Letters in Applied Microbiology

Letters in Applied Microbiology ISSN 0266-8254

Editor's Choice

ORIGINAL ARTICLE

How industrial bacterial cultures can be kept stable over time

T.M. Wassenaar[1] and K. Zimmermann[2]

1 Molecular Microbiology and Genomics Consultants, Zotzenheim, Germany
2 Symbiopharm GmbH, Herborn, Germany

Significance and Impact of Study: The questions addressed here are relevant for industries that depend on live bacteria for (manufacturing of) their products, as they have to guard their bacterial cultures that remain unchanged over time. The explanation why randomly selection of single colonies keeps a population stable can be of use in bacteriology courses. The limitations of whole-genome sequencing are relevant to legislators to avoid overinterpretation of those data.

Keywords
bacterial stocks, clonal propagation, *Escherichia coli*, genomic stability, industrial application, mutant frequency, mutation rate.

Correspondence
Trudy M. Wassenaar, Molecular Microbiology and Genomics Consultants, Tannenstrasse 7, 55576 Zotzenheim, Germany.
E-mail: trudy@mmgc.eu

2020/0527: received 23 March 2020, revised 28 April 2020 and accepted 29 April 2020

doi:10.1111/lam.13309

Abstract

The tremendous variation that exists between bacterial species illustrates the power of evolution, which is the continuous process of mutation and selection over time. Even within a bacterial species, individual members can harbour an impressive degree of genetic variation, depending on the species. The question then arises how similar the offspring of a given bacterial cell over time is, and how long it takes before differences are noticeable? Here we show that on the one hand one can expect random mutations to arise, as a result of various mechanisms. On the other hand, there are forces at play that keep the offspring of a cell genetically relatively constant, unless there is selection for a particular characteristic. The most common mechanisms behind mutations that can appear in a bacterial population are briefly introduced. Next, it is explained why nevertheless such mutations are rarely observed, as long as single colonies are randomly selected, unless selective pressures apply. Since quality control of industrial bacterial cultures is likely to depend heavily on genome sequencing in the near future, the accuracy of whole-genomic sequencing technologies is also discussed. It can be concluded that the bacteriologists who started picking single colonies from agar plates more than hundred years ago were unknowingly ingeneous, as their practice maintains a bacterial culture stable over time.

How do bacterial genomes mutate?

Multiple industries depend on large-scale microbial cultures that either end up in their product, for example probiotics (El Enshasy *et al.*, 2016; Wassenaar 2016; de Sousa *et al.* 2019), or are required for fermentation of products (as in the dairy or meat processing industry). In addition, an increasing number of biotechnological products, such as recombinant antibodies, enzymes or hormones depend on large-scale bacterial or yeast cultures (Nandy and Srivastava 2018; Tripathi and Shrivastava 2019). Such industries typically rely on stored bacterial stocks (although yeasts are not explicitly mentioned in this review, the text applies to those as well), from which the required large-scale cultures as produced. How can these industries be certain that their bacterial cultures stay stable over time? These days, the most definite answer would be to demonstrate stability of a whole genome sequence. Generating whole genome sequences has been suggested as a means for quality control during production of probiotics (Jackson *et al.* 2019), whereby the problem of genetic drift is noted by these authors. Indeed, the sequence of a bacterial cell's genome is only a snapshot in evolutionary time. Over a sufficiently long period of time,

the genome of the offspring of a bacterial cell will accumulate changes, which we call mutations. Changes in a bacterial genome can result from various mechanisms, even if that bacterial population would grow in complete isolation, so that foreign DNA is absent. Generally speaking, mutations can then arise from (i) single-nucleotide changes, (ii) duplications and deletions, and (iii) intra-genomic recombination, with examples shown in Fig. 1.

Single-nucleotide changes can arise from mistakes during replication (amongst other causes), provided such mistakes or damage is not corrected by the mismatch repair machinery of the cell. When such mutations arise within protein-coding genes they may affect protein sequences and the function of those proteins. That only applies when a mutated nucleotide changes the coded amino acid (resulting in a 'missense' or 'non-synonymous mutation'), but due to the highly redundant genetic code this is not always the case. When a mutation does not affect the amino acid sequence within a gene it is called 'silent' or 'synonymous' (the term 'nonsense' mutation is used for cases where a premature stop codon is introduced). Even when a single amino acid of a given protein is changed, this does not always result in changes in protein function. Synonymous mutations can still affect protein concentrations in the cell, since they change the codon used for translation of that amino acid (though they don't the amino acid itself) which may affect the translation efficiency. In addition, mutations in non-protein coding promoter regions may change transcription efficiency, which can also affect protein concentrations and may even have downstream effects, in case the change occurs in a regulatory protein.

Single-nucleotide mutations not only arise from nonrepaired mistakes produced during DNA replication, but also by mutagenic activity of external stimuli, such as UV-light, radioactive radiation or the presence of mutagenic substances. For a more extensive description of mutations the reader is referred to a recent review (Schroeder *et al*. 2018). In the absence of external mutagenic stimuli, single-nucleotide mutations are biased towards changes creating an AT basepair from a GC basepair (these are called transitions), which is caused by the chemical properties of DNA (Hershberg and Petrov 2010). This is not to say that all bacterial genomes become more AT rich over time, as other forces apply that restore the effect of the chemical mutational bias. In fact, bacterial genomes can vary in their GC content from < 15% to over 75%, depending on the species.

Intra-genomic duplications and deletions can arise from mistakes during DNA replication when the DNA polymerase machinery moves along the DNA template, or from erroneous repair of DNA strand breaks. Intra-genomic recombinations without insertions or duplications can change the order of particular sequences. Any of these events can affect protein sequences, when they introduce new combinations of coding sequences or affect the reading frames of genes. Duplicative insertions or short deletions can insert or delete a number of amino acids, as well as affecting the reading frame. Naturally, changes in reading frame, including premature stop codons, can have serious consequences to protein function. Furthermore, when the function of a protein is changed that interacts with, or regulates the activity of other proteins, a small change can have severe downstream effects.

When foreign DNA is present (for instance derived from bacteriophage DNA, incoming plasmids bearing mobile DNA elements, or DNA from other bacterial species), insertions introducing novel DNA sequences can be added as a fourth mechanism that can introduce mutations in a bacterial genome.

In the context of industrial large-scale production of a bacterial culture, the presence of foreign DNA is excluded, as utmost care is taken to avoid any contamination. Likewise, the activity of external mutagenic stimuli is considered to be absent. However, recombinations can also occur from incoming DNA that was derived from

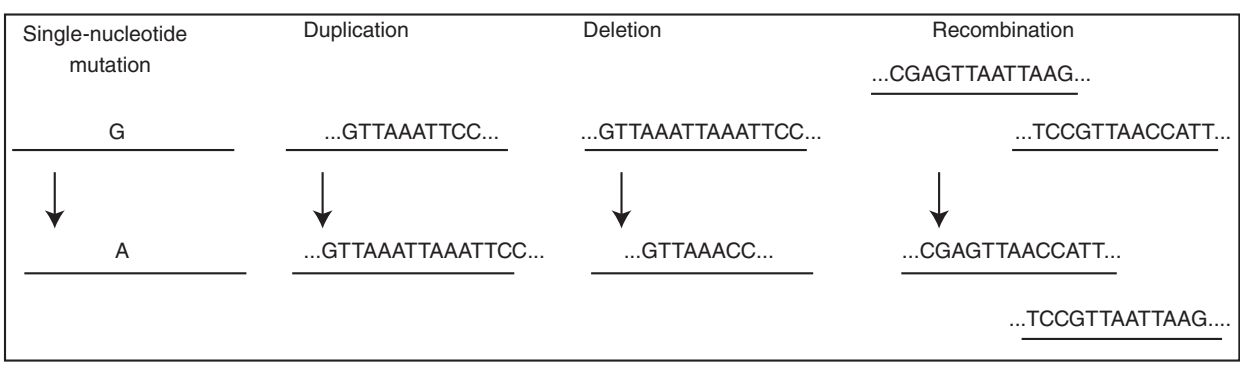


Figure 1 Examples of mutational events with (from left to right) a single-nucleotide mutation, a duplication, a deletion and a recombination.

cells of the same species. This can be considered to be bacterial 'sexual recombination' and occurs more or less frequently, depending on the species. As long as the incoming DNA exactly replaces the original sequence no changes will be introduced, however, that may not always be the case, as recombinations can result in duplications or deletions.

If bacterial genomes constantly mutate, why do we see so few changes over time for a given strain?

Mutations occur with a given frequency as single cells of a given species divide. Observed mutation rates are usually lower than the errors of that bacterial DNA polymerase makes during replication, since most of those errors are corrected by the cell's mismatch repair, whose identity varies between species (Lenhart *et al*. 2016). Such differences can result in differences in natural mutation rates between species. Nevertheless, in an early assessment a more or less constant mutation rate of 0·034 per genome per replication for 'DNA-based microbes' was calculated (Drake 1991). That calculation was based on an average from three locus-dependent *E. coli* observations, combined with data from *Saccharomyces cerevisiae*, *Neurospora crassa* and three *E. coli* bacteriophages (Drake 1991). Since the author expressed the average of mutations per genome, the variation in the genome size of these organisms was already taken into account.

Bacterial genomes can vary in size between < 0·5 megabasepairs (Mb) to over 15 Mb, but the smallest genomes are reserved for intracellular obligate bacteria that are not used in industrial applications. The shortest genomes of species relevant here would be those of Firmicutes such as *Lactobacillus* or *Enterococcus* species (around 1·5 Mb). The longest genomes are typically found in soil bacteria, some of which are relevant to antibiotic production, for instance *Streptomyces* species (12-15 Mb). The difference in genome length between the industrially important bacteria is roughly a factor of 10. However, the observed mutation rate of various bacterial species differs more than that. Table 1 summarizes data from the literature for a limited number of industrially relevant bacteria and fungi. A more extensive comparison across a wide variety of species was performed by Krašovec *et al*. (2017), who observed that the mutation frequency generally decreases as the population density of a given species increases. The spread they observed in published data for *E. coli* and *S. cerevisiae* is extensive (Table 1), due to inclusion of locus-dependent observations and data obtained from super-mutating strains in which mismatch repair systems are mutated (Krašovec *et al*. 2017). Even when these are ignored, the original assessment by Drake proved to be incorrect: the mutation frequency per nucleotide and per generation varies at least a factor of 100 between bacterial species of relevance here. When expressed as frequency per generation and per genome it varies by at least a factor of 200 between species (Table 1).

For *E. coli*, the natural mutation frequency (cultured without external influences) has been experimentally determined as one in every 10^{10} nucleotides (Lenski 2017). With a genome of approximately 5×10^6 basepairs (5 Mb), a population consisting of 10^6 cells (approximately a single colony) will be the result of 5×10^{12} nucleotides that have been joined together to form DNA, and that population will on average contain the result of 500 novel mutational events (Fig. 2). Without mutations, every dividing cell produces two identical offspring cells. A mutation arising early in a growing population will be passed on to all subsequent offspring, therefore building a larger proportion in the final population compared to a mutation that arises in a later incidence of cell division. Mutations that were lethal or deleterious will have been removed from the population, as those cells could no longer reproduce.

However, these mutations in the genomes of individual cells within a colony will not normally be detectable. Only when natural selection favours a mutated cell and its progeny over the original (non-mutated) population, will the mutation be fixed in the offspring and becomes detectable (Fig. 3). When the mutation is neutral, providing neither an advantage nor a disadvantage to the cell, it will remain present but the fraction of its progeny will not increase with respect to the total population, while a mutation that is slightly deleterious to the cell will be selected out of the population over time. Thus, natural selection to some degree counteracts mutational pressures. The degree of this counter-activity is mostly determined by two factors: the size of a population and sexual recombination. Fixation of a mutation is more likely in small bacterial populations than in large populations, which has been modelled mathematically (Crow 2010). When a population passes through a selective bottleneck (an event in which only a subpopulation survives and multiplies), mutations can become fixed in the surviving population. On the other hand, when a limited number of colonies are selected for propagation, the chances to pick a mutant are minute and the wild-type population is far more likely to be picked.

When the genes that a cell depends on to proofread and correct DNA mistakes are themselves malfunctioning, this results in a 10 or 100-fold increase in mutation rate. Such cells are called 'mutators' or even 'hypermutators' (Miller 1996). Mutator strains are the exception, not the rule in bacteria, as they live a 'dangerous' life: the adagio 'don't fix what isn't broken' also largely applies to Nature.

Table 1 Mutation rates of some industrially important bacteria and fungi

Species	Mutations per nucleotide per generation	Genome size*	Mutations per genome per generation
Lactobacillus gasseri	43×10^{-10} (Alexander *et al*. 2019)†	2.0 Mb	86×10^{-4}
Lactobacillus reuteri (new name: *Limosilactobacillus reuteri*)‡	8.7×10^{-10} (Alexander *et al*. 2019)†	2.2 Mb	19×10^{-4}
Lactobacillus salivarius (new name: *Ligilactobacillus salivarius*)‡	30×10^{-10} (Alexander *et al*. 2019)†	2.1 Mb	62×10^{-4}
Lactobacillus plantarum (new name: *Lactiplantibacillus plantarum*)‡	61×10^{-10} (Alexander *et al*. 2019)†	3.3 Mb	291×10^{-4}
Lactobacillus casei (new name: *Lacticaseibacillus casei*)‡	10×10^{-10} (Wang *et al*. 2017)†	3.0 Mb	10×10^{-4}
Lactococcus lactis	90×10^{-10} (Alexander *et al*. 2019)†	2.5 Mb	225×10^{-4}
Bacillus subtilis	$1.2–3.4 \times 10^{-10}$ (Schroeder *et al*. 2018)†	4.3 Mb	$1.5–14 \times 10^{-4}$
Escherichia coli	$0.4–2.0 \times 10^{-10}$ (Schroeder *et al*. 2018)†	4.6 Mb§	$1.8–9.2 \times 10^{-4}$
E. coli	$4.1–6.9 \times 10^{-10}$ (Drake 1991)†	4.7 Mb	$19–33 \times 10^{-4}$
E. coli	1.0×10^{-10} (Lenski 2017)§	4.6 Mb	4.6×10^{-4}
E. coli	$0.2–5000 \times 10^{-10}$ (Krašovec *et al*. 2017)¶	4.6–5.8 Mb	Not applicable
Saccharomyces cerevisiae	$0.79–5.1 \times 10^{-10}$ (Drake 1991)†,**		
Saccharomyces cerevisiae	$100.000–10 \times 10^{-10}$ (Krašovec *et al*. 2017)		
Aspergillus flavus, *Aspergillus nidulans*	0.4×10^{-10} (Álvarez-Escribano *et al*. 2019)**		
Aspergillus fumigatus	0.1×10^{-10} (Álvarez-Escribano *et al*. 2019)**		

*Genome size is taken from the Genome list at NCBI and does not necessarily correspond to the strain referred to in the litterature source.
†Reported mutation rates are for specific loci, which can vary considerably along a given genome.
‡New taxonomic descriptions according to Zheng *et al*. (2020).
§For *E. coli* K-12.
¶Data collected from the literature, with inclusion of supermutator strains.
**Mutations per genome per mitosis.

Can the accumulation of mutations be detected in bacterial cultures?

For mutations to have a selective advantage, a selective pressure must apply, otherwise the mutation would behave as 'neutral'. Under those conditions, over time the presence of spontaneous mutations in a given bacterial population will become detectable. This is most apparent when the advantage is a matter of staying alive where other cells die, for instance when it causes resistance to an antibiotic that is added to the culture. An antibiotic is strictly speaking a toxic compound that will kill (or inhibit growth of) bacteria, unless the bacteria are resistant to its toxicity. A number of antibiotic resistances observed in bacteria are due to a point mutation in the gene that codes for the target to which that antibiotic acts upon. Thus, a spontaneous point mutation in exactly that gene will have a tremendous advantage (live and multiply in presence of the antibiotic instead of stagnation or death), provided that point mutation counteracts the activity of the antibiotic. This is why in presence of an antibiotic, a resistant population can rapidly be selected: mutations that were already present in the population, but that were neutral in the absence of a selective antibiotic, will suddenly have an enormous advantage and will now selectively grow and multiply. The funnel depicting a bottleneck in Fig. 3 could be the application of an antibiotic, to which the drawn mutation (the white asterisk) provides resistance. Without the selective advantage, *i.e.* in absence of that antibiotic, the mutant will remain present (when in all other conditions it is neutral) or slowly being eliminated from the population (when it is slightly deleterious). There are other mechanisms that can lead to antibiotic resistance and that may depend on incoming foreign DNA, but that is not relevant in this context of this treatise.

When given enough time, even the accumulation of mutations in a bacterial culture without a selective pressure can be demonstrated. In the famous Long-Term Evolution Experiment by the research group of Prof. Richard Lenski that started in 1988 and is still continuing after more than 25 years (Lenski 2017), 12 cultures of two closely related *E. coli* strains (six cultures of each) were diluted 1: 100 every 24 h into fresh medium containing limited glucose. This allowed for approximately 6·7 generations per 24 h (Lenski 2017). Every 500 generations a sample was drawn and these samples were analysed with the tools available at the time, eventually by genome

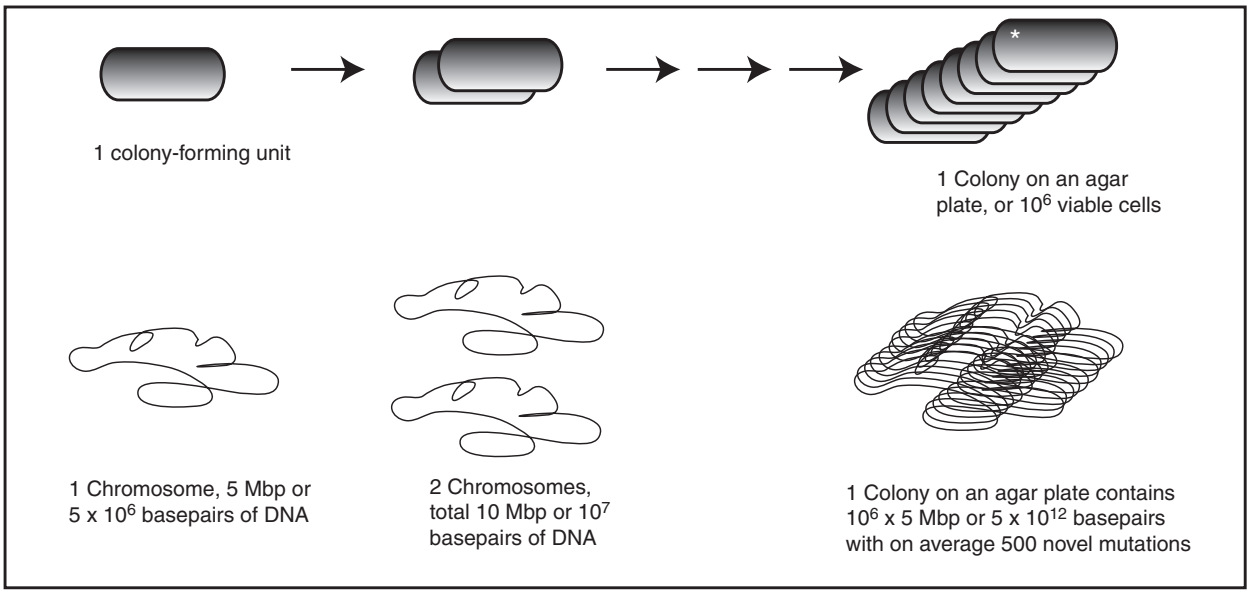


Figure 2 A single cell growing into a colony on an agar plate has produced approximately a million viable cells (10^6 CFU, top). The DNA of one *Escherichia coli* cell contains approximately 5 million basepairs of DNA. The colony contains a total of 5×10^{12} basepairs of DNA, that will have incorporated on average 500 novel spontaneous mutations, given the mutation frequency of 1 in very 10^{10} basepairs. A cell with a mutation in its genome is marked with a white asterisk (top right).

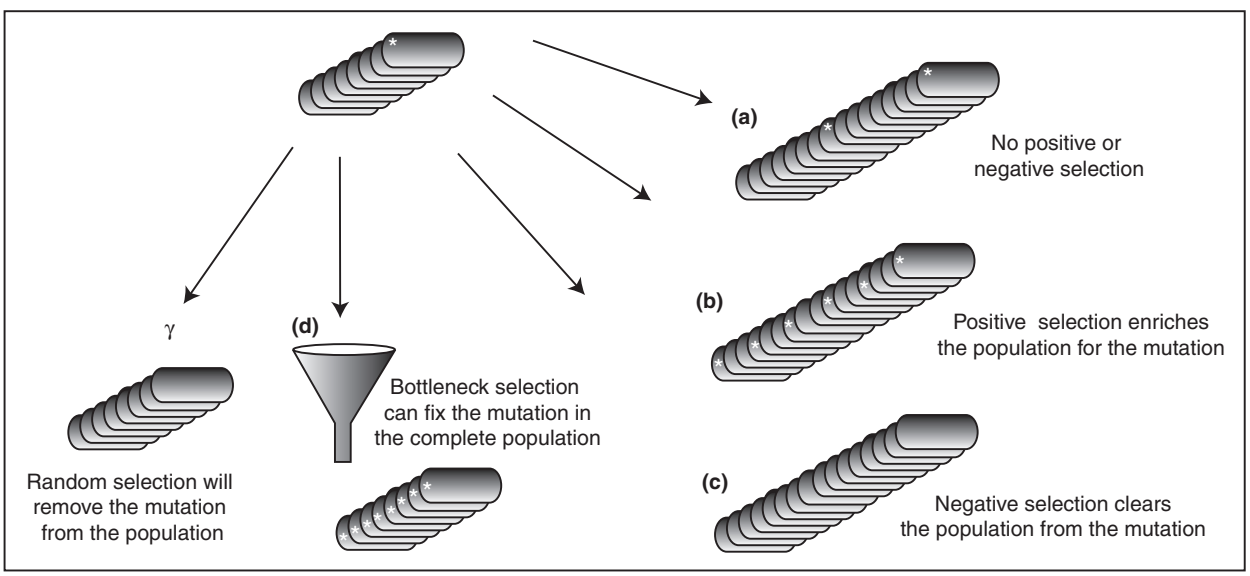


Figure 3 A mutant (shown with a white asterisk) can be maintained in an initial population (top left) when the mutation is neutral (a) without changing its relative abundance in the population. Under positive selection (b), when the mutation provides an advantage under the applied growth conditions, its proportion in the population will increase, while under negative selection, when the mutation is disadvantageous for growth, its proportion decreases (c). In a bottleneck selection (d), a mutant can disappear completely or take over completely (as shown here), depending on the selection of the bottleneck. When a bottleneck is provided by random selection (say, picking one or a few colonies), indicated by γ, the chances are that the mutation is removed from the population.

sequence analysis (e.g. Sniegowski *et al.* 1997; Tenaillon *et al.* 2016). Under these conditions, the *E. coli* population will slowly adapt to the growth medium (here, glucose limitation) and develop an improved fitness. After all, cells that mutate to become less fit will not be able to compete with their better-growing counterparts. Suppose

a mutation spontaneously arises that confers a 10% better growth. It would take approximately 250 generations for that mutation to become the majority in the population. For most of that time the mutant remains a tiny minority because, assuming a constant growth-rate differential, the ratio of the mutant to its progenitor changes exponentially. In the experiment, after approximately 300 generations, the fitness of the *E. coli* population started to increase (Lenski 2017), and continued to increase with two more steps over the next 60 000 generations or so (which took 24 years). This increased fitness to grow in the applied culture media coincided with the accumulation of five genomic mutations that could explain the observed changes in phenotype (Khan *et al*. 2011; Lenski 2017; Good *et al*. 2017). Parallel genetic changes were observed across the replicate populations that suggest presence of a common pool of adaptive mutations. That mutations can remain present in a population without taking over completely was demonstrated by the long-term presence of two genotypes that coexisted for over 50 000 generations (Lenski 2017).

Two things should be stressed about these experiments. First of all, the observations are based on extremely long-running experiments of continuous subculturing. Second, no selective pressure was applied, other than the culture conditions that were kept constant. The culture was kept alive by serial dilution into fresh medium, which is not normally practiced in a research laboratory or in industrial applications. In those settings, typically frozen or lyophilized stocks serve as a constant source from which cultures are repeatedly drawn and grown.

When a bacterial population is cultured from a stock, normally an agar plate is subbed, after which one or several single colonies (each containing approximately 10^6 cells that arose from a single cell after 20 generations of cell division) are subcultured on a fresh plate and then propagated to produce the final culture, which in industrial applications can easily reach a volume of 1000 l. A stationary culture of *E. coli* would contain approximately 10^9 viable cells per ml, so that 1000 l would amount to 10^{15} colony-forming units (CFU) in total. This population may have undergone approximately 55 generations since it was taken out of the freezer, even allowing a generous loss of viability of 10% of the population on average at every generation (Fig. 4). Note, that Lenski's experiment required at least one thousand generations before a change in phenotype was noticeable (Papadopoulos *et al*. 1999). Every time a new *E. coli* population is cultured from the same frozen stock, the number of generations required for production of the exemplified 1000 l culture will start over again, so that with this practice, never more than 55 or so generations will take place. Eventually, when stocks become low, a new series of stock vials will be produced, that add one or more bottleneck selection steps (picking single colonies) to the grand total, but that hardly affects the overall number of generations a 1000 l culture has undergone. When the culture conditions are kept constant, so that any selective pressure applied is constant, and when no foreign DNA is present, the accumulation of spontaneous, naturally occurring mutations will not detectable.

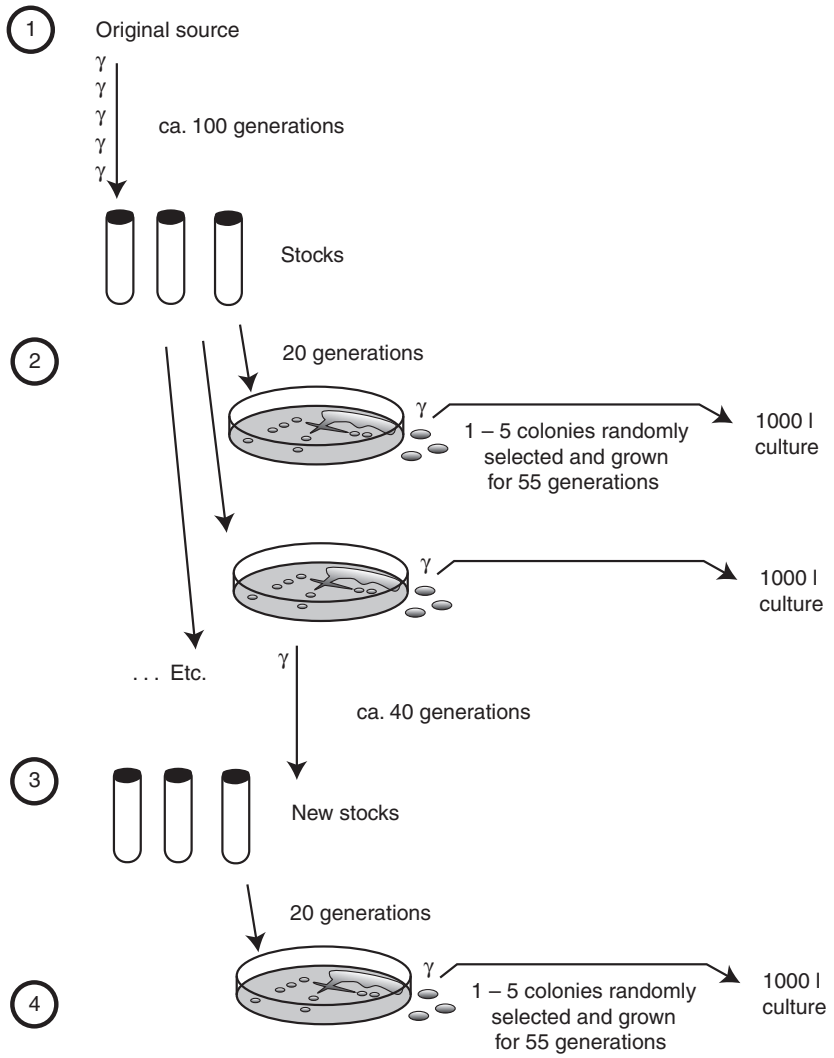
Presuming the relatively high locus-dependent mutation rate of *Lactococcus lactis* that was reported in the literature applies to its complete genome (which is most likely an over-estimate), this would mean that after 55 generations approximately 50 times more *de novo* mutations would be present compared to *E. coli*.

Will a genome sequence of a bacterial culture remain constant within the required number of generations?

Suppose a culture from the imaginary first batch of stock vials in Fig. 4 was used to isolate DNA from, and the complete genome sequence was established some 15 years ago. If that experiment would be repeated today, starting with the same frozen stock, would the result be the same genome sequence?

The determination of complete DNA sequences of a bacterial genome became a possibility after Sanger technology was automated, and in particular when short-read high-throughput sequencing became possible (Illumina, SOLiD, and other so-called Next Generation Sequencing technologies) (recently reviewed by Kozińska *et al*. 2019). However, none of these methods is completely free of errors. An accuracy of 99·992% that was achievable a decade ago (as reported by Pandya *et al*. 2007) may seem impressive, but implies that 8 in every 10^5 bases on average would be incorrect. That corresponds to 400 inaccurate bases per genome sequence of 5 Mb, a sequencing inaccuracy rate that was not unusual in the first decade of this century. Note, that this inherent degree of inaccuracy is in the same order of magnitude as the number of spontaneous mutations that may be present in a sequenced population. For all sequence methods that depend on multiple DNA molecules, the natural mutations will be averaged out to give a consensus sequence for the complete population, but the inaccuracy of the sequencing methodology remains an issue. Inaccuracies due to incomplete coverage and incorrect assembly of reads further contribute to sequence inaccuracies.

The situation is different for sequencing technology that uses single DNA molecules, such as the Oxford Nanopore Technology (ONT). This methodology could, in principle, detect the true diversity of individual genomes within a given population, provided that (i) the



Total nr. of bottle-necks	Total nr. of generations	Nr. of generations since last bottleneck
6	Approx. 175	55
6	Approx. 175	55
	. . . Etc.	
7	Approx. 195	55

Figure 4 Suppose a bacterial strain used for industrial purposes was subbed and single-colonies were picked five times since it had been originally isolated (step 1, top) before it was stored in stock vials. This would have involved five bottle necks of single colony selection (γ) and approximately 100 generations. The stock vials are used for each production round to first generate single colonies (adding 20 generations and a bottleneck γ, step 2) from which an industrial culture of 1000 l is being generated, which would require a further 55 generations. This product will have undergone approximately 175 generations since the original source was identified (table to the right), but, more importantly, only 55 generations since the last bottleneck of random colony selection. That applies for every product prepared from the stocks. When stocks become depleted, new vials are produced (step 3) which adds at least one extra bottleneck selection (single colony picking, γ), but the product produced from these new vials will once again only have undergone 55 generations since its last bottleneck selection (step 4).

sequence determination is 100% accurate; (ii) coverage is sufficient to determine mutations present at low abundance, and (iii) such determined mutations are reported and not averaged out by determining consensus sequences from individual reads. None of these requirements currently apply, so that even for the ONT technology, the true genetic diversity of a population of bacterial cells cannot be accurately captured.

When the complete genome sequences of particular bacterial species type strains were compared, a disturbing large number of inconsistencies between individual sequences was discovered (Dr. George Garrity, NamesForLife, personal communication). Those type strains are stored at national strain collections, and their stocks are being distributed to scientists all over the world. Ideally, every frozen stock should represent the same bacterial strain, as every species has only one defined type strain. This ideal situation is clearly not met, judged by the variation in obtained genome sequences from such type strains, although it can't be determined how much of the

variation was due to true biological variation between stocks, and how much was due to sequencing inaccuracies. Likewise, the genome sequence of the *E. coli* K12 strain MG1655 that was generated in 1997 (Blattner *et al.* 1997), and the genome sequence of the closely related *E. coli* K12 strain W311 (both are derived from the same parental strain) were compared (Hayashi *et al.* 2006). This revealed the presence of 251 locations that appeared to have mutations (short insertions, deletions or single-nucleotide changes), but 243 of these were not the result of mutations but rather due to inaccuracies in the genome seqeunce that was produced in 1997 (Hayashi *et al.* 2006).

These examples illustrate how difficult it is to define a bacterial strain as being genetically 'stable' over time and location, and to provide solid evidence that such genetic 'stability' exists.

Conclusion

A producer of a bacterial product relies on long-term stored stocks that are used to produce the final cultures. Those stocks are typically carefully maintained in redundancy, so that every batch of final product can be produced from a very limited number of stocks. As long as random selection of single colonies is incorporated in the production procedure, this limits the effect of genetic drift. It turns out that the pioneering bacteriologists chose exactly the right approach to keep a bacterial population as genetically constant as possible, long before the concept of clones, the discovery of DNA, or the insights of genetics, were available. Nevertheless it would be very difficult to produce solid evidence that a strain in question remains genetically completely unchanged over time. The odds are that a strain has not mutated beyond the naturally occurring variation that can be expected in a bacterial population. To demand evidence of this is practically impossible. The best whole-genome sequences can do is to provide an estimate of the general mutation rate of the strain in question. When mutational hotspots exist, they may be noticeable in the raw sequence reads, and one can report their frquency. However, the way genome sequences are reported as continuous texts of single-letter bases does not allow for direct reporting of mutation frequencies. Using non-cannonical IUPAC codes (e.g., Y for C or T, W for G or A, etc.) does not provide information on the relative frequency of those alternative bases. There are currently no methods to record mutations that are accounted in, say, only one of a thousand reproduced sequence reads. Reporting these would increase the data size of any given bacterial genome to impractical proportions. In conclusion, we will have to accept that a given genome sequence is an approximation of the most likely base found present in a given location, while it may, occasionally, be represented by an alternative base.

References

Alexander, L.M., Oh, J.H., Stapleton, D.S., Schueler, K.L., Keller, M.P., Attie, A.D. and van Pijkeren, J.P. (2019) Exploiting prophage-mediated lysis for biotherapeutic release by *Lactobacillus reuteri*. *Appl Environ Microbiol* **85**, e02335-18.

Álvarez-Escribano, I., Sasse, C., Bok, J.W., Na, H., Amirebrahimi, M., Lipzen, A., Schackwitz, W., Martin, J. *et al.* (2019) Genome sequencing of evolved aspergilli populations reveals robust genomes, transversions in *A. flavus*, and sexual aberrancy in non-homologous end-joining mutants. *BMC Biol* **17**, 88.

Blattner, F.R., Plunkett, G. 3rd, Bloch, C.A., Perna, N.T., Burland, V., Riley, M., Collado-Vides, J., Glasner, J.D. *et al.* (1997) The complete genome sequence of *Escherichia coli* K-12. *Science* **277**, 1453–1462.

Crow, J.F. (2010) Wright and Fisher on inbreeding and random drift. *Genetics* **184**, 609–611.

de Sousa, M.A., Rama, G.R., Volken de Souza, C.F. and Granada, C.E. (2019) Acid lactic lactobacilli as a biotechnological toll to improve food quality and human health. *Biotechnol Prog* **36**, https://doi.org/10.1002/btpr.2937.

Drake, J.W. (1991) A constant rate of spontaneous mutation in DNA-based microbes. *Proc Natl Acad Sci USA* **88**, 7160–7164.

El Enshasy, H., Malik, K., Malek, R.A., Othman, N.Z., Elsayed, E.A. & Wadaan, M. (2016) Anaerobic probiotics: The key microbes for human health. *Adv Biochem Eng Biotechnol*, **156**, 397–431.

Good, B.H., McDonald, M.J., Barrick, J.E., Lenski, R.E. and Desai, M.M. (2017) The dynamics of molecular evolution over 60,000 generations. *Nature* **551**, 45–50.

Hayashi, K., Morooka, N., Yamamoto, Y., Fujita, K., Isono, K., Choi, S., Ohtsubo, E., Baba, T. *et al.* (2006) Highly accurate genome sequences of *Escherichia coli* K-12 strains MG1655 and W3110. *Mol Syst Biol* **2**, 2006.0007.

Hershberg, R. and Petrov, D.A. (2010) Evidence that mutation is universally biased towards AT in bacteria. *PLoS Genet* **6**, e1001115.

Khan, A.I., Dinh, D.M., Schneider, D., Lenski, R.E. and Cooper, T.F. (2011) Negative epistasis between beneficial mutations in an evolving bacterial population. *Science* **332**, 1193–1196.

Kozińska, A., Seweryn, P. and Sitkiewicz, I. (2019) A crash course in sequencing for a microbiologist. *J Appl Genet* **60**, 103–111.

Krašovec, R., Richards, H., Gifford, D.R., Hatcher, C., Faulkner, K.J., Belavkin, R.V., Channon, A., Aston, E. *et al.*

(2017) Spontaneous mutation rate is a plastic trait associated with population density across domains of life. *PLoS Biol* **15**, e2002731.

Jackson, S.A., Schoeni, J.L., Vegge, C., Pane, M., Stahl, B., Bradley, M., Goldman, V.S., Burguière, P. *et al.* (2019) Improving end-user trust in the quality of commercial probiotic products. *Front Microbiol* **10**, 739.

Lenhart, J.S., Pillon, M.C., Guarné, A., Biteen, J.S. and Simmons, L.A. (2016) Mismatch repair in Gram-positive bacteria. *Res Microbiol* **167**, 4–12.

Lenski, R.E. (2017) Experimental evolution and the dynamics of adaptation and genome evolution in microbial populations. *ISME J* **11**, 2181–2194.

Miller, J.H. (1996) Spontaneous mutators in bacteria: insights into pathways of mutagenesis and repair. *Annu Rev Microbiol* **50**, 625–643.

Nandy, S.K. and Srivastava, R.K. (2018) A review on sustainable yeast biotechnological processes and applications. *Microbiol Res* **207**, 83–90.

Pandya, G.A., Holmes, M.H., Sunkara, S., Sparks, A., Bai, Y., Verratti, K., Saeed, K., Venepally, P. *et al.* (2007) A bioinformatic filter for improved base-call accuracy and polymorphism detection using the Affymetrix GeneChip whole-genome resequencing platform. *Nucleic Acids Res* **35**, https://doi.org/10.1093/nar/gkm918.

Papadopoulos, D., Schneider, D., Meier-Eiss, J., Arber, W., Lenski, R.E. and Blot, M. (1999) Genomic evolution during a 10,000-generation experiment with bacteria. *Proc Natl Acad Sci USA* **96**, 3807–3812.

Schroeder, J.W., Yeesin, P., Simmons, L.A. and Wang, J.D. (2018) Sources of spontaneous mutagenesis in bacteria. *Crit Rev Biochem Mol Biol* **53**, 29–48.

Sniegowski, P.D., Gerrish, P.J. and Lenski, R.E. (1997) Evolution of high mutation rates in experimental populations of *E. coli*. *Nature* **387**, 703–705.

Tenaillon, O., Barrick, J.E., Ribeck, N., Deatherage, D.E., Blanchard, J.L., Dasgupta, A., Wu, G.C., Wielgoss, S. *et al.* (2016) Tempo and mode of genome evolution in a 50,000-generation experiment. *Nature* **536**, 165–170.

Tripathi, N.K. and Shrivastava, A. (2019) Recent developments in bioprocessing of recombinant proteins: expression hosts and process development. *Front Bioeng Biotechnol* **7**, 420.

Wang, J., Dong, X., Shao, Y., Guo, H., Pan, L., Hui, W., Kwok, L.Y., Zhang, H. *et al.* (2017) Genome adaptive evolution of *Lactobacillus casei* under long-term antibiotic selection pressures. *BMC Genom* **18**, 320.

Wassenaar, T.M. (2016) Insights from 100 years of research with probiotic *E. coli*. *Eur J Microbiol Immunol (Bp)* **6**, 147–161.

Zheng, J., Wittouck, S., Salvetti, E., Franz, C.M.A.P., Harris, H.M.B., Mattarelli, P., O'Toole, P.W., Pot, B. *et al.* (2020) A taxonomic note on the genus *Lactobacillus*: description of 23 novel genera, emended description of the genus Lactobacillus Beijerinck 1901, and union of Lactobacillaceae and Leuconostocaceae. *Int J Syst Evol Microbiol* **70**, 2782–2858. https://doi.org/10.1099/ijsem.0.004107.