGI IMG under the GOLD study id Gs0142375. Raw reads of the eight strains are available at the NCBI SRA under the accession number PRJNA679699.

Code availability

The custom scripts used in this study have been deposited in the online repository (https://github.com/luolab-cuhk/Thermococcus-mut-genome-size).

Acknowledgements This research is supported by the National Key R&D Program of China (2018YFC0309800), National Nature of Science China (NSFC 41530967), China Ocean Mineral Resources R & D Association DY125-22-04, the Hong Kong Branch of Southern Marine Science and Engineering Guangdong Laboratory (Guangzhou) (SMSEGL20SC02), and the Hong Kong Research Grants Council Area of Excellence Scheme (AoE/M-403/16).

Author contributions HL conceptualized the work and strategy, directed the bioinformatics analyses, interpreted the data, and wrote the main manuscript. XX set up the experimental platform for deep-sea hyperthermophile studies, directed the experimental analyses and related writing, co-interpreted the data, provided comments to the main manuscript, and acquired the strains. JG performed all the experiments with contributions from XM, drafted the related methods in supplementary information and Fig. 1. XW performed all the bioinformatics analyses, co-interpreted the related results, drafted the related methods in supplementary information, Fig. 2 and all supplemental tables. YS contributed the bioinformatics tools for mutation detection and mutation rate calculation.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

1. Martin W, Baross J, Kelley D, Russell MJ. Hydrothermal vents and the origin of life. Nat Rev Microbiol. 2008;6:805–14.
2. Dick GJ. The microbiomes of deep-sea hydrothermal vents: distributed globally, shaped locally. Nat Rev Microbiol. 2019;17:271–83.
3. Zierenberg RA, Adams MWW, Arp AJ. Life in extreme environments: hydrothermal vents. Proc Natl Acad Sci USA. 2000;97:12961–2.
4. Friedman R, Drake JW, Hughes AL. Genome-wide patterns of nucleotide substitution reveal stringent functional constraints on the protein sequences of thermophiles. Genetics. 2004;167:1507–12.
5. Drake JW. Avoiding dangerous missense: thermophiles display especially low mutation rates. PLoS Genet. 2009;5:e1000520.
6. Grogan DW, Carver GT, Drake JW. Genetic fidelity under harsh conditions: analysis of spontaneous mutation in the thermoacidophilic archaeon *Sulfolobus acidocaldarius*. Proc Natl Acad Sci USA. 2001;98:7928–33.
7. Mackwan RR, Carver GT, Kissling GE, Drake JW, Grogan DW. The rate and character of spontaneous mutation in *Thermus thermophilus*. Genetics. 2008;180:17–25.
8. Long H, Miller SF, Strauss C, Zhao C, Cheng L, Ye Z, et al. Antibiotic treatment enhances the genome-wide mutation rate of target cells. Proc Natl Acad Sci USA. 2016;113:E2498–505.
9. Williams AB. Spontaneous mutation rates come into focus in *Escherichia coli*. DNA Repair. 2014;24:73–79.

41. Kashtan N, Roggensack SE, Rodrigue S, Thompson JW, Biller SJ, Coe A, et al. Single-cell genomics reveals hundreds of coexisting subpopulations in wild. Prochlorococcus Sci. 2014; 344:416–20.
42. Luo H, Huang Y, Stepanauskas R, Tang J. Excess of non-conservative amino acid changes in marine bacterioplankton lineages with reduced genomes. Nat Microbiol. 2017;2:1–9.
43. Luo H, Friedman R, Tang J, Hughes AL. Genome reduction by deletion of paralogs in the marine cyanobacterium *Prochlorococcus*. Mol Biol Evol. 2011;28:2751–60.
44. Mira A, Ochman H, Moran NA. Deletional bias and the evolution of bacterial genomes. Trends Genet. 2001;17:589–96.
45. Kuo C-H, Ochman H. Deletional bias across the three domains of life. Genome Biol Evol. 2009;1:145–52.