Title

Contributions of single-cell genomics to our understanding of planktonic marine archaea.

Journal

Philosophical Transactions of the Royal Society B: Biological Sciences, 374(1786)

Authors

Laperriere, S

Santoro, Alyson

Kellom, Matthew

Publication Date

2019-11-25

DOI

10.1098/rstb.2019.0096

Peer reviewed

PHILOSOPHICAL TRANSACTIONS B

royalsocietypublishing.org/journal/rstb

Review

Cite this article: Santoro AE, Kellom M, Laperriere SM. 2019 Contributions of single-cell genomics to our understanding of planktonic marine archaea. *Phil. Trans. R. Soc. B* **374**: 20190096.
http://dx.doi.org/10.1098/rstb.2019.0096

Accepted: 15 July 2019

One contribution of 18 to a discussion meeting issue 'Single cell ecology'.

Subject Areas:
microbiology, systems biology, molecular biology, evolution, ecology, environmental science

Keywords:
thaumarchaea, MGI archaea, MGII archaea, population biology

Author for correspondence:
A. E. Santoro
e-mail: asantoro@ucsb.edu

THE ROYAL SOCIETY PUBLISHING

Contributions of single-cell genomics to our understanding of planktonic marine archaea

A. E. Santoro, M. Kellom and S. M. Laperriere

Department of Ecology, Evolution, and Marine Biology, University of California, Santa Barbara, CA 93106-9620, USA

AES, 0000-0003-2503-8219; SML, 0000-0003-4691-8741

Single-cell genomics has transformed many fields of biology, marine microbiology included. Here, we consider the impact of single-cell genomics on a specific group of marine microbes—the planktonic marine archaea. Despite single-cell enabled discoveries of novel metabolic function in the marine thaumarchaea, population-level investigations are hindered by an overall lower than expected recovery of thaumarchaea in single-cell studies. Metagenome-assembled genomes have so far been a more useful method for accessing genome-resolved insights into the Marine Group II euryarchaea. Future progress in the application of single-cell genomics to archaeal biology in the ocean would benefit from more targeted sorting approaches, and a more systematic investigation of potential biases against archaea in single-cell workflows including cell lysis, genome amplification and genome screening.

This article is part of a discussion meeting issue 'Single cell ecology'.

1. Introduction

The ability to generate genomic sequence data from a single cell has greatly impacted many fields of biology [1–3]. Since its first applications to marine microbiology [4,5], single-cell genomics has yielded metabolic and evolutionary insights into many enigmatic and uncultured groups (e.g. [6–9]). Now, with the continued proliferation of single-cell genomics data, it is useful to reflect on which organismal systems single-cell genomics has been useful for, and where its promise remains unfulfilled. Here, we specifically consider its impact on a widespread and abundant group of microorganisms, the planktonic marine archaea [10].

2. The promise of single-cell genomics in marine microbiology

Applications of single-cell genomics in marine microbiology do not differ greatly from other fields, namely, a desire to understand cell-to-cell variability within populations, the ability to compare strain-level differences in metabolic potential, and metabolic function discovery in new lineages. A particularly attractive application for single-cell genomics is the ability to apply population genetic principles to uncultured marine microorganisms [11–13], which, as recently as 15 years ago, were limited to cultivable organisms (e.g. [14–16]) or hindered by the ability to study uncultivated marine microbes with only a single gene [17]. Early pioneers foresaw the eventual promise of marine metagenomics for population genomics [18]. Yet, though some attempts have been made to resolve population-level measures of recombination and divergence from metagenomic data [19,20], there remain many marine lineages for which estimates of drift, divergence and effective population sizes remain unknown.

No single microbial group in the ocean has been studied as extensively using single-cell genomics as the picocyanobacterium *Prochlorococcus* [21–24].

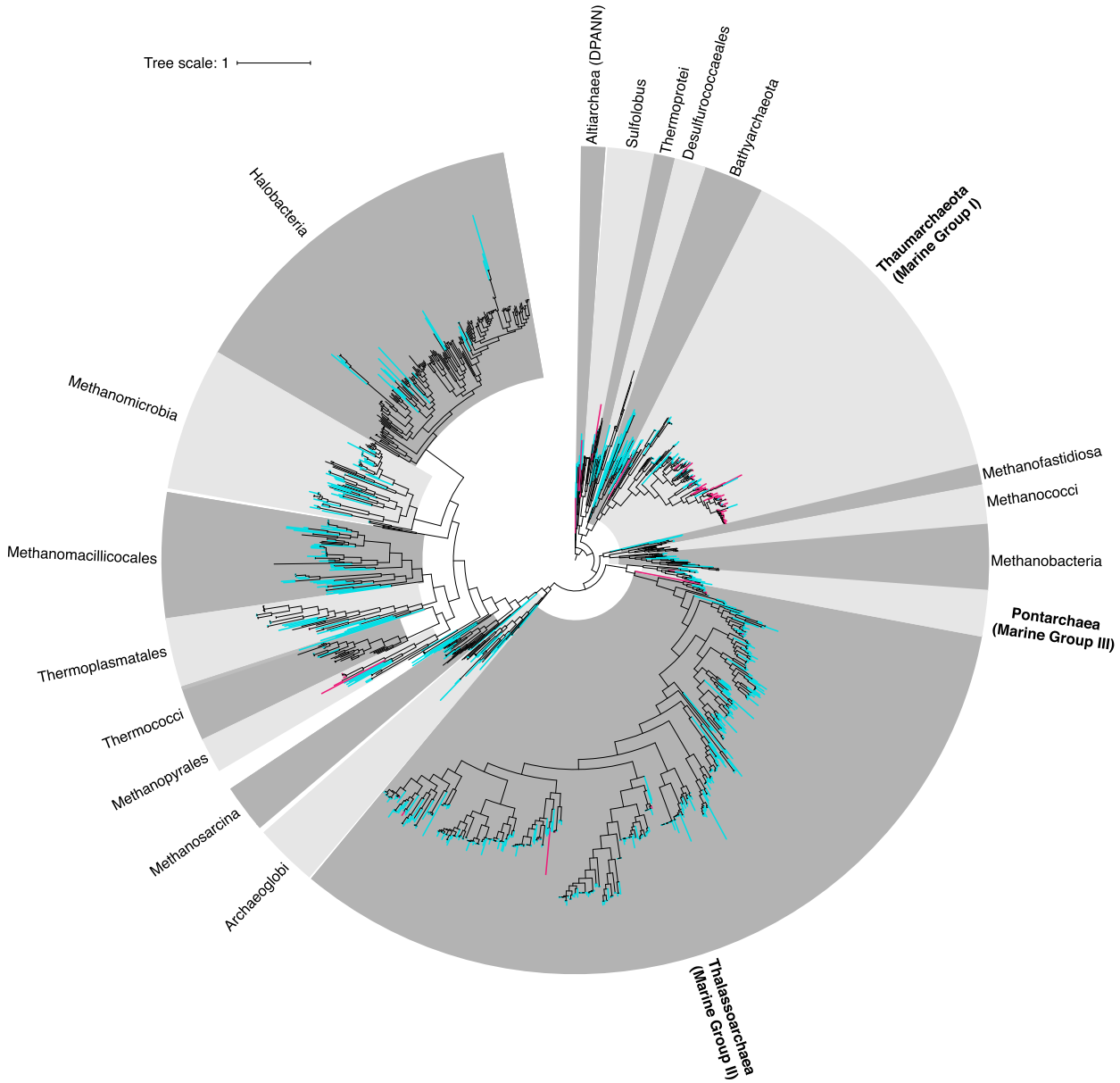

Figure 1. Phylogenetic distribution of SAGs (magenta) and metagenome-assembled genomes (MAGs) (teal) emphasizing the MGI thaumarchaea and the MGII euryarchaea. The presentation is meant to highlight the relative abundance of SAGs versus MAGs for the planktonic archaea and may not be complete with respect to other groups. Tree is a maximum-likelihood (RaxML) reconstruction built on a concatenated alignment of 32 proteins from 1360 archaeal genomes, including MAGs, single-cell genomes and cultivated isolates (shown in black), adapted from [10]. The tree was visualized using iTOL [30]. All data were obtained from NCBI Genbank or the Joint Genome Institute's Integrated Microbial Genomes database. Owing to the large number of marker genes included in the alignment, many partial SAGs available in public databases are not included. The alignment underlying the tree is available on FigShare (www.figshare.com); some taxa have been excluded from the final visualization for clarity.

Using over 1000 single-cell amplified genomes (SAGs) from a seasonal study at a single site, Kashtan *et al*. [22] identified hundreds of distinct *Prochlorococcus* subpopulations and identified population-resolved core and flexible genomes, calculating that these subpopulations diverged millions of years ago. More than any other, this study illustrates the power and promise of single-cell sequencing to realize true population biology in uncultivated marine microbes.

Yet, in failing to identify a study on a similar scale in other marine microbial groups, it is worth considering aspects of *Prochlorococcus* ecology and physiology that make it so amenable to single-cell approaches, and how those traits are not widespread among other marine microbes. *Prochlorococcus* is extremely abundant, reaching local population abundances of nearly $10^5\,\mathrm{ml}^{-1}$, dominating chlorophyll-containing photosynthetic cells and making up nearly 10% of total picoplankton cells [25]. *Prochlorococcus* is easily sorted based on its small cell size and unique fluorescence properties [26], and is apparently readily lysed using only alkaline lysis conditions [21]. The small genome sizes of *Prochlorococcus* cells should lead to high recovery rates [27], and indeed the average genome recovery is 70% [22].

One microbial group in particular that could benefit from such a comprehensive analysis is the marine archaea, which comprises four major subgroups spread across at least three phyla [10]. We focus here on the two most abundant groups of marine archaea—the marine thaumarchaea (also known as the Marine Group I (MGI) archaea) and the Marine Group II (MGII) euryarchaea (subdivided into two Orders provisionally known as the Thalassoarchaea [28]

and Poseidonarchaea [29]; figure 1). The contrasting ecology and genomic features of these two groups are useful for comparing genomic approaches to studying microbial population biology. The thaumarchaea are small cells of which all marine isolates have been shown to be chemolithoautotrophic ammonia oxidizers. They are consistently abundant members of the dark, mesopelagic (200–1000 m depth) microbial assemblage, where, at times, they make up more than 30% of DAPI-stainable cells [31]. By contrast, the MGII archaea are bigger cells, with larger, more metabolically versatile genomes [32]. They are transiently abundant in surface waters, but at times can reach high abundances, often following phytoplankton blooms [33,34].

Many questions remain regarding the ecology and evolution of planktonic archaea that single-cell genomics has the potential to answer—the discovery of new archaeal groups, the metabolic capability of archaeal groups for which only 16S rRNA sequences exist (e.g. MG-IV archaea, [35]), the metabolic basis of apparent depth-dependent niche partitioning [10], interactions with other microbes, either as prey or in symbioses, and rates of dispersal, genetic drift and mutation in the deep ocean.

3. Insights into the ecology and metabolic capability of planktonic archaea arising from single-cell genomics

To date, the primary utility of single-cell genomics in the study of planktonic archaea has been the discovery of previously unknown metabolic function. Single-cell genomics was used to generate one of the first near-complete thaumarchaeal genomes [36], demonstrating novel features not previously known in the thaumarchaea, including motility and chemotaxis. The water depth-differentiated 'shallow' and 'deep' ecotypes of marine thaumarchaea previously identified in single gene surveys [37,38] were identified in an analysis of single-cell thaumarchaeal genomes [39], which found a DNA photolyase gene in genomes clustering with the shallow water clade, suggesting this as an adaptation to their sunlit environment. An expanded analysis of the same dataset including additional single-cell genomes [40] did not identify any novel thaumarchaeal metabolic functions, but did find that 27% of the proteins identified in the SAGs were exclusive to SAGs and not found in cultured isolates, while recruiting metagenomic data from the ocean at a high (greater than 95%) identity. Unfortunately, the vast majority of these SAG-exclusive proteins were hypothetical proteins or proteins of unknown function.

A higher degree of metabolic novelty has come from applying single-cell genomics to more 'extreme' marine environments. For example, a single-cell genome of a thaumarchaeon from the hadal depths of the Puerto Rico trench [41] found genes for glycine cleavage, lipoic acid synthesis and fatty acid biosynthesis, suggesting a potentially mixotrophic lifestyle for hadal thaumarchaea. Similarly, combining SAGs and metagenome-assembled genomes (MAGs), a study of multiple hadal sites also identified genes for glycine cleavage that are not present in any cultivated thaumarchaea [42]. Although it is difficult to confirm the absence of genes in incomplete genomes, these deep sea thaumarchaea also apparently lack genes encoding for PolB, a DNA-repairing polymerase, suggesting a mechanism for generating genomic variation, perhaps to compensate for slow generation times in the deep ocean. In a unique application of single-cell genomics, a study of deep brines in the Red Sea [43] cross-assembled four SAGs to produce a high-quality draft genome of the dominant thaumarchaeon there. To investigate potential adaptations to osmotic stress, *in silico* isoelectric points were determined for the entire predicted proteome and for only proteins with transmembrane domains. A greater portion of acidic residues in the transmembrane domain-containing proteins was seen, as is the case with extreme halophiles, indicating a potential adaptation of these brine seep thaumarchaea to their high salt environment.

Though several MGII single-cell genomes have been deposited in public databases (e.g. [44]), we are not aware of a systematic analysis of their gene content. Beyond metabolic function discovery, using single-cell genomics to uncover interactions between archaea and other marine organisms has so far not been successful [45], which may simply reflect a lack of associations. Though MGII have been shown to bloom both coincident with and just after phytoplankton blooms [33,34], a targeted sorting approach to identify MGII associated with specific phytoplankton was not successful [46]. Single-cell methods have, however, been successfully applied to the detection of phage within archaeal genomes, a promising area for future development [47,48]. Archaeal SAGs have also been used in conjunction with metagenomic data to improve whole-genome assemblies. Using such a hybrid approach, Mende *et al*. [49] improved the quality of a thaumarchaeal SAG, largely with the effect of removing redundancy and contamination. The original SAG contained 32 contigs with an N50 of 79 kb, while the improved 'iSAG' contained 4 contigs with an ND50 of 313 kb, with no duplicated marker genes.

4. Challenges to applying single-cell genomics to marine archaea

Despite the insights presented above, a population-level view of marine archaea similar to that achieved for *Prochlorococcus* has yet to be achieved, which is particularly surprising for the thaumarchaea given their abundance in their mesopelagic habitat. Why might this be? Few papers report the details necessary to calculate the recovery efficiency of archaea in single-cell libraries from marine samples, namely, the number of cells sorted, the number of cells successfully amplified, the total number of archaea recovered and an independent estimate of the relative abundance of organisms in the original, sorted sample (such as fluorescent *in situ* hybridization, 16S rRNA amplicon analysis, or phylogenetic profiling of metagenomic data). Nonetheless, we have attempted to evaluate the recovery of archaeal genomes in published marine single-cell studies and our own unpublished data (table 1). As a per cent of identified SAGs, meaning cells that underwent successful whole-genome amplification (WGA) and 16S rRNA gene screening, thaumarchaea represent between 14 and 21% of cells. Taking the sample from Station ALOHA as an example, thaumarchaea were previously reported to be on average 30% of total cells at the sampling depth of 770 m [31], thus the archaeal recovery does not differ much from this percentage.

Table 1. Recovery of archaeal single-cell genomes from marine samples. cmbsf, centimetres below seafloor; MDA, multiple displacement amplification; MGI, Marine Group I thaumarchaea; SAG, single amplified genome.

location	water depth (m)	no. of sorted cells	MDA-positive wells	% of sorted cells MDA-positive	no. of SAGs identified[a]	no. of archaeal SAGs	no. of MGI SAGs	MGI as % of identified SAGs	archaea as a % of sorted cells	MGI as a % of sorted cells	reference
Aarhus Bay, marine sediments	16 m/10 cmbsf	630	not reported	not reported	71	6	0		1.0%		[6]
Eastern Tropical South Pacific	2000	315	228	72%	29	6	4	14%	1.9%	1.3%	A.E.S. 2013, unpublished data
North Atlantic	800	1252[b]	1252[b]	100%	257	76	67	26%	4.8%	4.3%	[40,44]
North Pacific Subtropical Gyre (Station ALOHA)	770	630[b]	630[b]	100%	245	55	52	21%	5.8%	5.5%	[40,44]
Ogasawara Trench	505	315	not reported	not reported	not reported	not reported	4			1.3%	[42]
Ogasawara Trench	2015	315	not reported	not reported	not reported	not reported	4			1.3%	[42]
Ogasawara Trench	5010	315	not reported	not reported	not reported	not reported	4			1.3%	[42]
Ogasawara Trench	9697	315	not reported	not reported	not reported	not reported	6			1.9%	[42]
Puerto Rico Trench	8219	2880	634	22%	70	1	1	1.43%	0.03%	0.03%	[41]

[a]SAGs phylogenetically identified using 16S rRNA gene sequencing.

[b]Numbers were originally reported as MDA-positive wells. However, given that the numbers are even multiples of 315 (the number of single cells sorted in a 384-well plate), we suspect this may actually be the number of cells sorted.

However, as a per cent of total sorted cells, the recovery of thaumarchaea is between 1 and 5%, vastly lower than their predicted abundance at these depths. At these recovery rates, one would have to sort 20 000–80 000 cells—more than 50 384-well plates—to achieve the number of SAGs used to study population biology of *Prochlorococcus* [22]. This suggests that typical sorting gates for marine bacterioplankton may miss the very small thaumarchaea [50,51], which have been estimated to be smaller than 0.2 µm when cell size was assessed using fluorescence-activated cell sorting [52].

What other factors might be behind this apparently low recovery of archaea from marine samples? Lysis is a critical consideration for single-cell workflows [1]. While various combinations of heat, surfactants, enzymes and alkaline lysis have been used for single-cell genomics, current workflows in use for high-throughput single-cell sorting and amplification of marine samples [52] use alkaline lysis and up to five freeze–thaw cycles. Proteinase K was added to the lysis procedure to lyse single cells of the thaumarchaeon '*Ca*. Nitrosoarchaeum limnia' for single-cell sequencing [36], and this may be an interesting improvement to consider in future archaeal-targeted workflows.

There does not appear to be a specific bias against archaea in the multiple displacement amplification (MDA) reaction used for WGA. While biases in MDA are frequently described [53–55], there does not appear to be a specific bias against archaea. One of the first single-cell isolations and MDA amplifications of uncultured microorganisms was targeted towards archaea from agricultural soil [56]. Of eight archaeal cells isolated from the environment, only two yielded DNA from which the 16S rRNA gene could be amplified. However, a parallel validation using a culture of the methanogenic archaeon *Methanothermobacter thermoautotrophicus* yielded successful amplifications in all 10 cells attempted, again suggesting there is no MDA bias against archaea, but highlighting the importance of lysis considerations for potentially difficult to lyse cells from the environment. In a separate study, sequencing of five single cells from a thaumarchaeal enrichment culture found that although each individual MDA reaction recovered only about 60% of the genome, each reaction recovered a different region of the genome, allowing for a total genome recovery of greater than 95% [36]. This suggests that, as reported for bacteria, MDA biases can be overcome by sequencing multiple cells from a single population. Given the low recovery of thaumarchaea in SAG libraries discussed above, however, the likelihood of retrieving multiple cells from the same population is very low except in highly clonal populations [43].

An unexplored barrier to archaeal genome amplification is the potential for histone binding on archaeal DNA. Histones are found in euryarchaeota and marine thaumarchaeota [57,58]. Archaeal histones mimic eukaryotic histone folding properties, but lack N- and C-terminal tails [58–61]. Archaeal histones have the ability to form polymer complexes that protect genome fragments from micrococcal nuclease (MNase) digestion [60], a process that is used in MNase-sequencing of nucleosome regions [62]. Histone genes have also been found in the emerging Asgard archaea phylum, which interestingly share residues with eukaryotic histones but not histones from euryarchaeota [63], and may be an additional barrier to targeting this group using WGA.

5. Are metagenome-assembled genomes a more tractable approach for genomic recovery of planktonic archaea?

There are dramatic differences between the genome-resolved diversity recovered from MGI and MGII using single-cell genomics (SAGs) versus metagenomics (MAGs) methods (figure 1). Genome resolution of MGII is almost completely dominated by MAGs, whereas MGI diversity is much more evenly covered between SAGs and MAGs. As we elaborate below, this pattern is potentially expected, given what we understand about the relative biases of each method.

While the phylogenetic distribution of bacterial SAGs recovered from marine environments generally reflects their relative abundance in metagenomes [24,44,64], MAGs generally represent a strikingly different phylogenetic distribution. Within large populations (such as *Prochlorococcus* and the abundant marine alphaproteobacterial clade SAR11), subpopulations of abundant organisms and their associated genomic heterogeneity mean that assembly algorithms struggle to assemble complete genomes for these groups [65,66]. For example, 85 MAGs generated from the central Baltic Sea contained just three alphaproteobacterial genomes [67], though alphaproteobacteria represented 20–40% of metagenomic reads in this region [68]. Massive cross-assemblies of the Tara Oceans dataset, the largest metagenomic dataset in the marine environment, recovered only 17 picocyanobacterial genomes [69]. Similarly, a separate assembly of just the Tara Oceans surface water data recovered no picocyanobacterial genomes and only 54 SAR11 genomes, out of nearly 1000 MAGs [70], when together these groups should represent nearly 50% of the surface ocean sequence data. By contrast, and perhaps not surprisingly, organisms with a low to intermediate relative abundance and low population-level diversity assemble well [69,71]. For example, five *Nitrospina* genomes (10% of MAGs) were assembled from the Gulf of Mexico hypoxic zone [72], although these bacteria represented just 0.5% of the total sequence data [73].

With respect to the planktonic archaea, specifically, in the same two studies, no thaumarchaeal genomes were assembled, despite comprising 26% of the 16S rRNA amplicon data. Only 13 thaumarchaeal genomes were recovered from the Tara dataset [69], and a cross-assembly of publically available metagenomes that yielded nearly 8000 MAGs recovered only 26 thaumarchaeal genomes. Yet, this same study recovered 206 MGII genomes [74]. Our suspicion is that organisms in low relative abundance are more likely to represent clonal populations, allowing for more successful metagenomic assembly. Detailed metagenomic recruitment analysis is necessary to understand the extent to which the thaumarchaeal MAGs recovered from the Tara dataset represent the diversity actually present in the environment. Previous studies have shown low recruitment of Global Ocean Sampling expedition metagenomic data to reference thaumarchaeal genomes when imposing a per cent identity threshold greater than 90%, suggesting that thaumarchaeal populations indeed do have high population-level diversity [51] that may not be covered by MAGs.

In conclusion, if what is desired is a genome-resolved analysis of archaeal populations in the ocean, it appears that while SAGs are better at recovering MGI genomes than MGII, there still appears to be an underrepresentation of

marine archaea in single-cell datasets, and that targeted sorting approaches would be needed to improve the recovery of archaeal cells. For the MGII, metagenomic assembly appears to be a much more tractable approach. It should be noted that SAGs and MAGs recover a fundamentally different scale of biological variability—as noted above, population-level variation is recovered within large SAG datasets, while MAGs represent a population 'average' masking individual variation. Nonetheless, a recent comparison of SAGs and MAGs of the same taxa from the same environment showed that nearly identical genomes can be recovered from environmental samples using both methods [75], thus the choice of method should be based on the specific question asked, and, given the additional costs associated with single-cell sequencing, the resources available.

6. Future research needs for archaeal single-cell biology

If single-cell sequencing is to continue to provide insights into archaeal population biology in the ocean, it is clear that methodological improvements are necessary to enrich the recovery of archaea from environmental samples. More systematic tests of lysis and WGA efficiency are needed to specifically evaluate the efficiency on archaeal genomes. Unfortunately, single-cell benchmarking studies rarely include archaea [52]. The newly developed WGA-X method improved recovery of high %GC genomes, but actually had lower recovery of thaumarchaea than the original MDA protocol. It is difficult to assess the efficiency of archaeal SAG generation from environmental samples, as it is uncertain what the appropriate benchmark is to compare against when calculating efficiency. Comparing archaeal SAG recovery against the relative abundance of archaea in a paired 16S rRNA gene amplicon library from the same sample is not enough, given ongoing biases in primer-based recovery of uncultivated microbial taxa [76].

As mentioned above, simply adjusting sorting gates to recover smaller cells may improve the overall recovery of thaumarchaea from marine samples. It may be possible to target planktonic thaumarchaea, in particular, based on the fluorescence of the cofactor F420, also found in methanogens, which has a characteristic blue fluorescence. Flow cytometric detection and sorting based on F420 fluorescence was recently used to quantify, isolate and sequence methanogenic archaea [77]. F420 fluorescence has been used to visualize soil thaumarchaea by microscopy [78], suggesting it may be possible to sort marine thaumarchaea based on this property. Coupling flow sorting with Raman spectroscopy could allow the targeted sorting of archaea based on the uptake of stable isotope-labelled substrates [79]; in the case of the MGII, they have been shown to take up stable isotope-labelled proteins [80].

A final consideration for improving the recovery of archaea in single-cell libraries is the choice of screening method for genome-amplified cells. It is clear that primer bias is an ongoing issue in the recovery of microbial diversity from the environment [76]. Commonly used PCR primers miss, for example, the newly discovered Asgard archaea [81], which remained 'hidden' from view despite 30 years of molecular surveys of the ocean. Low coverage genomic sequencing (LoCoS) as an alternative to PCR-based screening of amplified cells is a promising way to circumvent this problem [52].

Though we have focused here on single-cell genomics, other emerging single-cell technologies could be equally transformative for archaeal biology in the ocean. For example, a better leveraging of single-cell technologies could enable high-throughput cultivation of uncultured archaeal groups, a technique that has not been widely applied in the marine environment since its inception [82]. Droplet microfluidics, where individual cells are isolated in growth medium in an oil emulsion, could potentially allow for the set-up and screening of cultures at the throughput necessary to recover rare groups [83]. Combined with targeted single-cell sorting approaches for archaea, development of such techniques would improve our understanding of planktonic archaea and the factors that shape their ecology and evolution.

Data accessibility. The alignment underlying the phylogenetic tree in figure 1 is available on FigShare: https://doi.org/10.6084/m9.figshare.8411864.v1.

Authors' contributions. A.E.S. conceived the paper; all authors contributed to the data analysis, interpretation and writing of the paper.

Competing interests. We declare we have no competing interests.

Funding. A.E.S. and M.K. were supported by an award from the Gordon and Betty Moore Foundation's Marine Microbiology Initiative (MMI5514). A.E.S. was additionally supported by a Simons Foundation Early Career Investigator award in Marine Microbial Ecology (345889). S.M.L. was partially supported by National Academies Keck Futures Initiative award DBS17.

Acknowledgements. We thank Chris Dupont and R. Alexander Richter of the J. Craig Venter Institute for the alignment used to construct the phylogenetic tree in figure 1. We acknowledge helpful discussions with Karen Lloyd and Frank Stewart. Two anonymous reviewers provided helpful feedback on the manuscript.

References

1. Blainey PC. 2013 The future is now: single-cell genomics of bacteria and archaea. *FEMS Microbiol. Rev.* **37**, 407–427. (doi:10.1111/1574-6976.12015)
2. Gawad C, Koh W, Quake SR. 2016 Single-cell genome sequencing: current state of the science. *Nat. Rev. Genet.* **17**, 175–188. (doi:10.1038/nrg.2015.16)
3. Kalisky T, Blainey P, Quake SR. 2011 Genomic analysis at the single-cell level. *Annu. Rev. Genet.* **45**, 431–445. (doi:10.1146/annurev-genet-102209-163607)
4. Stepanauskas R, Sieracki ME. 2007 Matching phylogeny and metabolism in the uncultured marine bacteria, one cell at a time. *Proc. Natl Acad. Sci. USA* **104**, 9052–9057. (doi:10.1073/pnas.0700496104)
5. Woyke T *et al.* 2009 Assembling the marine metagenome, one cell at a time. *PLoS ONE* **4**, e5299. (doi:10.1371/journal.pone.0005299)
6. Lloyd KG *et al.* 2013 Predominant archaea in marine sediments degrade detrital proteins. *Nature* **496**, 215–218. (doi:10.1038/nature12033)
7. Cuvelier ML *et al.* 2010 Targeted metagenomics and ecology of globally important uncultured eukaryotic phytoplankton. *Proc. Natl Acad. Sci. USA* **107**, 14 679–14 684. (doi:10.1073/pnas.1001665107)
8. Thrash JC, Temperton B, Swan BK, Landry ZC, Woyke T, DeLong EF, Stepanauskas R, Giovannoni SJ. 2014 Single-cell enabled comparative genomics

of a deep ocean SAR11 bathytype. *ISME J.* **8**, 1440–1451. (doi:10.1038/ismej.2013.243)

9. Landry Z, Swan BK, Herndl GJ, Stepanauskas R, Giovannoni SJ. 2017 SAR202 genomes from the dark ocean predict pathways for the oxidation of recalcitrant dissolved organic matter. *MBio* **8**. (doi:10.1128/mBio.00413-17)
10. Santoro AE, Richter RA, Dupont CL. 2019 Planktonic marine archaea. *Annu. Rev. Mar. Sci.* **11**, 131–158. (doi:10.1146/annurev-marine-121916-063141)
11. Giovannoni SJ, Stingl U. 2005 Molecular diversity and ecology of microbial plankton. *Nature* **437**, 343–348. (doi:10.1038/nature04158)
12. Hanson CA, Fuhrman JA, Horner-Devine MC, Martiny JBH. 2012 Beyond biogeographic patterns: processes shaping the microbial landscape. *Nat. Rev. Microbiol.* **10**, 497–506. (doi:10.1038/nrmicro2795)
13. Polz MF, Hunt DE, Preheim SP, Weinreich DM. 2006 Patterns and mechanisms of genetic and phenotypic differentiation in marine microbes. *Phil. Trans. R. Soc. B* **361**, 2009–2021. (doi:10.1098/rstb.2006.1928)
14. Casteleyn G, Leliaert F, Backeljau T, Debeer A-E, Kotaki Y, Rhodes L, Lundholm N, Sabbe K, Vyverman W. 2010 Limits to gene flow in a cosmopolitan marine planktonic diatom. *Proc. Natl Acad. Sci. USA* **107**, 12 952–12 957. (doi:10.1073/pnas.1001380107)
15. Vergin KL, Tripp HJ, Wilhelm LJ, Denver DR, Rappé MS, Giovannoni SJ. 2007 High intraspecific recombination rate in a native population of *Candidatus Pelagibacter* ubique (SAR11). *Environ. Microbiol.* **9**, 2430–2440. (doi:10.1111/j.1462-2920.2007.01361.x)
16. Thompson JR *et al.* 2005 Genotypic diversity within a natural coastal bacterioplankton population. *Science* **307**, 1311–1313. (doi:10.1126/science.1106028)
17. Rodríguez-Valera F. 2002 Approaches to prokaryotic biodiversity: a population genetics perspective. *Environ. Microbiol.* **4**, 628–633. (doi:10.1046/j.1462-2920.2002.00354.x)
18. DeLong EF. 2004 Microbial population genomics and ecology: the road ahead. *Environ. Microbiol.* **6**, 875–878. (doi:10.1111/j.1462-2920.2004.00668.x)
19. Konstantinidis KT, DeLong EF. 2008 Genomic patterns of recombination, clonal divergence and environment in marine microbial populations. *ISME J.* **2**, 1052–1065. (doi:10.1038/ismej.2008.62)
20. Coleman ML, Chisholm SW. 2010 Ecosystem-specific selection pressures revealed through comparative population genomics. *Proc. Natl Acad. Sci. USA* **107**, 18 634–18 639. (doi:10.1073/pnas.1009480107)
21. Rodrigue S, Malmstrom RR, Berlin AM, Birren BW, Henn MR, Chisholm SW. 2009 Whole genome amplification and *de novo* assembly of single bacterial cells. *PLoS ONE* **4**, e6864. (doi:10.1371/journal.pone.0006864)
22. Kashtan N *et al.* 2014 Single-cell genomics reveals hundreds of coexisting subpopulations in wild *Prochlorococcus*. *Science* **344**, 416–420. (doi:10.1126/science.1248575)
23. Biller SJ, Berube PM, Lindell D, Chisholm SW. 2015 *Prochlorococcus*: the structure and function of collective diversity. *Nat. Rev. Microbiol.* **13**, 13–27. (doi:10.1038/nrmicro3378)
24. Berube PM *et al.* 2018 Single cell genomes of *Prochlorococcus, Synechococcus*, and sympatric microbes from diverse marine environments. *Sci. Data* **5**, 180154. (doi:10.1038/sdata.2018.154)
25. Malmstrom RR, Coe A, Kettler GC, Martiny AC, Frias-Lopez J, Zinser ER, Chisholm SW. 2010 Temporal dynamics of *Prochlorococcus* ecotypes in the Atlantic and Pacific oceans. *ISME J.* **4**, 1252–1264. (doi:10.1038/ismej.2010.60)
26. Chisholm SW, Olson RJ, Zettler ER, Goericke R, Waterbury JB, Welschmeyer NA. 1988 A novel free-living prochlorophyte abundant in the oceanic euphotic zone. *Nature* **334**, 340–343. (doi:10.1038/334340a0)
27. Dean FB, Nelson JR, Giesler TL, Lasken RS. 2001 Rapid amplification of plasmid and phage DNA using Phi 29 DNA polymerase and multiply-primed rolling circle amplification. *Genome Res.* **11**, 1095–1099. (doi:10.1101/gr.180501)
28. Martin-Cuadrado A-B, Garcia-Heredia I, Moltó AG, López-Úbeda R, Kimes N, López-García P, Moreira D, Rodriguez-Valera F. 2015 A new class of marine Euryarchaeota group II from the Mediterranean deep chlorophyll maximum. *ISME J.* **9**, 1619–1634. (doi:10.1038/ismej.2014.249)
29. Rinke C *et al.* 2019 A phylogenomic and ecological analysis of the globally abundant Marine Group II archaea (*Ca*. Poseidoniales ord. nov.). *ISME J.* **13**, 663–675. (doi:10.1038/s41396-018-0282-y)
30. Letunic I, Bork P. 2019 Interactive Tree Of Life (iTOL) v4: recent updates and new developments. *Nucleic Acids Res.* **47**, W256–W259. (doi:10.1093/nar/gkz239)
31. Karner MB, DeLong EF, Karl DM. 2001 Archaeal dominance in the mesopelagic zone of the Pacific Ocean. *Nature* **409**, 507–510. (doi:10.1038/35054051)
32. Iverson V, Morris RM, Frazar CD, Berthiaume CT, Morales RL, Armbrust EV. 2012 Untangling genomes from metagenomes: revealing an uncultured class of marine Euryarchaeota. *Science* **335**, 587–590. (doi:10.1126/science.1212665)
33. Needham DM, Fuhrman JA. 2016 Pronounced daily succession of phytoplankton, archaea and bacteria following a spring bloom. *Nat. Microbiol.* **1**, 16005. (doi:10.1038/nmicrobiol.2016.5)
34. Galand PE, Gutiérrez-Provecho C, Massana R, Gasol JM, Casamayor EO. 2010 Inter-annual recurrence of archaeal assemblages in the coastal NW Mediterranean Sea (Blanes Bay Microbial Observatory). *Limnol. Oceanogr.* **55**, 2117–2125. (doi:10.4319/lo.2010.55.5.2117)
35. Spang A, Ettema TJG. 2017 Archaeal evolution: the methanogenic roots of Archaea. *Nat. Microbiol.* **2**, 17109. (doi:10.1038/nmicrobiol.2017.109)
36. Blainey PC, Mosier AC, Potanina A, Francis CA, Quake SR. 2011 Genome of a low-salinity ammonia-oxidizing archaeon determined by single-cell and metagenomic analysis. *PLoS ONE* **6**, e16626. (doi:10.1371/journal.pone.0016626)
37. Francis CA, Roberts KJ, Beman JM, Santoro AE, Oakley BB. 2005 Ubiquity and diversity of ammonia-oxidizing archaea in water columns and sediments of the ocean. *Proc. Natl Acad. Sci. USA.* **102**, 14 683–14 688. (doi:10.1073/pnas.0506625102)
38. Hallam SJ, Mincer TJ, Schleper C, Preston CM, Roberts K, Richardson PM, Delong EF. 2006 Pathways of carbon assimilation and ammonia oxidation suggested by environmental genomic analyses of marine Crenarchaeota. *PLoS Biol.* **4**, e95. (doi:10.1371/journal.pbio.0040095)
39. Luo H, Tolar BB, Swan BK, Zhang CL, Stepanauskas R, Ann Moran M, Hollibaugh JT. 2014 Single-cell genomics shedding light on marine Thaumarchaeota diversification. *ISME J.* **8**, 732–736. (doi:10.1038/ismej.2013.202)
40. Swan BK *et al.* 2014 Genomic and metabolic diversity of Marine Group I Thaumarchaeota in the mesopelagic of two subtropical gyres. *PLoS ONE* **9**, e95380. (doi:10.1371/journal.pone.0095380)
41. León-Zayas R *et al.* 2015 Single cells within the Puerto Rico trench suggest hadal adaptation of microbial lineages. *Appl. Environ. Microbiol.* **81**, 8265–8276. (doi:10.1128/AEM.01659-15)
42. Wang Y *et al.* 2019 Genomics insights into ecotype formation of ammonia-oxidizing archaea in the deep ocean. *Environ. Microbiol.* **21**, 716–729. (doi:10.1111/1462-2920.14518)
43. Ngugi DK *et al.* 2015 Comparative genomics reveals adaptations of a halotolerant thaumarchaeon in the interfaces of brine pools in the Red Sea. *ISME J.* **9**, 396–411. (doi:10.1038/ismej.2014.137)
44. Swan BK *et al.* 2011 Potential for chemolithoautotrophy among ubiquitous bacteria lineages in the dark ocean. *Science* **333**, 1296–1300. (doi:10.1126/science.1203690)
45. Martinez-Garcia M, Brazel D, Poulton NJ, Swan BK, Gomez ML, Masland D, Sieracki ME, Stepanauskas R. 2012 Unveiling *in situ* interactions between marine protists and bacteria through single cell sequencing. *ISME J.* **6**, 703–707. (doi:10.1038/ismej.2011.126)
46. Orsi WD, Smith JM, Wilcox HM, Swalwell JE, Carini P, Worden AZ, Santoro AE. 2015 Ecophysiology of uncultivated marine euryarchaea is linked to particulate organic matter. *ISME J.* **9**, 1747–1763. (doi:10.1038/ismej.2014.260)
47. Chow C-ET, Winget DM, White 3rd RA, Hallam SJ, Suttle CA. 2016 Combining genomic sequencing methods to explore viral diversity and reveal potential virus–host interactions. *Front. Microbiol.* **6**, 265. (doi:10.3389/fmicb.2015.00265)
48. Labonté JM *et al.* 2015 Single-cell genomics-based analysis of virus–host interactions in marine surface bacterioplankton. *ISME J.* **9**, 2386–2399. (doi:10.1038/ismej.2015.48)
49. Mende DR, Aylward FO, Eppley JM, Nielsen TN, DeLong EF. 2016 Improved environmental genomes via integration of metagenomic and single-cell assemblies. *Front. Microbiol.* **7**, 143. (doi:10.3389/fmicb.2016.00143)

50. Könneke M, Bernhard AE, de la Torre JR, Walker CB, Waterbury JB, Stahl DA. 2005 Isolation of an autotrophic ammonia-oxidizing marine archaeon. *Nature* **437**, 543–546. (doi:10.1038/nature03911)
51. Santoro AE *et al.* 2015 Genomic and proteomic characterization of '*Candidatus* Nitrosopelagicus brevis': an ammonia-oxidizing archaeon from the open ocean. *Proc. Natl Acad. Sci. USA* **112**, 1173–1178. (doi:10.1073/pnas.1416223112)
52. Stepanauskas R *et al.* 2017 Improved genome recovery and integrated cell-size analyses of individual uncultured microbial cells and viral particles. *Nat. Commun.* **8**, 84. (doi:10.1038/s41467-017-00128-z)
53. van Dijk EL, Auger H, Jaszczyszyn Y, Thermes C. 2014 Ten years of next-generation sequencing technology. *Trends Genet.* **30**, 418–426. (doi:10.1016/j.tig.2014.07.001)
54. van Dijk EL, Jaszczyszyn Y, Thermes C. 2014 Library preparation methods for next-generation sequencing: tone down the bias. *Exp. Cell Res.* **322**, 12–20. (doi:10.1016/j.yexcr.2014.01.008)
55. Sabina J, Leamon JH. 2015 Bias in whole genome amplification: causes and considerations. *Methods Mol Biol.* **1347**, 15–41. (doi:10.1007/978-1-4939-2990-0_2)
56. Kvist T, Ahring BK, Lasken RS, Westermann P. 2007 Specific single-cell isolation and genomic amplification of uncultured microorganisms. *Appl. Microbiol. Biotechnol.* **74**, 926–935. (doi:10.1007/s00253-006-0725-7)
57. Cubonová L, Sandman K, Hallam SJ, Delong EF, Reeve JN. 2005 Histones in crenarchaea. *J. Bacteriol.* **187**, 5482–5485. (doi:10.1128/JB.187.15.5482-5485.2005)
58. Henneman B, van Emmerik C, van Ingen H, Dame RT. 2018 Structure and function of archaeal histones. *PLoS Genet.* **14**, e1007582. (doi:10.1371/journal.pgen.1007582)
59. Pereira SL, Grayling RA, Lurz R, Reeve JN. 1997 Archaeal nucleosomes. *Proc. Natl Acad. Sci. USA.* **94**, 12 633–12 637. (doi:10.1073/pnas.94.23.12633)
60. Mattiroli F *et al.* 2017 Structure of histone-based chromatin in archaea. *Science* **357**, 609–612. (doi:10.1126/science.aaj1849)
61. Gehring AM, Walker JE, Santangelo TJ. 2016 Transcription regulation in archaea. *J. Bacteriol.* **198**, 1906–1917. (doi:10.1128/JB.00255-16)
62. Ramani V, Qiu R, Shendure J. 2019 High sensitivity profiling of chromatin structure by MNase-SSP. *Cell Rep.* **26**, 2465–2476. (doi:10.1016/j.celrep.2019.02.007)
63. Henneman B, Dame RT. 2015 Archaeal histones: dynamic and versatile genome architects. *AIMS Microbiol.* **1**, 72–81. (doi:10.3934/microbiol.2015.1.72)
64. Eren AM, Murat Eren A, Esen ÖC, Quince C, Vineis JH, Morrison HG, Sogin ML, Delmont TO. 2015 Anvi'o: an advanced analysis and visualization platform for 'omics data. *PeerJ.* **3**, e1319. (doi:10.7717/peerj.1319)
65. Tringe SG *et al.* 2005 Comparative metagenomics of microbial communities. *Science* **308**, 554–557. (doi:10.1126/science.1107851)
66. Rusch DB *et al.* 2007 The Sorcerer II Global Ocean Sampling expedition: northwest Atlantic through eastern tropical Pacific. *PLoS Biol.* **5**, e77. (doi:10.1371/journal.pbio.0050077)
67. Hugerth LW, Larsson J, Alneberg J, Lindh MV, Legrand C, Pinhassi J, Andersson AF. 2015 Metagenome-assembled genomes uncover a global brackish microbiome. *Genome Biol.* **16**, 279. (doi:10.1186/s13059-015-0834-7)
68. Dupont CL *et al.* 2014 Functional tradeoffs underpin salinity-driven divergence in microbial community composition. *PLoS ONE* **9**, e89549. (doi:10.1371/journal.pone.0089549)
69. Tully BJ, Graham ED, Heidelberg JF. 2018 The reconstruction of 2,631 draft metagenome-assembled genomes from the global oceans. *Sci. Data* **5**, 170203. (doi:10.1038/sdata.2017.203)
70. Delmont TO, Quince C, Shaiber A, Esen ÖC, Lee ST, Rappé MS, McLellan SL, Lücker S, Eren AM. 2018 Nitrogen-fixing populations of Planctomycetes and Proteobacteria are abundant in surface ocean metagenomes. *Nat. Microbiol.* **3**, 804–813. (doi:10.1038/s41564-018-0176-9)
71. Tully BJ. 2019 Metabolic diversity within the globally abundant Marine Group II Euryarchaea offers insight into ecological patterns. *Nat. Commun.* **10**, 271. (doi:10.1038/s41467-018-07840-4)
72. Thrash JC, Baker BJ, Seitz KW, Temperton B, Campbell LG, Rabalais NN, Henrissat B, Mason OU. 2018 Metagenomic assembly and prokaryotic metagenome-assembled genome sequences from the Northern Gulf of Mexico 'Dead Zone'. *Microbiol. Resour. Announc.* **7**. (doi:10.1128/MRA.01033-18)
73. Gillies LE, Thrash JC, deRada S, Rabalais NN, Mason OU. 2015 Archaeal enrichment in the hypoxic zone in the northern Gulf of Mexico. *Environ. Microbiol.* **17**, 3847–3856. (doi:10.1111/1462-2920.12853)
74. Parks DH, Rinke C, Chuvochina M, Chaumeil P-A, Woodcroft BJ, Evans PN, Hugenholtz P, Tyson GW. 2017 Recovery of nearly 8,000 metagenome-assembled genomes substantially expands the tree of life. *Nat. Microbiol.* **2**, 1533–1542. (doi:10.1038/s41564-017-0012-7)
75. Alneberg J *et al.* 2018 Genomes from uncultivated prokaryotes: a comparison of metagenome-assembled and single-amplified genomes. *Microbiome* **6**, 173. (doi:10.1186/s40168-018-0550-0)
76. Lloyd KG, Steen AD, Ladau J, Yin J, Crosby L. 2018 Phylogenetically novel uncultured microbial cells dominate earth microbiomes. *mSystems* **3**. (doi:10.1128/mSystems.00055-18)
77. Lambrecht J, Cichocki N, Hübschmann T, Koch C, Harms H, Müller S. 2017 Flow cytometric quantification, sorting and sequencing of methanogenic archaea based on F_{420} autofluorescence. *Microb. Cell Fact.* **16**, 180. (doi:10.1186/s12934-017-0793-7)
78. Spang A *et al.* 2012 The genome of the ammonia-oxidizing *Candidatus* Nitrososphaera gargensis: insights into metabolic versatility and environmental adaptations. *Environ. Microbiol.* **14**, 3122–3145. (doi:10.1111/j.1462-2920.2012.02893.x)
79. Jing X *et al.* 2018 Raman-activated cell sorting and metagenomic sequencing revealing carbon-fixing bacteria in the ocean. *Environ. Microbiol.* **20**, 2241–2255. (doi:10.1111/1462-2920.14268)
80. Orsi WD *et al.* 2016 Diverse, uncultivated bacteria and archaea underlying the cycling of dissolved protein in the ocean. *ISME J.* **10**, 2158–2173. (doi:10.1038/ismej.2016.20)
81. Zaremba-Niedzwiedzka K *et al.* 2017 Asgard archaea illuminate the origin of eukaryotic cellular complexity. *Nature* **541**, 353–358. (doi:10.1038/nature21031)
82. Zengler K, Toledo G, Rappe M, Elkins J, Mathur EJ, Short JM, Keller M. 2002 Cultivating the uncultured. *Proc. Natl Acad. Sci. USA* **99**, 15 681–15 686. (doi:10.1073/pnas.252630999)
83. Guo MT, Rotem A, Heyman JA, Weitz DA. 2012 Droplet microfluidics for high-throughput biological assays. *Lab. Chip.* **12**, 2146–2155. (doi:10.1039/c2lc21147e)