

Gene duplications potentiate adaptive evolution of complex traits

(Short title ditto)

Matthew Andres Moreno^{1,2,3,4}^{*}, Cameron Haynes^{1,2,4,5}, Charles Ofria^{6,7,8}, Alexander Lalejini^{9,10}[†], 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, H. Lee Moffitt Cancer Center and Research Institute, Tampa, FL, USA
- 6** Program in Ecology, Evolution, and Behavior
- 7** Department of Computer Science and Engineering
- 8** Michigan State University, East Lansing, MI, United States
- 9** Department of Computer Science
- 10** 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

DRAFT

Abstract

Much of extant genetic material is known to have resulted from earlier genetic duplications. On this basis, gene duplications are widely recognized as a key factor in the evolution of organismal complexity — both in producing novel adaptations (i.e., neofunctionalization) and in accumulating functional specialization (i.e., subfunctionalization). However, across the tree of life, no ready counterfactual exists to directly test the aggregate impact of gene duplication on complexity over long evolutionary timescales. To this end, we turn to the Avida platform, using digital experiments to investigate two questions in specific: (1) Does gene duplication boost evolved (a) structural and (b) functional complexity? (2) If so, does evolved complexity rely on duplicated genetic information (i.e., versus simple genome expansion alone)? As to (1): we find clear evidence that duplicative inserts yield virtual organisms with higher complexity, in both measures. However, the causal role of genetic information differed between measures of complexity. Regarding (a) structural complexity: duplications grew genome size, on average, larger than randomized and empty-insert controls. By contrast, though, duplicated genetic information more weakly affected coding-site structure, with complexity here attributable to larger genome size alone. Regarding (b) functional complexity, duplications sped up evolutionary innovation compared to controls. For *simple* phenotypic traits, faster adaptation was attributable to increased genome size alone. For *complex* phenotypic traits, however, adaptation depended directly on duplicated genetic information. Such complex traits arose disproportionately from duplicated coding sites, and seldom arose in large-genome-only treatments. In sum, our findings show that genetic information in duplicative inserts can directly drive genome expansion and complex adaptation, but that simpler adaptations and structural complexity may follow from larger genome size alone. This perspective notably integrates adaptationist and neutralist framings, as duplicated information actively catalyzes complex functionality while certain aspects of structural complexity may instead accrue passively.

1 Introduction

A key objective of evolutionary biology is to understand how evolution gives rise to new genetic information and complex phenotypic traits, and gene duplication is one prime suspect [1–11]. While comparative studies provide increasingly detailed accounts of duplication events shaping specific evolved traits [12–25], work remains to solidify systematic understanding [26] of gene duplication’s broader role in both adaptive and neutral evolutionary processes [27–29]. In this paper, we use direct experimentation to test how gene duplication impacts structural and functional complexity, and the extent to which such impact depends on the actual genetic information content within duplicative inserts.

In recent years, experimental approaches to studying evolution — which can explicitly test causality — have become an important complement to observational studies of natural history [30,31]. In the context of gene duplication, microbial experiments have confirmed the adaptive significance of duplicative dose effects [32–34]. Another notable microbial experiment has tested how gene copy count affects evolutionary innovation [35]. Comparing evolution of fluorescence phenotypes between single- and double-copy coGFP populations, the double-copy populations exhibited greater robustness under mutation but no faster adaptation under selection. As the authors highlight, though, practical constraints required they impose mutation rates and selection pressures far removed from natural evolution.

More broadly, experimental work has also made notable headway in exploring how

evolution can produce complexity. Here, key contributions have relied on digital model systems [36–39]. Though substantially abstracting mechanistic aspects of biological organisms [40], such *in silico* experiments can bring within reach principled definitions of complexity [41], fast throughput spanning thousands of generations [42], and flexible exploration of arbitrary counterfactuals [43]. In this work, we rely on digital capabilities to impose wholesale suppression of duplicative mutations, as well as intermediate interventions partially disabling them. We also exploit the opportunity to trace duplicated sites across exact lineage histories [44], and thereby directly assess their role in later innovation.

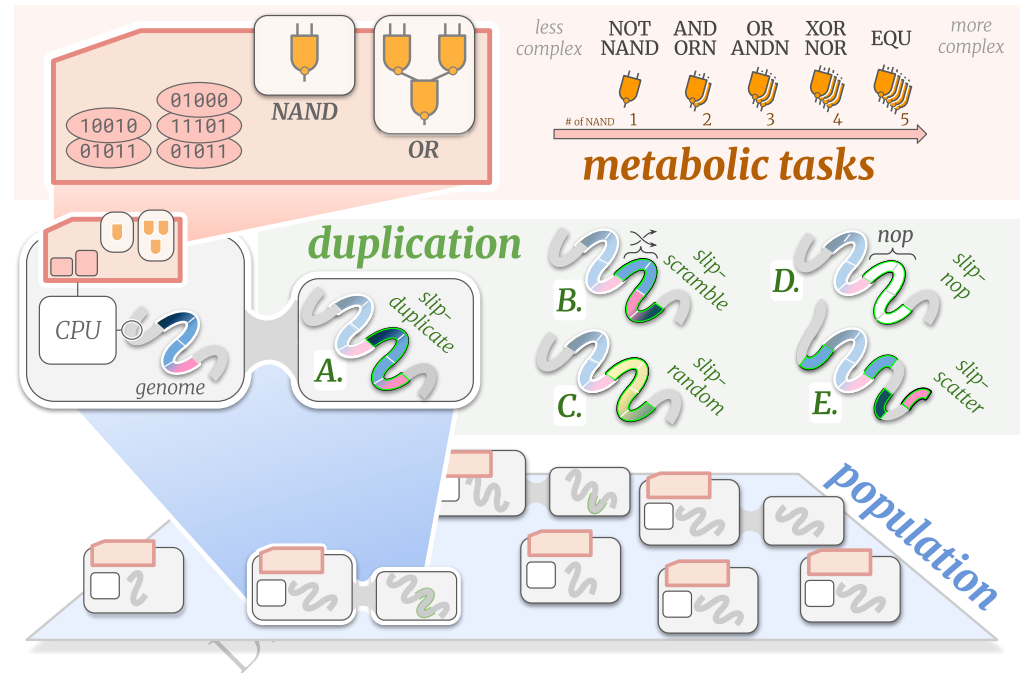


Fig 1. Replication, gene duplication, and task complexity in Avida.

Self-replicating computer programs serve as genomes of digital “model organisms” (middle left), competing to survive and reproduce in a fixed-capacity population (bottom). Available “metabolic” input/output tasks, varying in NAND-complexity, can accelerate replication (top). An organism’s metabolic “phenotype” arises from the expression of its genetic code. When copied from parent to offspring, this genetic code may be subject to slip insertions, which insert an exact copy of a target segment (A, middle left), or slip deletions, which excise a target segment. To test how information within duplicative inserts impacts evolution, we tested four *slip-duplicate* variants (middle right): B) *slip-scramble*, shuffled copy of target segment inserted; C) *slip-random*, segment of random instructions inserted; D) *slip-NOP*, segment of neutral nop-X instructions inserted; and E) *slip-scatter*, copy of target segment instructions are scattered throughout the genome.

We conducted digital experiments using Avida, an *in silico* platform which administers populations of self-replicating computer programs — providing, essentially, a digital “model organism” [45]. Because copy errors introduce heritable variation in how efficiently programs replicate, evolution by natural selection results [46].

In early work with Avida, Lenski *et al.* established a useful framework for studying phenotypic traits across a well-defined spectrum of complexity; input/output calculations can be treated as phenotypic traits, with the complexity of a given task determined according to its minimum necessary substeps. Applying this lens, Lenski *et al.* demonstrated that simpler traits could provide building blocks for complex ones — and, in fact, often amounted to necessary prerequisites for their evolution [47].

To complement the above phenotypic measure of functional complexity, we also

established a genetic measure of structural complexity. Here, we counted genome instructions that benefit an organism’s fitness [41]. These adaptive coding sites were identified through single-site knockout screens, disabling genome instructions one-by-one by substitution with a null placeholder.¹

With respect to both of these functional and structural complexity metrics, we find that gene duplication can systematically increase evolved complexity. Control treatments reveal, in particular, that genetic information in duplicated regions entrenches otherwise unlikely genome expansion. This expansion, in turn, can promote simple adaptive traits and accommodate more coding sites — in line with neutralist perspectives on biological complexity [50–52]. However, we also find that genetic information within duplicated regions catalyzes complex adaptive traits, beyond effects from increased genome size alone. Causality appears direct, as coding sites for these complex traits trace back to slip-duplicated regions. Thus, we also confirm that gene duplication can facilitate otherwise unlikely innovation in adaptation-driven scenarios [53].

Figure 1 details our experimental design. Digital organisms were well mixed, with population carrying capacity of 3,600. To supply adaptive potential, rewards were provided for each of Lenski *et al.*’s collection of nine phenotypic traits (different logical transforms on binary inputs) [47]. Under this scheme, organisms producing correct output received a metabolic boost to their genetic operations. Formally, complexity of each task can be described as a minimum quantity of necessary NAND operations [47].² 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 site-by-site. In addition to standard point mutation, experiments incorporated gene duplication using “slip mutation” events analogous to replication slippage [55]. 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 partial analogs for gene duplication, designed 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, just below the maximum 1,016-site genome size observed along extant lineages under slip duplication and near the upper extreme of fixed genome length, at 870 sites (Figure 5b, Supplementary Figure 8).

2 Results

2.1 Gene duplication facilitates adaptive evolution of complex traits

In a first set of experiments, we investigated whether gene duplication can promote *de novo* evolution of adaptive phenotypic traits. In evolutionary trials, *slip-duplicate* replicates allowing duplications did evolve more adaptive traits than under baseline conditions (median 9 vs. 7 traits; two-tailed Mann-Whitney U test, $W = 562.5$, Bonferroni-adjusted $p < 0.0001$; Figure 2a). In many cases, gene duplication drove substantial genome expansion — inflations up to 10-fold in size, mean 546 sites (Figure 5c). To test whether genome size alone could account for gene duplication’s adaptive benefit, we performed additional experiments incorporating a “long-genome” control

¹Although 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 [48, 49].

²Computer architecture theory considers NAND as a fundamental basis operation, since all other logic can be derived from compositions of NAND operations [54].

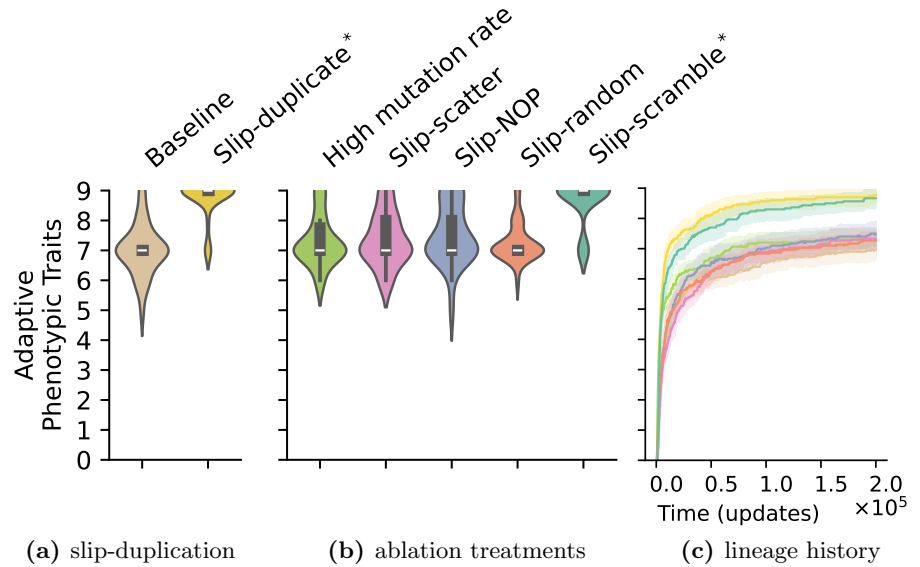


Fig 2. Genetic content of duplications facilitates adaptive evolution. Violin plots show number of adaptive traits evolved in final dominant genotypes. Asterisk (*) markers indicate treatments with greater adaptation compared to baseline (Mann-Whitney U test, $p < 0.0001$). Time series (2c) tracks acquisition of adaptive traits along lineages of final dominant genotypes; color-coding corresponds to violin plots.

treatment with 1,000-site genomes.

The long-genome control also outperformed adaptation under 100-site baseline conditions (Supplementary Figure 7). However, disaggregating by task complexity reveals that genome expansion alone primarily favors simple traits. For traits requiring 3 or fewer NANDs, adaptation in the long-genome control matched or exceeded gene duplication when compared generation-by-generation (Figure 3). By contrast, the larger genome control failed to match duplication in promoting complex 4- and 5-NAND traits (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). Thus, duplication specifically advantaged complex adaptation.

2.2 Gene content of duplicated inserts drives adaptation

Having observed that gene duplication can accelerate adaptive evolution beyond effects of increased genome size alone, we next sought to test whether this adaptive benefit stemmed, in specific, from genetic information propagated in duplications. For this purpose, we tested four variants of the *slip-duplication* operator, disrupting duplicated material in different ways (Figure 1), as well as an additional “high mutation rate” treatment where single-site insertion/deletion mutations were applied in lieu of duplication.

We found adaptation to strongly depend on the content of duplicated inserts. In all cases, partial analogs of gene duplication provided no adaptive benefit — except for *slip-scramble*, which randomized sequence order within duplicated regions (two-tailed Mann-Whitney U test, $p < 0.0001$; Figure 2b). However, *slip-scramble* still underperformed full-fledged *slip-duplicate*, tested via 100 additional trials to avoid multiple comparison (two-tailed Mann-Whitney U test, $W = 4305$; Bonferroni-adjusted $p < 0.02$). Sequence order of inserts notably affected genome expansion, with *slip-duplicate* yielding genomes nearly twice as large as *slip-scramble* ($491 \pm \text{SD } 223$ vs.

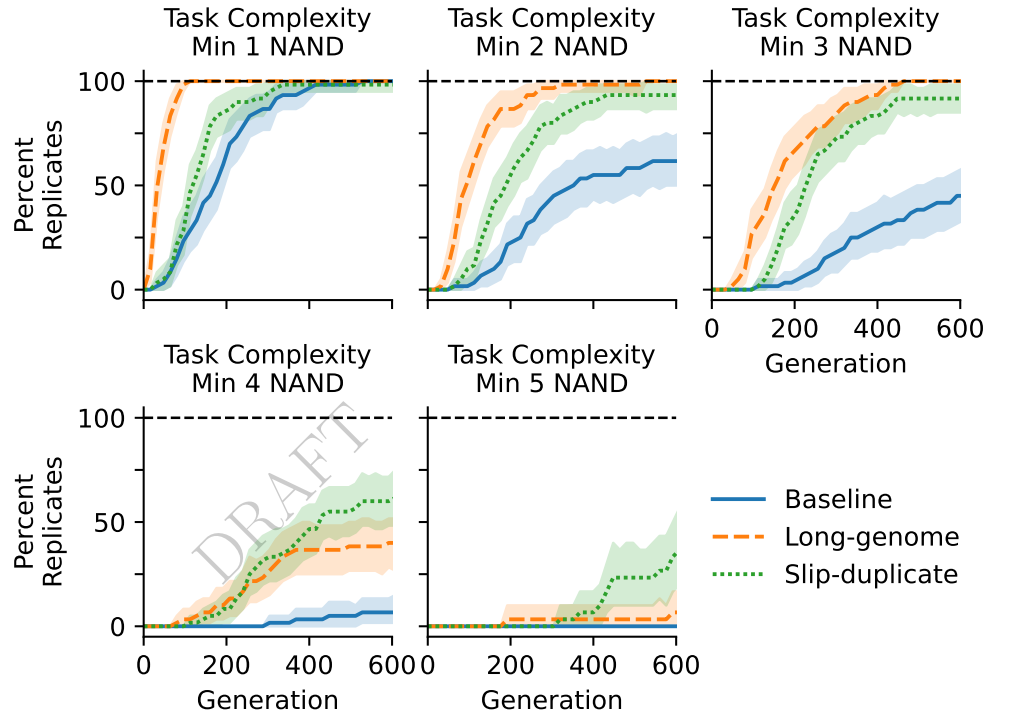


Fig 3. Gene duplication boosts adaptive evolution of complex phenotypic traits. Plots show fraction of replicates exhibiting available phenotypic traits, by generation from founding ancestor. Panels facet by trait complexity, measured by the minimum number of NAND operations required to complete the task. *Slip-duplicate*, which allows gene duplication mutations, produces faster adaptive evolution than the long-genome control only for complex tasks that require 4 or 5 NAND operations. Error bands give 95% CI, bootstrapped over 30 replicates per treatment.

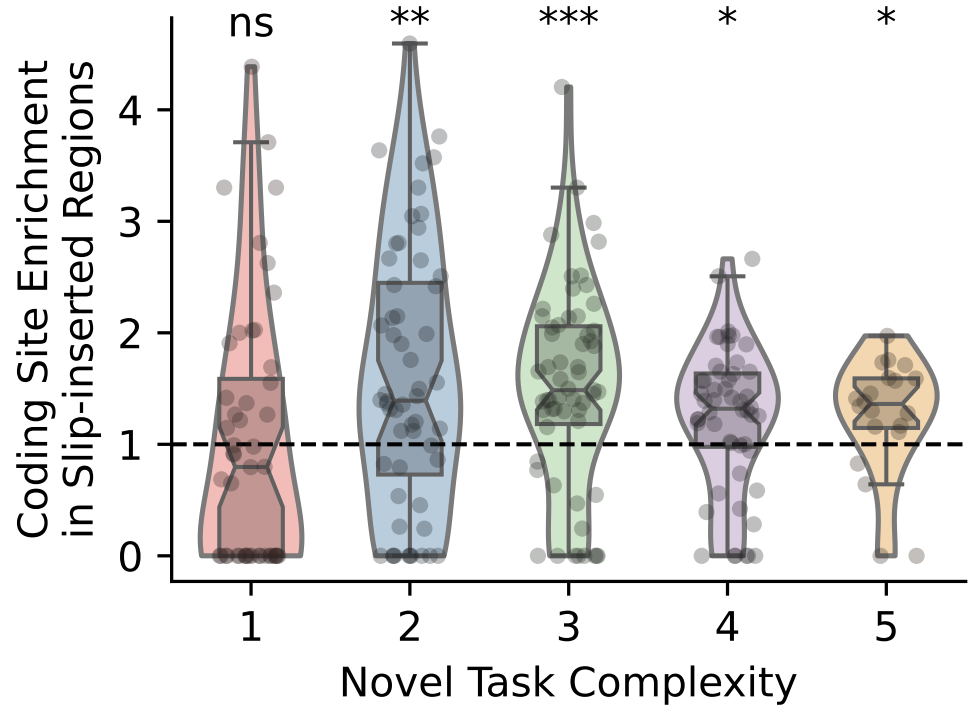


Fig 4. Duplications provide *de novo* coding sites for complex traits. Enrichment measures density of coding sites for phenotypic traits in duplicated regions, normalized to neutral expectation. Values greater than 1 indicate that coding sites are overrepresented in duplicated regions, compared to their background frequency. Each observation gives enrichment of coding site density within duplicated regions at the first completion of an available task; observation counts are 48, 56, 59, 52, and 20. Significance of deviation from null expectation of median value 1.0 is indicated with * ($p < 0.05$), ** ($p < 0.01$), or *** ($p < 0.001$) (one-tailed Wilcoxon signed-rank test).

252 $\pm$ SD 105 sites; two-tailed Mann-Whitney U test, $U = 734$, $p < 0.001$; Supplementary Figure 8).

2.3 Duplicated sites are potentiated for complex tasks

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

To address the question, we measured whether coding sites for novel adaptive traits were more likely to arise in genome regions that had previously been duplicated. High density for later coding sites would suggest that these regions are “potentiated” — that is, possessing latent gene content prone to facilitate new traits [56–58]. Figure 4 compares the involvement of sites having undergone duplication — versus those that had not — in coding for novel traits. For the simplest tasks, requiring only one NAND component, we found no potentiation by gene duplication. However, we found significant potentiation in duplicated regions for traits with two or more NAND components (one-sample Wilcoxon signed-rank tests; $n = 56, 59, 52, 20$ observations; Figure 4). Effect sizes on probability to code for novel traits were 1.6 $\times$, 1.6 $\times$, 1.2 $\times$, and 1.2 $\times$, respectively. Smaller effect sizes for 4- and 5-NAND tasks may in part reflect scarcity of unduplicated material at later generations (Supplementary Figure 10).

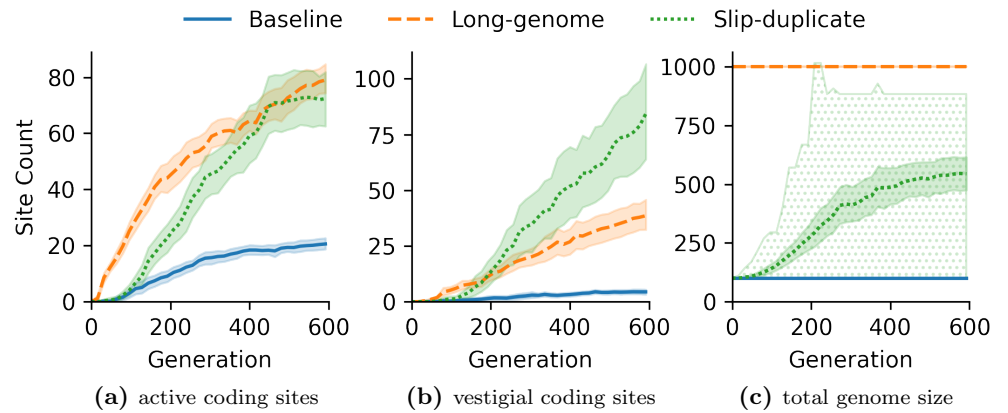


Fig 5. Gene duplication boosts accumulation of vestigial coding sites. Generation-by-generation counts of coding sites over evolutionary history, measured by single-site knockouts. Gene duplication yields active coding site counts comparable to long-genome control (panel 5a). Vestigial coding sites were labeled as active in an ancestor, but are no longer active coding sites. Vestigial coding site counts under slip-duplication treatment outpace control treatments (panel 5b). Panel 5c gives total genome size. Solid bands give 95% CI, bootstrapped over 30 replicates per treatment. Hatched band gives full range of genome sizes observed.

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

In addition to potentiating effects, gene duplications are also hypothesized to facilitate adaptation through direct effects — for example, by altering gene product dosage [59]. Indeed, we observed such direct effects in our system: a substantial fraction of novel gain-of-function steps on lineages coincided with duplications — 41 of 174, or 23.6%. However, in cases where duplication produced immediate gain-of-function, we still found evidence that sites coding for a new trait were more likely than chance to have been involved in earlier duplications (Supplementary Figure 10).

2.4 Duplication drives accumulation of vestigial coding sites

In a final set of analyses, we broadened our scope to test how gene duplication affects overall genome architecture, focusing on robustness. We quantified robustness by counting "critical" genome sites, where single-site knockouts disrupted adaptive traits or were lethal.

One conventional perspective on gene duplication is that copied genetic material can improve robustness [60]. To assess this hypothesis, we assayed the impact of synthetic gene duplication mutations on evolved genomes from *slip-duplicate* treatments. For each genome, we applied one randomly-sampled duplication and measured change in critical site count.

A slight majority of sampled duplications (56%) were fitness-neutral, producing no phenotypic changes affecting fitness. On average, we found that these fitness-neutral duplications decreased coding site count by 3% (bootstrapped 95% CI [6.4, 7.3] sites; Supplementary Figure 9). This effect remained robust in genomes with high phenotypic complexity; for instance, neutral duplications decreased coding site count by 3% and

5%, respectively, in genomes with 4- and 5-NAND traits (bootstrapped 95% CIs [8.5, 9.9] and [7.2, 9.3] sites; Supplementary Figure 9). Thus, short-term effects of gene duplication did, in fact, improve robustness.

Despite short-term robustness gains, however, gene duplication accrued many more critical coding sites over the long term — $3.5\times$ as many as baseline conditions by generation 600 (Mann-Whitney U test, $U = 854$, $p < 0.0001$; Figure 5a). At least in part, this effect likely arises from genome expansion. By generation 600, *slip-duplicate* accrued about as many active coding sites as the long-genome control (Mann-Whitney U test, $U = 361$, $p = 0.19$; Figure 5a). Despite comparable active coding sites, however, gene duplication accumulated $2.1\times$ more vestigial coding sites (i.e., previously, but no longer, coding) compared to the long-genome control (Mann-Whitney U test, $U = 630$, $p < 0.01$; Figure 5b). Thus, gene duplication appears to drive accumulation of unconstrained coding material, potentially recruitable for later innovation.

3 Discussion

In this work, we used an *in silico* model to experimentally test whether and how gene duplication can facilitate the evolution of novel adaptation and organismal complexity. To conduct these experiments, we adapted an established framework developed by Lenski *et al.* [47] to study adaptation across a well-defined spectrum of functional complexity. Though necessarily limited in fully reflecting natural organisms, such digital models can nonetheless provide useful footholds to biological science [36–39, 48, 61–70].

Our results demonstrate that gene duplications can promote the *de novo* evolution of adaptive traits. Leveraging the unique tractability of our study system, we were able to explore causation in detail, revealing: (1) gene duplication especially potentiated complex traits; (2) sequence content of duplications drove adaptation; and (3) though ultimately increasing total mutational sensitivity (despite initially buffering against it), duplicative genome expansion greatly accelerated the accumulation of neutral vestigial material.

3.1 Gene duplication facilitates adaptive evolution of complex traits

Adaptation by gene duplication in our study system corroborates findings across a wide variety of biological taxa and digital models [14, 28, 71]. Indeed, a striking example of gene duplication potentiating novel adaptation comes from the Long-Term Evolution Experiment in *E. coli* [72, 73], where a duplication in one population enabled expression of a key citrate transporter in aerobic conditions — resulting in a seven-fold increase in carrying capacity [74]. Among many other examples from natural history [19, 75–80], such observations underlie longstanding theory connecting gene duplication and adaptive traits, dating as far back as [53].

On the other hand, however, recent directed evolution experiments on *E. coli* notably do not find gene copy count to affect phenotypic adaptation rate [35]. Several factors may explain this contrast, including directed evolution’s shorter generational timescales, much higher mutation rates, and extreme selection strength. Another factor may be opportunity for open-ended, multi-gene composition in our study system.

A central objective of our experimental design is not only to test whether gene duplication can facilitate adaptation, but also to identify when and how it does so. By examining adaptive evolution across a spectrum of trait complexity, we find that gene duplication especially drives the evolution of complex traits (Figure 3). Comparison against a long-genome control confirms that this effect cannot be attributed to genome expansion alone.

A key aspect of Lenski *et al.*'s model of phenotypic complexity is functional overlap between defined traits, meant to reflect the tendency of biological complexity to arise by modification of existing structures and functions [47]. In this context, gene duplication's observed role in facilitating complex traits is broadly consistent with a “building block” model — where duplication facilitates novel assemblages of existing components [81]. Exploring how these dynamics are shaped by mechanistic complications, such as regulatory network structure [82, 83], is an exciting direction for future work.

3.2 Adaptation harnesses sequence information of duplications

In order to isolate how gene duplication facilitates adaptation, we tested a suite of partial analogs for full-fledged gene duplication (Figure 1). Among these, insertions that increase genome size without propagating genetic information provided no adaptive benefit. By contrast, both sequence duplications and shuffled duplications expedited adaptation, with the former (preserving sequence order) yielding the strongest effect (Figure 2). These results show that, rather than just side effects on genome expansion and mutation supply, adaptation can specifically harness information content of duplications — as in classical descriptions of subfunctionalization and neofunctionalization, where gene copies gain new functionality by specializing or diverging [50, 53].

In contrast to our long-genome control, no adaptive benefit arose from allowing neutral inserts (Figure 2). Despite opportunity to create increased genome length that advantages adaptation, relatively few neutral inserts accumulated. Gene duplication induced much greater genome expansion than neutral inserts — by a factor of nearly 4×, despite equivalent site-creation potential (Supplementary Figure 8). One possible explanation why few neutral inserts fixed is selection against mutational load arising from additional genome sites. (Indeed, in preliminary experiments we found that mutational load frequently drove genome shrinkage in founding ancestors.) Gene duplication's strong effect entrenching genome expansion, compared to partial analogs that do not propagate genetic information, supports theory that “lock-in” by dose effects or contingent interdependencies (e.g., “complexity ratchet” effects [84–87]) is essential to retention of new genome content [27, 88–93].

3.3 Duplicated sites are potentiated for later adaptation

Large-scale analyses of biological genomes report substantial content as having likely arisen by duplication [13, 14]. In our experiments, analysis of genome history allows us to draw direct connections between gene duplication and adaptation. Specifically, we found that (1) an outsized fraction of new adaptations arose in duplicated regions (Figure 4) and (2) many gain-of-function events (41 of 174 along analyzed lineages) directly coincided with slip duplications. Thus, adaptive outcomes appear to reflect a combination of both potentiation and immediate effects. Lastly, echoing how gene duplication especially benefited complex adaptation, lineage analysis finds that duplicated sites potentiate only complex adaptations — lending further credence to gene duplication as facilitating a “building block” mode of adaptation [81].

3.4 Gene duplications accumulate vestigial coding material

Consistent with other findings that duplication-associated redundancy can boost robustness [50], gene duplications tended to immediately reduce single points of failure detectable in knockouts (Supplementary Figure 9). Over time, however, gene duplication actually led to more mutationally-sensitive coding sites — similar in quantity to the long-genome control. One possible explanation is that, despite duplications boosting

redundancy in the short term, long-term balance between drift and mutational load drives convergence towards similar mutational sensitivity for large genomes. Notably, gene duplication diverges from control treatments in greatly increasing net supply of vestigial coding sites no longer constrained by selection (Figure 5). Thus, we observe gene duplication to reshape genome architecture by expanding size, but to also produce an outsized pool of defunct coding content — ripe to be co-opted. Here, as throughout this study, site-by-site genome histories and robust experimental controls highlight an apparent evolutionary role of genetic information carried within duplications.

4 Materials and Methods

4.1 Evolution trials

We conducted experiments using a custom version of Avida v2.14.0, extended to implement gene duplication [45]. 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 [94,95]. Simulation data is archived via the Open Science Framework at <https://osf.io/j5s4h/> [96,97], provided with a CC-BY Attribution 4.0 International license. Code used in this project incorporated numerous pieces of open-source scientific software [98–105]. Experiments reported here extend an earlier conference publication [106].

Populations were well mixed and configured to the Avida default size of 3,600 organisms across all trials. We configured available metabolic resources consisting of tasks NOT, NAND, OR-NOT, AND, OR, AND-NOT, NOR, XOR, and EQUALS. (This task set comprises all unique 1- and 2-input boolean logic functions.) Identical reward was provided for performing each task and kept consistent throughout trials. Rewards accrued independently for each task, allowing 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 tasks and receives no reward) to a maximum of 9 (*i.e.*, the organism performs all 9 available tasks and receives the maximum reward). As a metric of complexity, logic tasks were categorized by minimum required NAND operations (Figure 1; [47]).

Except where otherwise noted, experiments began with a 100-instruction ancestral self-replicator. In all cases, mutations reducing genome size below 100 instructions were disallowed to ensure sufficient genetic material for phenotypic traits. Run length was 200,000 updates across all experiments, where a single update processes on average 30 genome instructions per organism. This duration was sufficient to observe at least 600 generations of evolution in all trials. We configured Avida’s scheduler to ensure long genomes incurred no inherent disadvantage [45].

4.2 Mutation operators

Avida genomes are linear sequences of instructions, each encoding one of 27 available operations for basic computations, execution flow control, input and output, and self-replication. The instruction set is Turing complete and valid in any possible ordering. 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 [47].

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 could occur with a probability of 0.05 per offspring [47].

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 [55]. 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).

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.

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.

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 replicates. For this reason, first-phase experiments prioritized higher replicate counts, while second-phase experiments prioritized more detailed lineage tracking of individual genomic sites. We conducted 100 evolutionary replicates per treatment for first-phase experiments, and 30 replicates for second-phase experiments.

First-phase experiments consisted of seven treatments: one baseline treatment with slip-duplication disabled, five experimental treatments corresponding to the five surveyed slip-duplication mutation operators shown in Figure 1, and a high mutation rate control. 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 4 and 10), and analyses of slip-duplication outcome distributions (Figure 9). Adaptation outcomes 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 the long-genome controls elapsing on the order of $10\times$ fewer generations than baseline conditions, since self-replication is slower for longer genomes. 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 lengths. Thus, comparing by update remains conservative as faster adaptation under *slip-duplication* occurred without more generations elapsing.

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 assess phenotypic traits at each step in the lineage leading to the final dominant genotype.

For investigations involving the evolutionary history of individual genomic 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 [106]. 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 research was supported in part by the NSF Grant DEB-1813069 and NSF CAREER DEB-2540912. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

This material is based upon work supported by the Eric and Wendy Schmidt AI in Science Postdoctoral Fellowship, a Schmidt Sciences program.

This material is based upon work supported by the U.S. Department of Energy, Office of Science, Office of Advanced Scientific Computing Research (ASCR), under Award Number DE-SC0025634. This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark,

manufacturer, or otherwise does not necessarily constitute or imply its endorsement, 450
 recommendation, or favoring by the United States Government or any agency thereof. 451
 The views and opinions of authors expressed herein do not necessarily state or reflect 452
 those of the United States Government or any agency thereof. 453

References 454

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. 455
456
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. 457
458
3. Wagner A. Gene duplications, robustness and evolutionary innovations. *BioEssays*. 2008;30(4):367–373. doi:10.1002/bies.20728. 459
460
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. 461
462
463
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. 464
465
466
6. Metz CW. Duplication of Chromosome Parts as a Factor in Evolution. *The American Naturalist*. 1947;81(797):81–103. 467
468
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. 469
470
471
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. 472
473
474
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. 475
476
477
10. Lynch M, Conery JS. The Origins of Genome Complexity. *Science*. 2003;302(5649):1401–1404. doi:10.1126/science.1089370. 478
479
11. Birchler JA, Yang H. The multiple fates of gene duplications: Deletion, hypofunctionalization, subfunctionalization, neofunctionalization, dosage balance constraints, and neutral variation. *The Plant Cell*. 2022;34(7):2466–2474. doi:10.1093/plcell/koac076. 480
481
482
483
12. 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. 484
485
486
13. 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. 487
488
489
490

14. Teichmann SA, Babu MM. Gene regulatory network growth by duplication. *Nat Genet.* 2004;36(5):492–496. doi:10.1038/ng1340. 491 492
15. 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. 493 494 495
16. 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. 496 497 498 499
17. 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. 500 501 502 503
18. 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. 504 505 506
19. 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. 507 508 509 510
20. Tasnim M, Wahlquist P, Hill JT. Zebrafish: unraveling genetic complexity through duplicated genes. *Development Genes and Evolution.* 2024;234(2):99–116. doi:10.1007/s00427-024-00720-6. 511 512 513
21. Assis R, Bachtrog D. Neofunctionalization of young duplicate genes in *Drosophila*. *Proceedings of the National Academy of Sciences.* 2013;110(43):17409–17414. doi:10.1073/pnas.1313759110. 514 515 516
22. Deng C, Cheng CHC, Ye H, He X, Chen L. Evolution of an antifreeze protein by neofunctionalization under escape from adaptive conflict. *Proceedings of the National Academy of Sciences.* 2010;107(50):21593–21598. doi:10.1073/pnas.1007883107. 517 518 519 520
23. Monson R. Gene Duplication, Neofunctionalization, and the Evolution of C4Photosynthesis. *International Journal of Plant Sciences.* 2003;164(S3):S43–s54. doi:10.1086/368400. 521 522 523
24. Xue Q, Dweck HKM. Receptor sequence divergence, gain, loss, duplication, and neofunctionalization drive olfactory adaptation in *Drosophila suzukii*. *Proceedings of the National Academy of Sciences.* 2026;123(13). doi:10.1073/pnas.2529586123. 524 525 526 527
25. Lynch VJ. Inventing an arsenal: adaptive evolution and neofunctionalization of snake venom phospholipase A2 genes. *BMC Evolutionary Biology.* 2007;7(1):2. doi:10.1186/1471-2148-7-2. 528 529 530
26. Welch JJ. What’s wrong with evolutionary biology? *Biology & Philosophy.* 2016;32(2):263–279. doi:10.1007/s10539-016-9557-8. 531 532
27. 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. 533 534 535

28. Zhang J. Evolution by gene duplication: an update. *Trends in Ecology & Evolution*. 2003;18(6):292–298. 536
537
29. 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. 538
539
30. Kawecki TJ, Lenski RE, Ebert D, Hollis B, Olivieri I, Whitlock MC. Experimental evolution. *Trends in Ecology & Evolution*. 2012;27(10):547–560. 540
541
doi:10.1016/j.tree.2012.06.001. 542
31. 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. 543
544
545
32. 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. 546
547
548
33. 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. 549
550
551
552
34. Näsval J, Sun L, Roth JR, Andersson DI. Real-Time Evolution of New Genes by Innovation, Amplification, and Divergence. *Science*. 2012;338(6105):384–387. 553
554
doi:10.1126/science.1226521. 555
35. 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. 556
557
doi:10.7554/elife.97216.3. 558
36. Fortuna MA, Beslon G, Ofria C. Editorial: Digital evolution: Insights for biologists. *Frontiers in Ecology and Evolution*. 2022;10. 559
560
doi:10.3389/fevo.2022.1037040. 561
37. Adami C, Ofria C, Collier TC. Evolution of biological complexity. *Proceedings of the National Academy of Sciences*. 2000;97(9):4463–4468. 562
563
doi:10.1073/pnas.97.9.4463. 564
38. 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. 565
566
567
39. Fortuna MA, Zaman L, Ofria C, Wagner A. The genotype-phenotype map of an evolving digital organism. *PLOS Computational Biology*. 2017;13(2):e1005414. 568
569
doi:10.1371/journal.pcbi.1005414. 570
40. Peck SL. Simulation as experiment: a philosophical reassessment for biological modeling. *Trends in Ecology & Evolution*. 2004;19(10):530–534. 571
572
doi:10.1016/j.tree.2004.07.019. 573
41. Adami C. What is complexity? *BioEssays*. 2002;24(12):1085–1094. 574
575
doi:10.1002/bies.10192.
42. Moreno MA, Yang C, Dolson E, Zaman L. Trackable Agent-based Evolution Models at Wafer Scale. In: *The 2024 Conference on Artificial Life*. MIT Press; 2024. p. 87–98. Available from: https://doi.org/10.1162/isal_a_00830. 576
577
578

43. Langton CG. Artificial Life. 1st ed. Routledge; 1989. 579
44. Dolson E, Rodriguez-Papa S, Moreno MA. Phylotrack: C++ and Python libraries for in silico phylogenetic tracking; 2024. Available from: 580
<https://arxiv.org/abs/2405.09389>. 582
45. Ofria C, Bryson DM, Wilke CO. Avida: A Software Platform for Research in 583
Computational Evolutionary Biology. In: Artificial Life Models in Software. 584
Springer; 2009. p. 3–35. 585
46. Pennock RT. Models, simulations, instantiations, and evidence: the case of 586
digital evolution. Journal of Experimental & Theoretical Artificial Intelligence. 587
2007;19(1):29–42. doi:10.1080/09528130601116113. 588
47. Lenski RE, Ofria C, Pennock RT, Adami C. The evolutionary origin of complex 589
features. Nature. 2003;423(6936):139–144. doi:10.1038/nature01568. 590
48. Lenski RE, Ofria C, Collier TC, Adami C. Genome complexity, robustness and 591
genetic interactions in digital organisms. Nature. 1999;400(6745):661–664. 592
doi:10.1038/23245. 593
49. Moreno MA. Methods to Estimate Cryptic Sequence Complexity. In: The 2024 594
Conference on Artificial Life. MIT Press; 2024. Available from: 595
<https://arxiv.org/abs/2404.10854>. 596
50. Lynch M, Conery JS. The Evolutionary Fate and Consequences of Duplicate 597
Genes. Science. 2000;290(5494):1151–1155. doi:10.1126/science.290.5494.1151. 598
51. Beslon G, Liard V, Parsons DP, Rouzauud-Cornabas J. In: Of Evolution, 599
Systems and Complexity. Springer International Publishing; 2021. p. 1–18. 600
Available from: http://dx.doi.org/10.1007/978-3-030-71737-7_1. 601
52. Lynch M. The frailty of adaptive hypotheses for the origins of organismal 602
complexity. Proceedings of the National Academy of Sciences. 603
2007;104(suppl_1):8597–8604. doi:10.1073/pnas.0702207104. 604
53. Ohno S. Evolution by Gene Duplication. Springer Berlin Heidelberg; 1970. 605
Available from: <http://dx.doi.org/10.1007/978-3-642-86659-3>. 606
54. Mano MM, Kime CR. Logic and computer design fundamentals. Prentice-Hall, 607
Inc.; 1997. 608
55. Bzymek M, Lovett ST. Instability of repetitive DNA sequences: The role of 609
replication in multiple mechanisms. Proceedings of the National Academy of 610
Sciences. 2001;98(15):8319–8325. doi:10.1073/pnas.111008398. 611
56. Ferguson AJ, Ofria C. Potentiating Mutations Facilitate the Evolution of 612
Associative Learning in Digital Organisms. In: The 2023 Conference on 613
Artificial Life. Alife 2023. MIT Press; 2023. Available from: 614
http://dx.doi.org/10.1162/isal_a_00684. 615
57. Quandt EM, Gollihar J, Blount ZD, Ellington AD, Georgiou G, Barrick JE. 616
Fine-tuning citrate synthase flux potentiates and refines metabolic innovation in 617
the Lenski evolution experiment. eLife. 2015;4. doi:10.7554/elife.09696. 618
58. Blount ZD, Borland CZ, Lenski RE. Historical contingency and the evolution of 619
a key innovation in an experimental population of *Escherichia coli*. Proceedings 620
of the National Academy of Sciences. 2008;105(23):7899–7906. 621
doi:10.1073/pnas.0803151105. 622

59. 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. 623
624
625
60. 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. 626
627
628
61. Gupta A, LaBar T, Miyagi M, Adami C. Evolution of Genome Size in Asexual Digital Organisms. *Scientific Reports*. 2016;6(1). doi:10.1038/srep25786. 629
630
62. Goldsby HJ, Dornhaus A, Kerr B, Ofria C. Task-switching costs promote the evolution of division of labor and shifts in individuality. *Proceedings of the National Academy of Sciences*. 2012;109(34):13686–13691. 631
632
633
doi:10.1073/pnas.1202233109. 634
63. Goldsby HJ, Knoester DB, Ofria C, Kerr B. The Evolutionary Origin of Somatic Cells under the Dirty Work Hypothesis. *PLoS Biology*. 2014;12(5):e1001858. doi:10.1371/journal.pbio.1001858. 635
636
637
64. Misevic D, Ofria C, Lenski RE. Sexual reproduction reshapes the genetic architecture of digital organisms. *Proceedings of the Royal Society B: Biological Sciences*. 2005;273(1585):457–464. doi:10.1098/rspb.2005.3338. 638
639
640
65. Clune J, Misevic D, Ofria C, Lenski RE, Elena SF, Sanjuán R. Natural Selection Fails to Optimize Mutation Rates for Long-Term Adaptation on Rugged Fitness Landscapes. *PLoS Computational Biology*. 2008;4(9):e1000187. 641
642
643
doi:10.1371/journal.pcbi.1000187. 644
66. Clune J, Pennock RT, Ofria C, Lenski RE. Ontogeny Tends to Recapitulate Phylogeny in Digital Organisms. *The American Naturalist*. 2012;180(3):E54–e63. 645
646
doi:10.1086/666984. 647
67. Covert AW, Lenski RE, Wilke CO, Ofria C. Experiments on the role of deleterious mutations as stepping stones in adaptive evolution. *Proceedings of the National Academy of Sciences*. 2013;110(34). doi:10.1073/pnas.1313424110. 648
649
650
68. Elena SF, Wilke CO, Ofria C, Lenski RE. Effects Of Population Size And Mutation Rate On The Evolution Of Mutational Robustness. *Evolution*. 2007;61(3):666–674. doi:10.1111/j.1558-5646.2007.00064.x. 651
652
653
69. Wilke CO, Wang JL, Ofria C, Lenski RE, Adami C. Evolution of digital organisms at high mutation rates leads to survival of the flattest. *Nature*. 2001;412(6844):331–333. doi:10.1038/35085569. 654
655
656
70. Hazen RM, Griffin PL, Carothers JM, Szostak JW. Functional information and the emergence of biocomplexity. *Proceedings of the National Academy of Sciences*. 2007;104(suppl_1):8574–8581. doi:10.1073/pnas.0701744104. 657
658
659
71. 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;. 660
661
662
72. 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. 663
664
665
doi:10.1086/285289. 666

73. Lenski RE. Phenotypic and Genomic Evolution during a 20, 000-Generation Experiment with the Bacterium *Escherichia coli*; 2003. Available from: <http://dx.doi.org/10.1002/9780470650288.ch8>. 667
668
669
74. 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. 670
671
672
75. 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. 673
674
675
76. 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. 676
677
678
679
77. 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. 680
681
682
78. 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. 683
684
685
79. 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. 686
687
688
689
80. 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. 690
691
692
81. Jacob F. Evolution and Tinkering. *Science*. 1977;196(4295):1161–1166. doi:10.1126/science.860134. 693
694
82. Cussat-Blanc S, Harrington K, Banzhaf W. Artificial Gene Regulatory Networks—A Review. *Artificial Life*. 2019;24(4):296–328. doi:10.1162/artl_a_00267. 695
696
697
83. Lalejini A, Moreno MA, Ofria C. Tag-based regulation of modules in genetic programming improves context-dependent problem solving. *Genetic Programming and Evolvable Machines*. 2021;22(3):325–355. doi:10.1007/s10710-021-09406-8. 698
699
700
701
84. Liard V, Parsons DP, Rouzaud-Cornabas J, Beslon G. The Complexity Ratchet: Stronger than Selection, Stronger than Evolvability, Weaker than Robustness. *Artificial Life*. 2020;26(1):38–57. doi:10.1162/artl_a_00312. 702
703
704
85. Soyer OS, Bonhoeffer S. Evolution of complexity in signaling pathways. *Proceedings of the National Academy of Sciences*. 2006;103(44):16337–16342. doi:10.1073/pnas.0604449103. 705
706
707
86. Lukeš J, Archibald JM, Keeling PJ, Doolittle WF, Gray MW. How a neutral evolutionary ratchet can build cellular complexity. *IUBMB Life*. 2011;63(7):528–537. doi:10.1002/iub.489. 708
709
710

87. Gould SJ. Full House: The Spread of Excellence from Plato to Darwin. 1st ed. 711
New York, USA: Harmony Books; 1996. 712
88. Wagner A. The fate of duplicated genes: loss or new function? BioEssays. 713
1998;20(10):785–788. 714
doi:10.1002/(sici)1521-1878(199810)20:10;785::aid-bies2;3.0.co;2-m. 715
89. Tautz D. Problems and paradigms: Redundancies, development and the flow of 716
information. BioEssays. 1992;14(4):263–266. doi:10.1002/bies.950140410. 717
90. Force A, Lynch M, Pickett FB, Amores A, Yan YL, Postlethwait J. Preservation 718
of Duplicate Genes by Complementary, Degenerative Mutations. Genetics. 719
1999;151(4):1531–1545. doi:10.1093/genetics/151.4.1531. 720
91. Bergthorsson U, Andersson DI, Roth JR. Ohno’s dilemma: Evolution of new 721
genes under continuous selection. Proceedings of the National Academy of 722
Sciences. 2007;104(43):17004–17009. doi:10.1073/pnas.0707158104. 723
92. Conant GC, Wolfe KH. Turning a hobby into a job: How duplicated genes find 724
new functions. Nature Reviews Genetics. 2008;9(12):938–950. 725
doi:10.1038/nrg2482. 726
93. Francino MP. An adaptive radiation model for the origin of new gene functions. 727
Nature Genetics. 2005;37(6):573–578. doi:10.1038/ng1579. 728
94. Haynes C, Moreno MA, and AL. chaynes2019/AvidaGeneDupe; 2025. Available 729
from: <https://doi.org/10.5281/zenodo.14911295>. 730
95. European Organization For Nuclear Research, OpenAIRE. Zenodo; 2013. 731
Available from: <https://www.zenodo.org/>. 732
96. Moreno MA, Haynes C. Gene duplications potentiate adaptive evolution of 733
complex traits; 2025. Available from: osf.io/j5s4h. 734
97. Foster ED, Deardorff A. Open Science Framework (OSF). Journal of the 735
Medical Library Association. 2017;105(2). doi:10.5195/jmla.2017.88. 736
98. Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, 737
et al. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. 738
Nature Methods. 2020;17:261–272. doi:10.1038/s41592-019-0686-2. 739
99. Harris CR, Millman KJ, van der Walt SJ, Gommers R, Virtanen P, Cournapeau 740
D, et al. Array programming with NumPy. Nature. 2020;585(7825):357–362. 741
doi:10.1038/s41586-020-2649-2. 742
100. pandas developers. pandas-dev/pandas: Pandas. Zenodo. 743
2020;doi:10.5281/zenodo.3509134. 744
101. Wes McKinney. Data Structures for Statistical Computing in Python. In: 745
Stéfan van der Walt, Jarrod Millman, editors. Proceedings of the 9th Python in 746
Science Conference; 2010. p. 56–61. 747
102. Waskom ML. seaborn: statistical data visualization. Journal of Open Source 748
Software. 2021;6(60):3021. doi:10.21105/joss.03021. 749
103. Hunter JD. Matplotlib: A 2D graphics environment. Computing in Science & 750
Engineering. 2007;9(3):90–95. doi:10.1109/mcse.2007.55. 751

104. Moreno MA. mmore500/teeplot; 2023. Available from: <https://doi.org/10.5281/zenodo.10440670>. 752
753
105. Team RC. R: A language and environment for statistical computing; 2015. 754
Available from: <http://www.R-project.org/>. 755
106. Lalejini A, Wiser MJ, Ofria C. Gene duplications drive the evolution of complex 756
traits and regulation. In: Artificial Life Conference Proceedings. MIT Press; 757
2017. p. 257–264. 758

DRAFT

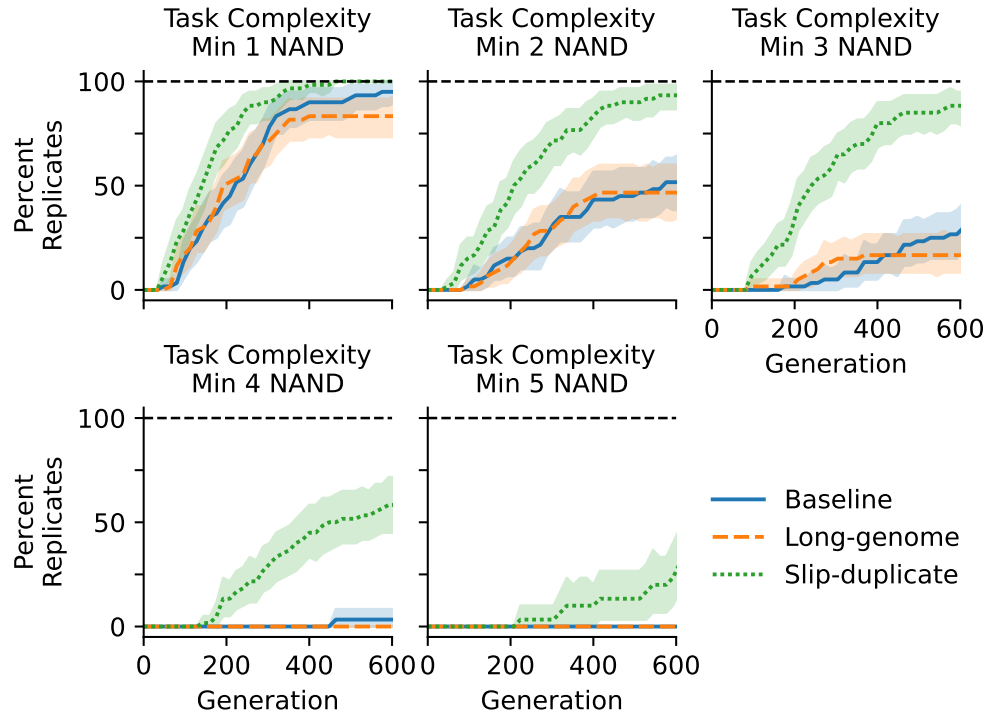


Fig 6. Mutational supply drives increased adaptation rate for long-genome control treatment. Results shown from supplemental control experiments using per-genome, rather than per-site, point mutation processes. In these trials, a count of point mutations was drawn from a Poisson distribution, to be applied 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

759

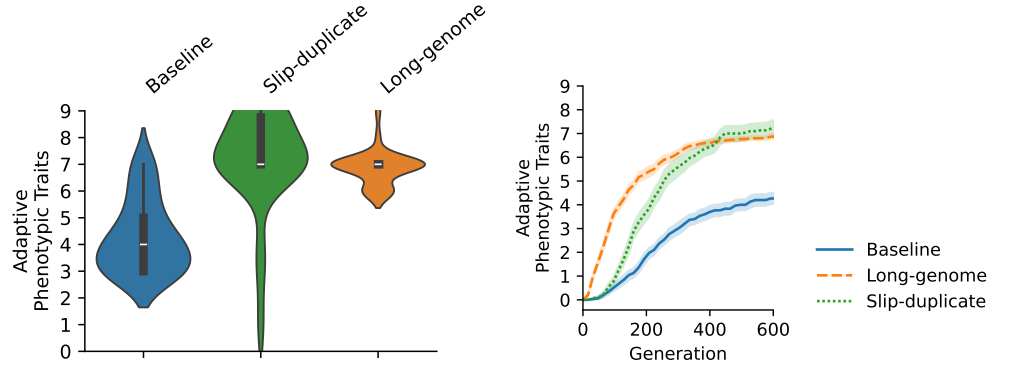


Fig 7. Aggregated adaptive evolution of phenotypic traits for long-genome control experiments. Violin plots show number of adaptive phenotypic traits evolved in final dominant genotypes. Time series (2c right) shows progression of adaptive phenotypic trait count along lineages of final dominant genotypes;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

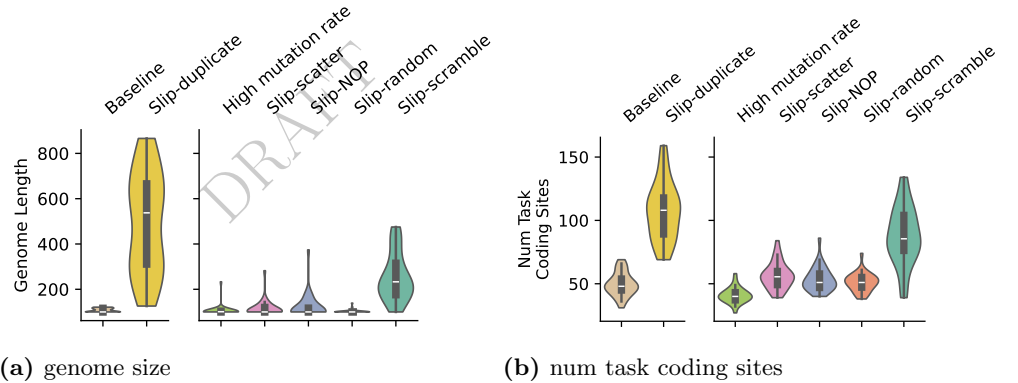
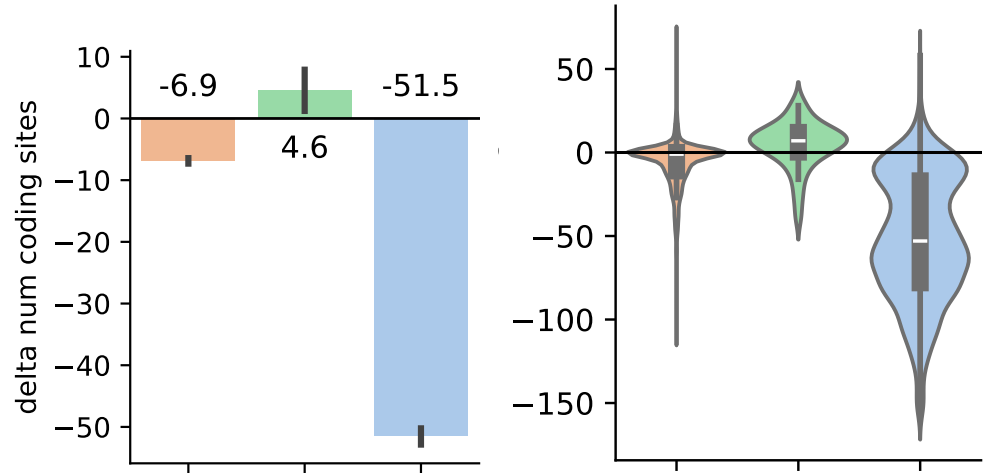
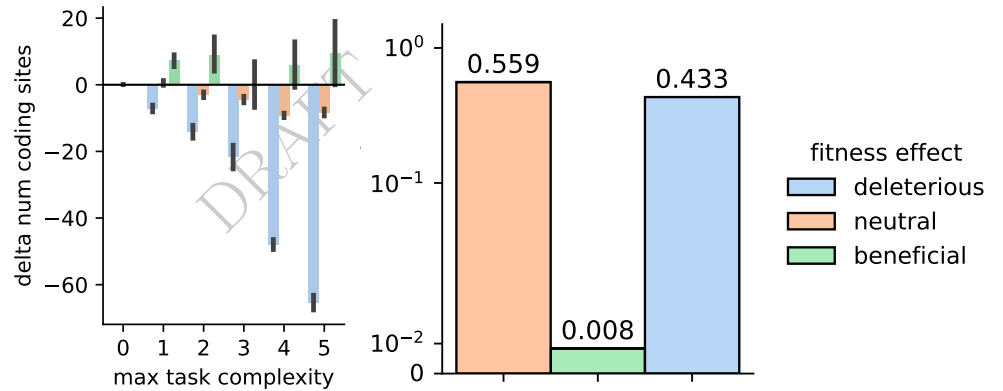


Fig 8. Genome size and task coding site count outcomes across slip-duplication ablation treatments. Violin plots show genome size (panel 8a) and number of task coding sites (panel 8b) in final dominant genotypes. Two-tailed Kruskal-Wallis tests indicate significant between-treatment variation in both 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 ($124 \pm \text{SD } 54$ vs. $252 \pm \text{SD } 105$ sites, $U = 814$, $p < 0.001$) and Slip-duplication treatments ($124 \pm \text{SD } 54$ vs. $491 \pm \text{SD } 223$ 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 ($54 \pm \text{SD } 10$ vs. $88 \pm \text{SD } 24$ sites, $U = 811$, $p < 0.001$) and Slip-duplication treatments ($54 \pm \text{SD } 10$ vs. $107 \pm \text{SD } 23$ 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 ($U = 654$, $p < 0.01$). To account for prolonged copy times, comparisons to long-genome treatment normalize for generations elapsed (Figure 5).



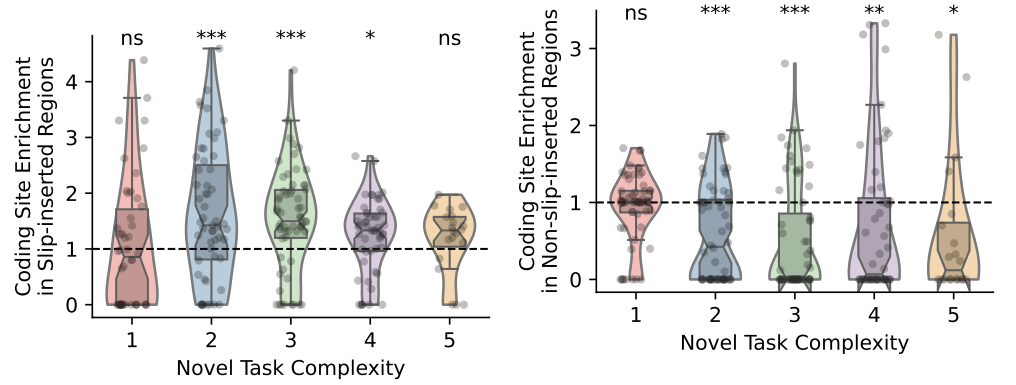
(a) change in coding site count by neutral, beneficial, and deleterious slip duplications; violin plots show count delta distributions and bar plots show mean count deltas



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

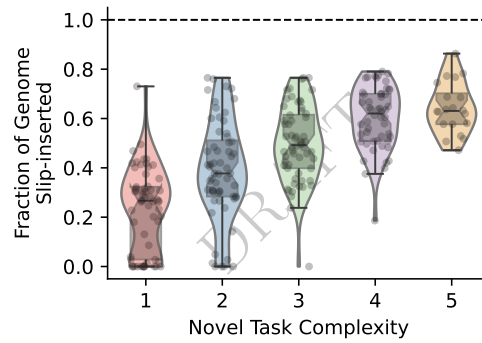
(c) proportion of sampled slip duplications with neutral, beneficial, and deleterious outcomes

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 nor 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 in 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.