





Positive Selection during Niche Adaptation Results in Large-Scale and Irreversible Rearrangement of Chromosomal Gene Order in Bacteria

Sha Cao ^{1,†} Gerrit Brandis ^{2,†} Douglas L. Huseby ¹ and Diarmaid Hughes ^{1,*}

¹Department of Medical Biochemistry and Microbiology, Uppsala University, Uppsala, Sweden

²Department of Cell and Molecular Biology, Uppsala University, Uppsala, Sweden

*Corresponding author: E-mail: diarmaid.hughes@imbim.uu.se.

[†]These authors contributed equally.

Associate editor: Heather Hendrickson

Abstract

Analysis of bacterial genomes shows that, whereas diverse species share many genes in common, their linear order on the chromosome is often not conserved. Whereas rearrangements in gene order could occur by genetic drift, an alternative hypothesis is rearrangement driven by positive selection during niche adaptation (SNAP). Here, we provide the first experimental support for the SNAP hypothesis. We evolved *Salmonella* to adapt to growth on malate as the sole carbon source and followed the evolutionary trajectories. The initial adaptation to growth in the new environment involved the duplication of 1.66 Mb, corresponding to one-third of the *Salmonella* chromosome. This duplication is selected to increase the copy number of a single gene, *dctA*, involved in the uptake of malate. Continuing selection led to the rapid loss or mutation of duplicate genes from either copy of the duplicated region. After 2000 generations, only 31% of the originally duplicated genes remained intact and the gene order within the *Salmonella* chromosome has been significantly and irreversibly altered. These results experientially validate predictions made by the SNAP hypothesis and show that SNAP can be a strong driving force for rearrangements in chromosomal gene order.

Key words: experimental evolution, chromosome rearrangements, *Salmonella* Typhimurium, SNAP hypothesis.

Introduction

The bacterial kingdom forms a coherent group evolved from a last common ancestor but is genetically extremely diverse with species and variants that occupy a huge variety of environmental niches. Our understanding of how bacterial species diverge is necessarily based on correctly understanding the diversifying mechanisms that fuel bacterial evolution. One feature that is currently difficult to explain is that the colinear organization of homologous genes on bacterial chromosomes of different species is highly variable and for most homologous genes, there is no long-range colinearity in gene order (Koonin et al. 1996, 2021; Puigbò et al. 2010). The standard interpretation for the low level of conservation is that selection to maintain linear gene order is weak and this allows changes in gene order to occur by genetic drift. However, there are problems with this interpretation: (1) the mechanisms of linear order rearrangement (inversion, transposition, deletion with reacquisition at a new location) each occur at low frequencies and without positive selection (or strong founder effects) would very rarely go to fixation and (2) the evidence suggests that many gene order rearrangements are deleterious, with purifying

selection operating (Rocha 2006). This paradox could be resolved if gene order rearrangements, including during speciation, could be driven by positive selection. We recently proposed selection during niche adaptation (SNAP) as a radical alternative to the drift hypothesis (Brandis and Hughes 2020). The major advantage of SNAP over the traditional model is that each step in the process occurs at a high frequency and each successive step can be driven by positive selection, potentially resulting in very rapid evolution and fixation of rearranged gene order. SNAP has the theoretical potential to contribute to the genetic separation of lineages, a vital step on the path to speciation, but has not previously been tested experimentally.

Here, we use *Salmonella enterica* serovar Typhimurium (*S. Typhimurium*) as a model organism to experimentally validate specific predictions made by the SNAP hypothesis. We adapted *S. Typhimurium* to growth in a novel environmental niche and monitored genetic changes to the bacterial chromosome at regular intervals over the course of 2000 generations. Our results were in full agreement with each prediction made by the SNAP hypothesis indicating that the SNAP mechanism could play a crucial role in the process of gene order diversification.

© The Author(s) 2022. Published by Oxford University Press on behalf of Society for Molecular Biology and Evolution.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (<https://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

Open Access

Results

The SNAP Hypothesis

Occurring ~500 Ma, the *Enterobacteriales* are the most recently diverged order within the *Gammaproteobacteria* (Marin et al. 2017). Closely related species within this order can display an almost identical linear order of genes across the chromosome. The chromosomal gene order of *S. Typhimurium* and *Escherichia coli* which diverged 100–200 Ma (Ochman and Wilson 1987; Doolittle et al. 1996; Baumbler et al. 2013) differs mainly in a 563 kb long section that is inverted between the two species (fig. 1c). Comparing the chromosomes of *S. Typhimurium* and *Proteus mirabilis* which diverged earlier in the evolution of the *Enterobacteriales*, about 300–500 Ma (Ochman and Wilson 1987; Brandis 2021), shows that this conservation of gene order is not maintained within the bacterial order (fig. 1c). Any observed long-range gene order conservation between species that diverged more than 500 Ma has been attributed to the existence of operons (Rocha 2006). It has been generally assumed that changes in chromosomal gene order are the result of genetic drift. The genetic drift model assumes that the linear organization of the bacterial chromosome is under weak selection and is subject to

rearrangements events involving inversion, transposition, deletion, and horizontal gene transfer (Hughes 2000; Skovgaard et al. 2011; Nourreen et al. 2019). Over time, a large number of successive rearrangement events could alter the linear order of genes on a chromosome.

We have recently proposed an additional mechanism that could contribute to rapid chromosomal reorganization. The SNAP hypothesis proposes that gene order rearrangements could emerge from positive SNAP (Brandis and Hughes 2020). The SNAP hypothesis proposes a potential evolutionary trajectory that bacteria can undergo when entering a novel growth environment (fig. 1a and b). Initially, duplication of a chromosomal region that provides a selective advantage, for example, by increased expression of a nutrient transporter gene, will be selected. This initial rapid genome expansion phase is followed by a reduction phase. In this phase, duplicate genes will be lost to reduce fitness disadvantages caused by genes that are not required in duplicate copies (due to waste of resources or interference with normal physiology). Inactivation of genes will occur randomly with respect to each copy of the duplication, which will ultimately result in a rearrangement of the linear order of genes within the duplicated region (fig. 1a and b).

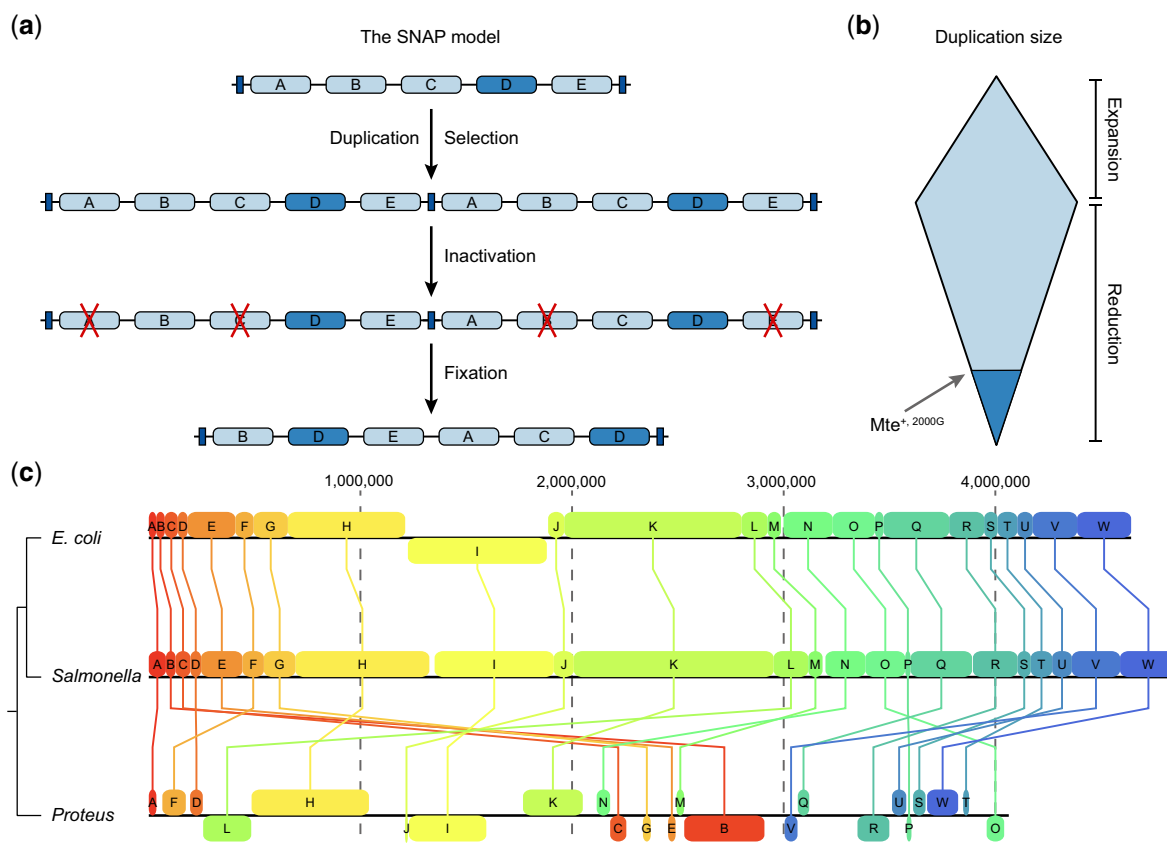


FIG. 1. The SNAP hypothesis. (a) Schematic overview of the SNAP model. The gene under selection for duplication is shown in dark blue and the regions of homology flanking the ends of the duplicated region are indicated by blue rectangles. Red crosses designate copies of genes that are deleted. (b) Change in duplication size during the stages of the SNAP model. The light blue area indicates the progress of the Mte⁺, 2000G isolate along the SNAP trajectory (31% of duplication remains). (c) Whole-genome alignment (with a 5 kb block size) of *Salmonella Typhimurium* (NC_003197) with *Escherichia coli* (NC_000913) and *Proteus mirabilis* (NC_010554). Homologous regions displayed below each black line are inverted relative *S. typhimurium*.

The SNAP Hypothesis Makes Specific Predictions on Multiple Steps Along the Evolutionary Trajectory

The evolutionary trajectory predicted by the SNAP hypothesis can be divided into four sequential stages: duplication, selection, inactivation, and fixation (fig. 1a and b). The fixation stage marks a point of no return at which the only viable path forward is the step-wise inactivation of duplicate genes that will ultimately result in an altered gene order on the bacterial chromosome (Brandis and Hughes 2020). To validate the SNAP hypothesis experimentally, we defined five specific predictions that need to be fulfilled:

- (i) Large chromosomal duplications arise frequently during growth in a novel niche.
- (ii) The duplications provide a selective benefit to bacteria in the novel niche due to the increase in copy number of a small number of duplicated genes.
- (iii) Continuing evolution within the niche leads to the inactivation of individual copies of some of the duplicate genes (including essential genes).
- (iv) Accumulation of mutations within the duplicated region will prevent segregation of the duplication.
- (v) Mutations will be randomly distributed between the two copies of the duplicated region.

Duplications are a Common Adaptation Mechanism to Novel Growth Conditions

The occurrence of duplications of large and small segments of the bacterial chromosome is common during growth in nonlethal selective environments (Straus and Hoffmann 1975; Sonti and Roth 1989; Näsvall et al. 2012; Hoegler and Hecht 2018; Hufnagel et al. 2021). The ability to utilize a novel carbon source has been shown to be a potential key event in the establishment of a population within a new environmental niche (Blount et al. 2012). Thus, we have used a minimal medium with malate as a sole carbon source to simulate a novel growth environment. Malate is a poor carbon source for wild-type *Salmonella* which forms only tiny colonies on minimal malate agar (~1 mm diameter after 48 h) (fig. 2a) and displays almost no increase in optical density after 24 h of growth in liquid minimal malate medium (fig. 2f). It was previously shown that large colony variants could be detected when plating *Salmonella* on minimal malate agar. Genetic mapping indicates a large chromosomal duplication to be present within the large colony variants but the isolates were never whole-genome sequenced (Straus and Hoffmann 1975). We repeated the selection of *Salmonella* on minimal malate agar and found that large colony variants (referred to as Mte^+) appeared at a frequency of 4×10^{-4} . Two independent Mte^+ colonies were isolated and whole-genome sequenced. The sequence analysis showed that both isolates contained an identical 1.66 Mb tandem duplication within the chromosome flanked by the duplicate *ccmABCDEFGHIH* operon which encodes cytochrome *c*-type biogenesis proteins (Thoöny-Meyer et al. 1995) (fig. 2b). The *ccm* operons

provide a 6.3 kb region of sequence homology which explains the high frequency at which Mte^+ colonies are selected (Gerstein et al. 1994; Brandis et al. 2018). This data show that duplications as large as a third of the bacterial chromosome can rapidly be selected for within a novel growth environment. One of the sequenced Mte^+ isolates was used for the further evolution experiment.

Selection for Increased Expression of a Single Gene is Responsible for a 1.66 Mb Duplication

The duplication in the Mte^+ strain contains 1,529 genes, of which six are involved in malate utilization. These genes encode two C4-dicarboxylate transporters (DctA and YhiT), three malate dehydrogenases (MaeB, Mdh, and STM3081), and the transcriptional regulator Crp which regulates *dctA* expression (Unden et al. 2016) (fig. 2b and c). We tested which of these six genes is the main driving force for selection of the duplication using two different approaches. First, we asked which of the six genes is necessary for the selection of the Mte^+ duplication. For this, we constructed a set of five strains with single gene deletions (excluding *crp* which is essential for growth) and selected large colony variants on minimal malate plates. As expected, bacteria with deletions of *dctA* or *mdh* were not able to grow with malate as a single carbon source (Davies et al. 1999; van der Rest et al. 2000). For the remaining three isolates ($\Delta maeB$, $\Delta STM3081$, and $\Delta yhiT$), large colony variants containing the Mte^+ duplication could be selected (fig. 2d). Thus, only three of the genes (*crp*, *dctA*, and *mdh*) are necessary for the selection of the Mte^+ duplication. Next, we asked if the duplication of any single gene is sufficient to explain the increased fitness of the Mte^+ strain in a minimal malate medium. Six strains with single-gene duplications were constructed and their growth rates in minimal malate media were measured (fig. 2e and f). Wild-type *Salmonella* shows almost no increase in optical density after 24 h of growth. This inability to grow with malate as a carbon source remains in four of the six strains that carry duplications of *maeB*, *mdh*, *STM3081*, and *yihT* and is only slightly elevated in the strain carrying the *crp* duplication (fig. 2f and supplementary fig. S1, Supplementary Material online). In stark contrast, the duplication of *dctA* significantly improves the growth characteristics of the strain leading to exponential doubling times close to those of the strain carrying the Mte^+ duplication (Mte^+ : 106 ± 5 min, *dctA*^{dup}: 139 ± 3 min) and indistinguishable optical densities after 24 h of growth (Mte^+ : OD_{600nm} 0.57 ± 0.01 , *dctA*^{dup}: OD_{600nm} 0.55 ± 0.07) (fig. 2f). Since *crp* is a regulator of *dctA* expression, it is most likely that the improved growth observed in the *crp* duplication strain is also the result of increased *dctA* expression. Taken together, these data show that the duplication of *dctA* is necessary and sufficient to explain the improved growth characteristics of the Mte^+ strain and that the duplication of a single gene can be the driving force for the selection of a 1.66 Mb duplication on the chromosome.

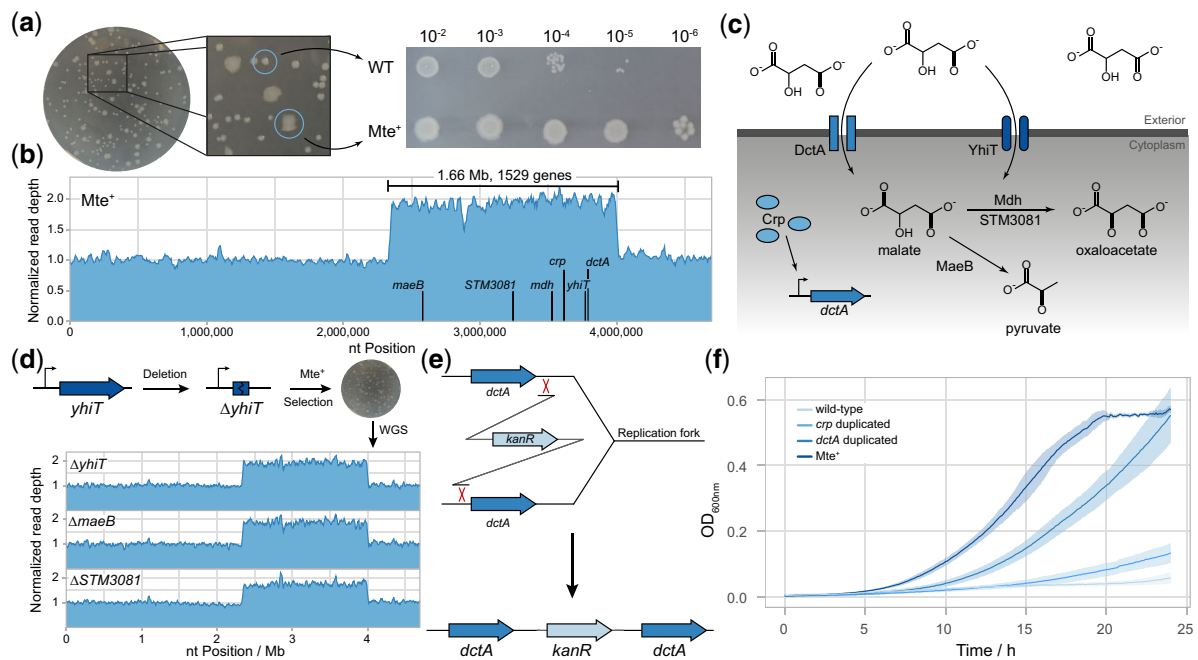


Fig. 2. Selection and analysis of the Mte⁺ duplication. (a) Comparison of wild-type (WT) and Mte⁺ colonies on minimal malate plates. (b) Normalized read depth coverage of the Mte⁺ isolate. The locations of the six genes involved in malate utilization within the duplicated region are indicated. (c) Schematic overview of the protein functions of the six malate utilization genes included in the duplicated region. (d) Individual genes involved in malate utilization were deleted and Mte⁺ colonies were selected. The read depth analysis of the Mte⁺ colonies carrying deletions of *yhiT*, *meaeB*, and *STM3081* is shown below. (e) Overview of the construction process of single-gene duplications using lambda-red recombineering. (f) Growth curves of strains in minimal malate medium. The curves show the average density \pm 95% confidence interval of five biological replicates.

Continuous Evolution within the New Niche Leads to Inactivation of Duplicate Genes

Our experiments show that selection pressure on increased expression of a single gene (*dctA*) can rapidly lead to the duplication of a large section of the chromosome. The SNAP hypothesis predicts that continuous evolution within the new niche will lead to the inactivation of duplicate genes that are not under selection and ultimately lead to a rearranged gene order within the duplicated section of the chromosome (Brandis and Hughes 2020) (fig. 1b and c). An alternative possibility is that mutations that increase *dctA* expression (e. g. within the *dctA* promoter region) will ultimately appear thus alleviating the selection on the Mte⁺ duplication. In that case, the duplication could segregate without changing the gene order within the chromosome.

To test these two possibilities, we evolved five independent lineages of the Mte⁺ strain in minimal malate media by serial passage (fig. 3a). After 500 generations of evolution, cultures were plated on minimal malate plates and a single clone was isolated from each lineage for whole-genome sequencing. The sequence analysis showed that none of the isolates had lost the entire Mte⁺ duplication but instead they had segregated a large segment within the duplicated region. Four of the five isolates segregated an \sim 1 Mb large segment containing 977 genes between two homologous transposon genes (*tnpA-3* and *tnpA-5*) (supplementary fig. S2, Supplementary Material online). The fifth isolate lost 1.1 Mb sequence containing

1,023 genes as a result of a nonhomologous recombination event between the genes *yfdZ* and *yheS*. This isolate also acquired an additional 23 kb duplication between the genes *yedI* and *nfo* outside the original duplication region (fig. 3b, supplementary fig. S2, Supplementary Material online). Each strain had acquired one mutation outside but none within the duplicated region (supplementary tables S1 and S2, Supplementary Material online). These results are consistent with the SNAP hypothesis and indicate that, at least in the short term, mutations that are sufficient to increase expression of the *dctA* gene to the required level are not selected for. To test if this observation holds true on a long term, we decided to accelerate the rate of evolution by inactivating the bacterial mismatch repair system thus increasing the mutation rate (Worth et al. 1994; Elez et al. 2007). We deleted the *mutL* gene within the five isolates which increased mutation rates more than 300-fold in minimal malate medium (fig. 3d) and continued the evolution experiment. After an additional 500 generations of growth, we plated the cultures and isolated a single clone of each lineage. Whole-genome sequence analysis showed that the remnants of the original Mte⁺ duplication remained stable within all five isolates (supplementary fig. S3, Supplementary Material online). Mutations had also started to accumulate within the duplicated region, with each of the five isolates carrying 3–8 mutations (supplementary tables S1 and S2, Supplementary material online). Most interestingly, one lineage (Mte⁺

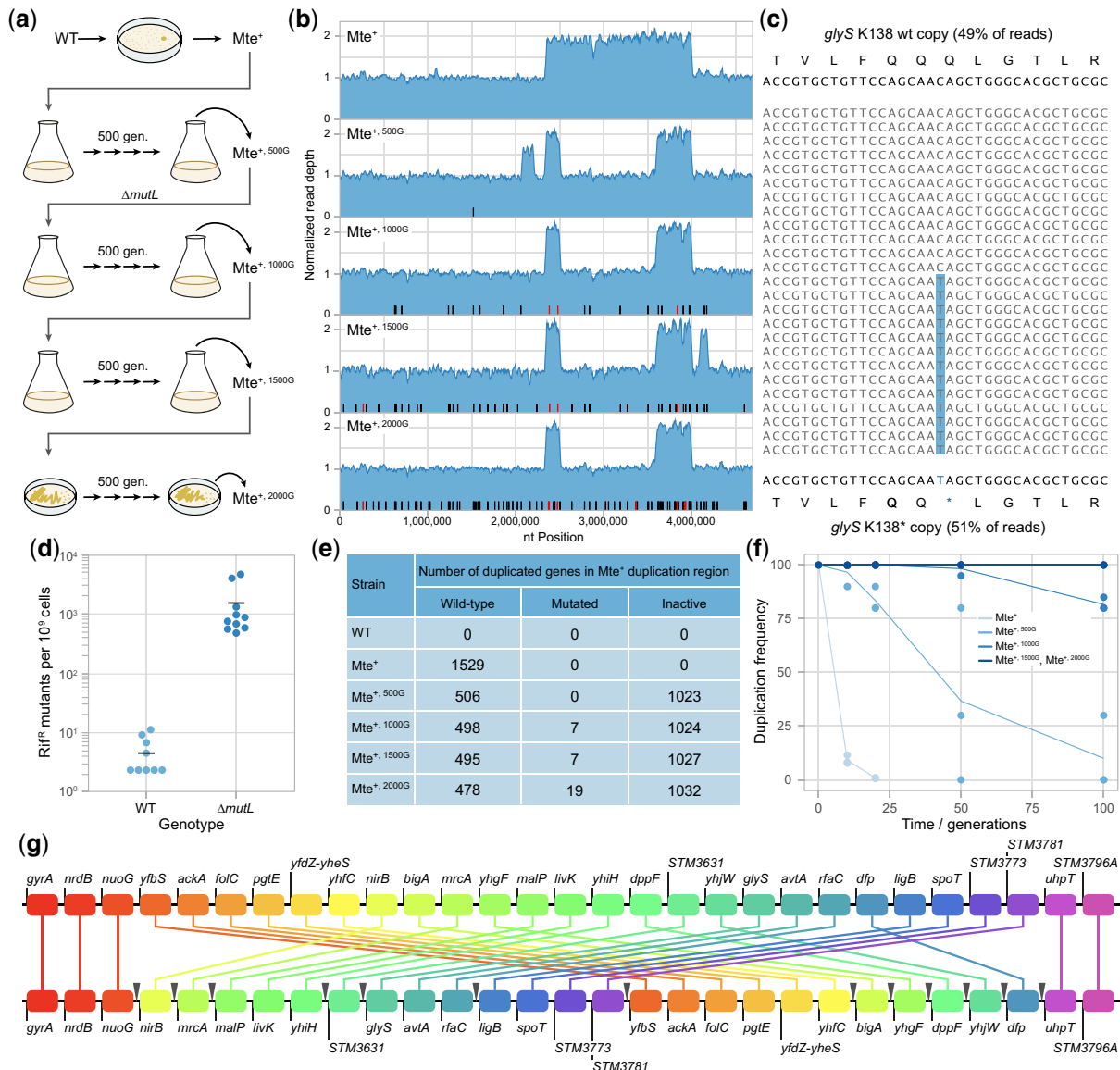


FIG. 3. Long-term evolution of the Mte⁺ isolate. (a) Schematic overview of the evolution experiment. (b) Read depth and mutation analysis of four strains along an evolutionary trajectory. Nonsynonymous mutations in the coding sequences of genes are indicated by black (nonessential genes) and red (essential genes) lines. (c) Example of mutation analysis in the duplicated region (26% of reads are shown). The consensus nucleotide and amino acid sequences are shown above (*glyS* wild-type reads) and below (*glyS* K138* reads). (d) The mutation rate of wild-type and $\Delta mutL$ *Salmonella* in minimal malate medium. (e) Overview of the number of duplicate genes within the Mte⁺ duplication region in the strains along an evolutionary trajectory. (f) Duplication stability assay in four strains along an evolutionary trajectory. Lines indicate the average of three biological replicates. The Mte⁺, 1500G and Mte⁺, 2000G isolates are both fully stabilized. (g) Comparison of gene order of wild-type genes between *Salmonella* (top) and the Mte⁺, 2000G isolate (bottom). Only genes with a mutated/deleted allele within the duplication region are shown. Changes in the linear order of genes are indicated by gray triangles.

1000G) acquired two mutations in essential genes, *glyS* Q357* and *nrdB* I147 T, in one of the two copies of these genes (fig. 3c). GlyS is the β subunit of glycine-tRNA ligase (Nagel et al. 1984; Goodall et al. 2018). The Q357* nonsense mutation truncates the protein resulting in the loss of 332 amino acids at the C-terminus which includes the anticodon binding domain (Finn et al. 2016). Thus, at least one of the eight mutations observed in this isolate inactivates a copy of a duplicated gene as predicted by the SNAP hypothesis (Brandis and Hughes 2020). The other

seven mutations are amino acid substitutions and could potentially reduce the activity of the respective proteins (supplementary table S1, Supplementary Material online).

After 1,000 generations of growth, including 500 generations with an accelerated mutation rate all lineages evolved according to the evolutionary trajectories predicted by the SNAP hypothesis. We were particularly interested in the strain (referred to as Mte⁺, 1000G) carrying the inactivating mutation in *glyS*. SNAP predicts that inactivation mutations in essential genes are a critical stepping

stone to permanent fixation of the original duplication (Brandis and Hughes 2020). We focused our evolution efforts on the $Mte^{+, 1000G}$ isolate that carries the *glyS* mutation and evolved 10 lineages for an additional 500 generations after which we isolated a single clone per lineage for whole-genome sequencing. As before, the remnants of the Mte^{+} duplication remained within all 10 isolates (fig. 3b, supplementary fig. S4, Supplementary Material online) and each strain had acquired 3–16 additional mutations within the duplicated region (supplementary tables S1 and S2, Supplementary Material online). The number of gene inactivation mutations (defined as nonsense and frameshift mutations) increased to up to four mutations per strain. One strain ($Mte^{+, 1500G}$) acquired an additional inactivating mutation within an essential gene, *folC* P360fs. *FolC* is the dihydrofolate synthetase and the mutation removes the last 62 amino acids of the C-terminus (Bognar et al. 1985; Goodall et al. 2018). Whereas this region does not contain a known essential domain a truncation of this size is expected to reduce protein functionality.

We conclude that evolution within the new niche leads to the stepwise deletion and inactivation of duplicate genes, including essential genes, as predicted by the SNAP hypothesis. After 1,500 generations of evolution, the number of intact duplicated genes within the $Mte^{+, 1500G}$ has been reduced from 1,529 genes to 495 genes corresponding to a reduction of 68% (fig. 3e).

Genetic Drift Increases the Rate of Gene Loss

The evolution experiments to this point were performed in liquid cultures with large population sizes ($\sim 1.5 \times 10^{11}$ cfu for the $Mte^{+, 1500G}$ strain). These conditions impose a constant competitive state on the evolving cells and restrict evolutionary trajectories to the most-fit ones. We asked if introducing genetic drift to the system could increase the loss rate of genes within the duplicated region by alleviating the requirement for optimal fitness. To address this, we evolved 10 lineages of the $Mte^{+, 1500G}$ on minimal malate agar plates. Serial restreaking of single colonies effectively introduces single-cell bottlenecks every 25 generations. After 500 generations of evolution, the final clone of each lineage was whole-genome sequenced. The results show that increasing the possibility for genetic drift increased the rate of accumulation of mutations within the duplicated region by 55% (Mann–Whitney test, 95% confidence level, $P = 0.035$). Accordingly, the average number of accumulated mutations increased from 6.4/lineage in generations 1,000–1,500 to 9.9/lineage in generations 1,500–2,000 (supplementary tables S1 and S2, Supplementary Materials online). One of the lineages ($Mte^{+, 2000G}$) carried a total of 28 mutations within the duplicated region of the chromosome out of which nine are predicted to be gene-inactivating mutations. In this strain, the total number of wild-type duplicated genes has been reduced from 1,529 genes to 478 genes (69% reduction).

Accumulation of Mutations Locks an Evolutionary Trajectory that Leads to a Novel Gene Order

Our data show that (1) adaptation to growth within a new niche can select for large chromosomal duplications and that (2) continuous evolution within the new niche results in the loss of duplicate genes. The next prediction made by the SNAP hypothesis is that the accumulation of mutations within the duplicated region will lead to the fixation of the duplication even in the absence of the initial selective pressure (Brandis and Hughes 2020). Every additional mutation has the ability to limit potential possibilities to segregate the duplication since the segregation might lead to a chromosome that lacks essential genes. Once a critical number of mutations has accumulated within the duplicated region, any viable segregation should become impossible. At this point, the only path to lose the remaining duplicated genes would be by stepwise inactivation/deletion. To test if this prediction is correct, we selected a set of four directly related strains that represent each of the steps along an evolutionary trajectory of the evolution experiment (Mte^{+} , $Mte^{+, 500G}$, $Mte^{+, 1000G}$, $Mte^{+, 1500G}$, and $Mte^{+, 2000G}$) and measured the stability of the duplication in rich medium (fig. 3f). The duplication is rapidly lost within the unevolved isolate (Mte^{+}) with a loss rate of 24.2% per generation. In contrast, and in agreement with the prediction made by the SNAP hypothesis, the duplication becomes increasingly stable with each additional evolutionary cycle. After 1,500 generations of evolution, the duplication is fully stabilized and no viable segregation is observed during the growth in the rich medium (fig. 3f and table 1).

These data confirm the prediction made by the SNAP hypothesis that the accumulation of mutations within the duplicated region will lead to a point of no return at which segregation of the duplication becomes inviable. At this point, an evolutionary trajectory is locked in place that inevitably leads to a rearranged gene order within the bacterial chromosome. Our data suggest that this point can be reached in as little as 1,500 generations of growth which would suggest that even relatively transient adaptation to a novel growth environment could lead to a rearrangement of gene order within the bacterial chromosome.

Adaptation Leads to a New Order of Genes on the Chromosome

The last prediction made by the SNAP hypothesis is that mutations will be randomly distributed among the two

Table 1. Stability of the Mte^{+} duplication along an evolutionary trajectory.

Strain	Loss rate (% per generation)
Mte^{+}	-24.2 ± 1.2
$Mte^{+, 500G}$	-1.6 ± 0.8
$Mte^{+, 1000G}$	-0.2 ± 0.0
$Mte^{+, 1500G}$	-0.0 ± 0.0
$Mte^{+, 2000G}$	-0.0 ± 0.0

copies of the duplication. Theoretically, it would be possible that all mutations arise within the same copy of the duplication which would not lead to a rearranged gene order. Unfortunately, the short-read sequencing used to identify duplications and mutations within this study does not reveal in which of the copies a mutation is located. To address this limitation, we chose three of the evolved strains: $Mte^{+, 500G}$ to identify the location of the initial large deletion and $Mte^{+, 1500G}$ as well as $Mte^{+, 2000G}$ to identify the locations of the point mutations. We introduced selective deletions within the duplicated region of these strains followed by whole-genome sequencing to identify which mutations were lost with the deletion and which remained (fig. 4). This method allowed us to identify the location of each of the acquired mutations (table 2). We found that the initial large deletion and 12 mutations occurred in the left copy, whereas 16 mutations occurred within the right copy of the duplication (based on the linear order of the *Salmonella* chromosome). These results show that mutations do indeed appear randomly within the two copies of the duplicated region and show that the gene order of the *Salmonella* strain evolved in minimal malate media is already significantly altered after 2,000 generations (fig. 3g).

Discussion

A low level of gene order conservation in bacterial genomes is well established (Koonin et al. 1996, 2021; Puigbò et al. 2010; Darmon and Leach 2014). Successive chromosomal inversions are one mechanism by which gene order could be rearranged and indeed inversions have been noted as a type of organizational variant in many bacterial species (Belda et al. 2005; Darling et al. 2008; Matthews et al. 2011; Scott and Ely 2016; Xu et al. 2016; Mao and Grogan 2017; Repar and Warnecke 2017; Ely et al. 2019; Shelyakin et al. 2019), sometimes shown to be associated with recombination between inverted repeats such as ribosomal RNA operons or mobile genetic elements including prophage (Matthews et al. 2011; Wang et al. 2017; Fitzgerald et al. 2021). However, it remains an open question to which degree the observed long-term evolutionary lack of gene order on bacterial chromosomes is due to the successive effects of overlapping inversions which in many cases are not expected to confer any immediate selective advantage on the affected bacterial strain. We have proposed a mechanism that could use the force of positive selection to drive rapid chromosomal rearrangements during the process of adaptation to new environmental niches: the SNAP hypothesis, involving duplication of chromosomal segments that carry genes where increased copy number provides a fitness advantage, followed by a process of random gene loss from each copy of the duplication, resulting in a rearranged gene order (Brandis and Hughes 2020) (fig. 1). Although the SNAP hypothesis is an attractive explanation for bacterial genome rearrangements, it has never been experimentally verified. Here we have tested the key predictions of the SNAP hypothesis, using *Salmonella*

Typhimurium as a model. We found that positive selection for adaptation to a novel environment resulted in the rapid and irreversible rearrangement of the bacterial chromosomal gene order.

Over the course of evolution, and associated with changes in the biological, chemical, and physical environment, bacteria will have frequently been under pressure to adapt to novel growth conditions such as shifts in temperature, pH, oxygen level, available carbon sources, the presence of antibacterial compounds and metals, and the evolving biological environment presenting new opportunities and challenges (Riehle et al. 2001; Kondrashov 2012; Dailey et al. 2017; Kjeldsen et al. 2019). According to evolutionary theory, the creation of new species is strongly associated with genetic isolation and adaptation to these new environmental niches (Baquero et al. 2021). Evolutionary adaptation by SNAP will result in rapid rearrangements in gene order as an inevitable by-product of this process. SNAP will contribute to genetically isolating the newly adapted population from their recent ancestors, by reducing the possibility of recombining DNA fragments acquired by horizontal genetic transfer, for example, conjugation or transformation. Importantly, this process of adaptation and genetic rearrangement can originate with a single founder cell, does not require large populations or the acquisition of rare mutations, and the entire process is driven rapidly by positive selection and through a series of high-frequency events.

Materials and Methods

Bacterial Strains and Growth Conditions

All strains are derived from *S. enterica* serovar Typhimurium strain LT2 (McClelland et al. 2001). See supplementary table S3, Supplementary Material online for a list of all strains used in this study. Bacteria were grown in a minimal malate medium (1 g/l K_2SO_4 , 13.5 g/l K_2HPO_4 , 4.7 g/l KH_2PO_4 , 0.1 g/l $MgSO_4 \cdot 7H_2O$, 10 mM NH_4Cl , and 0.2% L-malate) or Luria-Bertani (LB) broth (10 g/l tryptone, 5 g/l yeast extract, and 10 g/l NaCl [Oxoid, Basingstoke, UK]) at 37 °C for 24–48 h with aeration by shaking at 200 rpm. For solid plates, 15 g/l of improved essential agar-extra pure was supplied. All chemicals were purchased from Merck unless stated otherwise.

Selection of the Met^+ Duplication

Bacterial cultures were grown overnight in 1 ml LB and traces of the rich medium were removed by washing two times with 1 ml 0.9% NaCl. Cells were diluted in 0.9% NaCl, plated on minimal malate plates, and incubated at 37 °C for up to 48 h. Large colonies were restreaked on minimal malate plates and potential genetic changes were identified by whole-genome sequencing.

Strain Constructions

Genes involved in malate transport and metabolism were inactivated using DiRex (Nasvall 2017) or duplicated by

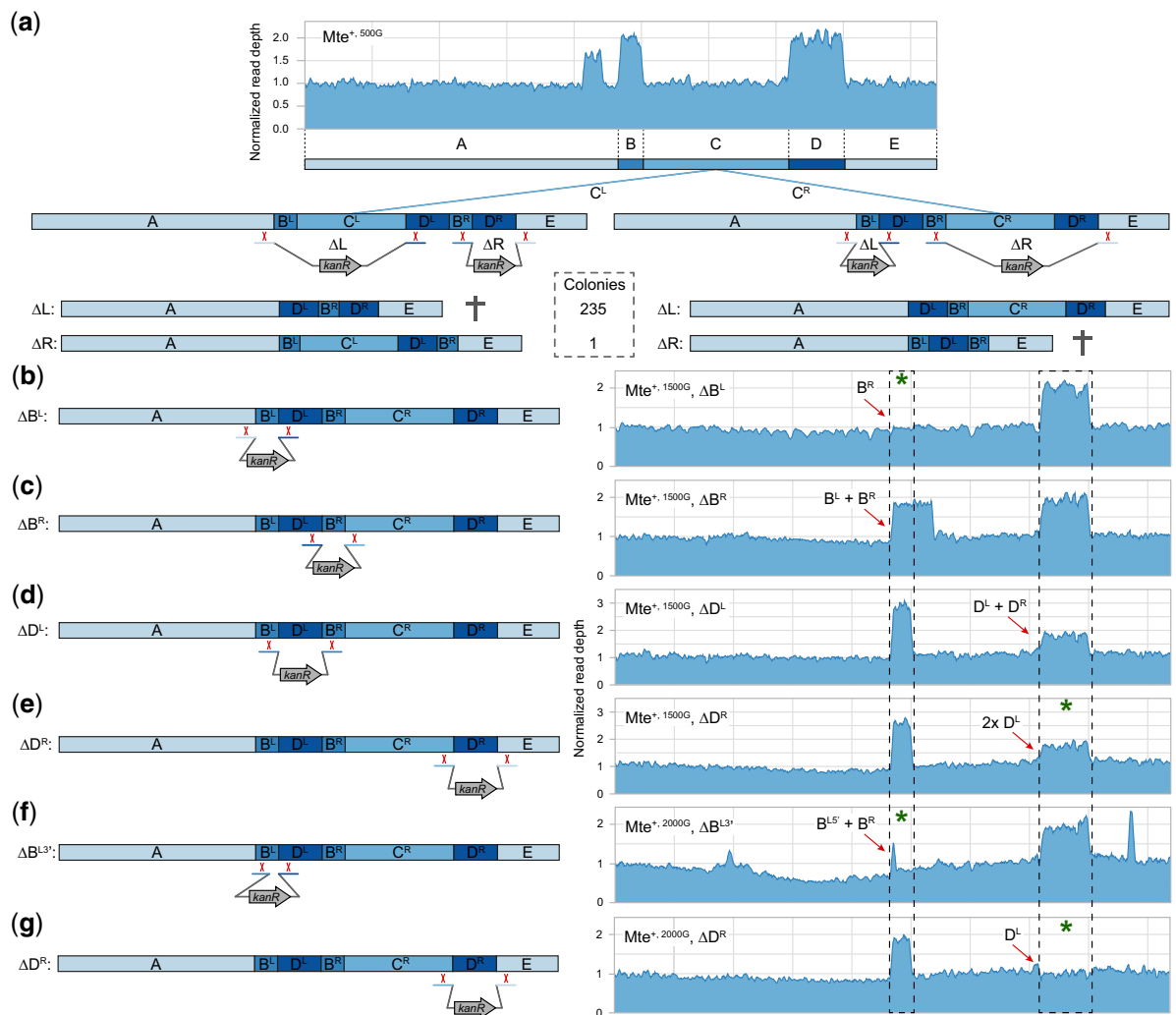


FIG. 4. Identification of the locations of the large deletion and the mutations in the evolved isolates. (a) The chromosome of the $Mte^{+, 500G}$ isolate was classified into five sections (a–e) and two deletions (ΔL and ΔR) were designed in order to identify the location of section C. The chromosomal order and theoretical result of the two deletions are shown for the case that section C is located in the left copy (left) or the right copy (right) of the duplication. The gray crosses indicate chromosomal structures that are nonviable due to the absence of essential genes. The dashed box shows the number of colonies acquired from the two transformation experiments. (b–g) Schematic overview of deletions designed to identify the location of mutations (left) and read depth coverage analysis of the resulting $Mte^{+, 1500G}$ (b–e) or $Mte^{+, 2000G}$ (f–g) isolates after transformation (right). The dotted black lines indicate the location of the originally duplicated section and the green stars indicate sequences that were used for the localization of the selected mutations.

inserting a kanamycin resistance gene (Nasvall et al. 2017). The *mutL* gene was inactivated by replacing it with a tetracycline resistance cassette using dsDNA lambda-red recombineering (Yu et al. 2000). The *mutL* deletion was moved into the evolved isolates using P22 phage transduction (Schmieger 1972). All oligonucleotides used in this study are shown in [supplementary table S4, Supplementary Material](#) online.

Whole-Genome Sequencing

Bacterial genomic DNA was prepared using MasterPure DNA Purification kit (Epicentre, Illumina Inc., CA, USA) according to the instruction of the manufacturer. DNA libraries were assembled using Nextera XT library preparation and index kits (Illumina Inc., CA, USA) according to manufacturer instructions. The libraries were

sequenced on a Miseq device using a 600-cycle V3 reagent kit (Illumina Inc., CA, USA). Reads were processed, aligned, and analyzed using CLC Genomics Workbench V9 (CLCbio, Qiagen, Denmark).

Mutation Frequency Measurements

The mutation frequency of wild-type and $\Delta mutL$ *Salmonella* was measured by plating $\sim 4 \times 10^8$ cells of overnight cultures on minimal malate plates containing 100 mg/l rifampicin. Each measurement was performed with 10 independent replicates.

Experimental Evolution

Evolution in liquid medium was performed with 100 ml minimal malate medium in 300 ml E-flasks with

Table 2. Location of mutations in the duplication region of the final $Mte^{+, 2000G}$ isolate.

Gene ^a	Left copy ^b	Right copy ^b
<i>gyrA</i> ^c	wt	Glu575*
<i>nrdB</i> ^c	wt	Ile147Thr
<i>nuoG</i>	wt	Ala411Val
<i>yfbS</i>	*609Gln	wt
<i>ackA</i>	Gln111Arg	wt
<i>folC</i> ^c	Pro360fs	wt
<i>pgtE</i>	Glu80Gly	wt
<i>yfdZ-yheS</i> ^d	ΔyfdZ-yheS	wt
<i>yhfC</i>	Gly356Asp	wt
<i>nirB</i>	wt	Val798Ala
<i>bigA</i>	His802fs	wt
<i>mrcA</i>	wt	ΔArg534-Met536
<i>yhgF</i>	Leu706Pro	wt
<i>malP</i>	wt	Thr16Met
<i>livK</i>	wt	ΔVal84-Ala85
<i>yhiH</i>	wt	Val18fs
<i>dppF</i>	Pro305Ser	wt
STM3631	Wt	Gly89fs
<i>yhjW</i>	Tyr268Cys	wt
<i>glyS</i> ^c	wt	Gln357*
<i>avtA</i>	wt	Met284fs
<i>rfaC</i>	wt	Val312Ile
<i>dfp</i> ^c	Leu146Pro	wt
<i>ligB</i>	wt	Gln345*
<i>spoT</i> ^c	wt	Glu444Lys
STM3773	wt	Thr109Ala
STM3781	wt	Pro70Leu
<i>uhpT</i>	Trp164*	wt
STM3796A.S	Gly141Ser	wt

*Indicates that the mutation creates a nonsense codon (a translation termination codon in the mRNA).

^aAll genes are listed in their linear order on the *Salmonella* chromosome.

^bLeft and right copy are defined based on the linear order of the *Salmonella* chromosome.

^cThe gene has been identified to be essential for growth in LB (Baba et al. 2006; Goodall et al. 2018).

^dThe segment contains genes that have been identified to be essential for growth in LB (Baba et al. 2006; Goodall et al. 2018).

continuously shaking at 37 °C. The lineages were serially passaged after each cycle of growth by transferring 100 μL of the grown cultures into 100 ml fresh medium (corresponding to 10 generations of growth per cycle). After 50 cycles (500 generations of growth), cultures were plated on minimal malate plates and a single colony of each lineage was isolated for further analysis. Single-cell bottleneck evolution was performed by serial restreaking on minimal malate plates with each cycle corresponding to ~25 generations of growth. Plates were incubated at 37 °C for 48 h. For each passage, the second last single colony of the streak on the plate was picked and restreaked on a fresh minimal malate plate. After 25 cycles (corresponding 500 generations of evolution), a single colony of each lineage was isolated for further analysis. A detailed scheme of the evolution experiment is shown in [supplementary fig. S5, Supplementary Material](#) online.

Growth Rate Measurements

Exponential growth rates were measured using a Bioscreen C machine (Oy Growth curves Ab Ltd). To prepare

bacteria, wild-type LT2 was streaked on LA plates, the strains with single-gene duplications were grown on LA with kanamycin at 50 mg/l and strains containing the Met^{+} duplication were grown on minimal malate plates. All strains were incubated at 37 °C until colonies of ~1 mm diameter were formed. Cells were scraped from each plate and resuspended in 0.9% NaCl to an OD_{600nm} of 0.8. The bacterial suspensions were then diluted 2000-fold in a minimal malate medium. 300 μl of each diluted culture were incubated at 37 °C with continuous shaking in honeycomb microtiter plates. Optical density (600 nm) was measured at 5 min intervals. Doubling times were calculated from the increase in optical density over a sliding 10 measurement points window. Maximum exponential growth rates were defined as the measurement window with the highest doubling time. Final optical densities were measured after 18 h of growth. All results are the average of five biological replicates.

Duplication Stability Measurements

Three independent cultures of the isolates Mte^{+} , $Mte^{+, 500G}$, $Mte^{+, 1000G}$, $Mte^{+, 1500G}$, and $Mte^{+, 2000G}$ were grown in 15 ml falcon tubes containing 1 ml LB medium. Each culture was serially passaged for 10 cycles by transferring 1 μl of overnight culture into 1 ml fresh LB (total of 100 generations). After cycles 1, 2, 3, 5, and 10, ~100 cfu of each lineage were plated on minimal malate plates (Mte^{+}) or LA plates (others) and incubated for 24 (LA) or 48 h (minimal malate) at 37 °C. For the Mte^{+} isolate on the minimal malate plates, the presence of the duplication was assessed by colony size. For the other isolates on LA plates, the presence of the duplication was tested by PCR across the unique junction present in the duplication. For each lineage and time point, 10 colonies were tested by PCR.

Supplementary Material

[Supplementary data](#) are available at *Molecular Biology and Evolution* online.

Acknowledgements

This work was supported by grants to D.H. from the Swedish Science Research Council (Vetenskapsrådet, grant numbers 2017-03953 and 2021-04814) and the Carl Trygger Foundation (grant numbers CTS20:190 and CTS21:1237). The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication. The authors declare no competing interests.

Data Availability

All genome sequences have been deposited in the Sequence Read Archive of the NCBI under the project accession number PRJNA783729. All data are provided with this paper and bacterial strains are available upon request.

Author Contributions

G.B. and D.H. designed the overall experimental plan. S.C. and D.L.H. performed the experiments and analyzed the data. All authors reviewed and edited the manuscript. D.H. acquired funding.

References

- Baba T, Ara T, Hasegawa M, Takai Y, Okumura Y, Baba M, Datsenko KA, Tomita M, Wanner BL, Mori H. 2006. Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol Syst Biol.* **2**:2006.0008.
- Baquero F, Coque TM, Galan JC, Martinez JL. 2021. The origin of niches and species in the bacterial world. *Front Microbiol.* **12**:657986.
- Baumler DJ, Ma B, Reed JL, Perna NT. 2013. Inferring ancient metabolism using ancestral core metabolic models of enterobacteria. *BMC Syst Biol.* **7**:46.
- Belda E, Moya A, Silva FJ. 2005. Genome rearrangement distances and gene order phylogeny in gamma-Proteobacteria. *Mol Biol Evol.* **22**:1456–1467.
- Blount ZD, Barrick JE, Davidson CJ, Lenski RE. 2012. Genomic analysis of a key innovation in an experimental *Escherichia coli* population. *Nature* **489**:513–518.
- Bognar AL, Osborne C, Shane B, Singer SC, Ferone R. 1985. Folylpolyl-gamma-glutamate synthetase-dihydrofolate synthetase. Cloning and high expression of the *Escherichia coli* folC gene and purification and properties of the gene product. *J Biol Chem.* **260**:5625–5630.
- Brandis G. 2021. Reconstructing the evolutionary history of a highly conserved operon cluster in *Gammaproteobacteria* and *Bacilli*. *Genome Biol Evol.* **13**:evab041.
- Brandis G, Cao S, Hughes D. 2018. Co-evolution with recombination affects the stability of mobile genetic element insertions within gene families of *Salmonella*. *Mol Microbiol.* **108**:697–710.
- Brandis G, Hughes D. 2020. The SNAP hypothesis: chromosomal rearrangements could emerge from positive selection during Niche Adaptation. *PLoS Genet.* **16**:e1008615.
- Dailey HA, Dailey TA, Gerdes S, Jahn D, Jahn M, O'Brian MR, Warren MJ. 2017. Prokaryotic heme biosynthesis: multiple pathways to a common essential product. *Microbiol Mol Biol Rev.* **81**:e00048-16.
- Darling AE, Miklos I, Ragan MA. 2008. Dynamics of genome rearrangement in bacterial populations. *PLoS Genet.* **4**:e1000128.
- Darmon E, Leach DR. 2014. Bacterial genome instability. *Microbiol Mol Biol Rev.* **78**:1–39.
- Davies SJ, Golby P, Omrani D, Broad SA, Harrington VL, Guest JR, Kelly DJ, Andrews SC. 1999. Inactivation and regulation of the aerobic C₄-dicarboxylate transport (*dctA*) gene of *Escherichia coli*. *J Bacteriol.* **181**:5624–5635.
- Doolittle RF, Feng DF, Tsang S, Cho G, Little E. 1996. Determining divergence times of the major kingdoms of living organisms with a protein clock. *Science* **271**:470–477.
- Elez M, Radman M, Matic I. 2007. The frequency and structure of recombinant products is determined by the cellular level of MutL. *Proc Natl Acad Sci U S A* **104**:8935–8940.
- Ely B, Wilson K, Ross K, Ingram D, Lewter T, Herring J, Duncan D, Aikins A, Scott D. 2019. Genome comparisons of wild isolates of *Caulobacter crescentus* reveal rates of inversion and horizontal gene transfer. *Curr Microbiol.* **76**:159–167.
- Finn RD, Coggill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, Potter SC, Punta M, Qureshi M, Sangrador-Vegas A, et al. 2016. The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res.* **44**:D279–D285.
- Fitzgerald SF, Lupolova N, Shaaban S, Dallman TJ, Greig D, Allison L, Tongue SC, Evans J, Henry MK, McNeilly TN, et al. 2021. Genome structural variation in *Escherichia coli* O157:H7. *Microb Genom.* **7**:000682.
- Gerstein M, Lesk AM, Chothia C. 1994. Structural mechanisms for domain movements in proteins. *Biochemistry* **33**:6739–6749.
- Goodall ECA, Robinson A, Johnston IG, Jabbari S, Turner KA, Cunningham AF, Lund PA, Cole JA, Henderson IR. 2018. The essential genome of *Escherichia coli* K-12. *mBio* **9**:e02096-17.
- Hoegler KJ, Hecht MH. 2018. Artificial gene amplification in *Escherichia coli* reveals numerous determinants for resistance to metal toxicity. *J Mol Evol.* **86**:103–110.
- Hufnagel DA, Choby JE, Hao S, Johnson AF, Burd EM, Langelier C, Weiss DS. 2021. Antibiotic-selected gene amplification heightens metal resistance. *mBio* **12**:e02994-20.
- Hughes D. 2000. Evaluating genome dynamics: the constraints on rearrangements within bacterial genomes. *Genome Biol.* **1**:REVIEWS0006.
- Kjeldsen KU, Schreiber L, Thorup CA, Boesen T, Bjerg JT, Yang T, Dueholm MS, Larsen S, Risgaard-Petersen N, Nierychlo M, et al. 2019. On the evolution and physiology of cable bacteria. *Proc Natl Acad Sci U S A* **116**:19116–19125.
- Kondrashov FA. 2012. Gene duplication as a mechanism of genomic adaptation to a changing environment. *Proc Biol Sci.* **279**:5048–5057.
- Koonin EV, Makarova KS, Wolf YI. 2021. Evolution of microbial genomics: conceptual shifts over a quarter century. *Trends Microbiol.* **29**:582–592.
- Koonin EV, Mushegian AR, Rudd KE. 1996. Sequencing and analysis of bacterial genomes. *Curr Biol.* **6**:404–416.
- Mao D, Grogan DW. 2017. How a genetically stable extremophile evolves: modes of genome diversification in the archaeon *Sulfolobus acidocaldarius*. *J Bacteriol.* **199**:e00177-17.
- Marin J, Battistuzzi FU, Brown AC, Hedges SB. 2017. The timetree of prokaryotes: new insights into their evolution and speciation. *Mol Biol Evol.* **34**:437–446.
- Matthews TD, Rabsch W, Maloy S. 2011. Chromosomal rearrangements in *Salmonella enterica* serovar Typhi strains isolated from asymptomatic human carriers. *mBio* **2**:e00060-11.
- McClelland M, Sanderson KE, Spieth J, Clifton SW, Latreille P, Courtney L, Porwollik S, Ali J, Dante M, Du F, et al. 2001. Complete genome sequence of *Salmonella enterica* serovar Typhimurium LT2. *Nature* **413**:852–856.
- Nagel GM, Cumberledge S, Johnson MS, Petrella E, Weber BH. 1984. The beta subunit of *E. coli* glycyl-tRNA synthetase plays a major role in tRNA recognition. *Nucleic Acids Res.* **12**:4377–4384.
- Nasvall J. 2017. Direct and Inverted Repeat stimulated excision (DIRex): simple, single-step, and scar-free mutagenesis of bacterial genes. *PLoS One* **12**:e0184126.
- Nasvall J, Knoppel A, Andersson DI. 2017. Duplication-Insertion Recombineering: a fast and scar-free method for efficient transfer of multiple mutations in bacteria. *Nucleic Acids Res.* **45**:e33.
- Näsval J, Sun L, Roth JR, Andersson DI. 2012. Real-time evolution of new genes by innovation, amplification, and divergence. *Science* **338**:384–387.
- Noureen M, Tada I, Kawashima T, Arita M. 2019. Rearrangement analysis of multiple bacterial genomes. *BMC Bioinform.* **20**:631.
- Ochman H, Wilson AC. 1987. Evolution in bacteria: evidence for a universal substitution rate in cellular genomes. *J Mol Evol.* **26**:74–86.
- Puigbò P, Wolf YI, Koonin EV. 2010. The tree and net components of prokaryote evolution. *Genome Biol Evol.* **2**:745–756.
- Repar J, Warnecke T. 2017. Non-random inversion landscapes in prokaryotic genomes are shaped by heterogeneous selection pressures. *Mol Biol Evol.* **34**:1902–1911.
- Riehle MM, Bennett AF, Long AD. 2001. Genetic architecture of thermal adaptation in *Escherichia coli*. *Proc Natl Acad Sci U S A* **98**:525–530.
- Rocha EP. 2006. Inference and analysis of the relative stability of bacterial chromosomes. *Mol Biol Evol.* **23**:513–522.
- Schmieger H. 1972. Phage P22-mutants with increased or decreased transduction abilities. *Mol Gen Genet.* **119**:75–88.

- Scott D, Ely B. 2016. Conservation of the essential genome among *Caulobacter* and *Brevundimonas* species. *Curr Microbiol.* **72**: 503–510.
- Shelyakin PV, Bochkareva OO, Karan AA, Gelfand MS. 2019. Micro-evolution of three *Streptococcus* species: selection, antigenic variation, and horizontal gene inflow. *BMC Evol Biol.* **19**:83.
- Skovgaard O, Bak M, Løbner-Olesen A, Tommerup N. 2011. Genome-wide detection of chromosomal rearrangements, indels, and mutations in circular chromosomes by short read sequencing. *Genome Res.* **21**:1388–1393.
- Sonti RV, Roth JR. 1989. Role of gene duplications in the adaptation of *Salmonella typhimurium* to growth on limiting carbon sources. *Genetics* **123**:19–28.
- Straus DS, Hoffmann GR. 1975. Selection for a large genetic duplication in *Salmonella typhimurium*. *Genetics* **80**:227–237.
- Thöny-Meyer L, Fischer F, Künzler P, Ritz D, Hennecke H. 1995. *Escherichia coli* genes required for cytochrome c maturation. *J Bacteriol.* **177**:4321–4326.
- Uden G, Strecker A, Kleefeld A, Kim OB. 2016. C4-dicarboxylate utilization in aerobic and anaerobic growth. *EcoSal Plus* **7**. doi: 10.1128/ecosalplus.ESP-0021-2015.
- van der Rest ME, Frank C, Molenaar D. 2000. Functions of the membrane-associated and cytoplasmic malate dehydrogenases in the citric acid cycle of *Escherichia coli*. *J Bacteriol.* **182**:6892–6899.
- Wang D, Li S, Guo F, Ning K, Wang L. 2017. Core-genome scaffold comparison reveals the prevalence that inversion events are associated with pairs of inverted repeats. *BMC Genom.* **18**:268.
- Worth LJ, Clark S, Radman M, Modrich P. 1994. Mismatch repair proteins MutS and MutL inhibit RecA-catalyzed strand transfer between diverged DNAs. *Proc Natl Acad Sci U S A* **91**:3238–3241.
- Xu T, Qin S, Hu Y, Song Z, Ying J, Li P, Dong W, Zhao F, Yang H, Bao Q. 2016. Whole genomic DNA sequencing and comparative genomic analysis of *Arthrospira platensis*: high genome plasticity and genetic diversity. *DNA Res.* **23**:325–338.
- Yu D, Ellis HM, Lee EC, Jenkins NA, Copeland NG, Court DL. 2000. An efficient recombination system for chromosome engineering in *Escherichia coli*. *Proc Natl Acad Sci U S A* **97**:5978–5983.