Title

Genomic Mysteries of Giant Bacteria: Insights and Implications

Journal

Genome Biology and Evolution, 15(9)

ISSN

1759-6653

Authors

Ionescu, Danny
Volland, Jean-Marie
Contarini, Paul-Emile
et al.

Publication Date

2023-09-04

DOI

10.1093/gbe/evad163

Genomic Mysteries of Giant Bacteria: Insights and Implications

Danny Ionescu [1,*], Jean-Marie Volland [2,3], Paul-Emile Contarini [2,4], and Olivier Gros [4]

[1]Department of Plankton and Microbial Ecology, Leibniz Institute of Freshwater Ecology and Inland Fisheries, Neuglobsow, Germany
[2]Laboratory for Research in Complex Systems, Menlo Park, California, USA
[3]Department of Energy Joint Genome Institute, Lawrence Berkeley National Laboratory, Berkeley, California, USA
[4]Institut de Systématique, Evolution, Biodiversité (ISYEB), Muséum National d'Histoire Naturelle, CNRS, Sorbonne Université, EPHE, Université des Antilles, Pointe-à-Pitre, France

*Corresponding author: E-mail: danny.ionescu@igb-berlin.de.

Accepted: September 01, 2023

Abstract

Bacteria and *Archaea* are traditionally regarded as organisms with a simple morphology constrained to a size of 2–3 µm. Nevertheless, the history of microbial research is rich in the description of giant bacteria exceeding tens and even hundreds of micrometers in length or diameter already from its early days, for example, *Beggiatoa* spp., to the present, for example, *Candidatus* Thiomargarita magnifica. While some of these giants are still being studied, some were lost to science, with merely drawings and photomicrographs as evidence for their existence. The physiology and biogeochemical role of giant bacteria have been studied, with a large focus on those involved in the sulfur cycle. With the onset of the genomic era, no special emphasis has been given to this group, in an attempt to gain a novel, evolutionary, and molecular understanding of the phenomenon of bacterial gigantism. The few existing genomic studies reveal a mysterious world of hyperpolyploid bacteria with hundreds to hundreds of thousands of chromosomes that are, in some cases, identical and in others, extremely different. These studies on giant bacteria reveal novel organelles, cellular compartmentalization, and novel mechanisms to combat the accumulation of deleterious mutations in polyploid bacteria. In this perspective paper, we provide a brief overview of what is known about the genomics of giant bacteria and build on that to highlight a few burning questions that await to be addressed.

Key words: genomics, giant bacteria, polyploidy, bacterial heterozygosity, size limitations.

Significance

Giant bacteria have been described for over a century, yet most of them remain understudied. We bring forth current knowledge on the genomics of giant bacteria and use this to postulate key questions that should be addressed to better understand these organisms and harness their large size and "bacterial simplicity" to gain insight into the subcellular organization of bacteria.

Meet the Giants

Bacteria and *Archaea* are typically regarded as single-celled microscopic organisms with diameters or lengths not exceeding approximately >2 µm and, while bacterial organelles are being more frequently recognized (Greening and Lithgow 2020), bacteria lack a membrane-bound nucleus and other complex organelles found in eukaryotic cells. Nevertheless, the history of environmental microbiology research is garnished with the occasional, yet constant, description of species that are far larger (fig. 1), starting with early

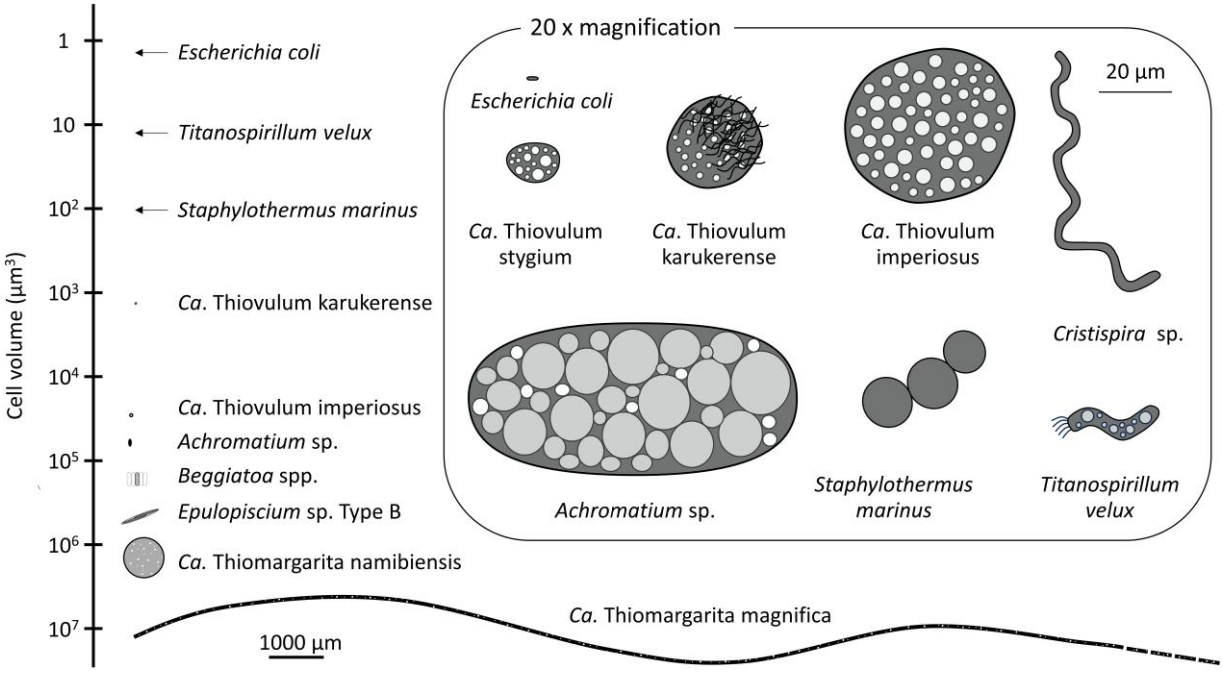

Fig. 1.—A graphical representation of giant bacteria across a seven-order magnitude biovolume scale. The giant bacteria are drawn to scale using the maximal dimensions as reported in the literature and compiled in table 1. The cell sizes and biovolumes in the figure should be considered as ranges and not as absolute numbers. The insert offers a closer look at the modestly sized giant bacteria in comparison with *Escherichia coli*, which is commonly used as a model organism.

descriptions of *Beggiatoa* sp. (Vaucher 1803; Trevisan 1842) and *Achromatium* sp. (Schewiakoff 1893), and continuing nowadays with, for example, *Candidatus* Thiomargarita magnifica (Volland et al. 2022), *Candidatus* Thiovulum stygium (Bizic et al. 2023), and *Candidatus* Thiovulum imperiosus (Sylvestre et al. 2022). Many microscopic bacteria form large multicellular structures, for instance, among cyanobacteria (Komárek and Johansen 2015), cable bacteria (Pfeffer et al. 2012), or magnetotactic bacteria (Keim et al. 2004). These are not regarded as true giants. Here, the term "giant bacteria" refers to bacterial cells, which may or may not form larger multicellular structures, but whose diameter or length exceeds that of "normal" bacteria being >10 µm in at least one dimension. In our discussion, we have also excluded differentiated cells such as heterocysts and akinetes of Cyanobacteria as in the case of heterocysts, this is an irreversible state, whereas under the appropriate conditions, akinetes will differentiate back into vegetative cells. What is noteworthy is that unicellular eukaryotes exhibit as well a large size variability, with some being the size of small bacteria (Lynch et al. 2022), yet as will be clarified, among bacteria, these have different consequences.

There are numerous descriptions of giant bacteria, yet phylogenetic data are not available for many of them because some descriptions preceded the advent of molecular phylogeny for example (Delaporte 1964, 1970). Nevertheless, it is clear that bacterial gigantism has evolved multiple times as giant cells are found in at least six phyla (fig. 2). A list of currently known giant bacteria is given in table 1, and a more detailed list is given in supplementary table S1, Supplementary Material online. As this list is based on a literature survey, it may not include taxa for which size was not reported explicitly. Also, specifically for Cyanobacteria, as this phylum harbors many cells that may exceed our definition of giants, noteworthy examples were given. An exhaustive resource for Cyanobacteria that includes size descriptions of most taxa can be found at http://www.cyanodb.cz (Hauer and Komarek 2022).

A high proportion of bacteria (up to 99.9%) have not been isolated in culture (Locey and Lennon 2016; Steen et al. 2019) and most of the bacterial biodiversity is recognized through molecular analysis of environmental samples that cannot provide information on bacterial sizes. It is, therefore, likely that some of the sequence data in databases correspond to unrecognized giant bacteria. It is also likely that bacterial biodiversity studies are biased in the way they sample cells in the first place. For example, aquatic bacterial communities are often filtered to exclude larger eukaryotic organisms and therefore exclude giant bacteria. All in all, there are likely more giant bacteria out there waiting to be discovered.

Giant bacteria were so far found in "energy rich" environments, that is, with ample organic matter (such as

digestive tracts as is the case for *Epulopiscium*, *Metabacterium*, and *Cristispira*) or ample electron donors for photo- or chemo-autotrophy as is the case for Cyanobacteria or large sulfur bacteria. This is likely linked to the energetic demands of being a giant. A large fraction of the known giant bacteria are free-living sulfur oxidizers and phototrophs (mainly cyanobacteria), and the rest for which information exists are heterotrophs mostly found in the microbiome of eukaryotic organisms as symbionts (fig. 2).

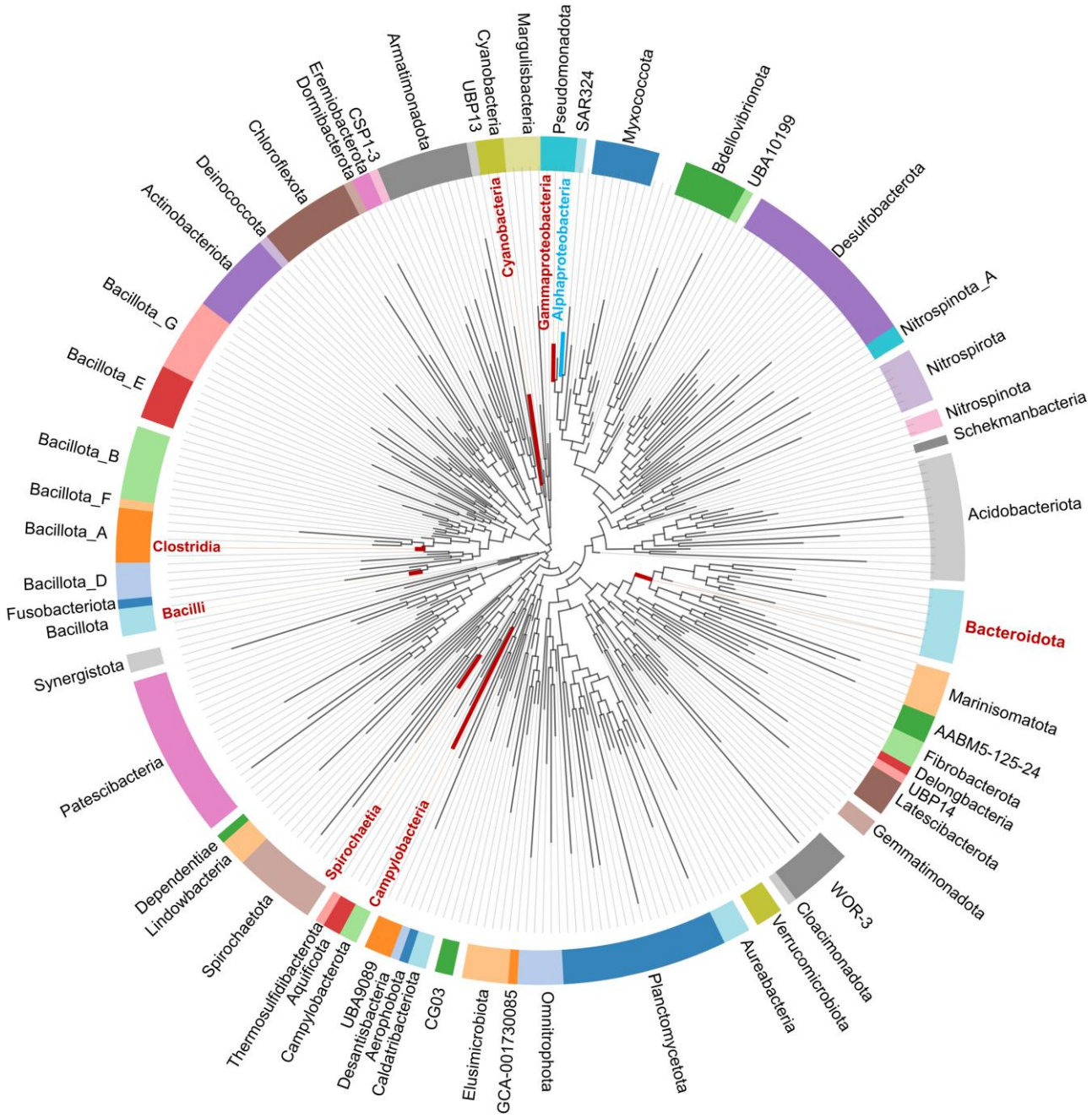

Fig. 2.—Phylogenetic tree of *Bacteria* highlighting the classes (inner circle) where giant bacteria occur. Taxa of giant bacteria for which the phylogenetic placement is inferred from genomics or molecular markers are shown in red. Taxa for which only morphological classification exists (the class *Alphaproteobacteria*) are marked in blue and assigned to the class based on the classification of associated taxa. The base tree was built using AnnoTree (http://annotree.uwaterloo.ca/app/; Mendler et al. 2019).

Table 1

Known Giant Bacteria and General Available Information

Taxon	Phylum	Maximal Cell Sizes	PhyM	Gen.	Chrom./Cell	Habitat	Features/Comments	References
S *Bacillus* spp.	Bacillota	Ø 7 μm; L 100 μm	N/A	N/A	N/A	Host associated		Delaporte (1964, 1970)
A *Epulopiscium* sp.	Bacillota	Ø 80 μm; L 800 μm	Av.	Av.	<500,000	Host associated	Spore forming	Arroyo et al. (2019)
A *Metabacterium* spp.	Bacillota	Ø 6 μm; L 25 μm	Av.	N/A	N/A	Host associated		Chatton and Perard (1913)
S *Lysinibacillus varians*	Bacillota	Ø 0.5 μm, L 1 cm	Av.	Av.	N/A	Sediment	Cable bacterium	Yang et al. (2021)
A *Karelsulcia muelleri*	Bacteroidota	Ø 5 μm, L 100 μm	Av.	Av.	200–900	Host associated	Also known as *Sulcia muelleri*	Moran et al. (2005)
M *Thiovulum* spp.	Campylobacterota	Ø 50 μm	Av.	Av.	>10 < 1,000	Fresh/brackish/marine sulfidic water		Sylvestre et al. (2022)
S *Chroococcus* spp.	Cyanobacteria	Ø 60 μm	N/A	N/A	N/A	Fresh/brackish/marine water		Wood et al. (2017)
S *Oscillatoria princeps*	Cyanobacteria	Ø 70 μm; L 10 μm	Av.	Av.	N/A	Cosmopolitan		Ward et al. (2021)
S *Stigonema* spp.	Cyanobacteria		Av.	Av.	N/A	Fresh water		Marter et al. (2021)
A *Arthrospira* spp	Cyanobacteria	Ø 12 μm	Av.	Av.	1	Fresh water		Borowitzka (2018)
A *Gomphosphaeria aponina*	Cyanobacteria	Ø 12 μm; L 15 μm	Av.	Av.	N/A	Fresh water		Dwivedi et al. (2010)
A *Porphyrosiphon* spp.	Cyanobacteria	Ø 15 μm; L 23 μm	Av.	N/A	1	Fresh water		Kim et al. (2022)
A *Achromatium* spp.	Pseudomonadota	Ø 35 μm; L 125 μm	Av.	Av.	<500	Aquatic sediments	$CaCO_3$ inclusions/ Heterozygous	Salman et al. (2016), Ionescu et al. (2017)
A *Beggiatoa* spp.	Pseudomonadota	Ø 5 μm; L 200 μm	Av.	Av.	N/A	Upper layer of sulfidic soil sediments, fresh water	Central vacuoles	Fomenkov et al. (2018)
A *Ca.* Marithrix	Pseudomonadota	Ø 100 μm; L 30 μm	Av.	Av.	N/A	Upper layer of sulfidic soil sediments		Salman-Carvalho et al. (2016)
A *Ca.* Parabeggiatoa sp.	Pseudomonadota	Ø 40 μm; L 14 μm	Av.	N/A	N/A	Upper layer of sulfidic soil sediments		Salman et al. (2011)
A *Ca.* Thiophysa spp.	Pseudomonadota	Ø 90 μm	Av.	N/A	N/A	Upper layer of sulfidic soil sediments		Salman et al. (2011)
A *Ca.* Thiopilula aggregata	Pseudomonadota	Ø 65 μm	Av.	N/A	N/A	Upper layer of sulfidic soil sediments		Salman et al. (2011)
S *Ca.* Thiosymbion spp.	Pseudomonadota	Ø 1.2 μm; L 120 μm	Av.	N/A	N/A	Host associated		Pende et al. (2014)
S *Chromatium* spp.	Pseudomonadota	Ø 6 μm; L 16 μm	Av.	Av.	N/A	Mud of lake		Luedin et al. (2019)
A *Isobeggiatoa* spp.	Pseudomonadota	Ø 40 μm; L 25 μm	Av.	N/A	N/A	Microbial mat, upper layer of sulfidic marine sediments	Central vacuoles	Jean et al. (2015)
A *Maribeggiatoa* spp.	Pseudomonadota	Ø 120 μm; L 40 μm	Av.	N/A	N/A	Microbial mat, upper layer of sulfidic marine sediments	Central vacuoles	Jean et al. (2015)
S *Spirillum* spp.	Pseudomonadota	Ø 7 μm; L 100 μm	N/A	N/A	N/A	Stagnant fresh water		Delaporte (1964, 1970)

(continued)

Table 1 Continued

Taxon	Phylum	Maximal Cell Sizes	PhyM	Gen.	Chrom./Cell	Habitat	Features/Comments	References
A *Thiomargarita* spp.	Pseudomonadota	Ø 50 μm; L 2 cm	Av.	Av.	10,000 to >1,000,000	Upper layer of sulfidic marine sediments	"Pepins", Central vacuoles	Volland et al. (2022)
A *Thioploca* spp.	Pseudomonadota	Ø 45 μm; L 35 μm	Av.	Av.	N/A	Mud of lake or marine benthos		Kojima et al. (2015)
A *Titanospirillum velox*	Pseudomonadota	Ø 5 μm; L 30 μm	N/A	N/A	N/A	Upper layer of sulfidic soil sediments		Guerrero et al. (1999)
S *Cristispira* sp.	Spirochaetes	Ø 3 μm; L 180 μm	Av.	N/A	N/A	Crystalline style of bivalve mollusks		Margulis and Hinkle (2013)
S *Spirochaeta plicatilis*	Spirochaetes	Ø 0.75 μm; L 250 μm	N/A	N/A	N/A	Fresh water and marine, sulfide-containing environments		Blakemore and Canale-Parola (1973)
S *Treponema* spp.	Spirochaetes	Ø 0.4 μm, L 20 μm	Av.	Av.	N/A	Host associated		Han et al. (2013)
M *Borrelia burgdorferi*	Spirochaetes	Ø 0.5 μm, L 26 μm	Av.	Av.		Host associated		Fraser et al. (1997)

NOTE.—S/M/A, some/most/all taxa in the genus are giants; Ø, cross-section or spherical diameter; L, cell length (nonspherical); PhylM, availability of phylogenetic markers, for example, SSU/LSU rRNA; Gen, availability of a genomic assembly for at least some species in the genus; Chrom./cell, reported ranges of number of chromosomes per single cell.

Cell sizes were taken as reported in publications where available or determined based on scale bars of photomicrographs. Details on species within genera as well as a detailed bibliography are given in supplementary table S1, Supplementary Material online. When possible, bibliography here points to the first genomic description.

What Are the Challenges of Being a Giant?

Bacterial gigantism is an interesting phenomenon not only because it challenges our traditional perception of bacteria as minuscule organisms but also because it challenges a series of theoretical biophysical and bioenergetic limitations on bacterial size, reflecting our gap in understanding the cellular biology and physiology of large bacteria. These limitations are predicted by models that are based on the morphology, ultrastructure, and metabolism of typical, small, model bacteria, and thus fail to encompass the unique traits of larger bacterial cells. Unfortunately, even studies developing new theorems on the drives of bacterial size do so within the size range of *Escherichia coli* and do not consider larger bacteria (Gallet et al. 2017). Obviously, the existence of multiple species of giant bacteria, their global distribution, and localized high abundance in nature (e.g., Beggiatoaceae [Teske and Salman 2014], *Achromatium* [Ionescu et al. 2020], and Thiotrichaceae [Ravin et al. 2022]) suggests that bacteria have found creative solutions to such theoretical problems.

Bacteria depend on diffusion, both for the transfer of substances from the environment into the cell and, lacking the internal active transport systems typical of eukaryotic cells, also for intracellular trafficking. In short, the larger the cell, the smaller the surface area-to-volume ratio is, dropping from a value of 6 for a spherical cell with a diameter of 1 µm to 0.6 and 0.06 for a cell with a diameter of 10 or 100 µm, respectively. Accordingly, the diffusion time of solutes from the environment changes from milliseconds to hours, and the internal trafficking time may theoretically reach even days. This topic has been reviewed previously and will not be discussed here (Schulz and Jørgensen 2002).

Trafficking of small solutes is one aspect of diffusion limitations, yet it is not the only one. Even if nutrient supply was not a limiting factor, a cell must generate and distribute enough macromolecules (such as structural and functional proteins, RNAs) to maintain its functionality. This becomes increasingly challenging as cell size grows (Scott et al. 2010). Synthesizing enough macromolecules from a single chromosome is likely insufficient to support a giant cell. Even if it were possible, the macromolecules produced in one part of the cell would still face diffusion limitations and would not reach their target sites in a meaningful time.

Building on the aforementioned, Kempes et al. (2016) suggested a second limitation on bacterial sizes driven by the number of ribosomes needed to maintain the transcriptional activity of a large cell. This so-called ribosome catastrophe predicts a maximum biovolume for bacterial cells at $1.39 \pm 0.03 \times 10^{-15}$ m^3. Above that, the volume of the ribosomes would exceed the total volume of the cell. The occurrence of cells as *Candidatus* Thiomargarita namibiensis, *Ca.* T. magnifica, and *Epulopiscium* sp. with biovolumes in the range of 4×10^{-12} m^3 (excluding the central vacuole), 2×10^{-11} m^3 (excluding the central vacuole; Volland et al. 2022), and 2.5×10^{-12} (entire cell), respectively, demonstrates that this concept does not necessarily apply.

Solutions and Consequences

Bacteria have evolved several adaptations to partially overcome diffusion limitations. These include: 1) shape shifting (Harris and Theriot 2018) and invaginating the cytoplasmic membrane (Angert 2006; Schulz 2006) to increase the surface-to-volume ratio; and, 2) localizing most of the metabolic activity close to the cytoplasmic membrane and having a large vacuole occupying at least 80% of the cell volume, thus reducing the active compartment of the cell (Schulz and Jørgensen 2002; Levin and Angert 2016; Volland et al. 2022).

A potential solution suggested for the "ribosome catastrophe" is to both lower the growth rate and cellular metabolism, to reduce the number of necessary ribosomes. While the growth rate of many giant bacteria cannot be accurately determined as these organisms are not available in culture, such a solution does not seem to be employed by *Epulopiscium* spp. which reproduce once in 24 h (Angert 2021).

One solution that appears to have been adopted by all giant bacteria is having multiple chromosomes distributed across the cell. Polyploidy in *Bacteria* and *Archaea* is defined as the presence of ten or more chromosomes per cell, even though the number itself does not convey any information on whether these chromosomes are identical or not. Polyploidy is more common among *Bacteria* and *Archaea* than previously thought (Oliverio and Katz 2014). So far, all investigated giant bacteria are polyploid, with cells containing a large number of chromosomes ranging from tens in smaller cells like *Ca.* T. stygium (Bizic et al. 2023; Ionescu, unpublished data), to hundreds in cells like *Achromatium* spp. (Ionescu et al. 2017) to hundreds of thousands in *Thiomargarita* and *Epulopiscium* (Mendell et al. 2008; Volland et al. 2022).

Polyploid bacteria should be vulnerable to the accumulation of deleterious mutations leading to extinction via a mechanism termed Müller's Ratchet (Takeuchi et al. 2014; Markov and Kaznacheev 2016). To overcome this, most small-sized polyploid bacteria make use of asymmetrical recombination (known as gene conversion; Soppa 2011; Takeuchi et al. 2014; Markov and Kaznacheev 2016). However, gene conversion requires physical interaction between the different chromosomes in the cells. In giant bacteria, where hundreds or thousands of chromosomes are distributed across a large cell, such interaction is impossible or unlikely. First, in some cases, the chromosomes are prevented by cellular architecture from interacting with distant chromosomes. In *Ca.* T. magnifica, one or more chromosomes are contained in membrane-bound

organelles (Volland et al. 2022). In the case of *Achromatium* spp., the chromosomes are in thin membranal stretches or pockets that are separated from each other by the cell's periplasmatic $CaCO_3$ crystals (Ionescu et al. 2017; Schorn et al. 2020). Second, even without this immobilization of chromosomes, the cytoplasmatic milieu, coupled with the cell size, makes it impossible for distant chromosomes to interact on time scales that are relevant for gene conversion. Goodsell (1991) provided an informative picture of the cytoplasmatic environment, changing previous concepts of "empty water bags" to an environment highly crowded by macromolecules. He further calculated using an example of a 160-kDa protein that the diffusion of macromolecules is approximately 1,000 times slower than in water. Bohrer and Xiao (2020) estimated the diffusion coefficient for DNA in a bacterial cell to be 0.001 $\mu m^2/s$. The time required for two molecules to meet inside a cell is given by the formula $t_{traffic} = L^3/DR$, where L is the cell radius, D is the sum of diffusion coefficients of the two molecules, and R is the molecule radius (Schulz and Jørgensen 2002). Trafficking time in giant bacteria was already suggested to be on the orders of hours for small solutes (Schulz and Jørgensen 2002), and calculating it for large molecules as chromosomes results in approximately 400 years (for a cell with a radius of 100 µm, and a packed chromosome radius of 40 nm; Takeyasu et al. 2004). The same calculation suggests that for any bacterium with a cytoplasmic diameter >4 µm two chromosomes would require >24 h to meet. While these are clearly rough estimates, they demonstrate the effect bacterial size has on the chances of two distant chromosomes meeting inside the cell. As already pointed out (Schulz and Jørgensen 2002), the cytoplasm of *Ca*. T. namibiensis is a thin peripheral film, thus limiting trafficking time even more. Similarly, the tens of thousands of chromosomes in cells of *Epulopiscium* sp. type B, are located at the cells' periphery, thus interaction between distant chromosomes within one generation is unlikely to take place.

The extreme polyploidy of most giant bacteria led to the discovery of novel phenomena and the suggestion of novel mechanisms for combatting Müller's ratchet. *Achromatium* spp. were the first to be recognized to harbor unprecedented genomic diversity in individual cells (Ionescu et al. 2017). Evidence for intracellular diversity also exists in the genomes of *Thiomargarita nelsonii* (Flood et al. 2016) and *Epulopiscium* sp., first in the form of very fragmented assemblies despite deep sequencing efforts, suggesting the presence of heterogeneity among sequences that impairs their assembly into longer scaffolds, similar to the case of *Achromatium* spp. (Ionescu et al. 2017). At least in the case of *Achromatium* sp. the use of long-read technology did not improve genome assemblies (Ionescu, unpublished data), suggesting that the sequencing approach is likely not the main cause of these observations. Intracellular diversity is further supported by analyses of strain heterogeneity in assemblies of single cells as reflected by divergent alleles of "single-copy" marker genes. In the case of *Epulopiscium* sp., the grounds for this diversity were recently attributed to an extraordinary mechanism to maintain genomic diversity, in which daughter cells take up naked DNA from the entire community of lysed mother cells (Angert 2021). In contrast to these examples, the chromosomes of the newly discovered *Ca.* T. magnifica are nearly identical copies of each other (Volland et al. 2022). Similarly, the chromosomes of the filamentous *Ca*. Marithrix sp. were suggested to be highly similar (Salman-Carvalho et al. 2016). Therefore, it is inevitable to ask, why are some (hyper)polyploid giant bacteria more heterozygous than others. At this point, both the degree of ploidy (number of chromosomes) and variability among the chromosomes, remain mostly unassessed for most giant bacteria.

Polyploid Giant Bacteria: An Open Field

The physiological properties and biogeochemical properties of several individual giant bacteria have been studied in the last century (Schulz-Vogt et al. 2007; Ionescu and Bizic 2019). For others, morphological and, to some extent, phylogenetic descriptions are available. However, no study has addressed giant bacteria as a group or has evaluated whether aside from their size and likely polyploidy, these organisms share other, genomic features. Below, we present selected open questions with regard to the genomic aspects of giant bacteria.

1. The recent discovery of novel giant bacteria from different phyla (Gros 2017; Sylvestre et al. 2022; Volland et al. 2022; Bizic et al. 2023) suggests that there are more out there to be discovered. Nevertheless, most recent biodiversity studies are based on metabarcoding or metagenomic analyses, generating data that do not hold any information on size. Data mining efforts can reveal new habitats (Bizic et al. 2023) or global distribution (Ionescu et al. 2020) of known giant bacteria. Yet, since no marker genes common to all giant bacteria have been recognized so far, it is impossible to identify new taxa of giant *Bacteria* and *Archaea* from sequence data. Thus, as a first step toward improving our understanding of the phylogeny, physiology, and genomics of giant bacteria, biodiversity studies should include field observations followed by microscopic analyses, specifically in areas where giant bacteria have been identified before as free-living bacteria (mangroves, deep sea, caves, etc.) or as symbionts in insects (Iida et al. 2000), bivalves (Margulis and Hinkle 2013), amphibians (Delaporte 1963, 1970), fishes (Montgomery and

Pollak 1988; Miyake et al. 2016), and mammals (Angert 2012).

2. To date, there is a consensus that giant bacteria are polyploid. Yet, this has been validated and quantified in merely a handful of species. Polyploidy, aside from setting the basis for an in-cell experimental genomics laboratory (Mendell et al. 2008; Oliverio and Katz 2014), serves to overcome the barriers to internal trafficking of proteins and the inability to synthesize sufficient enzymes from a single genome to serve the cell.

   Evaluating whether a correlation can be found between cell size and number of chromosomes could provide insights into the effective volume of a translational/transcriptional unit and help identify at which sizes polyploidy becomes an existential necessity. Thus far, data synthesis from *Epulopiscium* sp. (Mendell et al. 2008), *Achromatium* (Ionescu et al. 2017), and *Ca*. T. magnifica (Volland et al. 2022), reveals a ratio of approximately 1 chromosome per 8–10 μm^3, which is at least 10 times lower DNA per volume than reported (Mendell et al. 2008) or can be calculated (Pecoraro et al. 2011) for smaller bacteria. This information may further hint at whether polyploidy in organisms with smaller cells, such as *Synechocystis* sp. (Soppa et al. 2016), serves a similar function, or whether it has alternative purposes as is the case for *Deinococcus radiodurans* (Slade et al. 2009) or *Thermus thermophilus* (Li 2019).
3. Do giant bacteria have giant genomes? In microbiology, it is often the case that the term genome is used as a synonym to chromosome. However, a genome refers to the entire genetic material in an organism (e.g., Moss et al. 2020). Hence, whether giant bacteria have giant genomes should be split into two questions: a) Do they have larger-than-usual chromosomes? b) Do they contain more genetic information than other bacteria. At present, the data are insufficient to reach a definitive conclusion for either of these questions, howsoever one may speculate.

   Complete chromosomal sequences are available for a few giant bacteria only (table 1), like *Ca.* T. stygium (Bizic et al. 2023). For others, the size that can be assessed by the completeness level, and assembly size represents the shortest set of sequences with the highest completeness level. In a bacterium with a single chromosome, or in a bacterium with multiple identical chromosomes, this would be a good proxy for chromosome size. Yet, whether this is the case in giant hyperpolyploid bacteria, we do not know. For example, single cells of *Achromatium* sp. from the same population were predicted to have chromosomes of 4–12 Mb (Ionescu et al. 2017).

   However, using the available information, the chromosome size of giant bacteria falls within the ranges observed for their phyla (Rodríguez-Gijón et al. 2022). For example, *Ca.* T. stygium with a chromosome of 1.8 Mb (Bizic et al. 2023) is well within the range of other Campylobacterota (ca. 1.3–3.5 Mb; Rodríguez-Gijón et al. 2022). *Candidatus* Thiomargarita magnifica with an estimated chromosome size of approximately 12 Mb is within the upper limits of what has been observed for Gammaproteobacteria. Thus, it appears that the chromosome size of giant bacteria follows the overall lineage of their phylum.

   Assessing the entire genomic information in a cell is a much more complex task, given the amount of information that can possibly be stored on hundreds to tens of thousands of chromosomes. While functionally this information may converge to a finite set of functions, the allelic divergence is likely high, at least in some giant bacteria, resulting from continuous recombination events through different mechanisms (Ionescu et al. 2017; Ionescu et al. 2020; Angert 2021). In the genus *Achromatium*, the genomic information does not reflect any functional difference between distinct habitats (i.e., marine and fresh water), and it was proposed that the cells accumulate and store functions, thus using their multiple chromosomes to continuously expand their genome (Ionescu et al. 2020). Clearly, this is not the case for all giant bacteria and is likely linked to heterozygosity.
4. Heterozygosity, defined here as the occurrence of chromosomes with different genomic content within a single cell, is not equal in all giant bacteria. Why a 2-cm cell may contain almost a million nearly identical chromosomes (Volland et al. 2022) and a 50-μm cell contains a few hundred chromosomes with a diversity that exceeds that of a whole genus (Ionescu et al. 2017) is not understood. Heterozygosity or clonality was addressed only in a few giant bacteria (Mendell et al. 2008; Salman-Carvalho et al. 2016; Ionescu et al. 2017; Angert 2021; Volland et al. 2022). In the absence of reliable technology to isolate individual chromosomes from cells, assessment of heterozygosity can be done by inspecting the number of different variants of genes expected to occur in single copies. These results may be biased by adhering environmental DNA from closely related species which would pass bioinformatic filtering and qualify as the same cell. Nevertheless, repeated results from single *Achromatium* cells from which the surrounding layer of extracellular polymeric substances was removed and the cells individually washed (Ionescu et al. 2017, 2020) suggest that the results are not a methodological error. Based on the reported data, a pattern emerges, linking heterozygosity to cell architecture. First, filamentous bacteria are less heterozygous than some unicellular ones. Among the latter, heterozygosity is larger in cells where interaction between

chromosomes is unlikely, for example, *Epulopiscium* type B, and *Achromatium*. In the case of *Thiovulum* sp., the smallest species *Ca.* T. stygium displayed no heterozygosity (Bizic et al. 2023), followed by the larger *Candidatus* Thiovulum karukerense (Gros 2017), and the largest of the three, *Ca.* T. imperiosus (Sylvestre et al. 2022), suggesting that an increase in size and likely number of chromosomes, contributes as well to heterozygosity.

These patterns derived from merely a few cells remain to be validated across more taxa. For example, while the chromosomes of *Epulopiscium* type B are peripheral, those of type A are located in the center of the cell (Angert 2006). Does this translate into enhanced gene conversion in *Epulopiscium* type A and less heterozygosity, or does the same mechanism of DNA uptake from the lysed community determine the heterozygosity in these cells as well?

It further remains to be seen, how and if, the allelic divergence resulting from this heterozygosity is being put to use. Does it serve as an experimental laboratory to allow the cells to "test" and "adopt" new variants, as was suggested for polyploid bacteria (Mendell et al. 2008; Oliverio and Katz 2014)? Or, can the cell use the allelic divergence for fine-tuning their response to environmental changes as shown for a cold-water diatom (Mock et al. 2017)?

5. Polyploidy as a mechanism to overcome the limitation of internal protein and macromolecule trafficking in giant cells suggests that each chromosome may function independently. Can cells of giant bacteria regulate gene expression across hundreds and thousands of chromosomes? If so, how? Is this a function of the substrate gradient? Can different parts of the cells assume different roles, in a behavior paralogue to multicellular organisms? Can one chromosome express enhancers or repressors to control the transcriptional or translational activity in its surroundings? We have not yet begun to address such questions. Expression of different 16S rRNA alleles has been reported for *Achromatium* sp. (Ionescu et al. 2017), yet this likely reflects the distribution of these alleles in the cell. Acquiring further genomic information on giant bacteria will allow the application of tools such as in situ RNA sequencing (Ke et al. 2013) gene FISH and mRNA-FISH (Barrero-Canosa et al. 2017) to start understanding the distribution of activities. This information coupled with single-cell imaging metabolomics (Kompauer et al. 2017), Raman microspectroscopy (Du et al. 2020) and/or MALDI imaging (Feucherolles and Frache 2022), will couple gene expression with metabolic products and their distribution in the cell. While most of these methods do not have the necessary spatial resolution to resolve activity in "standard" bacteria, the large cell size of giant bacteria makes them ideal subjects for such methods.

Giant bacteria have been known for over a century, yet to date only a few have been genomically investigated, and to the best of our knowledge, no comparative genomics study has investigated giant bacteria as a group formed by a common phenotype—atypical sizes. Despite them being *Bacteria* (and potentially also *Archaea*), giant bacteria possess several traits typical of eukaryotes, such as heterozygosity (Ionescu et al. 2017; Angert 2021), cellular compartmentalization (Ionescu et al. 2017; Volland et al. 2022), and the existence of novel membrane-bound organelles (Volland et al. 2022), and may therefore be evolutionarily significant. The combination of their relatively "simpler" bacterial physiology, their relatively small genomes (compared with eukaryotes), and their large cell size, makes giant bacteria unique systems to study microbial cells at the subcellular level, applying tools initially developed for eukaryotic cells. We envision tools like single-cell metabolic imaging, flow-sorting of individual chromosomes (as done for viruses), mRNA-FISH, and in situ transcriptomics, being applied to such giant cells to improve our understanding of their genomic evolution, regulation, and functionality and opening new avenues of research into the evolution of bacterial gigantism.

The study of giant bacteria as free-living or symbionts in different habitats or hosts is of great scientific interest. These peaceful giants, if they can be cultivated one day, could represent unique study models in microbiology, allowing a better understanding of the adaptive capacities of bacteria. This is true both at the molecular level (DNA structure, polyploidy management, cell division, etc.) and at the metabolomic level (production of secondary metabolites with antimicrobial activity, production of specific proteins, etc.). Such advances will require a multidisciplinary approach combining visual observations by field researchers, development and application of recent and novel molecular tools in the laboratory, and subsequently genome-scale metabolic modeling approaches that may divulge the secrets to cultivating these organisms.

Supplementary material

Supplementary data are available at *Genome Biology and Evolution* online (http://www.gbe.oxfordjournals.org/).

Acknowledgments

D.I. is supported by DFG project IO 98/3-1. The work by J.-M.V. (proposal: 10.46936/10.25585/60001074) conducted by the US Department of Energy Joint Genome Institute (https://ror.org/04xm1d337), a Department of Energy (DOE) Office of Science User Facility, is supported

by the Office of Science of the DOE operated under contract no. DE-AC02-05CH11231. P.-E.C. is supported by a grant from the Collectivité Territoriale de Martinique.

Data Availability

There are no new data associated with this article.

Literature Cited

Angert ER. 2006. The enigmatic cytoarchitecture of *Epulopiscium* spp. In: Shively JM, editor. Complex intracellular structures in prokaryotes. Berlin, Heidelberg: Springer. p. 285–301.

Angert ER. 2012. DNA replication and genomic architecture of very large bacteria. Annu Rev Microbiol. 66:197–212.

Angert E. 2021. Challenges faced by highly polyploid bacteria with limits on DNA inheritance. Genome Biol Evol. 13:evab037.

Arroyo FA, Pawlowska TE, Choat JH, Clements KD, Angert ER. 2019. Recombination contributes to population diversification in the polyploid intestinal symbiont *Epulopiscium* sp. type B. ISME J. 13: 1084–1097.

Barrero-Canosa J, Moraru C, Zeugner L, Fuchs BM, Amann R. 2017. Direct-geneFISH: a simplified protocol for the simultaneous detection and quantification of genes and rRNA in microorganisms. Environ Microbiol. 19:70–82.

Bizic M, et al. 2023. Cave Thiovulum (*Candidatus* Thiovulum stygium) differs metabolically and genomically from marine species. ISME J. 17:340–353.

Blakemore RP, Canale-Parola E. 1973. Morphological and ecological characteristics of *Spirochaeta plicatilis*. Arch Mikrobiol. 89: 273–289.

Bohrer CH, Xiao J. 2020. Complex diffusion in bacteria. Adv Exp Med Biol. 1267:15–43.

Borowitzka MA. 2018. Biology of microalgae. In: Levine IA, Fleurence J, editors. Microalgae in health and disease prevention. London: Academic Press. p. 23–72.

Chatton E, Perard C. 1913. Schizophytes du caecum du cobaye. II *Metabacterium polyspora* n. g., n. s. C R Hebd Soc Biol (Paris). 74:1232–1234.

Delaporte B. 1963. Un phénomène singulier: des "spores mobiles" chez des grandes Bactéries. C R Hebd Soc Biol (Paris). 257: 1414–1417.

Delaporte B. 1964. Étude comparee de grands spirilles formant des spores: *Sporospirillum* (*Spirillum*) *praeclarum* (Collin) n. g., *Sporospirillum gyrini* n. sp. et *Sporospirillum bisporum* n. sp. Ann Inst Pasteur (Paris). 107:246–262.

Delaporte B. 1970. Etude de la structure, et plus specialement de l'appareil nucleaire, de tres grandes bacteries sporulantes. Z Allg Mikrobiol. 10:165–182.

Du J, et al. 2020. Raman-guided subcellular pharmaco-metabolomics for metastatic melanoma cells. Nat Commun. 11:4830.

Dwivedi VK, Tandon R, Tiwari GL. 2010. Polymorphic stages of the fresh water blue-green alga, *Gomphosphaeria aponina*. ALGAE 25:115–120.

Feucherolles M, Frache G. 2022. MALDI mass spectrometry imaging: a potential game-changer in a modern microbiology. Cells. 11: 3900.

Flood BE, et al. 2016. Single-cell (meta-)genomics of a dimorphic *Candidatus* Thiomargarita nelsonii reveals genomic plasticity. Front Microbiol. 7:603.

Fomenkov A, et al. 2018. Complete genome sequence of the freshwater bacterium *Beggiatoa leptomitoformis* strain D-401. Genome Announc. 6:e00311–e00318.

Fraser CM, et al. 1997. Genomic sequence of a Lyme disease spirochaete, *Borrelia burgdorferi*. Nature 390:580–586.

Gallet R, et al. 2017. The evolution of bacterial cell size: the internal diffusion-constraint hypothesis. ISME J. 11:1559–1568.

Goodsell DS. 1991. Inside a living cell. Trends Biochem Sci. 16: 203–206.

Greening C, Lithgow T. 2020. Formation and function of bacterial organelles. Nat Rev Microbiol. 18:677–689.

Gros O. 2017. First description of a new uncultured epsilon sulfur bacterium colonizing marine mangrove sediment in the Caribbean: *Thiovulum* sp. strain karukerense. FEMS Microbiol Lett. 364: fnx172.

Guerrero R, Haselton A, Solé M, Wier A, Margulis L. 1999. *Titanospirillum velox*: a huge, speedy, sulfur-storing spirillum from Ebro Delta microbial mats. Proc Natl Acad Sci U S A. 96: 11584–11588.

Han C, et al. 2013. Genome sequence of the thermophilic fresh-water bacterium *Spirochaeta caldaria* type strain (H1T), reclassification of *Spirochaeta caldaria*, *Spirochaeta stenostrepta*, and *Spirochaeta zuelzerae* in the genus *Treponema* as Treponema caldaria comb. nov., Treponema stenostrepta comb. nov., and Treponema zuelzerae comb. nov., and emendation of the genus Treponema. Stand Genomic Sci. 8:88–105.

Harris LK, Theriot JA. 2018. Surface area to volume ratio: a natural variable for bacterial morphogenesis. Trends Microbiol. 26:815–832.

Hauer T, Komarek J. 2022. CyanoDB 2.0: on-line database of cyanobacterial genera.—World-wide electronic publication. Univ. of South Bohemia & Inst. of Botany AS CR. [cited 2023 Jul 3]. Available from: http://www.cyanodb.cz.

Iida T, Ohkuma M, Ohtoko K, Kudo T. 2000. Symbiotic spirochetes in the termite hindgut: phylogenetic identification of ectosymbiotic spirochetes of oxymonad protists. FEMS Microbiol Ecol. 34:17–26.

Ionescu D, et al. 2020. Heterozygous, polyploid, giant bacterium, *Achromatium*, possesses an identical functional inventory worldwide across drastically different ecosystems. Mol Biol Evol. 38: 1040–1059.

Ionescu D, Bizic-Ionescu M, De Maio N, Cypionka H, Grossart H-P. 2017. Community-like genome in single cells of the sulfur bacterium *Achromatium oxaliferum*. Nat Commun. 8:455.

Ionescu D, Bizic M. 2019. Giant bacteria. In: eLS. Chichester: John Wiley & Sons, Ltd. p. 1–10.

Jean MRN, et al. 2015. Two new *Beggiatoa* species inhabiting marine mangrove sediments in the Caribbean. PLoS One. 10:e0117832.

Ke R, et al. 2013. In situ sequencing for RNA analysis in preserved tissue and cells. Nat Methods. 10:857–860.

Keim CN, et al. 2004. Multicellular life cycle of magnetotactic prokaryotes. FEMS Microbiol Lett. 240:203–208.

Kempes CP, Wang L, Amend JP, Doyle J, Hoehler T. 2016. Evolutionary tradeoffs in cellular composition across diverse bacteria. ISME J. 10:2145–2157.

Kim D-H, et al. 2022. *Porphyrosiphon annulatus* sp. nov. (Oscillatoriales, Cyanobacteria) isolated on moist soil in Suwon, Republic of Korea. Phytotaxa 532:288–300.

Kojima H, et al. 2015. Ecophysiology of *Thioploca ingrica* as revealed by the complete genome sequence supplemented with proteomic evidence. ISME J. 9:1166–1176.

Komárek J, Johansen JR. 2015. Filamentous Cyanobacteria. In: Wehr JD, Sheath RG, Kociolek JP, editors. Freshwater algae of North America. 2nd ed. Boston (MA): Academic Press. p. 135–235.

Kompauer M, Heiles S, Spengler B. 2017. Atmospheric pressure MALDI mass spectrometry imaging of tissues and cells at 1.4-μm lateral resolution. Nat Methods. 14:90–96.

Levin PA, Angert ER. 2016. Small but mighty: cell size and bacteria. Cold Spring Harb Perspect Biol. 7(7):a019216.

Li H. 2019. Random chromosome partitioning in the polyploid bacterium *Thermus thermophilus* HB27. G3 (Bethesda) 9:1249–1261.
Locey KJ, Lennon JT. 2016. Scaling laws predict global microbial diversity. Proc Natl Acad Sci U S A. 113:5970–5975.
Luedin SM, et al. 2019. Draft genome sequence of *Chromatium okenii* isolated from the stratified alpine Lake Cadagno. Sci Rep. 9:1936.
Lynch M, Trickovic B, Kempes CP. 2022. Evolutionary scaling of maximum growth rate with organism size. Sci Rep. 12:22586.
Margulis L, Hinkle G. 2013. Large symbiotic spirochetes: Clevelandina, Cristispira, Diplocalyx, Hollandina, and Pillotina. In: Dworkin M, Falkow S, Rosenberg E, Schleifer KH, Stackebrandt E, editors. The prokaryotes. New York (NY): Springer. p. 971–982.
Markov AV, Kaznacheev IS. 2016. Evolutionary consequences of polyploidy in prokaryotes and the origin of mitosis and meiosis. Biol Direct. 11:28.
Marter P, et al. 2021. Filling the gaps in the Cyanobacterial tree of life—metagenome analysis of *Stigonema ocellatum* DSM 106950, *Chlorogloea purpurea* SAG 13.99 and *Gomphosphaeria aponina* DSM 107014. Genes (Basel). 12:389.
Mendell JE, Clements KD, Choat JH, Angert ER. 2008. Extreme polyploidy in a large bacterium. Proc Natl Acad Sci U S A. 105: 6730–6734.
Mendler K, et al. 2019. Annotree: visualization and exploration of a functionally annotated microbial tree of life. Nucleic Acids Res. 47:4442–4448.
Miyake S, Ngugi DK, Stingl U. 2016. Phylogenetic diversity, distribution, and cophylogeny of giant bacteria (*Epulopiscium*) with their surgeonfish hosts in the Red Sea. Front Microbiol. 7:285.
Mock T, et al. 2017. Evolutionary genomics of the cold-adapted diatom *Fragilariopsis cylindrus*. Nature 541:536–540.
Montgomery LW, Pollak PE. 1988. Epulopiscium fishelsoni N. G., N. Sp., a protist of uncertain taxonomic affinities from the gut of an herbivorous reef fish. J Protozool. 35:565–569.
Moran NA, Tran P, Gerardo NM. 2005. Symbiosis and insect diversification: an ancient symbiont of sap-feeding insects from the bacterial phylum *Bacteroidetes*. Appl Environ Microbiol. 71:8802–8810.
Moss EL, Maghini DG, Bhatt AS. 2020. Complete, closed bacterial genomes from microbiomes using nanopore sequencing. Nat Biotechnol. 38:701–707.
Oliverio AM, Katz LA. 2014. The dynamic nature of genomes across the tree of life. Genome Biol Evol. 6:482–488.
Pecoraro V, Zerulla K, Lange C, Soppa J. 2011. Quantification of ploidy in proteobacteria revealed the existence of monoploid, (mero-)oligoploid and polyploid species. PLoS One. 6:e16392.
Pende N, et al. 2014. Size-independent symmetric division in extraordinarily long cells. Nat Commun. 5:4803.
Pfeffer C, et al. 2012. Filamentous bacteria transport electrons over centimetre distances. Nature 491:218–221.
Ravin NV, et al. 2022. History of the study of the genus *Thiothrix*: from the first enrichment cultures to pangenomic analysis. Int J Mol Sci. 23:9531.
Rodríguez-Gijón A, et al. 2022. A genomic perspective across earth's microbiomes reveals that genome size in *Archaea* and *Bacteria* is linked to ecosystem type and trophic strategy. Front Microbiol. 12:761869.
Salman-Carvalho V, Fadeev E, Joye SB, Teske A. 2016. How clonal is clonal? Genome plasticity across multicellular segments of a '*Candidatus* Marithrix sp.' filament from sulfidic, briny seafloor sediments in the Gulf of Mexico. Front Microbiol. 7:1173.
Salman V, et al. 2011. A single-cell sequencing approach to the classification of large, vacuolated sulfur bacteria. Syst Appl Microbiol. 34:243–259.
Salman V, et al. 2016. Insights into the single cell draft genome of 'Candidatus Achromatium palustre'. Stand Genomic Sci. 11:28.
Schewiakoff W. 1893. Über einen neuen bacterienähnlichen Organismus des Süsswassers [habilitation dissertation]. [Heidelberg]: University of Heidelberg.
Schorn S, et al. 2020. Cell architecture of the giant sulfur bacterium *Achromatium oxaliferum*: extra-cytoplasmic localization of calcium carbonate bodies. FEMS Microbiol Ecol. 96:fiz200.
Schulz-Vogt HN, Angert ER, Garcia-Pichel F. 2007. Giant bacteria. In: eLS. Chichester: John Wiley & Sons, Ltd.
Schulz HN. 2006. The genus Thiomargarita. In: Dworkin M, Falkow S, Rosenberg E, Schleifer KH, Stackebrandt E. editors. The Prokaryotes. New York (NY): Springer. p. 1156–1163.
Schulz HN, Jørgensen BB. 2002. Big bacteria. Annu Rev Microbiol. 55: 105–137.
Scott M, Gunderson CW, Mateescu EM, Zhang Z, Hwa T. 2010. Interdependence of cell growth and gene expression: origins and consequences. Science 330:1099–1102.
Slade D, Lindner AB, Paul G, Radman M. 2009. Recombination and replication in DNA repair of heavily irradiated *Deinococcus radiodurans*. Cell 136:1044–1055.
Soppa J. 2011. Ploidy and gene conversion in archaea. Biochem Soc Trans. 39:150–154.
Soppa J, Ludt K, Zerulla K. 2016. The ploidy level of *Synechocystis* sp. PCC 6803 is highly variable and is influenced by growth phase and by chemical and physical external parameters. Microbiology (Reading) 162:730–739.
Steen AD, et al. 2019. High proportions of bacteria and archaea across most biomes remain uncultured. ISME J. 13:3126–3130.
Sylvestre M-N, et al. 2022. *Candidatus* Thiovulum sp. strain imperiosus: the largest free-living Epsilonproteobacteraeota Thiovulum strain lives in a marine mangrove environment. Can J Microbiol. 68:17–30.
Takeuchi N, Kaneko K, Koonin EV. 2014. Horizontal gene transfer can rescue prokaryotes from Muller's Ratchet: benefit of DNA from dead cells and population subdivision. G3 (Bethesda) 4:325–339.
Takeyasu K, et al. 2004. Genome architecture studied by nanoscale imaging: analyses among bacterial phyla and their implication to eukaryotic genome folding. Cytogenet Genome Res. 107:38–48.
Teske A, Salman V. 2014. The family Beggiatoaceae. In: Rosenberg E, DeLong EF, Lory S, Stackebrandt E, Thompson F. editors. The prokaryotes: Gammaproteobacteria. Vol. 9783642389. Berlin: Springer. p. 93–134.
Trevisan VBA. 1842. PROSPETTO DELLA FLORA EUGANEA. Alla Sezione di Botanica e Fisiologia vegetabile della Quarta Riunione degli Scienziati Italiani. Padova: Coi Tipi del Seminario.
Vaucher J-P. 1803. Histoire des conferves d'eau douce, contenant leurs différents modes de reproduction, et la description de leurs principales espèces. Genève: J.J. Paschoud.
Volland J-M, et al. 2022. A centimeter-long bacterium with DNA contained in metabolically active, membrane-bound organelles. Science 376:1453–1458.
Ward RD, et al. 2021. Metagenome sequencing to explore phylogenomics of terrestrial Cyanobacteria. Microbiol Resour Announc. 10:e0025821.
Wood SA, et al. 2017. Phylogenetic characterisation of marine *Chroococcus*-like (Cyanobacteria) strains from the Pacific region. N Z J Bot. 55:5–13.
Yang Y, et al. 2021. Long-distance electron transfer in a filamentous Gram-positive bacterium. Nat Commun. 12:1709.

Associate editor: Rebecca Zufall