Non-equivalent genomes in polyploid prokaryotes

Many bacteria and archaea are polyploid. Here, the means by which some of these prokaryotes carry genomes that are not always equivalent in sequence and/or function are described, and the importance of such non-equivalent genomes is discussed.

Jörg Soppa

For decades it has been believed that bacteria and archaea contain a single copy of their circular chromosome and thus that they are monoploid (often erroneously called haploid). Several polyploid species, such as *Deinococcus radiodurans, Azotobacter vinelandii* and certain cyanobacteria, were known but were thought to be rare exceptions. However, during the last decade, it has become clear that many prokaryotic species are polyploid. This is true for species of various phylogenetic groups of prokaryotes, including proteobacteria, cyanobacteria, Deinococcales, low-guanine–cytosine (GC) Gram-positive bacteria, halophilic archaea and methanogenic archaea[1]. In addition, all giant bacteria appear to be polyploid, irrespective of their phylogenetic position[2]. In fact, it may well be that the majority of prokaryotes are oligo- or polyploid and that monoploid species are exceptional, contrary to common belief.

Polyploidy offers a variety of putative evolutionary advantages as compared to monoploidy[1]. For example, two polyploid species of haloarchaea and cyanobacteria have been shown to sustain growth in the absence of an external phosphate source, which would be impossible for monoploid species lacking other storage polymers, such as polyphosphate. Additional examples include resistance to conditions that induce double-strand breaks, survival through geological times and survival under simulated martian conditions[1,3]. It seems that polyploidy has evolved at different times and for different reasons in many species of various phylogenetic groups.

Currently, it is assumed that all genome copies of polyploid prokaryotes are identical in sequence and function. One indication for sequence identity is that genome sequencing projects yielded unequivocal sequences for polyploid species. In addition, for halophilic and methanogenic archaea, it was shown that intermolecular gene conversion operates and leads to the equalization of genome copies[4,5]. In contrast to the sequence equivalence, the functional equivalence of all genome copies has just been tacitly assumed. In this Comment, a few counterexamples that exemplify the lack of sequence identity or/and the absence of functional equivalence will be discussed.

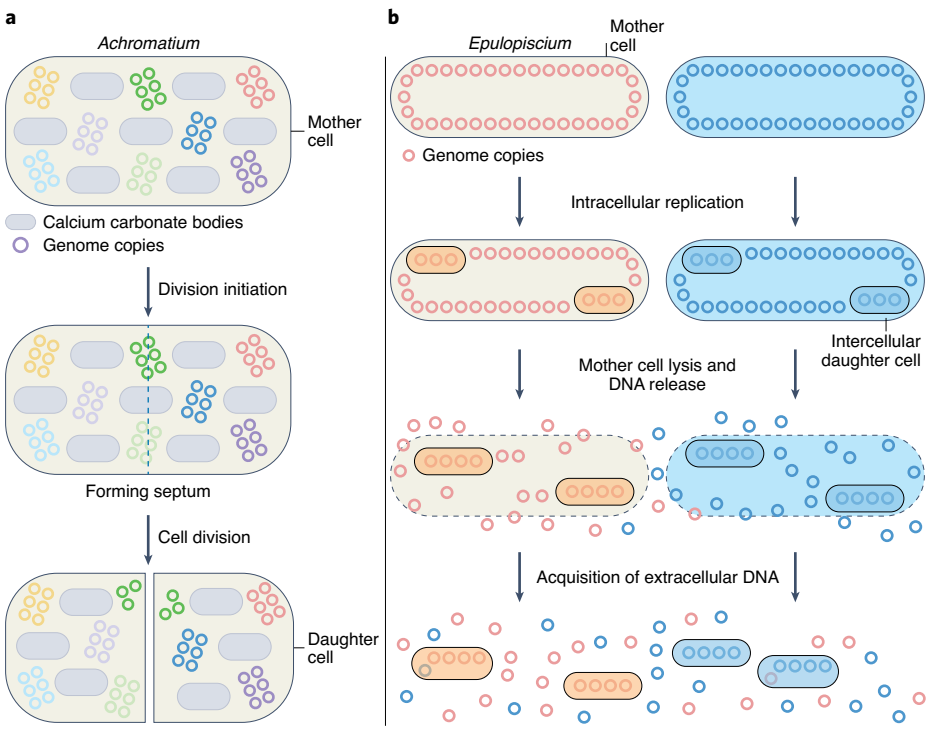

Fig. 1 | Schematic overview of non-equivalent genomes in *Azobacter ocaliferum* and in *Epulopiscium*. a, A schematic overview of cell division for the giant bacterium *A. oxaliferum*, showing that the daughter cells are genetically not identical. Genomes of different sequences are indicated by circles of different colours. **b**, A schematic overview of the formation of living offspring within the mother cell (viviparity) in *Epulopiscium*. Two cells with non-identical genome sequences are shown. The genomes are indicated by pink and blue circles, respectively. Only very few germline genomes are inherited by the daughter cells, while the majority of somatic genomes are degraded or released when the mother cell lyses. The daughter cells take up DNA from the environment, leading to genetic variability.

The first example is the as yet uncultured species *Achromatium oxaliferum*. *Achromatium* is a very large (>100 µm) sulfur bacterium characterized by the accumulation of calcium carbonate bodies. It was recently shown that a membrane network stretches throughout the cells such that the calcium carbonate bodies reside in the periplasm[6]. This intracellular membrane network results in a drastic reduction of the cytoplasmic volume, which is confined to a thin layer underneath the cell wall and to 'pockets' throughout the cell interior. Staining with DNA dyes resulted in the visualization of many DNA spots, showing

Box 1 | Heterozygous strains generated via laboratory selection

Laboratory selection has led to the isolation of heterozygous strains in species of various archaeal and bacterial groups, including *Azotobacter*, *Zymomonas*, different species of cyanobacteria and streptomyces, and halophilic and methanogenic archaea. Attempts to delete essential genes often result in heterozygous cells, as the presence of the desired genome is enforced by laboratory selection (for example, the addition of an antibiotic and use of a resistance gene as a selection marker); while essential genes, by definition, cannot be totally removed from the cell. In polyploid species, even when deletions of non-essential genes are eventually successful, many generations under selection conditions are typically required before the intermediate heterozygous strains become homozygous for the desired deletion copy. The ease of generating heterozygous strains via laboratory selection may indicate that the formation of heterozygous cells also occurs in nature under specific conditions. Gene conversion results in the equalization of genome copies, and the formation of homozygous cells from heterozygous intermediates in the absence of selection has been reported for various species. Recently, this was exploited to establish an efficient method for genome design in *Synechococcus sp*. PCC 7002, which includes intermediate heterozygosity of genes for DNA-modifying enzymes and essential genes. The term 'heterozygous' is used (although prokaryotes do not have a zygote), as well as the term 'merodiploid'; thus, in literature searches, both terms should be used together with the prokaryotic groups of interest (for example, 'cyanobact*' or 'archae*').

that *Achromatium* is highly polyploid. Surprisingly, it has been shown that the genomes are not identical throughout the cell but rather that *Achromatium* harbours various different genomes[7]. This unexpected result was confirmed by single-cell genomics, metagenomics and single-cell amplicon sequencing, and directly by fluorescence in situ hybridization (FISH) with genome-specific probes[7]. These results showed that the various cytoplasmatic pockets are more or less isolated genetically so that they can harbour different genomes (schematically shown in Fig. 1a). The (dis-)similarities between various proteins encoded by different genomes within one cell differ widely, for example, 75% of the proteins have Dayhoff distances exceeding 1 (ref. [7]). In fact, it has been argued that each *Achromatium* cell contains a degree of genetic diversity that is otherwise only found between species of the same genus[7]. Cell division of an *A. oxaliferum* cell with non-identical genomes leads to two daughter cells that are not genetically identical (Fig. 1a). A recent study showed that *Achromatium* is widespread around the world in ecosystems that differ in pH, temperature, salinity and depth[8]. Currently, *Achromatium* is the only prokaryotic species that is known to be natively heterozygous; however, a considerable number of heterozygous species have been generated via laboratory selection (see Box 1).

The second example is *Epulopiscium* sp. type B — another giant bacterium that is approximately 100–300 µm in length[2]. *Epulopiscium* is an endosymbiont that lives in the digestive tract of the omnivorous surgeonfish *Naso tonganus*. As yet, no axenic cultures of *Epulopiscium* could be established. A recent study used FISH to visualize replication origins[9]. Between 600 and 9,400 origins per cell were detected, and there was a good correlation between the numbers of origins and cell size. The genomes were not evenly distributed throughout the cell but were regularly spaced beneath the cytoplasmic membrane (shown schematically in Fig. 1b).

A unique feature of *Epulopiscium* is its reproduction mechanism[2]. The offspring is formed intracellularly. Two daughter cells grow within the mother cell (Fig. 1b) until the mother cell eventually decomposes and the offspring is released. Thus, *Epulopiscium* reproduces via viviparity. Only a minor fraction of genomes (about 1%) is inherited by the next generation, while the vast majority are confined to the mother cell. *Epulopiscium* therefore has 'somatic' and 'germline' genomes, which are clearly non-equivalent in fate and function. During early stages of intracellular growth of the daughter cells, the somatic genomes in the mother cell cytoplasm still replicate; later, they are degraded or released to the environment (Fig. 1b).

A recent study revealed that the *Epulopiscium* cells within a single fish are not uniform but instead exhibit considerable genetic diversity, indicating that bacteria are horizontally transmitted between fish[10]. The nucleotide diversity in six analysed housekeeping genes was 0.00109–0.00543. A model for genetic exchange between different *Epulopiscium* cells has been put forward, which rests on the diurnally synchronized life cycle of *Epulopiscium*[10]. According to this model, daughter cells take up extracellular DNA that originates from various decayed mother cells with different genomes, generating genetically non-identical offspring (Fig. 1b).

A second prokaryotic group that reproduces via the generation of intracellular offspring is segmented filamentous bacteria (SFB), which are found in the digestive tract of many vertebrates, including mice and, probably, humans[11]. Reproduction starts with an asymmetric division, and the resulting small (daughter) cell is then engulfed by the large (mother) cell. The daughter cell divides within the mother cell so that two living offspring cells are generated per mother cell. Finally, the offspring cells are released while the mother cell decays. In this case, only the genomes of the offspring cells are passed on to the next generation. Both *Epulopiscium* and SFBs are related to Firmicutes; therefore, it seems that both forms of bacterial viviparity evolved from the process of endospore formation.

While the examples discussed above concern giant bacteria, which may be regarded as exceptional, the following examples represent normal-sized bacteria and archaea and concern functional non-equivalence. *Synechococcus elongatus* PCC 7942 is an oligoploid bacterium with about 3–6 genome copies. An earlier study had already indicated that not all copies are replicated synchronously, but only one copy is replicated at any given time[12]. In a recent study, the replication and transcription activities of the genomes were visualized using green fluorescent protein (GFP) fused to the single-stranded binding protein (SSB) and to the β-subunit of RNA polymerase (RpoC2)[13]. It was revealed that all genome copies were actively being transcribed, while, in stark contrast, only a single copy was replicated. The small fraction of replicating chromosomes was attributed to the rather high ATPase activity of DnaA, and this was proven by the construction of a DnaA point mutant[13].

The non-equivalence of genomes with respect to replication was shown not only for the oligoploid *S. elongatus* PCC 7942 but also for the polyploid cyanobacterium *Synechocystis* PCC 6803, which contains 10–20 genome copies. In this species, all copies of the chromosome were active in transcription, which was visualized using a RpoC2–GFP fusion[13]. However, also in this species, only a single copy was replicated, which was shown using a SSB–GFP fusion and through the incorporation of bromodeoxyuridine

Box 2 | Mero-oligoploid species with intragenomic gene dosage differences

Under optimal conditions, many species, including the Gram-negative model species *E. coli* and the Gram-positive model species *Bacillus subtilis*, have a doubling time that is shorter than the time needed to replicate and segregate the chromosome. Under these conditions, a new round of replication is initiated before the previous round has been terminated, and two or three rounds of replication thus become intertwined. As a consequence, fast-growing *E. coli* cells contain 8 copies of the genes near the origin, but only 1–2 copies of the genes near the terminus. This replication-associated differential gene dosage effect has led to an enrichment of highly expressed genes near the replication origin. An above-average concentration of highly expressed genes near replication origins has even been observed for the archaeon *Sulfolobus solfataricus*, which contains three replication origins on its chromosome. Although, strictly speaking, 'mero-oligoploid' is the correct term, most publications have used the term 'merodiploid' for this phenomenon, and both terms should therefore be used for literature searches.

(BrdU)[13]. The combination of a single replicating chromosome and a total of 10–20 chromosomes represents evidence that *S.* PCC 6803 does not have a short S phase but synthesizes genomes asynchronously during either a major fraction of or the whole cell cycle. Therefore, at any given time in the cell cycle, the chromosomes are equivalent with transcription but non-equivalent with replication.

The same phenomenon has also been observed in a totally different phylogenetic group of prokaryotes, the halophilic archaea. Cultures of *Halobacterium salinarum* can be synchronized with high fidelity, and they have been used by my research group to reveal cell-cycle-dependent variation in numerous processes, including differential cyclic expression of several groups of genes, short-time increase of the cyclic adenosine monophosphate (cAMP) concentration and intracellular transport of genomic DNA. Very surprisingly, BrdU incorporation in synchronized cultures showed that replication is constitutive throughout the whole cell cycle and occurs with identical efficiency before, during and after cell division[14]. The doubling time of *H. salinarum* is about four hours; therefore, the constitutive replication throughout the whole cell cycle represents direct evidence for the absence of an S phase and indirect evidence that not all chromosomes are replicated simultaneously, but rather that only a small fraction of chromosomes is replicated at any given time. Direct evidence for this interpretation has been obtained with another halophilic archaeon, *Haloferax volcanii*. As with the two cyanobacteria discussed prior, replication sites were visualized using a SSB–GFP fusion[15]. It was revealed that only 1–3 chromosomes replicate at any given time during exponential growth, while *H. volcanii* contains a total of appoximately 25 chromosomes. The result that only about 10% of the chromosomes replicate simultaneously was corroborated by visualizing replication sites via the incorporation of BrdU[15].

Until now, the characterization of species representing four genera of cyanobacteria and halophilic archaea revealed that replication of their multiple chromosomes does not occur simultaneously in an S phase, but rather that replication of only a small fraction of chromosomes occurs at any given time and thus that replication of all chromosomes happens continuously throughout the cell cycle. This is in stark contrast to the cell cycle of slow-growing monoploid species with doubling times that exceed the replication and segregation times, which is characterized by a short S phase and long G1 and G2 phases that are devoid of DNA synthesis. The observed asynchronous continuous DNA synthesis offers several putative advantages; thus, I predict that it will turn out to be typical and widespread in polyploid prokaryotes. The first advantage is that only a small fraction of the genome biosynthesis capacity is needed when the synthesis is not concentrated on a short S phase. This does not only concern DNA polymerase, SSB, helicase, topoisomerase and all other proteins related to the replication fork, but also all enzymes involved in the biosynthesis of purines, pyrimidines, deoxyribose and deoxynucleoside triphosphate (dNTP) substrates for replication. The second putative advantage is that every new chromosome is generated in the presence of an abundant number of old chromosomes. This greatly enhances the chance that replication errors are directly overwritten by native sequences via gene conversion. The third putative advantage of continuous replication is that there is a strict correlation between the generation of new genomes and the mass increase of the cell. Therefore, the gene dosage per cell volume remains constant throughout the cell cycle, which could be advantageous for the regulation of gene expression to guarantee a stable cellular composition. However, it should be noted that so-called mero-oligoploid species make ample use of just the opposite — that is, an uneven gene dosage around the chromosome (see Box 2).

In summary, this Comment predicts another change of concept in microbial genetics owing to the discovery that the majority of prokaryotes are oligoploid or polyploid. In recent years, evidence has accumulated to indicate that the genomes in polyploid prokaryotes are not always equivalent in sequence and/or function, and the examples discussed may well represent only the tip of the iceberg. Differential transcription of genome copies at different cellular sites is an additional non-equivalent feature that has been predicted to occur in polyploid prokaryotes, especially in giant bacteria, but experimental evidence for this is currently lacking. At present, only a few examples of naturally heterozygous cells — which have somatic and germline genomes,and asynchronous, constitutive replication — are known in prokaryotes. I expect that additional studies will identify further examples and will unravel how widespread these phenomena are. ❐

Jörg Soppa ✉
Biocentre, Institute for Molecular Biosciences, Goethe-University, Frankfurt, Germany.
✉*e-mail: soppa@bio.uni-frankfurt.de*

Published online: 23 December 2021
https://doi.org/10.1038/s41564-021-01034-3

References

1. Soppa, J. *J. Molec. Microbiol. Biotech.* **24**, 409–419 (2014).
2. Angert, E. R. *Genome Biol. Evol.* **13**, evab037 (2021).
3. Ludt, K. & Soppa, J. *Biochem. Soc. Trans.* **47**, 933–944 (2019).
4. Wasser, D. et al. *Front. Microbiol.* **12**, 680854 (2021).
5. Hildenbrand, C. et al. *J. Bacteriol.* **193**, 734–743 (2011).
6. Schorn, S. et al. *FEMS Microbiol. Ecol.* **96**, fiz200 (2020).
7. Ionescu, D. et al. *Nat. Commun.* **8**, 455 (2017).
8. Ionescu, D. et al. *Molec. Biol. Evol.* **38**, 1040–1059 (2021).
9. Hutchison, E. et al. *Molec. Microb.* **107**, 68–80 (2018).
10. Arroyo, F. A. et al. *ISME J.* **13**, 1084–1097 (2019).
11. Schnupf, P. et al. *Nature* **520**, 99–103 (2015).
12. Chen, A. H. et al. *PLoS ONE* **7**, e47837 (2012).
13. Ohbayashi, R. et al. *Mbio* https://doi.org/10.1128/mBio.00510-19 (2019).
14. Zerulla, K., Baumann, A. & Soppa, J. in *Halophiles: Genetics and Genomes* (eds Papke, T. & Oren, A.) 145–166 (Caister Academic Press, 2014).
15. Delpech, F. et al. *Nucleic Acids Res.* **46**, 10757–10770 (2018).

Acknowledgements

I thank the German Research Council (DFG) for funding the ploidy research of my group through projects So264/16 and So264/24.

Competing interests

The author declares no competing interests.