Received: 13 June 2023 | Revised: 20 November 2023 | Accepted: 24 November 2023

DOI: 10.1111/mec.17234

ORIGINAL ARTICLE

MOLECULAR ECOLOGY WILEY

Forward-in-time simulation of chromosomal rearrangements: The invisible backbone that sustains long-term adaptation

Paul Banse[1] | Juliette Luiselli[1] | David P. Parsons[1] | Théotime Grohens[2] | Marco Foley[1] | Leonardo Trujillo[1] | Jonathan Rouzaud-Cornabas[1] | Carole Knibbe[3] | Guillaume Beslon[1]

[1]Université de Lyon, INSA-Lyon, Inria, CNRS, Université Claude Bernard Lyon 1, ECL, Université Lumière Lyon 2, LIRIS UMR5205, Lyon, France

[2]Centre for Genomic Regulation (CRG), The Barcelona Institute of Science and Technology, Barcelona, Spain

[3]Université de Lyon, INSA-Lyon, Inria, Université Claude Bernard Lyon 1, Inserm, INRAE, CarMeN laboratory, Pierre-Bénite, France

Correspondence
Guillaume Beslon, Université de Lyon, INSA-Lyon, Inria, CNRS, Université Claude Bernard Lyon 1, ECL, Université Lumière Lyon 2, LIRIS UMR5205, Lyon, F-69621, France.
Email: guillaume.beslon@inria.fr

Funding information
Agence Nationale de la Recherche, Grant/Award Number: ANR19-CE45-0010

Handling Editor: Kay Lucek

Abstract

While chromosomal rearrangements are ubiquitous in all domains of life, very little is known about their evolutionary significance, mostly because, apart from a few specifically studied and well-documented mechanisms (interaction with recombination, gene duplication, etc.), very few models take them into account. As a consequence, we lack a general theory to account for their direct and indirect contributions to evolution. Here, we propose Aevol, a forward-in-time simulation platform specifically dedicated to unravelling the evolutionary significance of chromosomal rearrangements (CR) compared to local mutations (LM). Using the platform, we evolve populations of organisms in four conditions characterized by an increasing diversity of mutational operators—from substitutions alone to a mix of substitutions, InDels and CR—but with a constant global mutational rate. Despite being almost invisible in the phylogeny owing to the scarcity of their fixation in the lineages, we show that CR make a decisive contribution to the evolutionary dynamics by comparing the outcome in these four conditions. As expected, chromosomal rearrangements allow fast expansion of the gene repertoire through gene duplication, but they also reduce the effect of diminishing-returns epistasis, hence sustaining adaptation on the long-run. At last, we show that chromosomal rearrangements tightly regulate the size of the genome through indirect selection for reproductive robustness. Overall, these results confirm the need to improve our theoretical understanding of the contribution of chromosomal rearrangements to evolution and show that dedicated platforms like Aevol can efficiently contribute to this agenda.

KEYWORDS
chromosomal rearrangements, evolution, InDels, modelling, simulation

Paul Banse and Juliette Luiselli are joint first author.

1 | INTRODUCTION

Genomic structural variations occur in all domains of life, including viruses, prokaryotes and the full range of eukaryotic taxa (Alkan et al., 2011; Cao et al., 2022; Darling et al., 2008; Gao et al., 2017). These structural variations include insertions of transposable elements, recombinations, and chromosomal rearrangements. Although the precise definition of chromosomal rearrangements varies across references (Alkan et al., 2011; Audrézet et al., 2004; Mérot et al., 2020), they generally refer to inversions, translocations, duplications, and deletions of DNA segments. Chromosomal rearrangements have classically been a blind spot of molecular evolution, mainly due to technical issues linked to short-reads sequencing but also due to their strong deleterious effects that can rapidly eliminate them from the population (Campo et al., 2004; Connallon & Olito, 2022; Kara et al., 2014; Rocha, 2006). Nevertheless, recent improvements in sequencing techniques have strongly increased our ability to detect them (Hanlon et al., 2022; Ho et al., 2020; Wala et al., 2018), and more and more data is being accumulated regarding their decisive impact on evolution, as highlighted in the 2019 special issue published by *Molecular Ecology* (Wellenreuther et al., 2019). It appears that duplications and deletions are far from rare in eukaryotes. In some cases, the per locus gene duplication rate can be higher than the per nucleotide substitution rate (Katju & Bergthorsson, 2013), resulting in one gene duplication per haploid genome every 50 generations in the yeast *S. Cerevisiae* (Lynch et al., 2008), and every 500 generations in the fruit fly *D. Melanogaster* (Schrider et al., 2013). In the human genome, many duplications and large deletions have been identified as causes of genetic diseases or cancers (Nattestad et al., 2018). In prokaryotes, Richard Lenski's Long Term Evolution Experiment (LTEE) has shown the importance of large scale rearrangements as drivers of genomic plasticity (Raeside et al., 2014) and innovation (Blount et al., 2012).

While new sequencing techniques and discoveries have shed a new light on chromosomal rearrangements (Ho et al., 2020; Quandt et al., 2015), theoretical frameworks have been slow to adapt. Indeed, the effect of chromosomal rearrangements is generally not addressed in theoretical articles and textbooks. In most models of evolution, substitutions are still the sole source of variation, with recombination merely expected to shuffle these variations among individuals (Weissman et al., 2010). In the rare cases where ectopic recombination is considered in evolutionary models, its effect is generally limited to gene permutations or variation of copy number, excluding a priori any effect on gene sequences themselves (Bhatia et al., 2018; Yancopoulos et al., 2005). Similarly, inversions are often viewed as just an evolutionary pathway that prevents recombination, hybridization and introgression (Noor et al., 2001), thus keeping specific alleles together (Hoffmann et al., 2004; Kirkpatrick, 2010). Nevertheless, the ubiquity of these rearrangements (Raeside et al., 2014; Wellenreuther & Bernatchez, 2018) calls for more in-depth studies of their potential other effects.

There are several reasons for chromosomal rearrangements not to be accounted for in classical evolutionary models. First, contrary to substitutions and InDels that act at the allelic scale, chromosomal rearrangements are multi-scale events that can modify both the micro- and the macro-structure of the genome (i.e. the allelic sequences and the global organization of the genome), while most models simulate genes as unbreakable units, with different alleles but no explicit sequences (Bhatia et al., 2018; Weissman et al., 2010; Yancopoulos et al., 2005).

Second, chromosomal rearrangements entail a wide diversity of complex effects, notably due to their length distribution which spans several orders of magnitude, from a few base pairs to a substantial fraction of the genome (Darling et al., 2008), contrary to for example, InDels, which length distribution is narrower. As a consequence, rearrangements can significantly modify the genome size, thus changing the overall probability of another rearrangement, as bigger chromosomes generally undergo more rearrangements (Jensen-Seaman et al., 2004; Kaback et al., 1992). As a consequence, successive chromosomal rearrangements should not be considered independent: the occurrence of a rearrangement is likely to change the rate and Distribution of Fitness Effects (DFE) of upcoming events.

The variety and complexity of chromosomal rearrangements makes it challenging to build a theoretical understanding of their effect on evolution. In this context, forward-in-time simulations are a promising tool to observe the effect of rearrangements and unravel their importance in adaptation to new environments (Mérot et al., 2020). However, in forward-in-time models—like the well-known SLiM (Haller & Messer, 2017), the effect of mutations is often either an allelic change, drawn from a predefined DFE, or a positional change of the gene. This prevents these models from considering any combination of small- and large-scale effects, and makes it difficult to account for non-independent events (where some kinds of events modify the DFEs of others). To overcome these difficulties, a model designed to study rearrangements should not rely on explicit a priori DFEs. On the opposite, the mutations should affect the pre-existing genome sequence, without regards for the phenotypic effect, which is computed after the mutation. In this way, the effect of a mutation depends on its characteristics (type, location, length), but also on the current genomic structure, the environment and the genotype-to-phenotype map.

Hence, a model designed to study chromosomal rearrangements should provide an explicit genome with both coding and non-coding regions, in which rearrangements can happen blindly and have both direct (when altering coding regions) and indirect (when modifying the DFE of the different mutational operators—including rearrangements themselves) effects on fitness.

In this article, we use Aevol, a model addressing these requirements. Aevol is a forward-in-time simulation platform that emulates the evolution of prokaryotic-like organisms and enables repeated evolution experiments with adjustable parameters (Knibbe et al., 2007). Although the model has been presented before (Batut et al., 2013; Liard et al., 2020; Parsons, 2011; Rutten et al., 2019), recent computational and methodological improvements have opened up a wide range of new possibilities for the

software. Aevol allows for both local mutations and chromosomal rearrangements of the genetic sequence, without an a priori DFE. We propose a use-case of the software to highlight the importance of chromosomal rearrangements in genome evolution. To this end, we simulate evolution under multiple mutational scenarios of increasing complexity: with substitutions only, with local mutations only (mutations that can only alter the sequence at the allelic scale: substitutions, small Insertions and small Deletions), and with a full range of mutational operators, including local mutations and chromosomal rearrangements (duplications, deletions, and inversions). Also, in order to test whether chromosomal rearrangements can generate enough diversity on their own to enable efficient adaptation, we added a fourth scenario where only chromosomal rearrangements are present, without any kind of local mutation. These scenarios are repeated with two types of populations, one starting far from the fitness optimum and one starting close to it.

Our simulations first show that, when far from the optimum, chromosomal rearrangements are an essential component of evolution, and even more important than local mutations. Indeed, by the end of the simulation, populations evolved with solely chromosomal rearrangements are far better adapted than populations evolved with local mutations or substitutions only. Moreover, the simulations also show that the evolution of genetic structure—including the genome size—is very different when rearrangements are allowed, emphasizing their role in the regulation of the amount of DNA (Knibbe et al., 2007). Simulations starting close to the fitness optimum confirm the latter effect but also demonstrate that, on the long term, chromosomal rearrangements reduce the effect of diminishing-returns epistasis, defined as the speed at which the marginal improvement of beneficial mutations decreases at each improvement (Wiser et al., 2013). Taken together, these simulations emphasize the decisive contribution of chromosomal rearrangements to long-term evolution and show the potential of the Aevol platform to study their evolutionary impact.

2 | MATERIALS AND METHODS

2.1 | Aevol: A forward-in-time evolutionary simulator with complex mutations

Aevol (https://www.aevol.fr) is a forward-in-time evolutionary simulator that simulates the evolution of a population of haploid organisms through a process of variation and selection (Batut et al., 2013; Beslon et al., 2010; Frenoy et al., 2013; Knibbe et al., 2007; Parsons et al., 2010). Each artificial organism, similarly to prokaryotes, is asexual, haploid and owns a single circular chromosome. The design of the model focuses on the realism of the genome structure and of the mutational process. Aevol can therefore be used to decipher the effect of chromosomal rearrangements on genome evolution, including their interactions with other types of mutational events.

In short, Aevol is made of three components (Figure 1a):

- A mapping that decodes the genomic sequence of an individual into a phenotype. The genomic sequence of each organism is a double-stranded circular binary sequence. Reading this sequence enables us to identify start-stop locus of transcription and translation, thus delimiting open-reading frames. These are genes that are then decoded into proteins, represented by mathematical functions which sum represents the phenotype. Finally, the phenotype is compared to an environmental target, and their difference is used to compute the individual's fitness value.
- A population of organisms, each owning its own genome, hence its own phenotype and fitness. These individuals are located on a grid with one individual per grid cell. At each generation, the organisms are selected according to their fitness to populate the next generation. By default the competition is local (each organism competing with its neighbours), although other selection modes are possible.
- A genome replication process during which genomes can undergo several kinds of mutational events. These include chromosomal rearrangements and local mutations, but no recombination in the current version. The seven modelled types of mutation are depicted in Figure 1b and comprise three local mutations: substitutions, small insertions and small deletions; two balanced rearrangements (which conserve the genome size), inversions and translocations; and two unbalanced rearrangements, duplications and deletions. This allows the user to study the effect of chromosomal rearrangements and their interaction with other kinds of events such as substitutions and InDels. The position of the mutations and the breakpoints of the rearrangements are chosen uniformly along the genome. Hence, longer chromosomes can undergo longer rearrangements. By contrast, InDels have a predefined length distribution (1 to 6 bp by default).

A detailed presentation of the model is available in the Figure S1.

2.2 | In silico experimental set-up: Evolution with limited mutations

2.2.1 | Experiment starting from naive individuals

We run 11 replicate simulations for four types of conditions: substitutions only (SUB), local mutations only (LM—substitutions and InDels), chromosomal rearrangements only (CR—duplications, deletions and inversions), and both chromosomal rearrangements and local mutations (CRLM). Note that translocations, although possible in Aevol, are excluded here to have as many local mutations as chromosomal rearrangements, and so a constant per base mutation rate in our different set-ups. The median (in terms of final fitness) CRLM run will be used to start the second set of simulations. The simulations begin with naive individuals owning a single gene. We want to study lineages for 1,000,000 generations, which is enough to reach

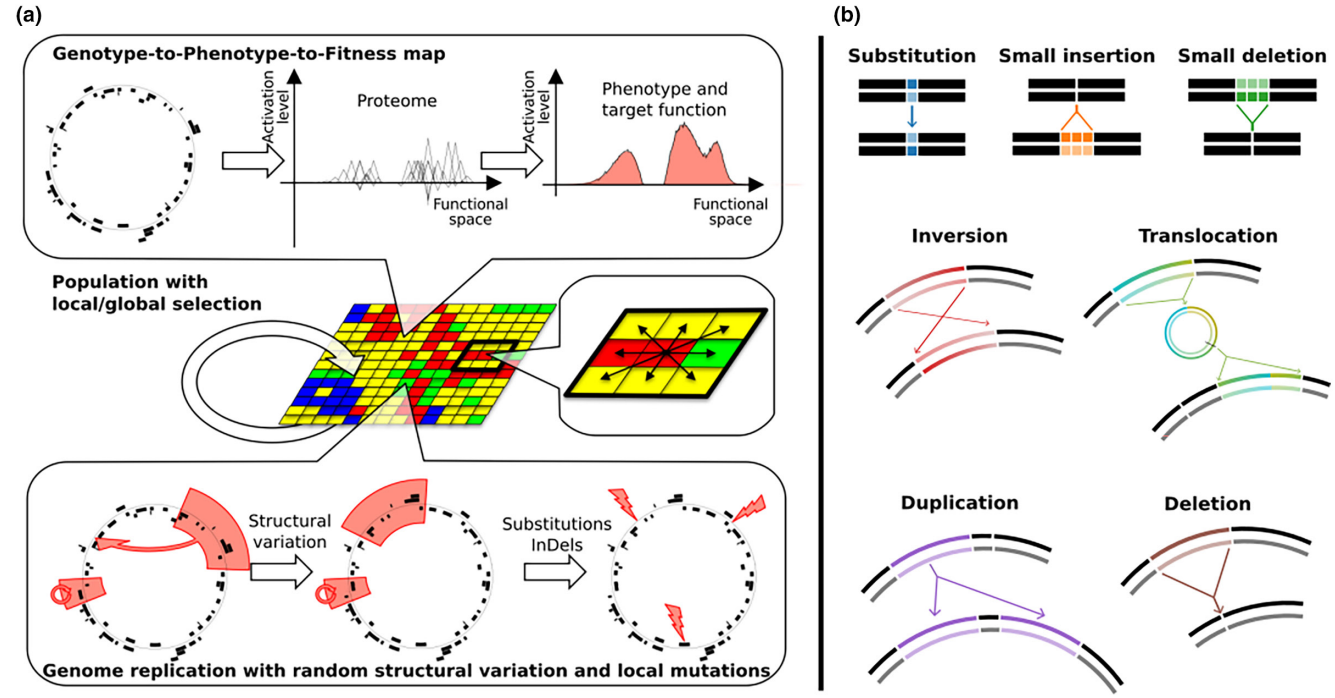

FIGURE 1 The Aevol model. The left panel (a) shows all steps of a generation in Aevol. (*top*) Overview of the genotype-to-phenotype map. Note that the organism shown here is a real organism evolved within Aevol for 1,000,000 generations with a typical target. It contains many Open-Reading Frames on both strands, a large proteome (the set of proteins), and it is well adapted to its environment (i.e. its phenotypic function — black curve — is very close to the target function — light red area). (*middle*) Population on a grid is fully renewed every generation. Example of a local selection process occurring with a 3×3 neighbourhood. (*bottom*) Mutation operators include chromosomal rearrangements (duplications, deletions, translocations and inversions – here a translocation and an inversion are shown) and local mutations (substitutions and InDels). These mutations are described more precisely in the right panel (b): (*top*) Local mutations: substitution (one base pair is mutated to another), small insertion and small deletion (a few base pairs are inserted or deleted). (*middle*) Balanced chromosomal rearrangements: inversion (two points are drawn and the segment in between is rotated) and translocation (a segment is excised, circularized, re-cut and inserted elsewhere in the genome). (*bottom*) Unbalanced chromosomal rearrangements: duplication (copy-paste of a segment in the genome) and deletion (suppression of a segment of the genome).

a stable genome with no more large variations in genome size and structure—although there is still room for adaptation. To this end, we run the simulations for 1,100,000 generations, the last 100,000 being used to ensure the survival of the lineage we retrieve.

All replicates share the same population size (1024 individuals on a 32×32 square grid), the same environment (a sum of three Gaussian lobes, see Figure 2 and Figure S3) and the same selection mode (local competition against the direct neighbours). The only difference lays in the mutation rates, as shown in Table 1. Importantly, for each condition, mutation rates are equally balanced between all mutation types and adjusted such that the overall mutation probability per locus is constant throughout all experiments. An example parameter file for the CRLM set-up is provided in the Figure S3.

For every simulation, we reconstruct the final lineage by tracking the ancestry of an individual from the final generation. We then retrieve the fitness, genome size, coding and non-coding sizes, and number of genes of all the individuals in this lineage. We also extract all mutations along the lineage, and record their type and effect on fitness.

Finally, along the line of descent of the 11 CRLM experiments, we extracted the 11 individuals at generation 1,000,000 and selected the median one (in term of fitness) to estimate its distribution of fitness effect (DFE) for each type of mutation. This allows to better understand the differences between local mutations and chromosomal rearrangements in terms of impact on the fitness and chances of fixation. Note that this individual is the same that to one used to initiate the second run of experiments (see below).

2.2.2 | Evolution from wild types

After 1,000,000 generations, individuals are well-adapted to their environment, especially in the CRLM experiments. They can be used as wild types to start new experiments. Here, the median CRLM experiment (in terms of final fitness) is used to initialize new clonal populations to test evolution from a well-adapted genome in the four mutational scenarios (SUB, LM, CR and CRLM). These populations are then evolved for another 3,100,000 generations to study the impact of chromosomal rearrangements when individuals are already well adapted to the environmental conditions. The same processing as for the first part of the experiments is then performed: reconstruction of the lineage for 3,100,000 generations and analysis of the genomes and mutations from generation 0 to 3,000,000 along this lineage (generations 3,000,001 to 3,100,000 being removed to ensure coalescence).

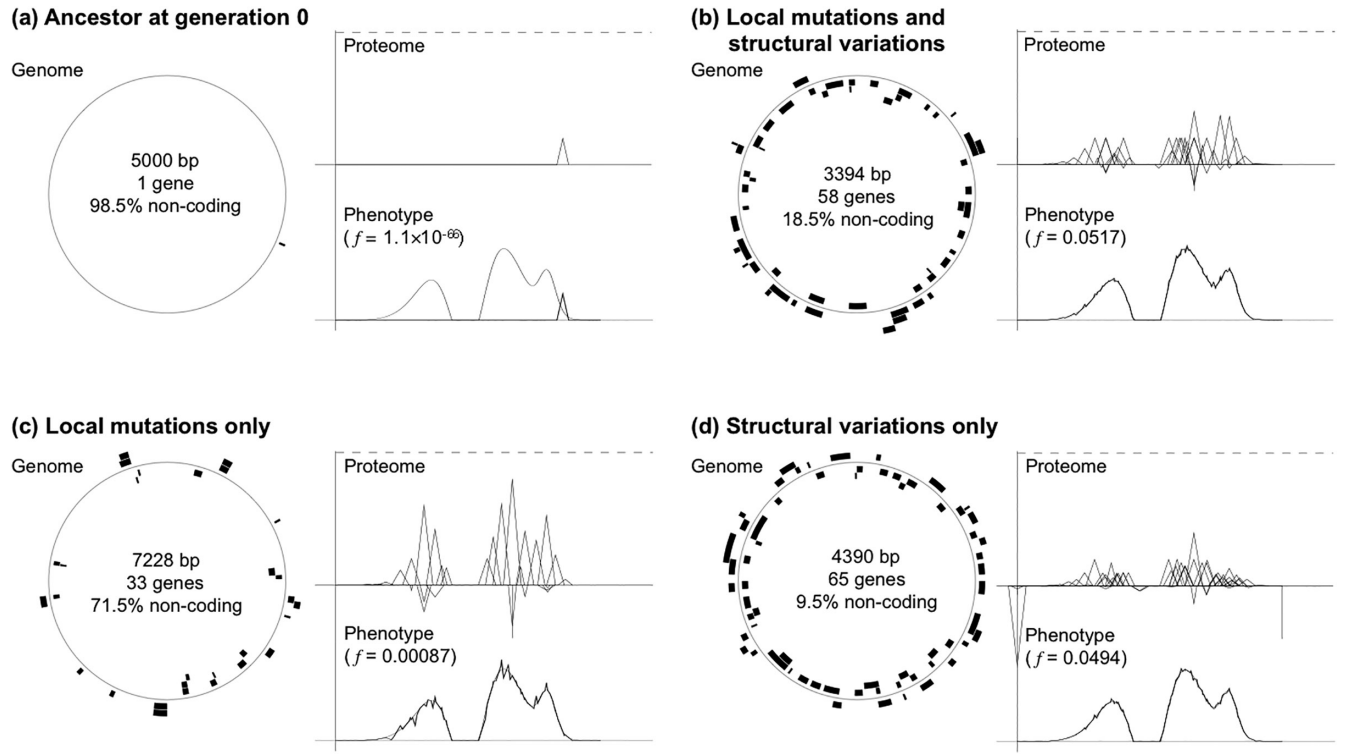

FIGURE 2 Initial ancestor (a) and examples of evolved organisms in the CRLM (b), LM (c) and CR (d) conditions after 1,000,000 generations. The organism presented in (b) corresponds to the Wild Type used for the second step of the experiments. For each organism, there is on the left a visualization of its genes localized on the genome. On the right, the proteome shows all the single proteins, and the phenotype (black curve) is their sum. The grey curve plotted in addition to the phenotype is the environmental target function, a sum of 3 Gaussian lobes (2 positives and 1 negative) – see Figure S3. Finally, *f* is the absolute fitness value computed from the difference between the phenotype and the target function.

TABLE 1 Mutation rates per base pair per generation for the four mutational scenarios: SUB, LM, CR and CRLM.

	SUB	LM	CR	CRLM
Local mutations				
Substitutions (per bp)	3×10^{-5}	1×10^{-5}	0	5×10^{-6}
Small insertions (per bp)	0	1×10^{-5}	0	5×10^{-6}
Small deletions (per bp)	0	1×10^{-5}	0	5×10^{-6}
Chromosomal rearrangements				
Duplications (per bp)	0	0	1×10^{-5}	5×10^{-6}
Deletions (per bp)	0	0	1×10^{-5}	5×10^{-6}
Inversions (per bp)	0	0	1×10^{-5}	5×10^{-6}
Total per base pair per generation event rate	3×10^{-5}	3×10^{-5}	3×10^{-5}	3×10^{-5}

Note: For mutations affecting subsequences (i.e. all mutations but substitutions), this rate corresponds to the probability to initiate an event at a given locus. Note that the total mutation rate (per base pair, per generation) is constant across experiments. An additional scenario (CRLMx2) has been tested to have equal mutation rates for all kind of events (1×10^{-5}) between CR, LM and CRLMx2 (see Figure S4).

2.2.3 | Fitting fitness trajectories

In order to estimate diminishing-returns epistasis, that is, how fast the advantage provided by each new beneficial mutation reduces over time, for each mutational condition, we fit the mean fitness values along the 11 lines of descent with power laws of type $f = (bt+1)^a$, where *f* is the fitness, *t* is the time in generations (Wiser et al., 2013). *a* and *b* are the parameters to be fitted with *a* corresponding to the diminishing-returns epistasis when $0 \leq a \leq 1$ ($a=1$ corresponding to linear fitness growth without diminishing-returns epistasis) and *b* corresponding to an initial fitness growth parameter.

To compute the fit, we use the `lmfit` Python package with the least squares method. In order to ease the fitting process, the data points were sampled once every 1000 generations.

3 | RESULTS

To investigate the contribution of chromosomal rearrangements to evolutionary innovation, we compare the evolutionary dynamics of four sets of runs: SUB, with only substitutions; LM, with only local mutations; CRLM, with both local mutations and chromosomal rearrangements; and CR, with only chromosomal rearrangements. As we suspect that the relative contribution of chromosomal rearrangements versus local mutations depends on the distance to the fitness optimum, we repeated these experiments in two conditions: starting with naive individuals (see Section 3.1) or with pre-evolved ones (WT—see Section 3.2).

3.1 | Local mutations are dispensable when far from the optimum

As shown in Figures 2 and 3, the evolutionary trajectories in terms of fitness, genome size and number of genes without local mutations (CR) are similar to the evolutionary trajectories with both rearrangements and local mutations (CRLM), whereas the simulations without rearrangements (SUB and LM) produce significantly less adapted organisms, with fewer genes and a smaller coding genome size despite a greater total genome size.

Strikingly, the end fitness in the CRLM set-up is not statistically different from the CR setup (Mann–Whitney U test, p-value = .65), while both values are highly different from those in the cases without chromosomal rearrangements (Mann–Whitney U test, $p = 5 \times 10^{-4}$). This result is surprising, given that local mutations are usually thought to be a major evolutionary force and would therefore be expected to provide a boost in fitness when present.

There are also structural differences in the genomes depending on the set of allowed mutations. First, the dynamics of gene creation is much slower in the SUB and LM simulations, as could be expected in the absence of gene duplication. Indeed, in the CRLM setup, a fixed duplication adds on average 2.58 genes to the genome (for a total across repetitions of 1241 new genes), while all other mutations stand below 0.05 per fixed mutation (for a total of 388 new genes for all other mutations). However, we observe that the genomes evolved in the CRLM setup achieve a similar fitness but with fewer genes than the ones in the CR setup, highlighting that local mutations are better than chromosomal rearrangements

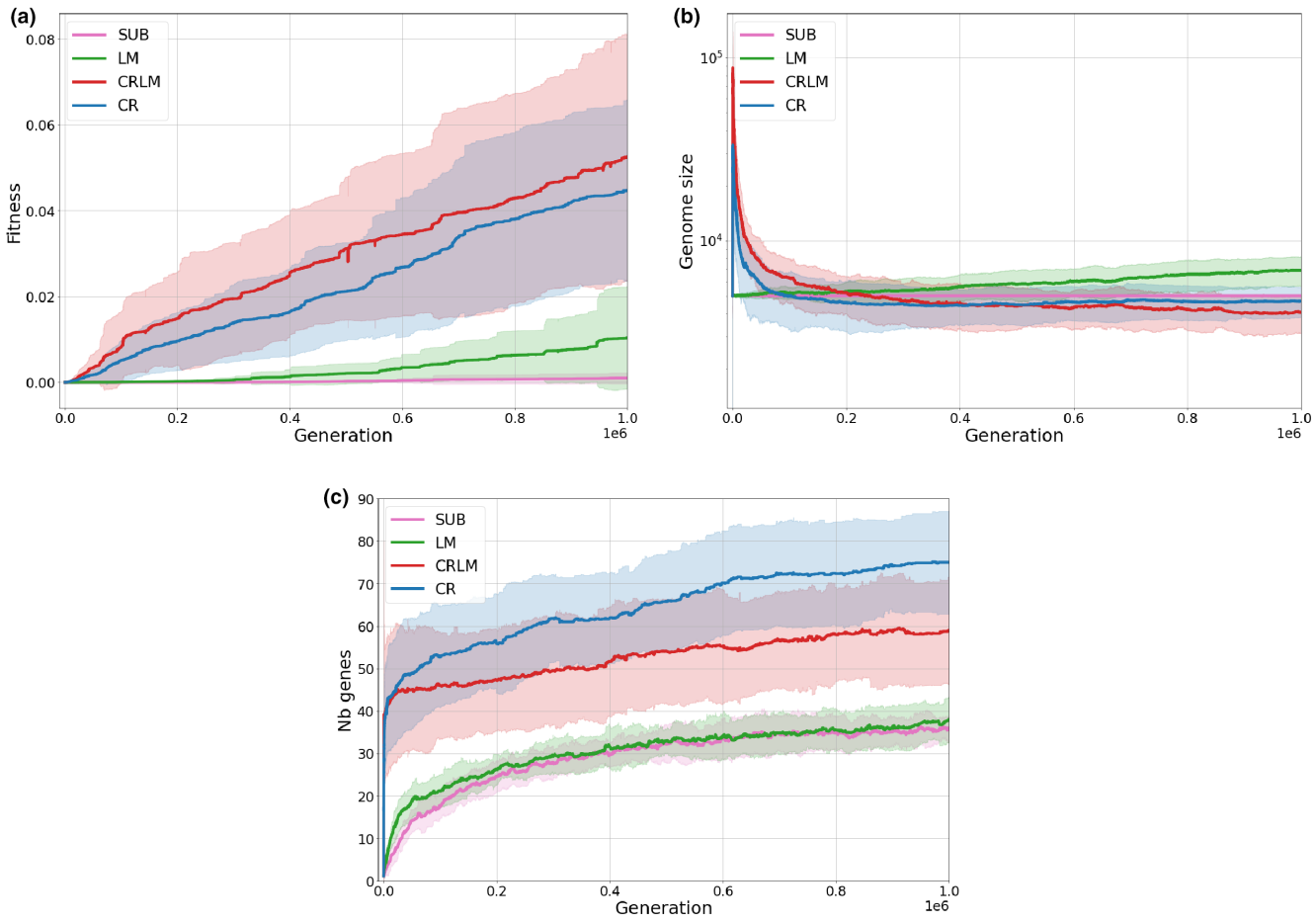

FIGURE 3 Mean variation of fitness (a), genome size (b) and gene number (c) on the line of descent of the final populations, starting from a naive individual for the four mutational scenarios. The shaded areas indicate the variability across the 11 repetitions (SD).

at fine-tuning existing genes. Chromosomal rearrangements and local mutations also have different effects on genome size. Indeed, in the presence of chromosomal rearrangements (CR and CRLM), genome size sharply increases at first, before slowly reducing and stabilizing around 3000 bp. On the contrary, in the LM set-up, genome size never ceases to grow all along the experiment, although at a slow pace. This is caused by the fixation of more small insertions than small deletions (see Figure 4b). Ultimately, genomes evolved under the LM set-up are longer than genomes evolved under the CR and CRLM setups but they contain much fewer genes, resulting in a larger proportion of non-coding DNA (see Figure 2).

Finally, comparing the SUB and LM setups shows that the dynamics of de novo gene creation is similar in both conditions, but that the fitness of the LM simulations increases much faster than the fitness of the SUB ones. This shows that InDels do not facilitate de novo gene creation but that once a gene is present on the genome, they facilitate its evolution, hence reaching higher fitness.

To better understand the origin of these differences, we first look at the contribution of each mutation type to the end fitness. We computed the total gain of fitness per mutation type along the ancestral lineage during the 1,000,000 generations of each experiment (Figure 4a). Interestingly, although CRLM are much fitter than LM, it is still the local mutations that contribute the most to the overall fitness gain in CRLM. Local mutations are crucial to evolution, and it is not surprising that they are the most impactful. However, the difference between SUB, LM and CRLM shows that their potential is only fully unleashed when chromosomal rearrangements are also present and create a substrate that local mutations can then finely tune.

The number of non-neutral mutations fixed along the line of descent (Figure 4b) shows that rearrangements, although rarely fixed compared to local events and hence almost invisible in the phylogeny, favour the fixation of beneficial local mutations. This is consistent with the dynamics of gene number shown on Figure 3c: by allowing for the recruitment of more genes, rearrangements increase the number of potential mutational targets on which local events can have an effect, hence favouring the fixation of more favourable local events.

The very rare fixation of rearrangements compared to the fixation rate of local mutations can be better understood by looking at the distribution of fitness effects (DFE) for each type of mutation (see Figure 5). Duplications and deletions have a very broad effect and can disturb, delete or imbalance essential genes: they are therefore very often lethal (in approximately 95% of cases here). Local mutations, on the other hand, have a smaller chance of disrupting an essential gene, as they affect a restricted section of the genome. They are more often neutral or "simply" deleterious, and lethal only in <40% of cases. Finally, inversions have two

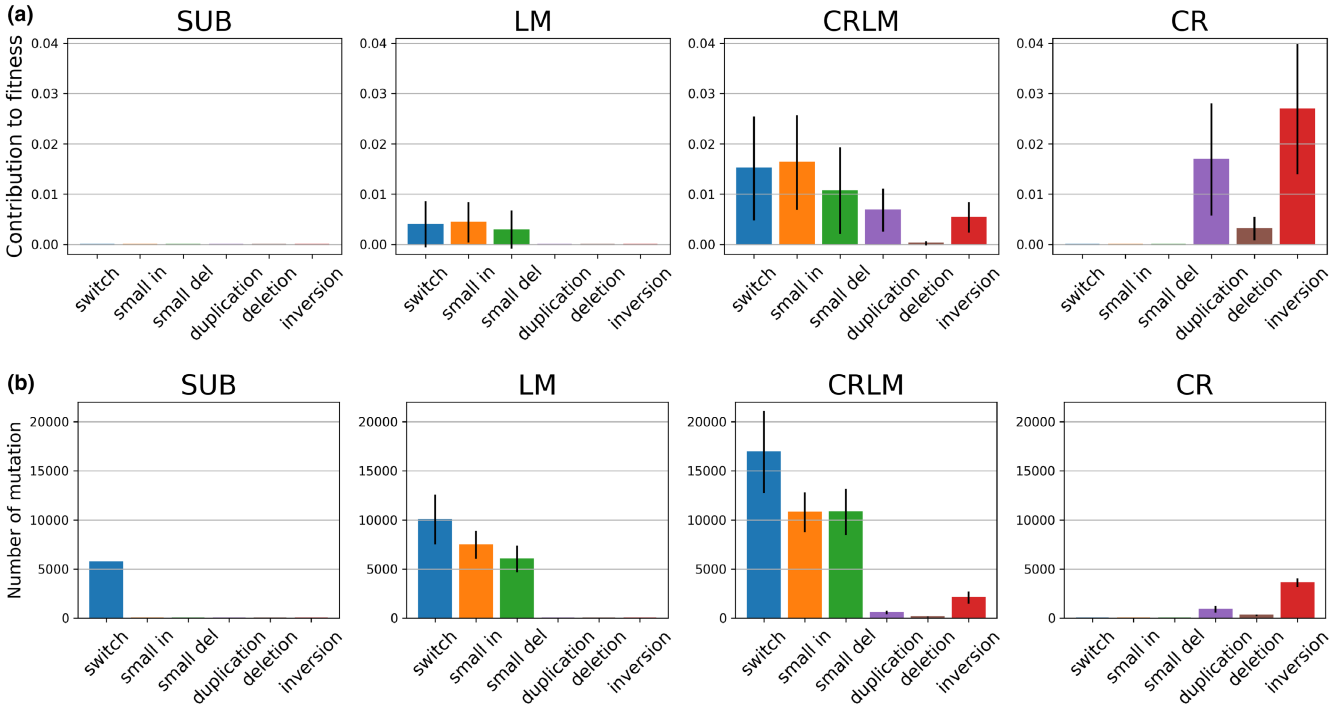

FIGURE 4 Fitness contribution and number of mutations fixed during the initial evolution from naive individuals (a) Contribution of each type of mutation to the total fitness gains, measured as the sum of the change in fitness of each mutation on the line of descent of the final best individuals, starting from naive individuals. Histograms show the mean values across the 11 repetitions, and the bars show their standard deviation. Reverted mutations (mutations which effect on fitness was exactly compensated by the following one) were filtered out to reduce noise. Fitness increase in the SUB simulations are negligible at this scale. (b) Number of non-neutral and non-reverted mutations fixed for the different mutation types and for the four conditions, normalized by the number of mutations occurring ($L \times \mu$, with L the genome size), on the line of descent of the final best individuals, starting from naive individuals. Histograms show the mean values across the 11 repetitions, and the bars show their standard deviation.

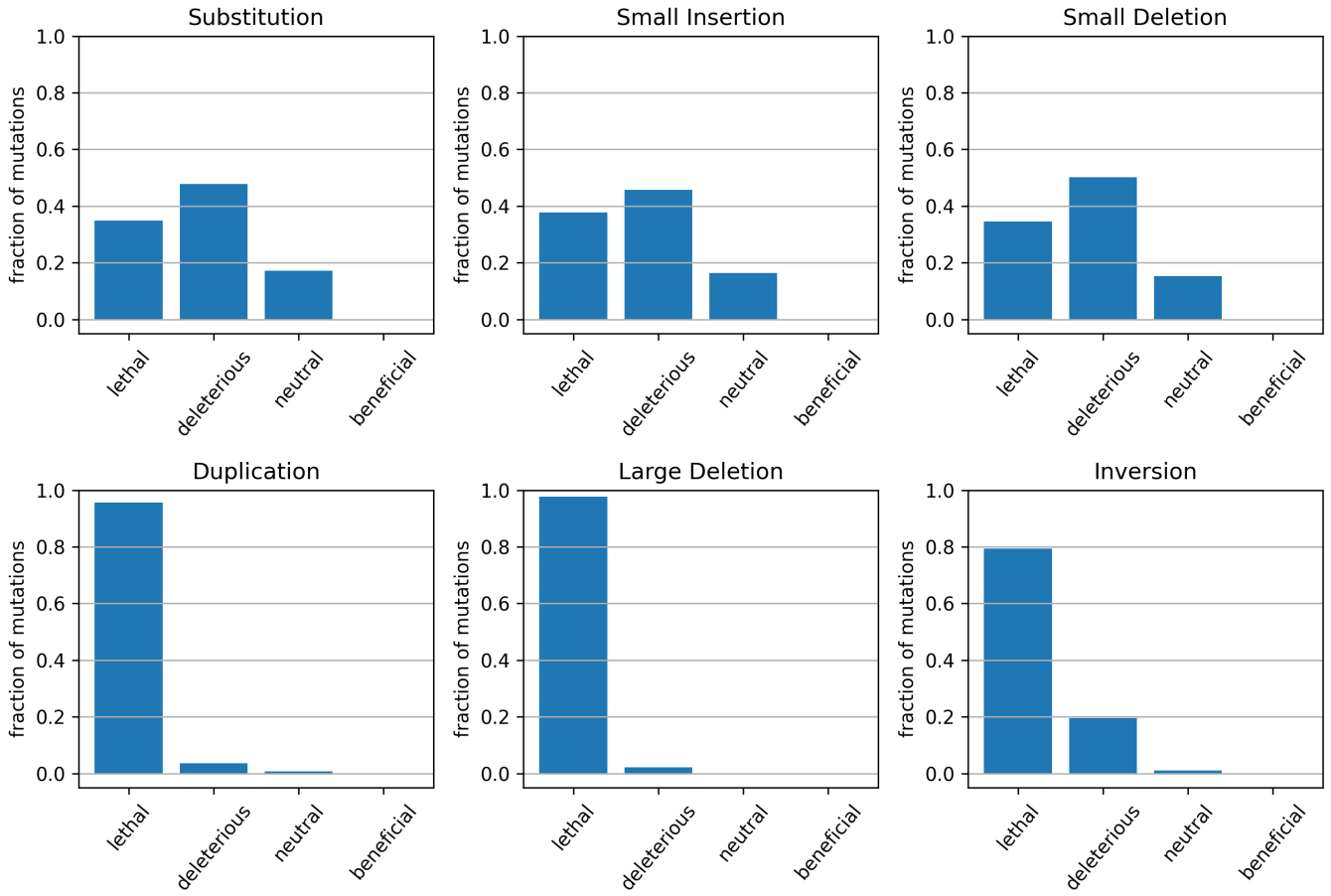

FIGURE 5 Distribution of fitness effect of the different types of mutation, on the median individual of the CRLM experiment, after 1,000,000 generations when starting from a naive individual. For each mutation type, 1,000,000 mutants were generated, except for the substitution, which were exhaustively tested. The selection coefficient is computed as $s = \frac{f_{mutant}}{f_{parent}} - 1$. Lethality is defined as $s < -0.999$, and neutrality as $s \in [-0.001, 0.001]$. The detailed Distribution of Fitness Effect (DFE) is presented in Figure S2. Interestingly, there is no advantageous substitutions available, showing that the population has reached a local fitness optimum for these mutations. However, as shown by Figure S2, a few beneficial InDels and a few beneficial segmental duplications are available, although they are not frequent enough to be visible here.

breakpoints while local mutations have only one and are therefore more lethal than local mutations (80%), but as inversions are balanced rearrangements, they are less likely to be deleterious than duplications or deletions.

3.2 | Chromosomal rearrangements sustain long-term adaptation

When starting from a wild-type individual, whose gene repertoire has already evolved, the advantage of gene duplication over de novo gene creation vanishes, and we can study more subtle interactions between local mutations and chromosomal rearrangements. Here we initiate experiments from clonal populations of the median CRLM individual evolved in the previous set of experiments and follow their evolution for 3,000,000 generations in SUB, LM, CRLM, and CR conditions.

Figure 6a shows that the four conditions result in very different dynamics of genome size. While the genome size of CR and CRLM experiments is quite stable, as observed at the end of the previous experiments, in LM conditions the genome size increases continuously during the 3 million generations of the experiment. At first sight, this result may seem contradictory, as the genome size is much more likely to vary in the presence of long segmental duplications/deletions than in the sole presence of small InDels. This shows the complex effect of chromosomal rearrangements in regulating genome size and highlights the difference between InDels and rearrangements in doing so.

As expected, when looking at the fitness gain along the 3 million generations of the experiment (Figure 6b) the difference between the mutational scenarios is not as marked as what was observed when far from the optimum, at least for the LM, CR and CRLM scenarios. Yet, the SUB scenario still clearly lags behind in terms of fitness, showing again that substitutions alone are not sufficient in fine-tuning genes. In the four conditions, fitness improves all along the experiment, albeit with a clear diminishing-returns epistasis in the SUB, LM and CRLM conditions. Following Wiser et al. (2013), we used power-law curve fitting to estimate the amount of diminishing-returns epistasis

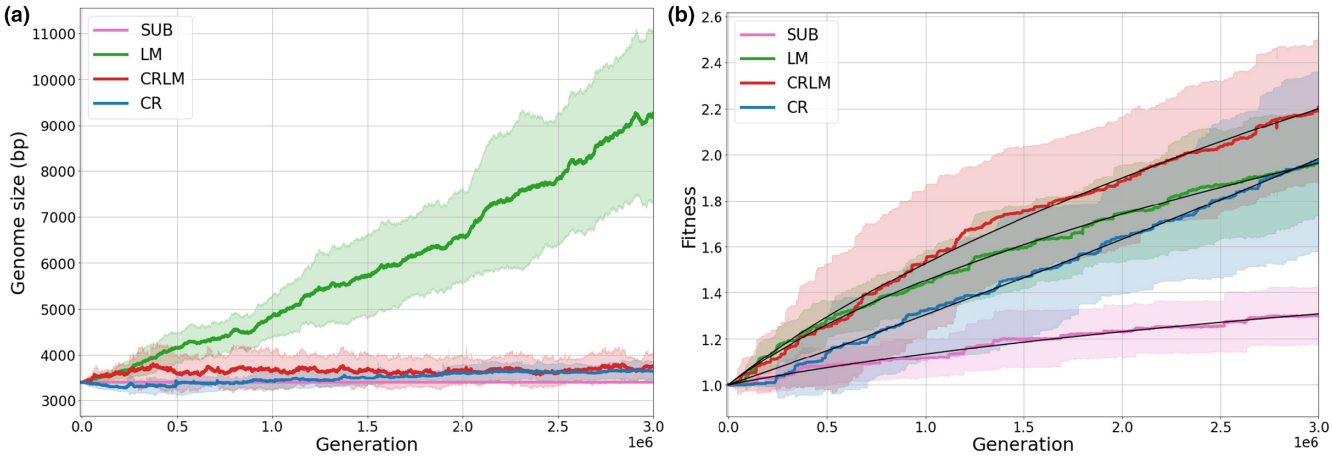

FIGURE 6 Temporal changes in genome size and fitness in evolution started from the WT. (a) Mean change in genome size on the line of descent of the final populations, for the 11 repetitions and the three conditions. All simulations started from the same wild type with a genome length of 3394 bp (Figure 2b) and evolved for 3,000,000 generations. The shaded areas indicate the variability across repetitions (standard deviation). (b) Relative fitness variation on the line of descent of the final population, starting from a wild type. The shaded areas indicate the variability across repetitions (standard deviation). Black curves show the fitted power laws for the mean fitness values of the four sets of simulations (see Methods, Section 2.2.3). The fitted parameters are: $a_{SUB} = 0.2$, $b_{SUB} = 7.0 \times 10^{-7}$, $a_{LM} = 0.4$, $b_{LM} = 1.8 \times 10^{-6}$; $a_{CR} = 1.5$, $b_{CR} = 2.0 \times 10^{-7}$, $a_{CRLM} = 0.5$, $b_{CRLM} = 1.5 \times 10^{-6}$.

in the four conditions (black lines on Figure 6b). Results show that diminishing-returns epistasis is higher in the SUB and LM conditions than in the CRLM conditions ($a_{SUB} = 0.2$; $a_{LM} = 0.4$; $a_{CRLM} = 0.5$—see Methods, Section 2.2.3) which, in the long run, advantages the CRLM over the other scenarios. Strikingly, when evolving only with chromosomal rearrangements (CR scenario), populations show no diminishing-returns epistasis throughout the duration of the experiment ($a_{CR} = 1.5 > 1$). This contrasts with the other conditions and allows the CR populations to catch up with the SUB and LM ones, despite an initial disadvantage.

As previously, we measured the total fitness effect and the number of non-neutral mutations fixed along the lineage for the different types of mutation and for the four mutational scenarios (Figure 7a,b respectively). As already noticed when starting far from the optimum, this shows that chromosomal rearrangements, although very rarely fixed in the lineage, have a dual contribution to fitness. While, in the CRLM, fixed rearrangements have a small impact on fitness on their own (Figure 7a), they contribute to increasing the number of favourable substitutions. Indeed, substitutions and InDels are more likely to be favourable and fixed in the CRLM populations than in the LM populations and almost as likely—for the substitutions—as in the SUB ones (Figure 7b). This leads to a sustained evolutionary dynamics, despite rearrangements being almost invisible in the phylogeny owing to their very low fixation probability.

4 | DISCUSSION

It is widely admitted that genomes evolve under the combined pressure of a large variety of mutational operators, including of course substitutions and InDels but also chromosomal rearrangements (Berdan, Blanckaert, Slotte, et al., 2021; Mérot et al., 2020). However, models of genome evolution almost exclusively focus on the former, the latter being generally ignored owing to their difficult modelling and their apparent low frequency in phylogenies that could suggest a moderate impact compared to other events. A direct consequence is that the contribution of chromosomal rearrangements to the evolutionary dynamics is largely overlooked. Indeed, while substitution-based epistasis is largely recognized and quantified in several model systems (Bank et al., 2015; Diss & Lehner, 2018; Olson et al., 2014; Starr & Thornton, 2016), the epistatic effect of rearrangements is, with very few exceptions (Blount et al., 2012), *terra incognita*.

Here, we used Aevol to simulate genome evolution under several conditions characterized by an increased mutational diversity but a constant overall mutational rate (see Table 1). We completed these experiments by testing evolution under the exclusive pressure of chromosomal rearrangements, in order to estimate their capacity to generate enough variation to allow sustained evolution. This enables an experimental (though simulated) exploration of the consequences of chromosomal rearrangements on the evolutionary dynamics. Specifically, we analysed the results of the simulations with a focus on two levels: genome structure, which is likely to be largely impacted by rearrangements and individuals' fitness.

Regarding the evolution of genome structure, our results show two clear differences when genomes evolve with (CRLM and CR simulations) or without (SUB and LM simulations) chromosomal rearrangements. First, they confirm the well-established theory of evolution by gene duplication (Kalhor et al., 2023; Zhang, 2003): in our simulations, rearrangements are essential for the rapid acquisition of a large gene repertoire and duplications are the main cause of increase in gene number (see Section 3.1). Indeed, gene number rapidly increases in the very first thousands of generations for CR and CRLM (Figure 3c), and this process of gene recruitment is maintained throughout the simulation, though at a lower pace. On the

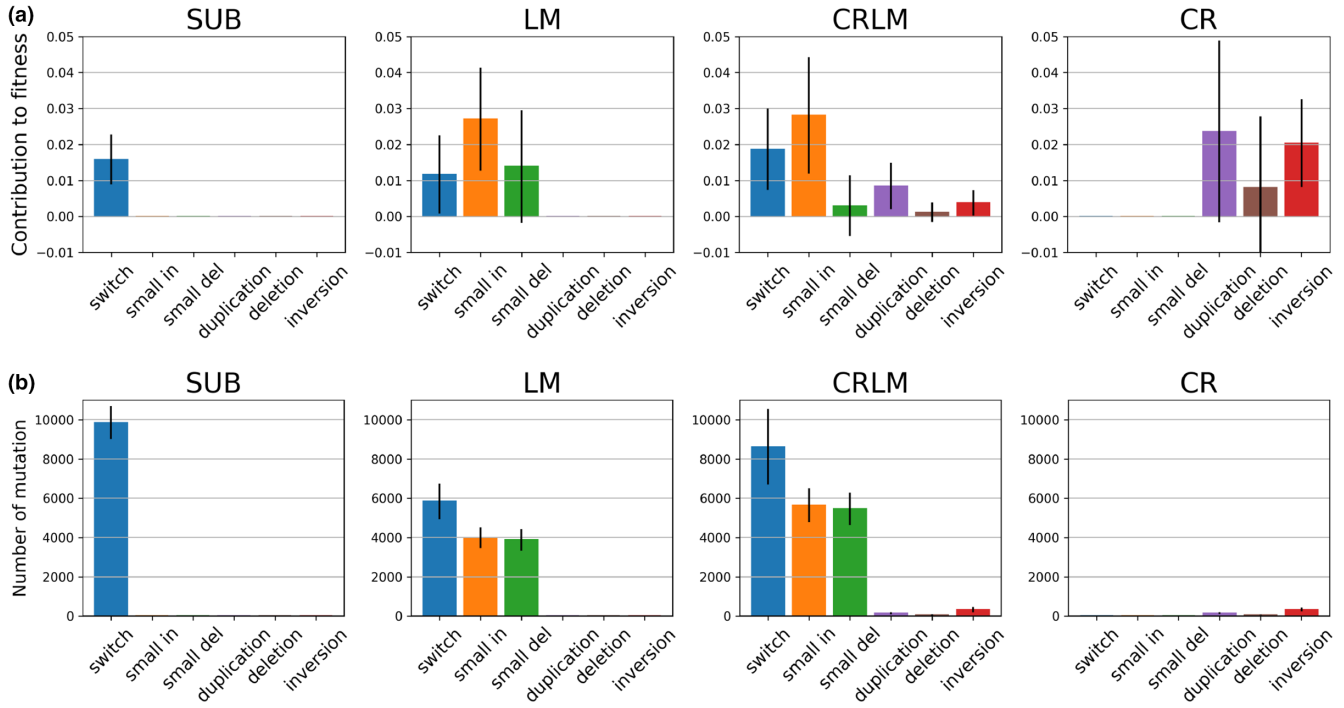

FIGURE 7 Fitness contribution and number of mutations during the evolution from WT individuals (a) Contribution of each type of mutation to the total fitness gains, measured as the sum of the change in fitness of each mutation on the line of descent of the final best individuals, starting from WT individuals. Histograms show the mean values across the 11 repetitions, and the bars show their standard deviation. Reverted mutations (mutations which effect on fitness was exactly compensated by the following one) were filtered out to reduce noise. (b) Number of fixed non-neutral and non-reverted mutations per generation for the different mutation types per million generation, normalized by the number of mutations occurring ($L \times \mu$, with L the genome size), on the line of descent of the final best individuals, starting from WT individuals. Histograms show the mean values across the 11 repetitions, and the bars show their standard deviation.

opposite, lineages evolving without rearrangements only acquire a limited gene repertoire (see Figure 3c).

In a less intuitive way, our simulations show an important contribution of chromosomal rearrangements to the stabilization of genome length during evolution. Indeed, Figures 3b and 6a show that, after an initial burst of genome size at the very beginning of the evolution (corresponding to the phase of fast gene acquisition through duplications), CR and CRLM lineages quickly undergo a reduction of their genome size (while preserving their gene repertoire – see Figure 3b,c). Continuing the simulation for 3 million generations, we see that genome size varies very little thereafter (Figure 6a). This dynamic contrasts sharply with that of the LM lineages, which show a steady increase in genome size, both when starting far or close to the optimum. This sustained growth of genome size under the sole pressure of InDels advocates in favour of the mechanism of border-induced selection, which has been recently conceptualized by Loewenthal et al. (2022). Indeed, despite their spontaneous mutation rates being equal, the probability of fixation of neutral insertions is slightly higher than the probability of fixation of neutral deletions, due to interference with gene borders (Loewenthal et al., 2022): a small insertion close to a gene is most often harmless, while a small deletion at the same point can impact a gene if the size of the deletion is larger than the distance to this gene. In the absence of other constraints on the genome size, this bias leads to a steady genome growth, as we observe on Figures 3b and 6a. Strikingly, in the presence of chromosomal rearrangements, this bias is not visible anymore, showing that rearrangements generate an evolutionary pressure that prevents genome growth. As already proposed by Knibbe et al. (2007), deleterious chromosomal rearrangements lead to selection for robustness, favouring smaller genomes as these undergo fewer rearrangements than longer ones. This hypothesis is sustained by the low rate of fixation of chromosomal rearrangements (Figure 7b): they are largely filtered-out by purifying selection, suggesting that they have a strong robustness effect. The low number of fixed rearrangements, due to their high lethality, (Figure 5) questions the concept of mutation rate. Indeed, by measuring mutation rates on a live population, a bias is introduced towards non-lethal mutations. This bias has been observed in the case of substitutions (Wang et al., 2012) but we hypothesize that this could be even more important in the case of genome rearrangements, and models should take into account that spontaneous mutation rates could be very different from observed and fixed ones.

The influence of chromosomal rearrangements on fitness evolution is also very different depending on whether the simulations start far from the optimum (hence requiring them to acquire new genes) or close to the optimum (with a gene pool already acquired but that can still be optimized). In the former situation, lineages evolving in the presence of chromosomal rearrangements have a much higher fitness than those evolving with only substitutions or even with all local mutations (Figure 3a). This confirms that,

in such a situation, gene duplication has a decisive contribution (Kalhor et al., 2023; Zhang, 2003), enabling both the CR and CRLM lineages to largely overcome the LM and the SUB lineages. Strikingly, lineages evolving with chromosomal rearrangements only (CR) perform almost as well as those evolving with both chromosomal rearrangements and local mutations (CRLM). This illustrates the multi-scale nature of chromosomal rearrangements that can both enlarge the gene repertoire through large duplications but also optimise gene sequences by reorganizing them through, for example inversions. This is coherent with the work of Trujillo et al. (2022), which modelled inversions in simpler evolutionary setting and showed that, given enough time, inversions allow reaching higher fitness peaks than substitutions. Interestingly, the fitness of the SUB lineages (that evolved under the sole pressure of substitutions) is much lower than the fitness of the LM lineages (that evolved through substitutions and InDels) despite a very similar dynamic of gene recruitment. This confirms that small insertion and small deletions are decisive operators when the evolution of protein sequence is concerned, as they can add/remove codons when substitutions can only mutate existing ones (Leushkin et al., 2012; Vakhrusheva et al., 2011).

When starting close to the fitness optimum, the differences between the experiments are more subtle, except when substitutions are the sole mutational operator, in which case fitness gains are much lower than in the three other conditions (SUB curve on Figure 6b), highlighting the importance of the diversity of mutational operators (Berdan, Blanckaert, Slotte, et al., 2021). In all experiments, the dynamics of fitness is similar to what can be observed in vitro, for example in experimental evolution with bacteria (Wang et al., 2016; Wiser et al., 2013), or yeast strains (Wei & Zhang, 2019): simulations show a sustained fitness gain all along the experiment albeit with a more or less pronounced diminishing-returns epistasis. Inspired by Wiser et al. (2013), we estimated the diminishing-returns epistasis in these different conditions, and showed that, in the long run, chromosomal rearrangements reduce diminishing-returns epistasis, hence enabling sustained evolutionary dynamics. It is known that clonal interference could also induce diminishing return (Wiser et al., 2013). However, as the population size and global mutation rates are the same in all our simulations (CR, CRLM, LM and SUB), we assumed clonal interference had similar effect in all simulations. Moreover, as shown by Figure 7a, the effect of rearrangements is mainly indirect: they have a small effect by themselves but potentiate other factors. Indeed, in the CRLM lineage, substitutions have a larger impact than in the SUB and LM lineage. This suggests that rearranged sequences open new targets to substitutions, hence increasing the probability to fix beneficial local events (Figure 7b). Finally, as Figure 7b also shows, this effect is due to a very low number of fixed rearrangements. Hence, while rearrangements sustain long-term adaptation by reducing the effect of diminishing-returns epistasis, they are almost invisible in the phylogeny.

When quantifying the diminishing return, a striking result was the apparent accelerating evolution in the CR populations ($a_{CR} > 1$). We hypothesize that this is due to the low fixation rate of chromosomal rearrangements (Figure 7b). As CR populations undergo only rearrangements, fitness comparatively evolve by bigger steps but with longer waiting times between mutations, and this creates an initial lag in the fitness gain (Figure 6b), hence the appearance of acceleration. Now, the number of possible rearrangements for a given genome is much larger than the number of possible local events (it is indeed mainly linked to the number of breakpoints to be chosen for a given type of event: one for local mutations, two for inversions and deletions, three for duplications—see Figure 1b). A direct consequence is that, contrary to substitutions and InDels, rearrangements neighbourhood cannot be explored in a reasonable time, hence the lower diminishing-returns epistasis observed on the duration of our simulations when rearrangements are allowed. Further, exploring this question, for example by estimating the contribution of each type of rearrangement to the phenomenon, is a very promising research direction opened by our results.

Overall, our simulations show that chromosomal rearrangements have both a direct (through gene duplications) and an indirect (by potentiating the effect of local mutations) contribution to the evolutionary dynamics. They seem to also act as regulators of genome size, due to purifying selection against long genomes which undergo too many mutational events, as already proposed by Knibbe et al. (2007). This inverse correlation between mutation rates and genome size has already been observed in prokaryotes (Drake, 1991; Lynch, 2010), but for substitutions only. Our results suggest that its main determinant could be the rearrangement rates. Interestingly, this hypothesis implies that the regulation of genome size is due to the events that *do not* go to fixation in the winning lineage. Hence, despite them being almost invisible in the phylogeny, chromosomal rearrangements act as a major player of evolution by regulating genome size, limiting the effect of diminishing-returns epistasis, and sustaining long-term adaptation. Our results also illustrate the potential power of forward-in-time simulators like Aevol to unravel the effect of "non-conventional" mutational operators. Despite their artificial nature, models mimicking genome structures and the genotype-to-phenotype map allow deciphering the impact of the different types of mutation with a limited set of a priori hypotheses.

All models rely on simplifying assumptions, and ours makes no exception. However, the interest of modelling is precisely to reduce the complexity of the system to be studied. Here, studying only a limited number of mutational operators has enabled us to identify effects that could have been blurred in a more complex setting. Indeed, our experimental strategy, which relies on a progressive complexification of the mutational repertoire, has enabled us to uncover profound differences between chromosomal rearrangements and small InDels, both in the evolution of genome size and in the adaptation of organisms. Both kinds of events may seem rather similar at first sight, but they differ on two important aspects: first, contrary to duplications that copy pre-existing genomic sequences, small insertions add random sequences to the genome. Hence, they cannot duplicate genes, while this process is central in evolution (Zhang, 2003). Second, even though both types of mutation add/

remove genomic segments to the chromosome, the distribution of the size of these segments is different: in the case of InDels, this distribution is fixed while in the case of rearrangements, the distribution depends on the size of the genome. A direct consequence of this property is that larger genomes undergo more deleterious rearrangements, leading to a lower robustness (Knibbe et al., 2007). In our simulations, large duplications and deletions, far from randomly shuffling the genome size as could have been expected, impose a tight constraint on it.

In the development of the model, we chose to stay close to prokaryotic genomics. This means that genomes are haploid and circular, and undergo no recombination. This obviously prevents us from studying the interplay between structural variation and recombination and its potential effect on speciation and on the fate of chromosomal rearrangements (Berdan, Blanckaert, Butlin, & Bank, 2021). We also chose to study a limited set of chromosomal rearrangements (duplications, deletions, and inversions), while many other types of events could be added to the model (e.g., transposable elements, horizontal gene transfer, etc.). As for the rearrangements we model, breakpoints are chosen uniformly on the chromosome, leading to a uniform distribution of rearrangement lengths. This distribution is difficult to estimate in real organisms, as a large fraction of chromosomal rearrangements are likely to be lethal (Rocha, 2006). However, experimental studies show that the rearranged segments can reach lengths of the same order of magnitude as the size of the genome (Raeside et al., 2014), hence supporting our simplifying hypothesis, although the shape of the distribution in more likely to be geometric (Darling et al., 2008). However, we choose the simplest hypothesis of random breakpoints so as not to add additional parameters. We conjecture that our main results hold even with a geometric distribution of rearrangements, as the tail of the distribution will indeed grow with genome length. Yet, this could partly relax the robustness constraints, as they are mostly due to the longest rearrangements. We therefore expect that the effect of chromosomal rearrangements on genome size would hold, although it might be less pregnant with another distribution.

Our conclusions are drawn from the comparison of the evolutionary trajectories of different experiments and open up several interesting perspectives. For example, Aevol also includes several analysis tools, such as the computation of the distribution of fitness effect for all mutation types and for all genomes along a lineage, as illustrated by Figure 5. Taking advantage of the perfect record of the mutational events, these measures help quantify the evolutionary forces at work, as well as the relative contribution of the different types of mutation to these forces. As exemplified on Figures 4a and 7a, the impact of the different types of mutation on the fitness can easily be quantified, allowing to estimate the direct contribution of each type of mutation. Although it would be very computationally demanding, it could be interesting to also quantify the consequences of each mutation type on robustness and evolvability as this could allow to estimate their indirect effect and explain how the different types of mutations interact. Finally, as long as chromosomal rearrangements are concerned, an obvious prospect is to extend the model to diploid eukaryote-like genomes with recombination. This would enable exploring the interplay between rearrangements and recombination (Berdan, Blanckaert, Butlin, & Bank, 2021).

The experiments we presented here only scratch the surface of what can be done with Aevol. Indeed, as Table S1 of the Supplementary Material shows, many other experiments can be done, including testing the effect of mutation rates, mutation biases or population size. Aevol is available to any team that would like to test hypotheses regarding the effect of these parameters on the evolutionary dynamics and on genome structure. Moreover, as the code is open and freely available, any team can modify it to test some specific mutation type that would not already be implemented (see Data availability and Benefit-Sharing statement). Notably, there are many ways to be far from the optimum. Here, we choose to start with naive individuals but another approach would be to force environmental changes. In Aevol, this could easily be done by moving the target function after having adapted organisms to a first environment. This would enable studying the contribution of rearrangements to evolutionary rescue. Indeed, a previous study with Aevol has shown that, in the case of an environmental change, the frequency of gene duplications is positively correlated with the distance to the optimum (Kalhor et al., 2023), but the impact of all chromosomal rearrangements could be studied more in details by limiting the number of possible mutation types, as we do in the present study. The role of chromosomal rearrangements when organisms are confronted to a perpetually moving target, and so always relatively far from the optimum, could also be further studied.

Despite the highly artificial nature of our model, our simulations are consistent with the classical view of evolution: among the variety of mutational operators, substitutions and small InDels are by far the most visible adaptive events both in terms of their number (Figures 4b and 7b) and their contribution to the fitness (Figures 4a and 7a). However, our simulations also show that the scarcity of rearrangements that we observe in the phylogenies masks an important contribution to adaptation. While the vast majority of models and simulators of molecular evolution still implements a solely allelic view of evolution, where rearrangements can modify gene organization but cannot create new gene sequences, our results suggest that the innovative potential of rearrangements is not marginal and that it is essential to integrate them into population genetics models.

AUTHOR CONTRIBUTIONS

GB, CK, JRC, DP, TG and MF developed the model. GB, PB and JL conceived, realized and analysed the experiments. All authors discussed the results and contributed to the final manuscript.

ACKNOWLEDGEMENTS

M.F. is funded by the French Agence Nationale pour la Recherche (Evoluthon grant). L.T. thanks the Institut National des Sciences Appliquées de Lyon (INSA-Lyon) as well as the Laboratoire d'InfoRmatique en Image et Systèmes d'information (LIRIS) for hospitality while part of this research was done. J.L. and G.B. would like to thank the Rhône-Alpes Institute for Complex Systems (IXXI)

for funding. All authors thank the Grid'5000 testbed, supported by a scientific interest group hosted by Inria and including CNRS, RENATER and several Universities as well as other organizations (see https://www.grid5000.fr), for computational support. J.L and P.B. thank Lisa Chabrier for instructive discussions.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest.

DATA AVAILABILITY AND BENEFIT-SHARING STATEMENT

The code of the Aevol software is available on gitlab (https://gitlab.inria.fr/aevol/aevol). More documentation is available on the website https://www.aevol.fr. All relevant final data, as well as parameters files to redo the simulations, are available on Zenodo (https://zenodo.org/record/8307916; Banse et al., 2023).

ORCID

Paul Banse https://orcid.org/0000-0003-2373-6785
Juliette Luiselli https://orcid.org/0000-0002-7854-3545
David P. Parsons https://orcid.org/0000-0003-0511-0703
Théotime Grohens https://orcid.org/0000-0002-2954-2677
Leonardo Trujillo https://orcid.org/0000-0001-9995-4135
Jonathan Rouzaud-Cornabas https://orcid.org/0009-0006-6324-0716
Carole Knibbe https://orcid.org/0000-0002-2026-2580
Guillaume Beslon https://orcid.org/0000-0001-8562-0732

REFERENCES

Alkan, C., Coe, B. P., & Eichler, E. E. (2011). Genome structural variation discovery and genotyping. *Nature Reviews Genetics*, *12*(5), 363–376.

Audrézet, M.-P., Chen, J.-M., Raguénes, O., Chuzhanova, N., Giteau, K., Maréchal, C. L., Quéré, I., Cooper, D. N., & Férec, C. (2004). Genomic rearrangements in the cftr gene: Extensive allelic heterogeneity and diverse mutational mechanisms. *Human Mutation*, *23*(4), 343–357.

Bank, C., Hietpas, R. T., Jensen, J. D., & Bolon, D. N. (2015). A systematic survey of an intragenic epistatic landscape. *Molecular Biology and Evolution*, *32*(1), 229–238.

Banse, P., Luiselli, J., Parsons, D. P., Grohens, T., Foley, M., Trujillo, L., Rouzaud-Cornabas, J., Knibbe, C., & Beslon, G. (2023). Forward-in-time simulation of chromosomal rearrangements: The invisible backbone that sustains long-term adaptation: Simulated data. *Zenodo*. https://doi.org/10.5281/zenodo.8307916

Batut, B., Parsons, D., Fischer, S., Beslon, G., & Knibbe, C. (2013). In silico experimental evolution: A tool to test evolutionary scenarios. *BMC Bioinformatics*, *14*(Suppl 15), S11.

Berdan, E. L., Blanckaert, A., Butlin, R. K., & Bank, C. (2021). Deleterious mutation accumulation and the long-term fate of chromosomal inversions. *PLoS Genetics*, *17*(3), e1009411.

Berdan, E. L., Blanckaert, A., Slotte, T., Suh, A., Westram, A. M., & Fragata, I. (2021). Unboxing mutations: Connecting mutation types with evolutionary consequences. *Molecular Ecology*, *30*(12), 2710–2723.

Beslon, G., Parsons, D., Sanchez-Dehesa, Y., Peña, J.-M., & Knibbe, C. (2010). Scaling laws in bacterial genomes: A side-effect of selection of mutational robustness? *Biosystems*, *102*(1), 32–40.

Bhatia, S., Feijão, P., & Francis, A. R. (2018). Position and content paradigms in genome rearrangements: The wild and crazy world of permutations in genomics. *Bulletin of Mathematical Biology*, *80*, 3227–3246.

Blount, Z. D., Barrick, J. E., Davidson, C. J., & Lenski, R. E. (2012). Genomic analysis of a key innovation in an experimental Escherichia coli population. *Nature*, *489*(7417), 513–518.

Campo, N., Dias, M. J., Daveran-Mingot, M. L., Ritzenthaler, P., & Le Bourgeois, P. (2004). Chromosomal constraints in gram-positive bacteria revealed by artificial inversions: Experimental genome inversions in Streptococcaceae. *Molecular Microbiology*, *51*(2), 511–522.

Cao, S., Brandis, G., Huseby, D. L., & Hughes, D. (2022). Positive selection during niche adaptation results in large-scale and irreversible rearrangement of chromosomal gene order in bacteria. *Molecular Biology and Evolution*, *39*(4), msac069.

Connallon, T., & Olito, C. (2022). Natural selection and the distribution of chromosomal inversion lengths. *Molecular Ecology*, *31*(13), 3627–3641.

Darling, A. E., Miklós, I., & Ragan, M. A. (2008). Dynamics of genome rearrangement in bacterial populations. *PLOS Genetics*, *4*(7), e1000128.

Diss, G., & Lehner, B. (2018). The genetic landscape of a physical interaction. *eLife*, *7*, e32472.

Drake, J. W. (1991). A constant rate of spontaneous mutation in dna-based microbes. *Proceedings of the National Academy of Sciences*, *88*(16), 7160–7164.

Frenoy, A., Taddei, F., & Misevic, D. (2013). Genetic architecture promotes the evolution and maintenance of cooperation. *PLoS Computational Biology*, *9*(11), e1003339.

Gao, Y., Zhao, H., Jin, Y., Xu, X., & Han, G.-Z. (2017). Extent and evolution of gene duplication in dna viruses. *Virus Research*, *240*, 161–165.

Haller, B. C., & Messer, P. W. (2017). Slim 2: Flexible, interactive forward genetic simulations. *Molecular Biology and Evolution*, *34*(1), 230–240.

Hanlon, V. C. T., Lansdorp, P. M., & Guryev, V. (2022). A survey of current methods to detect and genotype inversions. *Human Mutation*, *43*(11), 1576–1589.

Ho, S. S., Urban, A. E., & Mills, R. E. (2020). Structural variation in the sequencing era. *Nature Reviews Genetics*, *21*(3), 171–189.

Hoffmann, A. A., Sgrò, C. M., & Weeks, A. R. (2004). Chromosomal inversion polymorphisms and adaptation. *Trends in Ecology & Evolution*, *19*(9), 482–488.

Jensen-Seaman, M. I., Furey, T. S., Payseur, B. A., Lu, Y., Roskin, K. M., Chen, C.-F., Thomas, M. A., Haussler, D., & Jacob, H. J. (2004). Comparative recombination rates in the rat, mouse, and human genomes. *Genome Research*, *14*(4), 528–538.

Kaback, D. B., Guacci, V., Barber, D., & Mahon, J. W. (1992). Chromosome size-dependent control of meiotic recombination. *Science*, *256*(5054), 228–232.

Kalhor, R., Beslon, G., Lafond, M., & Scornavacca, C. (2023). Classifying the post-duplication fate of paralogous genes. In K. Jahn & T. Vinař (Eds.), *Comparative genomics*, *lecture notes in computer science* (pp. 1–18). Springer Nature.

Kara, E., Kiely, A. P., Proukakis, C., Giffin, N., Love, S., Hehir, J., Rantell, K., Pandraud, A., Hernandez, D. G., Nacheva, E., Pittman, A. M., Nalls, M. A., Singleton, A. B., Revesz, T., Bhatia, K. P., Quinn, N., Hardy, J., Holton, J. L., & Houlden, H. (2014). A 6.4 mb duplication of the α-synuclein locus causing frontotemporal dementia and parkinsonism: Phenotype-genotype correlations. *JAMA Neurology*, *71*(9), 1162–1171.

Katju, V., & Bergthorsson, U. (2013). Copy-number changes in evolution: Rates, fitness effects and adaptive significance. *Frontiers in Genetics*, *4*, 273.

Kirkpatrick, M. (2010). How and why chromosome inversions evolve. *PLoS Biology*, *8*(9), e1000501.

Knibbe, C., Coulon, A., Mazet, O., Fayard, J.-M., & Beslon, G. (2007). A long-term evolutionary pressure on the amount of noncoding dna. *Molecular Biology and Evolution*, *24*(10), 2344–2353.

Leushkin, E. V., Bazykin, G. A., & Kondrashov, A. S. (2012). Insertions and deletions trigger adaptive walks in drosophila proteins. *Proceedings of the Royal Society B: Biological Sciences*, *279*(1740), 3075–3082.

Liard, V., Parsons, D. P., Rouzaud-Cornabas, J., & Beslon, G. (2020). The complexity ratchet: Stronger than selection, stronger than evolvability, weaker than robustness. *Artificial Life*, *26*(1), 38–57.

Loewenthal, G., Wygoda, E., Nagar, N., Glick, L., Mayrose, I., & Pupko, T. (2022). The evolutionary dynamics that retain long neutral genomic sequences in face of indel deletion bias: A model and its application to human introns. *Open Biology*, *12*(12), 220223.

Lynch, M. (2010). Evolution of the mutation rate. *Trends in Genetics*, *26*(8), 345–352.

Lynch, M., Sung, W., Morris, K., Coffey, N., Landry, C. R., Dopman, E. B., Dickinson, W. J., Okamoto, K., Kulkarni, S., Hartl, D. L., & Thomas, W. K. (2008). A genome-wide view of the spectrum of spontaneous mutations in yeast. *Proceedings of the National Academy of Sciences*, *105*(27), 9272–9277.

Mérot, C., Oomen, R. A., Tigano, A., & Wellenreuther, M. (2020). A roadmap for understanding the evolutionary significance of structural genomic variation. *Trends in Ecology & Evolution*, *35*(7), 561–572.

Nattestad, M., Goodwin, S., Ng, K., Baslan, T., Sedlazeck, F. J., Rescheneder, P., Garvin, T., Fang, H., Gurtowski, J., Hutton, E., Tseng, E., Chin, C.-S., Beck, T., Sundaravadanam, Y., Kramer, M., Antoniou, E., McPherson, J. D., Hicks, J., McCombie, W. R., & Schatz, M. C. (2018). Complex rearrangements and oncogene amplifications revealed by long-read DNA and RNA sequencing of a breast cancer cell line. *Genome Research*, *28*(8), 1126–1135.

Noor, M. A. F., Grams, K. L., Bertucci, L. A., & Reiland, J. (2001). Chromosomal inversions and the reproductive isolation of species. *Proceedings of the National Academy of Sciences*, *98*(21), 12084–12088.

Olson, C. A., Wu, N. C., & Sun, R. (2014). A comprehensive biophysical description of pairwise epistasis throughout an entire protein domain. *Current Biology*, *24*(22), 2643–2651.

Parsons, D. (2011). *Indirect selection in Darwinian evolution: Mechanisms and implications*. PhD Thesis. INSA Lyon.

Parsons, D. P., Knibbe, C., & Beslon, G. (2010). Importance of the rearrangement rates on the organization of transcription. *Proceedings of Artificial Life XII*, *20*, 479–486.

Quandt, E. M., Gollihar, J., Blount, Z. D., Ellington, A. D., Georgiou, G., & Barrick, J. E. (2015). Fine-tuning citrate synthase flux potentiates and refines metabolic innovation in the lenski evolution experiment. *eLife*, *4*, e09696.

Raeside, C., Gaffé, J., Deatherage, D. E., Tenaillon, O., Briska, A. M., Ptashkin, R. N., Cruveiller, S., Médigue, C., Lenski, R. E., Barrick, J. E., & Schneider, D. (2014). Large chromosomal rearrangements during a long-term evolution experiment with escherichia coli. *MBio*, *5*(5), e01377.

Rocha, E. P. C. (2006). Inference and analysis of the relative stability of bacterial chromosomes. *Molecular Biology and Evolution*, *23*(3), 513–522.

Rutten, J. P., Hogeweg, P., & Beslon, G. (2019). Adapting the engine to the fuel: Mutator populations can reduce the mutational load by reorganizing their genome structure. *BMC Evolutionary Biology*, *19*(1), 191.

Schrider, D. R., Houle, D., Lynch, M., & Hahn, M. W. (2013). Rates and genomic consequences of spontaneous mutational events in *Drosophila melanogaster*. *Genetics*, *194*(4), 937–954.

Starr, T. N., & Thornton, J. W. (2016). Epistasis in protein evolution. *Protein Science*, *25*(7), 1204–1218.

Trujillo, L., Banse, P., & Beslon, G. (2022). Getting higher on rugged landscapes: Inversion mutations open access to fitter adaptive peaks in nk fitness landscapes. *PLoS Computational Biology*, *18*(10), e1010647.

Vakhrusheva, A. A., Kazanov, M. D., Mironov, A. A., & Bazykin, G. A. (2011). Evolution of prokaryotic genes by shift of stop codons. *Journal of Molecular Evolution*, *72*, 138–146.

Wala, J. A., Bandopadhayay, P., Greenwald, N. F., O'Rourke, R., Sharpe, T., Stewart, C., Schumacher, S., Li, Y., Weischenfeldt, J., Yao, X., Nusbaum, C., Campbell, P., Getz, G., Meyerson, M., Zhang, C. Z., Imielinski, M., & Beroukhim, R. (2018). Svaba: Genome-wide detection of structural variants and indels by local assembly. *Genome Research*, *28*(4), 581–591.

Wang, J., Fan, H. C., Behr, B., & Quake, S. R. (2012). Genome-wide single-cell analysis of recombination activity and de novo mutation rates in human sperm. *Cell*, *150*(2), 402–412.

Wang, Y., Diaz Arenas, C., Stoebel, D. M., Flynn, K., Knapp, E., Dillon, M. M., Wünsche, A., Hatcher, P. J., Moore, F. B.-G., Cooper, V. S., & Cooper, T. F. (2016). Benefit of transferred mutations is better predicted by the fitness of recipients than by their ecological or genetic relatedness. *Proceedings of the National Academy of Sciences*, *113*(18), 5047–5052.

Wei, X., & Zhang, J. (2019). Patterns and mechanisms of diminishing returns from beneficial mutations. *Molecular Biology and Evolution*, *36*(5), 1008–1021.

Weissman, D. B., Feldman, M. W., & Fisher, D. S. (2010). The rate of fitness-valley crossing in sexual populations. *Genetics*, *186*(4), 1389–1410.

Wellenreuther, M., & Bernatchez, L. (2018). Eco-evolutionary genomics of chromosomal inversions. *Trends in Ecology & Evolution*, *33*(6), 427–440.

Wellenreuther, M., Mérot, C., Berdan, E., & Bernatchez, L. (2019). Going beyond snps: The role of structural genomic variants in adaptive evolution and species diversification. *Molecular Ecology*, *28*(6), 1203–1209.

Wiser, M. J., Ribeck, N., & Lenski, R. E. (2013). Long-term dynamics of adaptation in asexual populations. *Science*, *342*(6164), 1364–1367.

Yancopoulos, S., Attie, O., & Friedberg, R. (2005). Efficient sorting of genomic permutations by translocation, inversion and block interchange. *Bioinformatics*, *21*(16), 3340–3346.

Zhang, J. (2003). Evolution by gene duplication: An update. *Trends in Ecology & Evolution*, *18*(6), 292–298.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Banse, P., Luiselli, J., Parsons, D. P., Grohens, T., Foley, M., Trujillo, L., Rouzaud-Cornabas, J., Knibbe, C., & Beslon, G. (2023). Forward-in-time simulation of chromosomal rearrangements: The invisible backbone that sustains long-term adaptation. *Molecular Ecology*, *00*, 1–14. https://doi.org/10.1111/mec.17234