Gene duplications potentiate adaptive evolution of complex traits

(Short title ditto)

Matthew Andres Moreno[1,2,3,4☯*], Cameron Haynes[5,6☯], Charles Ofria[7,8,9], Alex Lalejini[10,11‡], Luis Zaman[1,2,4‡]

1 Department of Ecology and Evolutionary Biology
2 Center for the Study of Complex Systems
3 Michigan Institute for Data and AI in Society
4 University of Michigan, Ann Arbor, MI, United States
5 Integrated Mathematical Oncology
6 Moffitt Cancer Center, Tampa, FL, United States
7 Program in Ecology, Evolution, and Behavior
8 Department of Computer Science and Engineering
9 Michigan State University, East Lansing, MI, United States
10 Department of Computer Science
11 Grand Valley State University, Grand Rapids, MI, United States

☯These authors contributed equally to this work.
‡These authors also contributed equally to this work.
* morenoma@umich.edu

Abstract

Gene duplication events are widely recognized as a key factor in the evolution of organismal complexity. Mirroring broader contrasts between adaptation- versus contingency-driven hypotheses for the evolutionary origins of biological complexity, gene duplication outcomes are typically framed in terms of neo- and sub-functionalization scenarios. In the former, duplicated genetic material catalyzes novel functionality; in the latter, it is co-opted to elaborate existing functionality. Examples of both scenarios are widespread in natural history, but practical constraints have limited direct experimental investigation of the relationship between gene duplication and organismal complexity. Using the Avida platform for digital evolution, we show that while increased genome size can promote the emergence of simple adaptive traits, gene duplication uniquely facilitates the *de novo* evolution of complex adaptive phenotypes. Tracing the ancestry of individual genetic sites, we find that slip duplication of a site increases its subsequent likelihood to code for novel phenotypic traits. We then harness the unique *in silico* capabilities of our model system to compare evolutionary outcomes across degraded variants of full-fledged gene duplication. This ablative analysis confirms that the observed adaptive potentiation indeed arises from the duplication of existing genetic information. In contrast to purely neutral framings of biological complexity, our results support gene duplication events as a contributing factor in adaptive origins of complex traits.

1 Introduction

A fundamental objective of evolutionary biology is to understand how complex phenotypic traits and new genetic information originate. Among other factors, theory identifies gene duplication as a major contributor to biological diversity and complexity [1–10]. While comparative genomic studies have provided an increasingly detailed account of duplication events shaping genetic and phenotypic traits in nature [11–17], ambiguities remain in bridging these findings with explicit, process-based models of evolutionary dynamics [18]. In particular, it remains difficult to gauge the importance of adaptive processes like neo-functionalization and dose effects versus contingent processes like sub-functionalization [19–21].

In modern evolutionary biology, experimental approaches have become an important complement to retrospective studies, helping to establish causality in evolutionary dynamics [22, 23]. Notable experimental evidence, employing *in vivo* models, has been established for the adaptive significance of duplicative dose effects within evolving populations [24–26]. Also notable is recent *in vivo* experimental work by [27], who apply a directed evolution approach using fluorescence-activated cell sorting to compare the rate of adaptive evolution for between *E. coli* strains differing in copy count for genes encoding fluorescing protein coGFP. Over five rounds of mutagenesis, amplification, and selection, Mihajlovic et al. demonstrate that fluorescence phenotypes in double-copy populations exhibit greater robustness under mutation, but they do not find evidence of faster adaptation in double-copy populations vis-a-vis fluorescence intensity. As the authors highlight, though, achieving the work within their model system required trade-offs in using mutation rates and selection pressures differing significantly from those typical in natural evolution.

The theme of evolution of complexity has drawn notable contributions, in particular, from experiments incorporating digital model systems [28]. While lacking in realism and richness compared to work with biological organisms, *in silico* approaches offer complementary capabilities in supporting principled definitions of complexity [29], fast throughput spanning thousands of generations, and flexible exploration of arbitrary

counterfactuals [30]. In this work, digital systems allow us to systematically enable or disable duplications, supporting direct comparisons of evolutionary outcomes. Insight may also be gained from tracing the fate of duplicated sites, given the availability of exhaustive mutational histories.

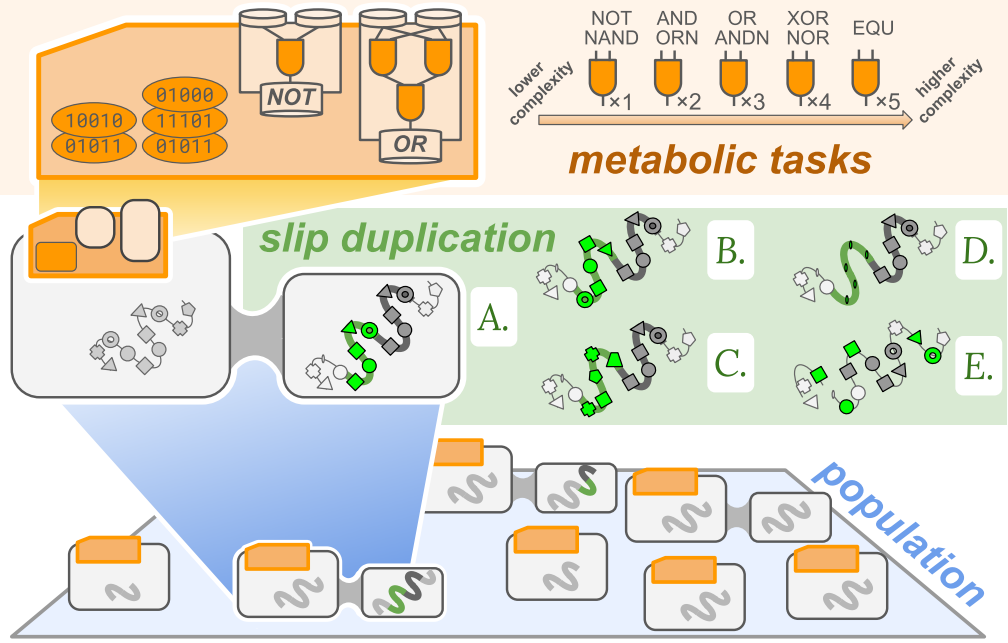


Fig 1. Genome replication and phenotypic traits in Avida. Self-replicating computer programs serve as digital model organisms (bottom panel). Organisms comprise virtual stacks and registers used to store binary values and pointers within a genome of program instructions used to track instruction execution and copying. Competition to survive and reproduce occurs within a limited-capacity population. Replication activity can be accelerated by carrying out available "metabolic" input/output tasks (top panel). These tasks vary in complexity with respect to the number of NAND operations required to perform them. An organism's metabolic "phenotype" arises from the expression of its genetic code. Genetic code copied from parent to offspring may be subject to point mutations, which change the individual instruction values, and slip mutations, which introduce or remove many instructions all at once (slip inserts shown as bright green). Reported experiments compare five alternate variants of slip mutation: A) *Slip-duplicate*, an exact duplication is inserted adjacent to the target segment; B) *Slip-scramble*, shuffled duplication is inserted directly after the target segment; C) *Slip-random*, random instructions are inserted directly after the target segment; D) *Slip-NOP*, neutral nop-X instructions are inserted directly after the target segment; and E) *Slip-scatter*, randomly-drawn instructions are inserted at random throughout the genome.

Here, we investigate the role of adaptive innovation in the relationship between gene duplication and the evolution of complexity using the Avida digital evolution platform. This system allows evolution experiments to be conducted using populations of self-replicating computer programs, operating as digital "model organisms" [31]; owing to heritable variation in replication rate introduced by copy errors, evolution by natural selection unfolds within these populations [32]. Among other early work with Avida, Lenski *et al.* established a useful framework for quantifying the complexity of phenotypic traits evolved by Avida organisms, by counting the minimum number of necessary substeps to accomplish a task of interest [33]. Applying this lens to a phenotype ("EQU") comprising a large number of substeps, controlled experiments and step-by-step lineage analyses demonstrated that simple traits could provide building blocks for EQU — and, in fact, amounted to necessary prerequisites for the evolution of complex traits. Other work with Avida has yielded wide-ranging contributions shedding light on selection's influence on information-theoretic measures of genome

complexity [34], co-evolutionary pressures toward complex phenotypic traits [35], and the abundance of genetic encodings for complex phenotypes [36].

To determine the role of novel adaptations in complexity arising from gene duplication, we independently assessed the influence of slip-duplication mutations on both genetic and phenotypic complexity. We measured genetic complexity by counting the program instructions contributing to an organism's fitness. For phenotypic complexity, we used the minimal quantity of computational substeps necessary to recreate exhibited behaviors [33]. We find that gene duplication facilitates adaptive evolution of phenotypic traits, with this adaptive benefit appearing exclusively for complex traits. Moreover, we show genome sites encoding complex phenotypic traits to be disproportionately localized within slip-duplicated regions. Genetic complexity, by comparison, did not appreciably respond to gene duplications. Our results therefore contrast with perspectives on biological complexity according lesser significance to the role of adaptive innovation [37–39].

Figure 1 provides a schematic overview of the experiments conducted in our work. All experiments comprised well-mixed populations with a carrying capacity of 3,600 Avida organisms. To supply adaptive potential, we adopted an framework developed by Lenski [33] to define a set of nine advantageous phenotypic traits. Each trait corresponds to a possible logical transform on available binary inputs. Under this scheme, organisms producing correct output values for a task benefit by accelerating their genome evaluation (and, thus, also their self-replication), analogously to a metabolic process. Formally, "substeps" necessary to carry out each task may be quantified in terms of the minimum number of NAND operations necessary to carry it out [33].[1] As seen in Figure 1, the most complex function (EQU) requires five NAND operations while the simplest (NAND and NOT) require only one.

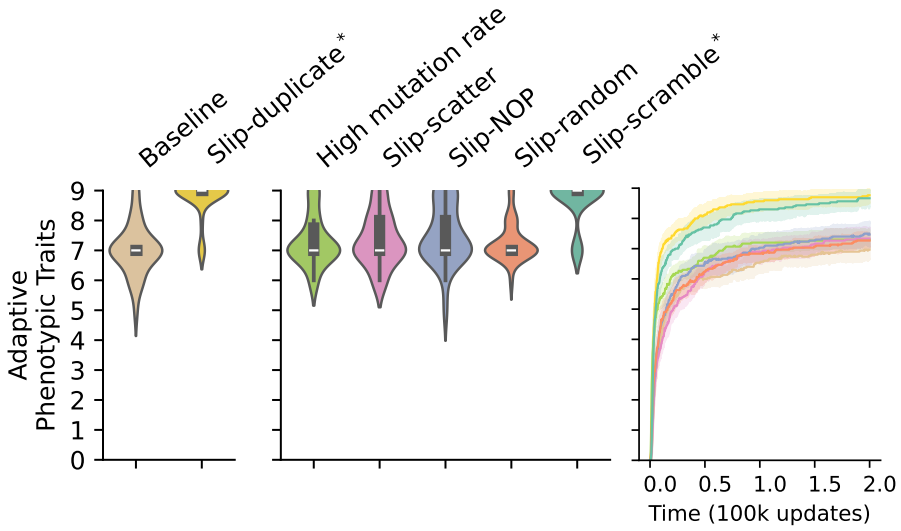
Avida organisms replicate asexually by copying their genome one locus at a time. In addition to a baseline point mutation rate, our experiments modeled gene duplication events using "slip mutation" processes analogous to replication slippage [41]. Under baseline conditions, these slip mutations allow arbitrary segments of program content to be duplicated or excised. As overviewed in Figure 1, we also tested a series of mutation operators isolating particular functional effects of gene duplication, in order to tease apart causality in greater detail. Experiments began with 100-site genomes. To control for the effects of genome length, we also included trials with longer 1,000-site genomes, which was near the upper extreme of genome lengths observed over the course of slip-duplication experiments.

2 Results

2.1 Gene duplication facilitates adaptive evolution of complex traits

In a first set of experiments, we investigated the impact of slip duplications on *de novo* evolution of adaptive phenotypic traits. Comparing the acquisition of phenotypic traits between slip-duplicate-enabled and baseline treatments, we found slip duplications to yield significantly more evolved adaptive phenotypic traits (two-tailed Mann-Whitney U test, $W = 562.5$, Bonferroni-adjusted $p << 0.0001$; Figure 2a). Investigating further, most replicates saw substantial growth in genome length from the initial length of 100 sites — in some cases, a tenfold increase, reaching sizes slightly more than 1,000 sites. To test whether adaptive benefit was attributable to genome length, we performed

[1] Computer architecture theory considers NAND as a fundamental basis operation, since all other logic can be derived from compositions of NAND operations [40].



(a) slip-duplication **(b)** ablation treatments **(c)** lineage history

Fig 2. Treatments preserving slip-duplicated content facilitate adaptive evolution. Violin plots show number of adaptive traits evolved in final dominant genotypes. Time series (2c right) shows progression of adaptive phenotypic trait counts along lineages of final dominant genotypes; color-coding corresponds to violin plots. Asterisk (*) markers indicate treatments with significantly more adaptive phenotypic traits compared to baseline, comparison across both 2a and 2b panels. Simulation time unit is "updates," corresponding to evaluation of 30 genome sites per organism.

additional experiments incorporating a "long-genome" control treatment initialized using genomes extended to 1,000 sites with neutral inserts.

Comparing the long-genome and baseline treatments, we observed increased genome length alone to also significantly boost task acquisition. Disaggregating by task complexity, though, reveals impact of genome length as most prominent in the acquisition of simple tasks. Figure 3 compares acquisition rates for tasks across task complexity classes — with and without slip duplication, including the long-genome control. The long-genome control matched or exceeded the performance of slip-duplication in evolving simple traits with 3 or fewer components. However, slip duplication evolved more complex 4- and 5-component traits within a significantly higher fraction of replicates compared to the long-genome control (Fisher's exact tests; 36/60 vs. 24/60, $p < 0.05$ [4 components]; 10/30 vs. 2/30, $p < 0.03$ [5 components]; Figure 3).

2.2 Information content of duplications provides adaptive benefit

Having observed that slip-duplicate mutations accelerate evolution of adaptive phenotypic traits, we next sought to isolate the aspects of slip duplication contributing to adaptation. For this purpose, we tested four variants of the slip-duplication operator, disabling or replacing a particular aspect of slip duplication (overviewed in Figure 1), as well as an additional "high mutation rate" treatment where single-site insertion/deletion mutations were applied in lieu of slip mutation.

As shown in Figure 2, we detected benefits to adaptive evolution only for the follow-up slip-scramble treatment — which randomized sequence order within duplicated regions (two-tailed Mann-Whitney U test; Figures 2a and 2b). All other

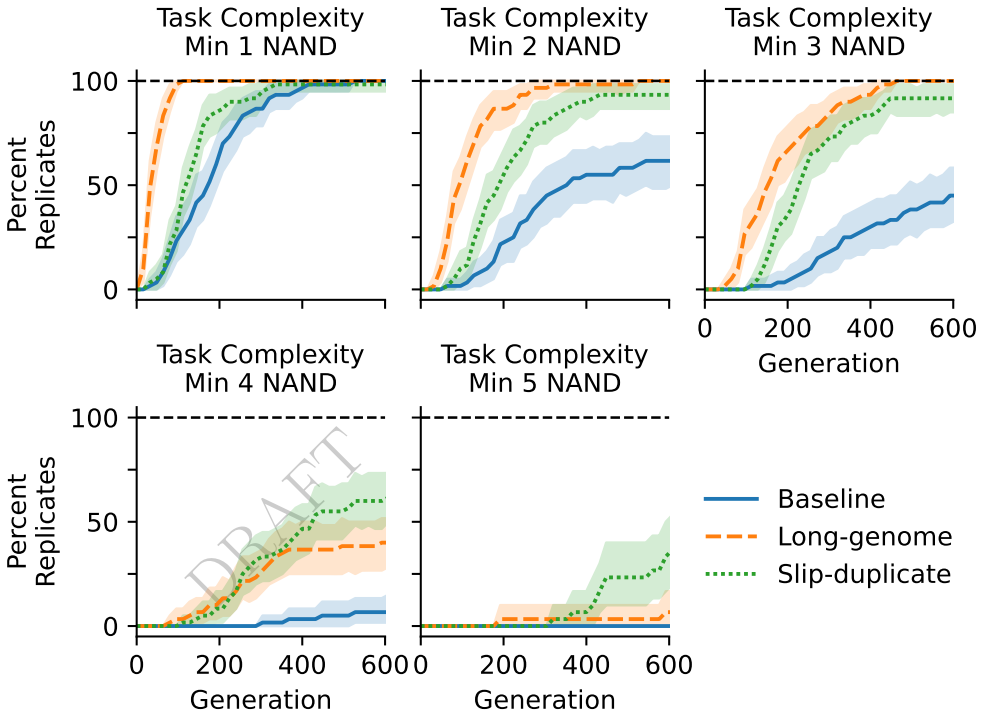


Fig 3. Gene duplication boosts adaptive evolution of complex phenotypic traits. Plots show fraction of replicates exhibiting available phenotypic traits, by generation from founding ancestor. Panels facet by trait complexity, measured by the minimum number of NAND operations required to complete the task. Simple tasks (top left) require only one NAND operation. More complex tasks (bottom right) require up to five NAND operations as shown in Figure 1. Slip-duplication treatment facilitates significantly faster adaptive evolution than long-genome treatment for the more complex tasks that require 4 or 5 subcomponents. Error bands give 95% CI, bootstrapped over 30 replicates per treatment.

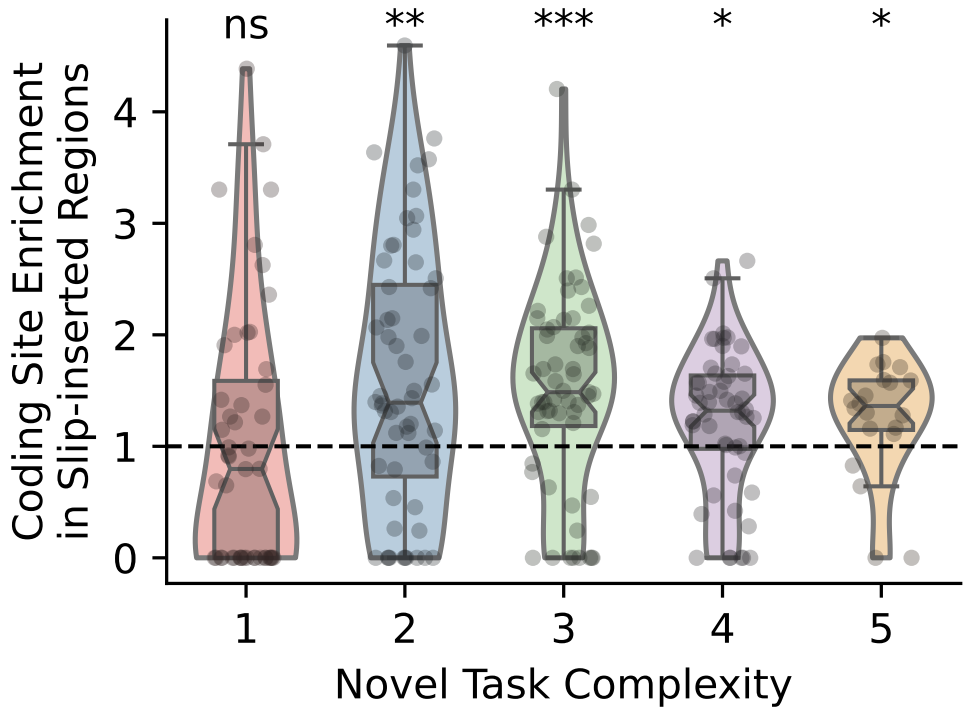


Fig 4. *De novo* coding sites for complex traits are overrepresented in slip-duplicated regions. Distributions compare enrichment of coding sites for novel logic-9 traits in slip-duplicated regions, normalized to neutral expectation. Values greater than 1 indicate that coding sites of novel traits occur more often in slip-duplicated regions compared to their background frequency. Each observation is enrichment of coding sites density within slip-duplicated regions at the first occurence of each observed trait; observation counts are 48, 56, 59, 52, and 20. Significance of deviation from null expectation median value of 1.0 is indicated with * ($p < 0.05$), ** ($p < 0.01$), or *** ($p < 0.001$) (one-tailed Wilcoxon signed-rank test).

slip-duplicate variants were indistinguishable or slower-adapting compared with the baseline treatment.

Given the efficacy of the slip-scramble treatment in facilitating adaptation, we additionally tested whether phenotypic adaptation differed between the slip-scramble and full-fledged slip-duplicate operators. To prevent issues with multiple comparisons, we ran 100 new trials under both treatments for this test. We found that the slip-duplicate treatment did, in fact, yield higher task counts compared to the slip-scramble treatment (two-tailed Mann-Whitney U test, $W = 4305$ respectively, Bonferroni-adjusted $p < 0.02$). As shown in Supplementary Figure 8, full slip-duplication was also associated with significantly larger genome size compared to the slip-scramble treatment — by a factor of approximately 94% (two-tailed Mann-Whitney U test, $U = 734$, $p < 0.001$).

2.3 Duplicated sites are potentiated for complex traits

Thus far, we have established that slip duplication can promote evolution of novel traits, with this effect biasing toward complex traits. We next sought to understand whether duplicated genetic material itself exhibits elevated potential to code for novel traits.

To address the question, we assessed the density of coding sites for novel adaptive traits in genome regions that had previously been slip duplicated. Coding sites for these traits arising more frequently than expected by chance in duplicated regions would

suggest that these regions are "potentiated" — that is, possessing latent gene content predisposed to produce a new trait. Figure 4 compares the involvement of sites having undergone slip duplication — versus those that had not — in coding for novel traits. For the simplest tasks, requiring only one NAND component, we found no significant difference in the likelihood of duplicated sites participating in coding regions for new tasks. However, we found significant associations for traits with two or more NAND components (one-sample Wilcoxon signed-rank tests; $n = 56, 59, 52, 20$ observations). Effect sizes on probability to code for novel traits were 1.6×, 1.6×, 1.2×, and 1.2×, respectively, for 2, 3, 4, and 5 task components. Smaller effect sizes at 4- and 5-component tasks may be due to a larger portion of the genome becoming comprised of slip-duplicated sites (Supplementary Figure 10), thus lowering the upper ceiling on deviation from expected.

One possible confounding factor in coding site analysis is evolutionary constraint at genome sites involved in organisms' self-replication loop. These sites are critical to viability, with lethal outcomes when knocked out. We found that these critical sites were less likely to be involved in slip duplication and also less likely to be involved in coding for *de novo* traits. However, after excluding such fitness-critical sites from analysis, we still found generally similar potentiation signatures from slip duplication (Supplemental Figure 10).

In addition to neofunctionalization, gene duplications are also hypothesized to directly facilitate adaptation by directly producing beneficial mutational changes, for example, through dose effects [42]. In line with this possibility, we observed that a substantial fraction of gain-of-function steps on lineages directly coincided with slip duplications — 41 of 174, or 23.6%. However, in these cases, we still found evidence that sites coding for a new trait directly were more likely than chance to have been involved in earlier slip duplications (Supplemental Figure 10).

2.4 Genome length drives genetic complexity

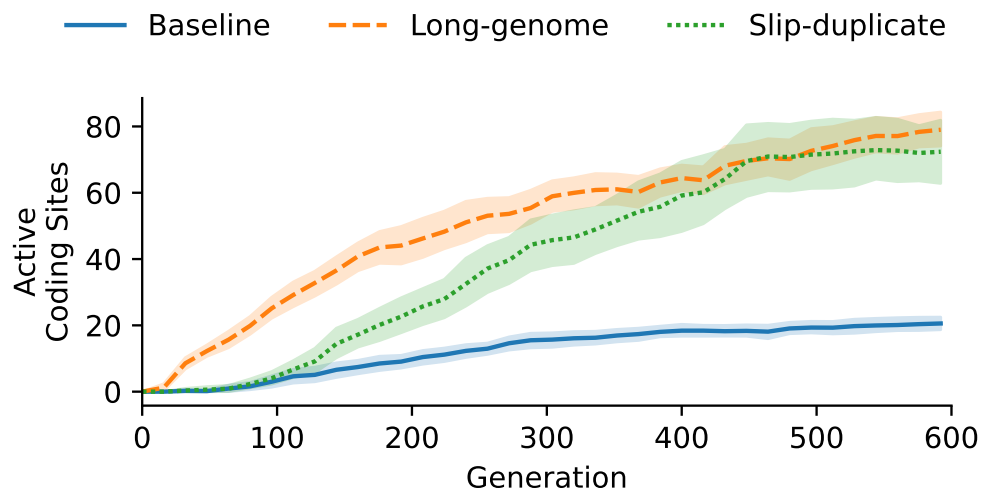
In a final set of analyses, we broadened our scope to assess consequences of gene duplication on whole-genome architecture with respect to genome robustness, which we defined in terms of sensitivity of fitness to mutation [43]. We quantified robustness by counting the number of "critical" genome sites, where a single-site knockout disrupted replicator viability or one or more adaptive phenotypic traits.[2]

One conventional perspective on gene duplication is *vis-a-vis* neutral dynamics, wherein copied genetic material reduces brittleness by introducing redundancy [45]. To assess the relevance of this model within our study system, we performed slip-duplicate mutational assays to quantify the baseline effect of slip insertion mutations on robustness. We applied our assay to final-dominant genome lineages evolved with slip duplication, sampling one slip duplication per genome and measuring change in critical site counts between corresponding wildtype and mutated variants.

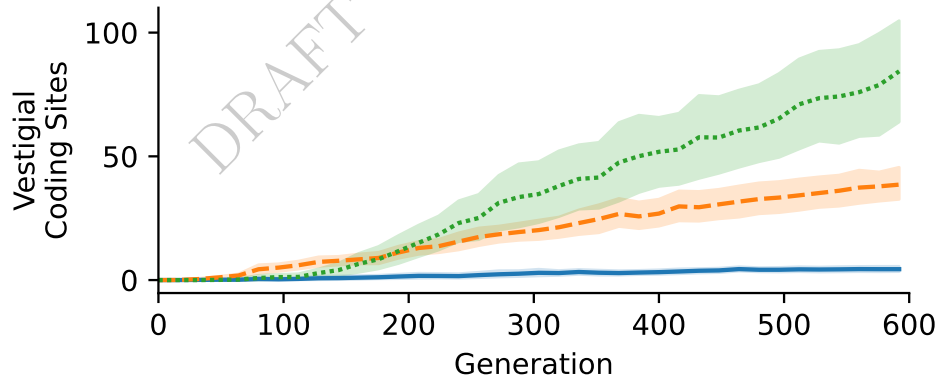
On average, we found that fitness-neutral slip insertions decreased coding site count by 6.8 sites (bootstrapped 95% CI 6.4 to 7.3; median 3% of coding sites). This effect was strongest in genomes with high complexity; for instance, neutral insertion mutations decrease coding site count by 9.2 and 8.3 sites on average in genomes that encode 4- and 5-component complexity tasks, respectively (bootstrapped 95% CIs 8.5 to 9.9 and 7.2 to 9.3; median 3% and 5% of coding sites). Supplementary Figure 9 presents these results.

To assess evolutionary consequences of redundancies introduced by slip-duplication, we next analyzed coding site accumulation within genomes over the course of evolution. Counter to naive expectation, we found that the slip duplication treatment accrued

[2] Although sufficiently representative for our purposes, limitations exist in detecting Avida genome functionality through single-site knockouts; such an approach can underestimate aspects of genome sequence complexity involving small effects or redundancy [43, 44].



(a) active coding sites



(b) vestigial coding sites

Fig 5. Gene duplication boosts accumulation of vestigial coding sites. Generation-by-generation counts of coding sites over evolutionary history. Here, "active" coding sites refer to genome instructions determined through knockout to contribute to fitness with respect to self-copy viability or a rewarded phenotypic trait. As shown in panel 5a, gene duplication yields active coding site counts comparable to long-genome control. Vestigial coding site count, by contrast, reports the number of sites determined to have contributed to fitness in an ancestor, but are no longer active coding sites. As shown in panel 5a, vestigial coding site count under slip-duplication treatment outpace control treatments. Error bands give 95% CI, bootstrapped over 30 replicates per treatment.

fitness-critical sites at a generation-on-generation rate comparable to the long-genome baseline treatment; Mann-Whitney test, $U = 361$, $p = 0.19$). Despite this similarity, however, when measurements were taken inclusive of vestigial coding sites (those which had *previously* been fitness-critical earlier within a lineage), we found a significantly increased coding site count associated with the slip-duplicate treatment (Mann-Whitney test, $U = 630$, $p < 0.01$). Figure 5 compares growth in active versus vestigial coding site counts for baseline, long-genome, and slip-duplication treatment.

3 Discussion

In this work, we used *in silico* evolution experiments to investigate how gene duplication influences the evolution of biological complexity. Specifically, we examined the hypothesis that gene duplication acts in shaping adaptive evolution of complex traits. To conduct these experiments, we adapted a framework developed by Lenski et al. [33] to introduce adaptive potential for phenotypic traits across a well-defined spectrum of functional complexity (Figure 1).

Overall, our results support the premise that gene duplication can promote the *de novo* evolution of adaptive traits. Leveraging the unique tractability of our study system, we were able to explore these dynamics in greater detail, providing key insights into: (1) sensitivity of potentiation effects to trait complexity, (2) contributions of sequence information in duplicated material to adaptive outcomes, and (3) a nuanced role of duplications in enhancing short-term robustness while predominantly promoting accumulation of vestigial coding material rather than neutral increases in complexity.

3.1 Gene duplication facilitates adaptive evolution of complex traits

While we find that gene duplication can facilitate adaptive evolution, we detect this effect only for phenotypic traits with greater functional complexity; adaptive evolution of simpler traits exhibits no benefit beyond the effect of increased genome size alone (Figure 3). Given the compositional nature of trait functionality within our study system [33], this outcome would be consistent with a "building block" model of adaptation, where duplication facilitates discovery of novel medleys of existing components.

In broad strokes, adaptive significance of gene duplication within our study system aligns with existing findings across a wide variety of biological taxa and digital models that slip-duplication of genetic material can facilitate evolution of adaptive traits [13, 20, 46]. Indeed, a striking example of gene duplication potentiating novel adaptation comes from the Long-Term Evolution Experiment in *E. coli* [47, 48], where a duplication in one population broadened expression of a key citrate transporter to aerobic conditions — resulting in a seven-fold increase in carrying capacity within experimental conditions [49]. Among many other examples in nature [50–56], such observations provide a basis for longstanding connections drawn between gene duplications and adaptive traits, dating as far back as [57].

Present work, however, contrasts with recent directed evolution experiments investigating the effect of gene copy count on adaptation rate [27]. This work finds no effect of gene copy count on phenotypic adaptation rate, comparing single- vs. dual-copy coGFP *E. coli* strains under selection for increased fluorescence. Several factors may explain this discrepancy, including our use of longer generational timescales and more naturalistic selection coefficients and mutational processes. An contributing aspect may also be the opportunity for more open-ended, multi-gene composition of phenotypic traits in our study system. Importantly, our work broadens the conversation

by providing systematic and controlled experimental evidence supporting the role of gene duplication in adaptive evolution. In particular, our findings suggest that the adaptive benefits of gene duplications may be especially important for the evolution of complex traits. However, future studies will be necessary to determine whether similar constraints and patterns hold in natural systems, where additional mechanisms such as structured regulatory interactions may significantly influence trait evolution.

3.2 Sequence information of duplications contributes to adaptive outcomes

Our digital evolution study system allowed us to further disentangle which characteristics of gene duplications increase evolutionary potential by imposing variants of the slip mutation operator, isolating individual aspects of duplicative processes (Figure 1). We find that only duplications of existing genetic information — either as-is or in a "scrambled" order — provide adaptive benefit, with those preserving both content and order ultimately performing best. In contrast, slip mutation variants that insert neutral or random genetic material provide no observable adaptive benefit over a baseline control (Figure 2). These results indicate that both the content and structure of duplicated genetic material contribute to facilitating adaptive evolution, beyond the impact of side effects such as an increase in genome size or mutational supply. This finding is consistent with theoretical expectations that gene duplications generate both raw material and combinatorial novelty, enabling subfunctionalization and neofunctionalization [37, 57].

Interestingly, unlike our long-genome control, we did not observe an adaptive benefit from the NOP-insert slip-duplication treatment (Figure 2). One possible explanation for this difference is selection against mutational load associated with increased genome length. Indeed, in preliminary experiments we found that deleterious mutational load associated with larger genome size frequently drove extreme genome shrinkage, necessitating a lower bound on genome size. As shown in Supplementary Figure 8, we found genome size to grow significantly larger under full slip-duplication compared to NOP-insert slip-duplication — by a factor of almost 4×. This pattern aligns with established theory that mechanistic factors such as dose effects or epistatic drift are necessary in ensuring preservation of new genome content [19, 58–61].

3.3 Duplicated sites are potentiated for complex traits

Through step-by-step analysis of lineage histories, we found that duplicated content was disproportionately likely to contribute to novel complex traits (Figure 4). An outsized fraction of coding sites for newly gained adaptive traits could be traced back to duplicated regions, and a substantial fraction of gain-of-function mutations (41 of 174 along analyzed lineages) directly coincided with slip duplications. Thus, the adaptive characteristics of slip duplication observed in our system likely result from a combination of direct facilitation and potentiation of subsequent neo-functionalization. This pattern concords with large-scale analyses of biological genomes, which have revealed that a high fraction of genes show evidence of having arisen from duplications [12, 13]. As such, our results support long-standing hypotheses about the role of gene duplications in evolutionary innovation and highlight the mechanistic basis for the observed increase in evolvability, particularly for traits requiring multiple interacting components.

3.4 Genome length drives genetic complexity

Consistent with evidence that duplication-associated redundancy can boost robustness [37], we found that slip insertions with neutral fitness effects tended to

reduce the number of genome sites detectable as a single point of failure for metabolic tasks or reproductive viability (Supplemental Figure 9). Within evolutionary trials in our study system, slip duplication appears to increase the net supply of coding material in the genome available to neutral processes, but does not significantly affect accumulation of genetic brittleness (Figure 5). One possible explanation is that, despite slip duplication boosting redundancy in the short term, mutational erosion and selective pressures against mutational fragility drive both the slip-duplication and long-genome treatments to converge on a similar number of critical sites. On the other hand, observed proliferation of vestigial coding material by gene duplications aligns with adaptive potentiation effects discussed earlier. Given that comparative studies have linked ancient duplication events with subsequent gains in genetic robustness and evolutionary innovation [3], these findings suggest directions for future experimental work investigating how such dynamics unfold over extended timescales to shed light on possible connections to larger-scale evolutionary patterns.

4 Materials and Methods

4.1 Evolution trials

We conducted experiments using a custom version of Avida v2.14.0, extended to support our slip mutation ablation treatments [31]. Supporting software and executable notebooks for this work are freely available under the GNU GPL license, via GitHub at `https://github.com/chaynes2019/AvidaGeneDupe/` and archived via Zenodo [62, 63]. Simulation data is archived via the Open Science Framework at `https://osf.io/j5s4h/` [64, 65], provided with a CC-By Attribution 4.0 International license. Some components of our experiments were adapted from an earlier version of this work [66]. Code used in this project incorporated numerous pieces of open-source scientific software [67–74].

Population size was configured to the Avida default of 3,600 organisms across all trials, with well-mixed arrangement analogous to chemostat conditions. We configured available metabolic resources consisting of tasks NOT, NAND, OR-NOT, AND, OR, AND-NOT, NOR, XOR, and EQUALS. Identical reward was provided for performing each task and kept consistent throughout trials. Rewards accrued independently for each task, enabling fitness benefits to be compounded by performing multiple tasks. In some analyses, we report a count of adaptive phenotypic traits, ranging from a minimum of 0 (*i.e.* the organism performs no metabolic tasks and receives no reward) to a maximum of 9 (*i.e.* the organism performs all 9 available metabolic tasks and receives the maximum reward). Phenotypic traits were also described in terms of minimum required NAND operations as a metric of complexity, as given in Figure 1 [33].

Except where otherwise noted, experiments were seeded with a 100-instruction ancestral self-replicator. In all cases, mutations reducing genome size below 100 instructions were disallowed to ensure sufficient raw genetic material to encode phenotypic traits. Run length was 200,000 "updates" across all experiments, a unit of simulation time roughly equivalent to the amount of time required for an organism to execute 30 instructions. This duration was sufficient to observe at least 600 generations of evolution in all trials.

4.2 Mutation operators

Genomes in our experiments comprised linear sequences of instructions, providing 27 operations for basic computations, execution flow control, input and output, and self-replication. This instruction set is Turing complete and syntactically robust,

meaning that any ordering of instructions is syntactically valid. When an organism produced an offspring, we imposed extrinsic mutational operations on the offspring genome: substitution, insertion, deletion, and slip mutations. *Copy mutations* introduce an erroneous *substitution* where an instruction is written into the offspring genome instead of the intended instruction. In our experiments, substitutions occurred with a per-site probability of 0.0025 [33].

Insertion mutations, by contrast, insert an arbitrary instruction at random, increasing the length of the offspring's genome by one. In a similar vein, *deletion* mutations act by removing a random instruction from the offspring's genome. In our experiments, each was allowed to occur with a probability of 0.05 per offspring [33].

For the purposes of this study, we augmented Avida with an additional mutation operator: *slip mutation*, designed to act analogously to gene duplications and deletions caused by replication slippage events [41]. When a slip mutation occurs, two sites in the offspring genome are randomly selected, defining the target segment for the operation. If the first site is upstream of the second, the slip mutation results in an insertion — as if the organism's replication machinery had slipped backward during replication and copied a segment twice. If the second site is upstream of the first, the slip mutation results in a deletion — as if the organism's replication machinery skipped over a genetic segment. As implemented, slip-insertions and slip-deletions occur with equal probability; thus, absent selection, this mutational process increases genome length variation but does not introduce an inherent bias on mean genome length.

Across trials, we assessed full-fledged slip-duplicate mutation in comparison with four variant mutation operators:

1. **slip-scramble**, where duplicated code was shuffled to test effects of sequence order,
2. **slip-random**, where duplicated code was replaced with random instructions to test effects of sequence content,
3. **slip-NOP**, where duplicated code was replaced with neutral ("no-operation") instructions to test the effect of neutral increase in genome size, and
4. **slip-scatter**, where duplicated code was dispersed across the genome to test the effect of insertion locality.

Figure 1 illustrates example outcomes across surveyed slip-mutation variants.

We also applied a **high mutation rate** control treatment, for which we increased the single instruction insertion/deletion rate to 0.0075, to result in approximately the same number of mutations per divide as under slip mutation.

When insertion occurs, all slip mutation operators add a number of instructions equal to the length of the target segment. However, the composition and location of inserted instructions vary according to the slip mutation operator schema. Deletions acted identically across slip mutation variants, except slip-scatter, which randomized deletions uniformly across the genome. Where enabled, slip mutations occurred with 5% probability per divide event.

4.3 Experimental design

Experiments were conducted in two phases. The first phase tested aggregate differences in evolutionary outcomes between surveyed slip duplication operator variants across runs, while the second focused more heavily on teasing apart evolutionary history and genetic structure within runs. For this reason, first-phase experiments prioritized higher replicate counts, while second-phase experiments prioritized more detailed lineage

tracking of individual genome sites. Replicate count was 100 independent trials for first-phase experiments, and 30 replicates for second-phase experiments.

First-phase experiments consisted of six treatments: one baseline treatment with slip-duplication disabled and five experimental treatments corresponding to the five surveyed slip-duplication mutation operators shown in Figure 1. These treatments differed only in the available mutation operators and the rates at which those operators were applied. First-phase data was used for analyses comparing evolved adaptive phenotypic traits between slip mutation operators.

For second-phase experiments, we included an additional **long-genome baseline treatment** to directly test the adaptive role of large genome size associated with slip-duplicators. Genomes in this treatment operated identically to those in the baseline treatment above, except that they were initialized with a 1,000-site ancestor. The 1,000-site length was chosen to approximate the upper bound of genome sizes observed in first-phase slip-duplication treatment experiments. Second-phase data was used for analyses of task gain by complexity class (Figure 3), analyses of coding site count trajectories (Figure 5), potentiation analyses (Figures figs. 4 and 10 and analyses of slip-duplication outcome distributions (Figure 9). Phenotypic adaption scores shown in Figure 2 were also derived from second-phase experiments.

For experiments incorporating the long-genome baseline control, timecourses of adaptive evolution were compared in terms of generations, rather than in terms of updates (simulation timesteps). This strategy was necessary to account for substantially fewer generations elapsing per thousand updates under the long-genome controls, because of extended copy loop durations. In other experiments, control treatments generally elapsed at least as many generations as the slip-duplication experimental treatment, on account of comparable or shorter genome length. Thus, even when comparing by update in these other experiments, elevated adaptive rates under slip duplication could not be have been caused by more generations elapsing under the slip-duplication treatment.

4.4 Lineage analyses

Time series assessments of evolutionary history were conducted using the ancestral lineage of the most abundant end-state genotype ("final dominant"). In a postprocessing step, we applied Avida's "analyze mode" to identify each ancestor's tasks.

For investigations involving the evolutionary history of individual genome sites, we used mutational metadata saved with lineage files to identify corresponding sites between parent and offspring. Due to memory constraints, these analyses were conducted from population save files recorded at update 50,000 rather than the end-state population at update 200,000. This timepoint was chosen to encompass the phase in which the bulk of adaptive evolution had already transpired.

Acknowledgment

This project benefited from substantial contributions from Mike Wiser, who co-authored earlier work related to this project [66]. We also would like to acknowledge Jeffrey E. Barrick and David M. Bryson for laying the groundwork for this research. Additional thanks to Rosangela Canino-Koning and Emily Dolson for their comments and the Digital Evolution Laboratory for thoughtful discussions, ideas and support. This research was supported in part by the BEACON Center for the Study of Evolution in Action and by Michigan State University through the computational resources provided by the Institute for Cyber-Enabled Research. This material is based upon work

supported by the Eric and Wendy Schmidt AI in Science Postdoctoral Fellowship, a Schmidt Sciences program.

References

1. Zhang J. Evolution by gene duplication: an update. Trends in Ecology & Evolution. 2003;18(6):292–298. doi:10.1016/s0169-5347(03)00033-8.

2. Otto SP, Whitton J. Polyploid Incidence And Evolution. Annual Review of Genetics. 2000;34(1):401–437. doi:10.1146/annurev.genet.34.1.401.

3. Wagner A. Gene duplications, robustness and evolutionary innovations. BioEssays. 2008;30(4):367–373. doi:10.1002/bies.20728.

4. Crow KD, Wagner GP. What is the role of genome duplication in the evolution of complexity and diversity? Molecular biology and evolution. 2006;23(5):887–892.

5. Magadum S, Banerjee U, Murugan P, Gangapur D, Ravikesavan R. Gene duplication as a major force in evolution. Journal of genetics. 2013;92(1):155–161.

6. Metz CW. Duplication of Chromosome Parts as a Factor in Evolution. The American Naturalist. 1947;81(797):81–103.

7. Hu T, Banzhaf W. Evolvability and Speed of Evolutionary Algorithms in Light of Recent Developments in Biology. Journal of Artificial Evolution and Applications. 2010;2010(4):1–28.

8. Castellanos MdP, Wickramasinghe CD, Betrán E. The roles of gene duplications in the dynamics of evolutionary conflicts. Proceedings of the Royal Society B: Biological Sciences. 2024;291(2024). doi:10.1098/rspb.2024.0555.

9. Lynch M, Force AG. The Origin of Interspecific Genomic Incompatibility via Gene Duplication. The American Naturalist. 2000;156(6):590–605. doi:10.1086/316992.

10. Lynch M, Conery JS. The Origins of Genome Complexity. Science. 2003;302(5649):1401–1404. doi:10.1126/science.1089370.

11. Zhang G, Li C, Li Q, Li B, Larkin DM, Lee C, et al. Comparative genomics reveals insights into avian genome evolution and adaptation. Science. 2014;346(6215):1311–1320. doi:10.1126/science.1251385.

12. Teichmann SA, Park J, Chothia C. Structural assignments to the Mycoplasma genitalium proteins show extensive gene duplications and domain rearrangements. Proceedings of the National Academy of Sciences. 1998;95(25):14658–14663.

13. Teichmann SA, Babu MM. Gene regulatory network growth by duplication. Nat Genet. 2004;36(5):492–496. doi:10.1038/ng1340.

14. DuBose JG, de Roode JC. The link between gene duplication and divergent patterns of gene expression across a complex life cycle. Evolution Letters. 2024;8(5):726–734. doi:10.1093/evlett/qrae028.

15. Voordeckers K, Brown CA, Vanneste K, van der Zande E, Voet A, Maere S, et al. Reconstruction of Ancestral Metabolic Enzymes Reveals Molecular Mechanisms Underlying Evolutionary Innovation through Gene Duplication. PLoS Biology. 2012;10(12):e1001446. doi:10.1371/journal.pbio.1001446.

16. Pougach K, Voet A, Kondrashov FA, Voordeckers K, Christiaens JF, Baying B, et al. Duplication of a promiscuous transcription factor drives the emergence of a new regulatory network. Nature Communications. 2014;5(1). doi:10.1038/ncomms5868.

17. Moore RC, Purugganan MD. The early stages of duplicate gene evolution. Proceedings of the National Academy of Sciences. 2003;100(26):15682–15687. doi:10.1073/pnas.2535513100.

18. Welch JJ. What's wrong with evolutionary biology? Biology & Philosophy. 2016;32(2):263–279. doi:10.1007/s10539-016-9557-8.

19. Innan H, Kondrashov F. The evolution of gene duplications: classifying and distinguishing between models. Nature Reviews Genetics. 2010;11(2):97–108. doi:10.1038/nrg2689.

20. Zhang J. Evolution by gene duplication: an update. Trends in Ecology & Evolution. 2003;18(6):292–298.

21. Kuzmin E, Taylor JS, Boone C. Retention of duplicated genes in evolution. Trends in Genetics. 2022;38(1):59–72. doi:10.1016/j.tig.2021.06.016.

22. Kawecki TJ, Lenski RE, Ebert D, Hollis B, Olivieri I, Whitlock MC. Experimental evolution. Trends in Ecology & Evolution. 2012;27(10):547–560. doi:10.1016/j.tree.2012.06.001.

23. Elena SF, Lenski RE. Evolution experiments with microorganisms: the dynamics and genetic bases of adaptation. Nature Reviews Genetics. 2003;4(6):457–469. doi:10.1038/nrg1088.

24. Tong K, Datta S, Cheng V, Haas DJ, Gourisetti S, Yopp HL, et al. Genome duplication in a long-term multicellularity evolution experiment. Nature. 2025;639(8055):691–699. doi:10.1038/s41586-025-08689-6.

25. Dhar R, Bergmiller T, Wagner A. Increased Gene Dosage Plays A Predominant Role In The Initial Stages Of Evolution Of Duplicate Tem-1 Beta Lactamase Genes: Experimental Evolution Of Duplicate Genes. Evolution. 2014;68(6):1775–1791. doi:10.1111/evo.12373.

26. Näsvall J, Sun L, Roth JR, Andersson DI. Real-Time Evolution of New Genes by Innovation, Amplification, and Divergence. Science. 2012;338(6105):384–387. doi:10.1126/science.1226521.

27. Mihajlovic L, Iyengar BR, Baier F, Barbier I, Iwaszkiewicz J, Zoete V, et al. A direct experimental test of Ohno's hypothesis. eLife. 2025;13. doi:10.7554/elife.97216.3.

28. Fortuna MA, Beslon G, Ofria C. Editorial: Digital evolution: Insights for biologists. Frontiers in Ecology and Evolution. 2022;10. doi:10.3389/fevo.2022.1037040.

29. Adami C. What is complexity? BioEssays. 2002;24(12):1085–1094. doi:10.1002/bies.10192.

30. Langton CG. Artificial Life. 1st ed. Routledge; 1989.

31. Ofria C, Bryson DM, Wilke CO. Avida: A Software Platform for Research in Computational Evolutionary Biology. In: Artificial Life Models in Software. Springer; 2009. p. 3–35.

32. Pennock RT. Models, simulations, instantiations, and evidence: the case of digital evolution. Journal of Experimental & Theoretical Artificial Intelligence. 2007;19(1):29–42. doi:10.1080/09528130601116113.

33. Lenski RE, Ofria C, Pennock RT, Adami C. The evolutionary origin of complex features. Nature. 2003;423(6936):139–144. doi:10.1038/nature01568.

34. Adami C, Ofria C, Collier TC. Evolution of biological complexity. Proceedings of the National Academy of Sciences. 2000;97(9):4463–4468. doi:10.1073/pnas.97.9.4463.

35. Zaman L, Meyer JR, Devangam S, Bryson DM, Lenski RE, Ofria C. Coevolution Drives the Emergence of Complex Traits and Promotes Evolvability. PLoS Biology. 2014;12(12):e1002023. doi:10.1371/journal.pbio.1002023.

36. Fortuna MA, Zaman L, Ofria C, Wagner A. The genotype-phenotype map of an evolving digital organism. PLOS Computational Biology. 2017;13(2):e1005414. doi:10.1371/journal.pcbi.1005414.

37. Lynch M, Conery JS. The Evolutionary Fate and Consequences of Duplicate Genes. Science. 2000;290(5494):1151–1155. doi:10.1126/science.290.5494.1151.

38. Beslon G, Liard V, Parsons DP, Rouzaud-Cornabas J. In: Of Evolution, Systems and Complexity. Springer International Publishing; 2021. p. 1–18. Available from: http://dx.doi.org/10.1007/978-3-030-71737-7_1.

39. Lynch M. The frailty of adaptive hypotheses for the origins of organismal complexity. Proceedings of the National Academy of Sciences. 2007;104(suppl_1):8597–8604. doi:10.1073/pnas.0702207104.

40. Mano MM, Kime CR. Logic and computer design fundamentals. Prentice-Hall, Inc.; 1997.

41. Bzymek M, Lovett ST. Instability of repetitive DNA sequences: The role of replication in multiple mechanisms. Proceedings of the National Academy of Sciences. 2001;98(15):8319–8325. doi:10.1073/pnas.111008398.

42. Kondrashov FA. Gene duplication as a mechanism of genomic adaptation to a changing environment. Proceedings of the Royal Society B: Biological Sciences. 2012;279(1749):5048–5057.

43. Lenski RE, Ofria C, Collier TC, Adami C. Genome complexity, robustness and genetic interactions in digital organisms. Nature. 1999;400(6745):661–664. doi:10.1038/23245.

44. Moreno MA. Methods to Estimate Cryptic Sequence Complexity. In: The 2024 Conference on Artificial Life. MIT Press; 2024.Available from: https://arxiv.org/abs/2404.10854.

45. Wagner A. Genetic redundancy caused by gene duplications and its evolution in networks of transcriptional regulators. Biological Cybernetics. 1996;74(6):557–567. doi:10.1007/bf00209427.

46. Koza JR. Gene duplication to enable genetic programming to concurrently evolve both the architecture and work-performing steps of a computer program. Ijcai (1). 1995;.

47. Lenski RE, Rose MR, Simpson SC, Tadler SC. Long-Term Experimental Evolution in Escherichia coli. I. Adaptation and Divergence During 2, 000 Generations. The American Naturalist. 1991;138(6):1315–1341. doi:10.1086/285289.

48. Lenski RE. Phenotypic and Genomic Evolution during a 20, 000-Generation Experiment with the BacteriumEscherichia coli; 2003. Available from: `http://dx.doi.org/10.1002/9780470650288.ch8`.

49. Blount ZD, Barrick JE, Davidson CJ, Lenski RE. Genomic analysis of a key innovation in an experimental Escherichia coli population. Nature. 2012;489(7417):513–518. doi:10.1038/nature11514.

50. Zhang J, Zhang Yp, Rosenberg HF. Adaptive evolution of a duplicated pancreatic ribonuclease gene in a leaf-eating monkey. Nature Genetics. 2002;30(4):411–415. doi:10.1038/ng852.

51. Gaines TA, Zhang W, Wang D, Bukun B, Chisholm ST, Shaner DL, et al. Gene amplification confers glyphosate resistance in Amaranthus palmeri. Proceedings of the National Academy of Sciences. 2009;107(3):1029–1034. doi:10.1073/pnas.0906649107.

52. Perry GH, Dominy NJ, Claw KG, Lee AS, Fiegler H, Redon R, et al. Diet and the evolution of human amylase gene copy number variation. Nature Genetics. 2007;39(10):1256–1260. doi:10.1038/ng2123.

53. Hittinger CT, Carroll SB. Gene duplication and the adaptive evolution of a classic genetic switch. Nature. 2007;449(7163):677–681. doi:10.1038/nature06151.

54. Benderoth M, Textor S, Windsor AJ, Mitchell-Olds T, Gershenzon J, Kroymann J. Positive selection driving diversification in plant secondary metabolism. Proceedings of the National Academy of Sciences. 2006;103(24):9118–9123. doi:10.1073/pnas.0601738103.

55. Dulai KS, von Dornum M, Mollon JD, Hunt DM. The Evolution of Trichromatic Color Vision by Opsin Gene Duplication in New World and Old World Primates. Genome Research. 1999;9(7):629–638. doi:10.1101/gr.9.7.629.

56. Ogino Y, Kuraku S, Ishibashi H, Miyakawa H, Sumiya E, Miyagawa S, et al. Neofunctionalization of Androgen Receptor by Gain-of-Function Mutations in Teleost Fish Lineage. Molecular Biology and Evolution. 2015;33(1):228–244. doi:10.1093/molbev/msv218.

57. Ohno S. Evolution by Gene Duplication. Springer Berlin Heidelberg; 1970. Available from: `http://dx.doi.org/10.1007/978-3-642-86659-3`.

58. Wagner A. The fate of duplicated genes: loss or new function? BioEssays. 1998;20(10):785–788. doi:10.1002/(sici)1521-1878(199810)20:10¡785::aid-bies2¿3.0.co;2-m.

59. Tautz D. Problems and paradigms: Redundancies, development and the flow of information. BioEssays. 1992;14(4):263–266. doi:10.1002/bies.950140410.

60. Force A, Lynch M, Pickett FB, Amores A, Yan Yl, Postlethwait J. Preservation of Duplicate Genes by Complementary, Degenerative Mutations. Genetics. 1999;151(4):1531–1545. doi:10.1093/genetics/151.4.1531.

61. Bergthorsson U, Andersson DI, Roth JR. Ohno's dilemma: Evolution of new genes under continuous selection. Proceedings of the National Academy of Sciences. 2007;104(43):17004–17009. doi:10.1073/pnas.0707158104.

62. Haynes C, Moreno MA, and AL. chaynes2019/AvidaGeneDupe; 2025. Available from: `https://doi.org/10.5281/zenodo.14911295`.

63. European Organization For Nuclear Research, OpenAIRE. Zenodo; 2013. Available from: `https://www.zenodo.org/`.

64. Moreno MA, Haynes C. Gene duplications potentiate adaptive evolution of complex traits; 2025. Available from: `osf.io/j5s4h`.

65. Foster ED, Deardorff A. Open Science Framework (OSF). Journal of the Medical Library Association. 2017;105(2). doi:10.5195/jmla.2017.88.

66. Lalejini A, Wiser MJ, Ofria C. Gene duplications drive the evolution of complex traits and regulation. In: Artificial Life Conference Proceedings. MIT Press; 2017. p. 257–264.

67. Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, et al. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. Nature Methods. 2020;17:261–272. doi:10.1038/s41592-019-0686-2.

68. Harris CR, Millman KJ, van der Walt SJ, Gommers R, Virtanen P, Cournapeau D, et al. Array programming with NumPy. Nature. 2020;585(7825):357–362. doi:10.1038/s41586-020-2649-2.

69. pandas developers. pandas-dev/pandas: Pandas. Zenodo. 2020;doi:10.5281/zenodo.3509134.

70. Wes McKinney. Data Structures for Statistical Computing in Python. In: Stéfan van der Walt, Jarrod Millman, editors. Proceedings of the 9th Python in Science Conference; 2010. p. 56–61.

71. Waskom ML. seaborn: statistical data visualization. Journal of Open Source Software. 2021;6(60):3021. doi:10.21105/joss.03021.

72. Hunter JD. Matplotlib: A 2D graphics environment. Computing in Science & Engineering. 2007;9(3):90–95. doi:10.1109/mcse.2007.55.

73. Moreno MA. mmore500/teeplot; 2023. Available from: `https://doi.org/10.5281/zenodo.10440670`.

74. Team RC. R: A language and environment for statistical computing; 2015. Available from: `http://www.R-project.org/`.

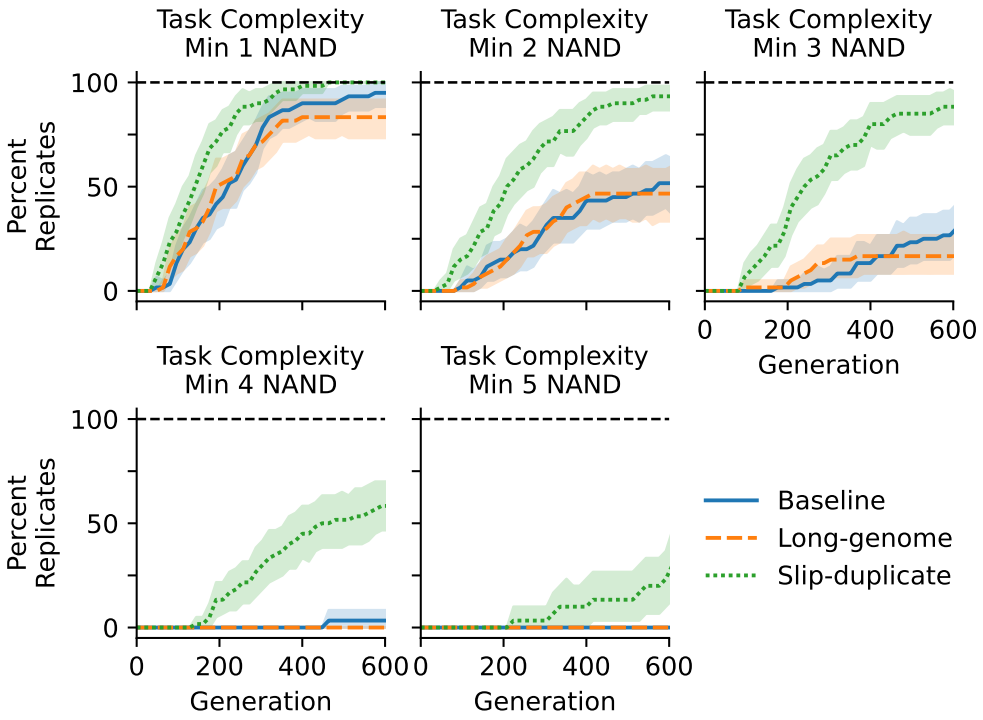


Fig 6. Mutational supply drives increased adaptation rate for long-genome control treatment. Results shown from supplemental control experiments using per-genome, rather than per-site, point mutation processes. In these trials, a count of point mutations was drawn from a Poisson distribution, to be applied at random positions on the genome. Mean per-genome mutation rate was configured corresponding with that under per-site mutation for a 100-site genome. Plots show fraction of replicates exhibiting available phenotypic traits, by generation from founding ancestor. Panels facet by complexity of a trait, measured by the minimum number of NAND operations required to complete the task. Simple tasks (top left) require only one NAND operation. More complex tasks require up to five NAND operations (bottom right). Compared to Figure 3, which shows results with per-site mutation rates, the relative rate of adaptive evolution in long-genome control is diminished under Poisson-distributed treatment, where mutation count is not proportional to genome length. Error bands give 95% CI, bootstrapped over 30 replicates per treatment.

5 Supporting Information

665

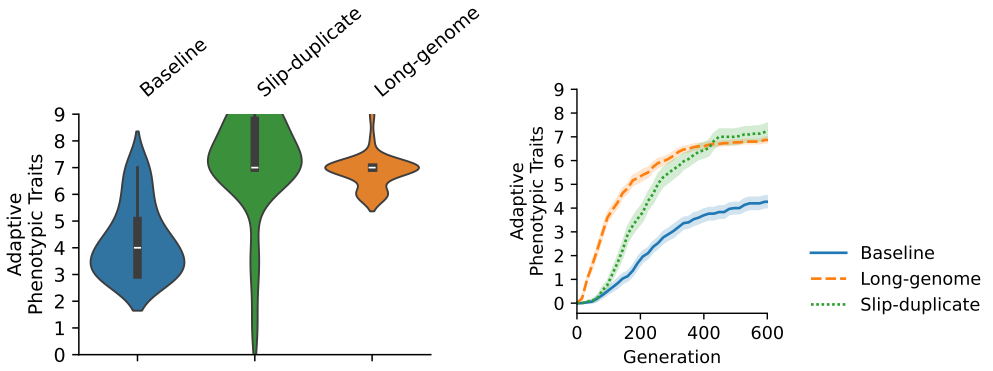
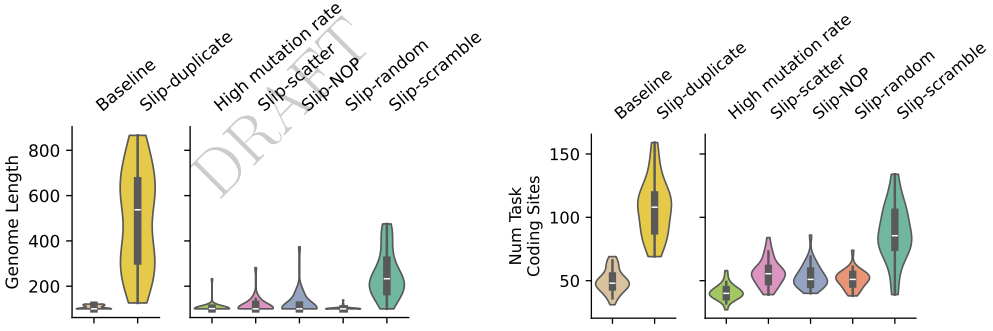
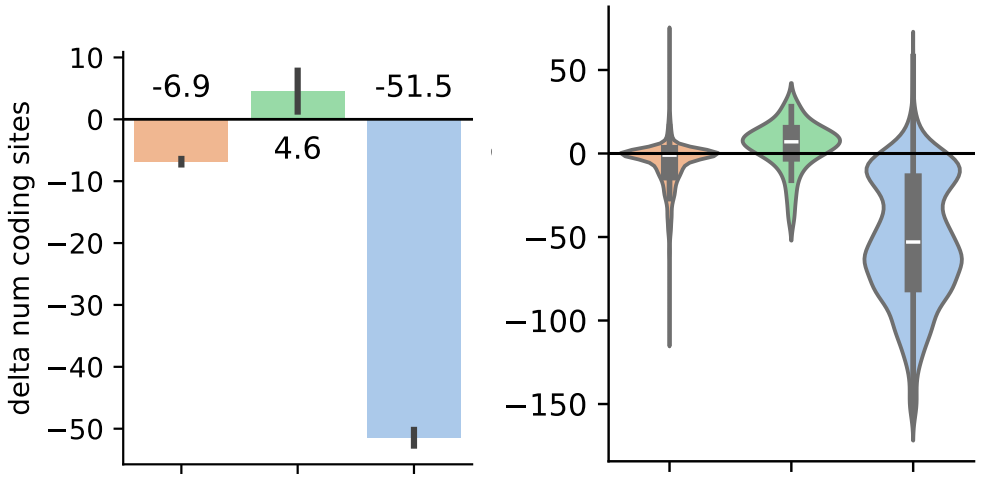


Fig 7. Aggregated adaptive evolution of phenotypic traits for long-genome control experiments. Violin plots show number of adaptive phenotypic traits evolved in final dominant genotypes. Time series (2c right) shows progression of adaptive phenotypic trait count along lineages of final dominant genotypes; color-coding corresponds to violin plots. Error bands give 95% CI, bootstrapped over 30 replicates per treatment.

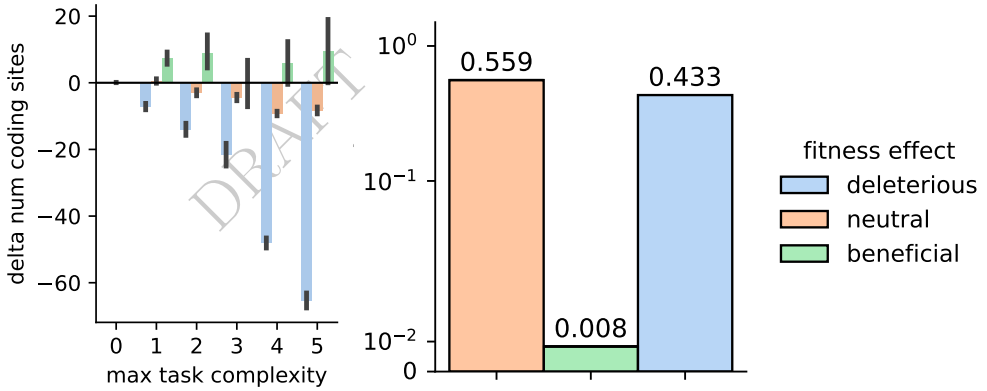


(a) genome size **(b)** num task coding sites

Fig 8. Genome size and task coding site count outcomes across slip-duplication ablation treatments. Violin plots show genome size (panel 8b) and number of task coding sites (panel 8b) in final dominant genotypes. Two-tailed Kruskal-Wallis tests indicate significant between-treatment variation in both genome size ($H = 130$, $p << 0.001$) and number of task coding sites ($H = 135$, $p << 0.001$). After applying Bonferroni correction for three comparisons, two-tailed Mann-Whitney tests confirm that genome sizes are significantly smaller under the Slip-NOP treatment compared to Slip-scramble (124 ± SD 54 vs. 252 ± SD 105 sites, $U = 814$, $p < 0.001$) and Slip-duplication treatments (124 ± SD 54 vs. 491 ± SD 223 sites, $U = 877$, $p < 0.001$); genome sizes are also significantly smaller under the Slip-scramble treatment compared to Slip-duplication ($U = 734$, $p < 0.001$). Similarly, the number of coding sites for metabolic tasks is significantly smaller under the Slip-NOP treatment compared to Slip-scramble (54 ± SD 10 vs. 88 ± SD 24 sites, $U = 811$, $p < 0.001$) and Slip-duplication treatments (54 ± SD 10 vs. 107 ± SD 23 sites, $U = 894$, $p < 0.001$); the number of task coding sites is also significantly smaller under the Slip-scramble treatment compared to Slip-duplication ($U = 654$, $p < 0.01$).



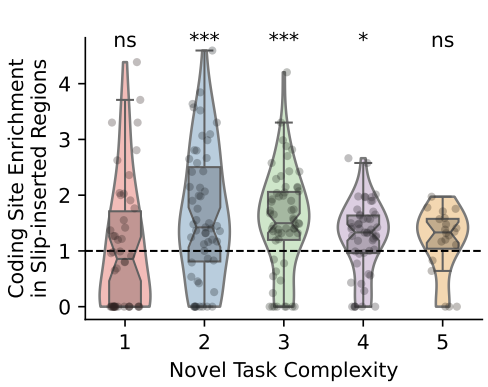
(a) change in coding site count by neutral, beneficial, and deleterious slip duplications; violin plots show count delta distributions and bar plots show mean count deltas



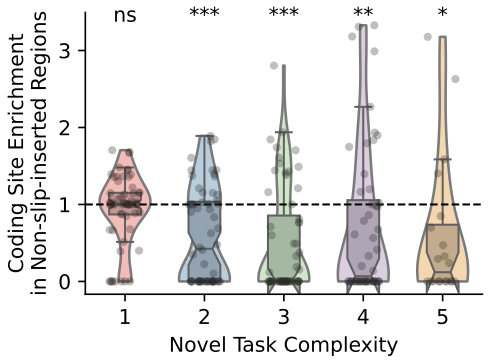
(b) mean change in coding site count by neutral, beneficial, and deleterious slip duplications, disaggregated by maximum task complexity of derived genome

(c) proportion of sampled slip duplications with neutral, beneficial, and deleterious outcomes

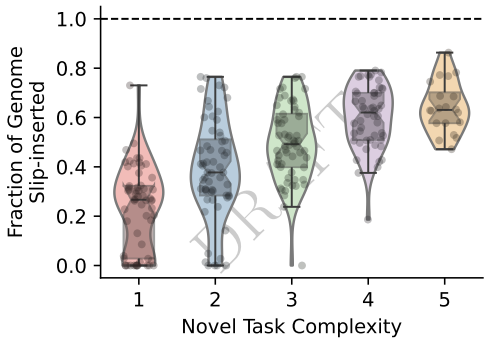
Fig 9. Distribution of slip-insertion mutation outcomes. Outcomes were measured by applying random slip-insertion to genomes sampled from along line-of-descent for final-dominant genotype over slip-duplication treatment lineage histories for slip-duplicate trials. Notably, insertion mutations that neither add or lose tasks tend to increase robustness by reducing the number of task-critical coding sites — particularly for genomes that have acquired complex tasks. Unsurprisingly, deleterious mutations tend to greatly decrease coding site count and beneficial mutations, which add new tasks, tend to increase them. Error bars give bootstrapped 95% CI.



(a) excluding sites critical self-copy loop



(b) coding site enrichment in *non-slip-duplicated* regions



(c) fraction of genome that has been slip-duplicated

Fig 10. Slip-duplication potentiation analysis detail. Panel 10a shows enrichment in slip-duplicated regions for coding sites associated with *de novo* discovery of a phenotypic trait, excluding sites assessed by knockout analysis as critical to self-copy loop viability. Plot composition follows 4, with values greater than 1 indicating that coding sites of novel traits occur more often in slip-duplicated regions compared to their background frequency. Significance of deviation from null expectation median value of 1.0 is indicated with * ($p < 0.05$), ** ($p < 0.01$), or *** ($p < 0.001$) (one-tailed Wilcoxon signed-rank test). Panels figs. 10b and 10c provide additional context, showing coding site enrichment values for *non*-duplicated sites and the overall fraction of the genome that has been slip-duplicated.