Evolution of Drift Robustness in Small Populations of Digital Organisms

Thomas LaBar[1,2,3] & Christoph Adami[1,2,3,4,⋆]

[1] Department of Microbiology & Molecular Genetics
[2] BEACON Center for the Study of Evolution in Action
[3] Program in Ecology, Evolutionary Biology, and Behavior
[4] Department of Physics and Astronomy
Michigan State University, East Lansing, MI 48824

Most mutations are deleterious and cause a reduction in population mean fitness known as the mutational load. In small populations, weakened selection against slightly-deleterious mutations results in an additional reduction in fitness: the drift load. Many studies have established that populations can evolve a reduced mutational load by evolving mutational robustness, but it is uncertain whether populations can evolve a reduced drift load. Here, using digital experimental evolution, we show that small populations do evolve reduced drift loads, that is, they evolve robustness to genetic drift, or "drift robustness". We find that, compared to genotypes from large populations, genotypes from small populations have a decreased likelihood of small-effect deleterious mutations, thus causing small-population genotypes to be drift-robust. We further show that drift robustness is not under direct selection, but instead arises because small populations preferentially adapt to drift-robust fitness peaks. These results have implications for genome evolution in organisms with small effective population sizes.

One consequence of the power of adaptation is that the majority of mutations reduce their bearer's fitness [1]. The recurring nature of these deleterious mutations results in an equilibrium reduction of population fitness at mutation-selection balance. At the population level, this reduction in fitness is known as the genetic or mutational load [2–5]. As selection generally acts to increase a population's mean fitness, one avenue for selection to increase mean fitness is to reduce the mutational load by altering mutation-selection balance and increasing mutational robustness [6, 7]. The evolution of mutational robustness has been demonstrated using theoretical modeling [8–11], digital experimental evolution [12–14], and microbial experimental evolution [15–17].

Recurring deleterious mutations are not the only strain on fitness. In small populations, genetic drift leads to the fixation of slightly-deleterious mutations that bring about a reduction in fitness called the drift load [18, 19]. Over time, genetic drift can lead to continual fitness declines and ultimately population extinction [20, 21]. In asexual populations, this

phenomenon of fitness decline is known as Muller's Ratchet [22] and is thought to play a role in the evolution of mitochondria [23], bacterial endosymbionts [24], the Y chromosome [25], and other obligate asexual lineages. Muller's Ratchet may explain why there are few long-lived obligate asexual species and may provide a selection pressure for the evolution of sexual recombination [26]. However, it was recently shown that small populations do not continuously decline in fitness, but only do so until they reach drift-selection balance when the fixation of beneficial mutations counteracts the fixation of slightly-deleterious mutations [18, 27–29]. Furthermore, Muller's ratchet may be limited in strength if small populations can alter drift-selection balance and evolve drift robustness. However, it is unknown if populations can evolve drift robustness and a reduced drift load, or what genetic and evolutionary mechanisms could cause a reduced drift load.

To test whether populations could evolve drift robustness, we used the digital experimental evolution system Avida [30]. As with microbial experimental evolution, digital evolution makes it possible to study evolution as it occurs [31]. In Avida, a population of self-replicating computer programs ("avidians") compete for the resources necessary for reproduction. During self-replication, random mutations occur, potentially altering the new avidian's reproduction speed. When an avidian successfully reproduces, its offspring replaces a random individual in the population, resulting in genetic drift. As avidians that replicate faster will produce more offspring per unit time than avidians with slower replication speeds, faster replicators are selected for and spread mutations that enable faster replication. Because Avida populations undergo selection, mutation, and drift, they represent a digital model system to study fundamental questions concerning evolutionary dynamics. Avida has been previously used to study both the evolution of mutational robustness [12, 13] and the role of population size on evolutionary outcomes [14, 32].

Results

Here, we evolved 100 replicate populations at small (10^2 individuals) and 100 populations at large (10^4 individuals) population sizes. Small populations evolved for 10^6 generations and large populations for 10^4 generations in order to equalize the number of mutations each population experienced during the experiment. By keeping the experimental mutation supply constant, we aimed to evolve all populations to similar levels of fitness (Fig. S1) in order to reduce the effect of fitness differences on robustness. We hypothesized that small populations would evolve robustness to genetic drift because these populations experienced the stress of strong genetic drift during their initial adaptation. At the same time, we surmised that large populations would not evolve drift robustness as drift is not a strong factor in such environments.

To test for the evolution of drift robustness, we took the most abundant genotype from each population (100 small-population genotypes and 100 large-population genotypes) and measured these genotypes' change in fitness when placed in an environment with strong genetic drift (i.e., low population size). First, we measured the decrease in fitness after 10^3 generations of evolution in a population of 50 individuals (Fig. 1a). Small-population genotypes decreased in fitness a median of 1%, while large-population genotypes decreased in fitness a median of 6%. We next repeated the same test, except we evolved the genotypes

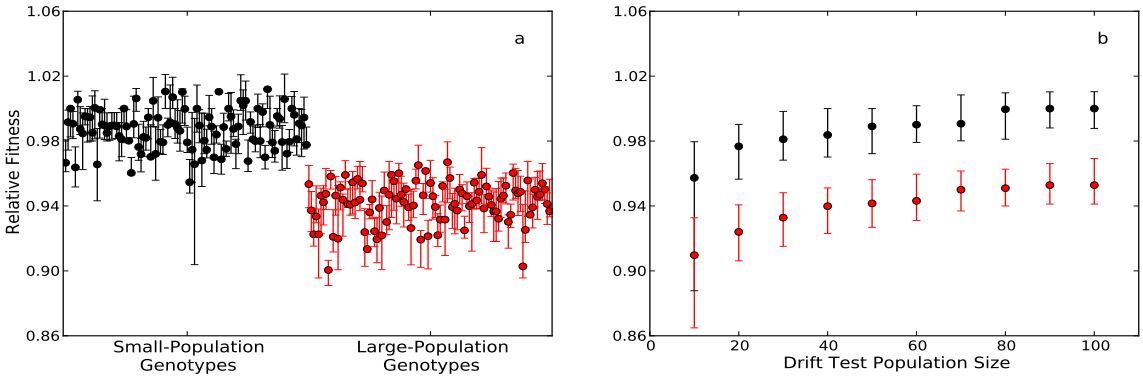

Figure 1: Change in fitness during the drift robustness test for small-population and large-population genotypes. Black (red) markers are for small-population (large-population) genotypes. a) Relative fitness before and after 10^3 generations of evolution with test population size of 50 individuals. Circles represent the median value of 10 replicates for one genotype; error bars are 1st and 3rd quartile. b) Same experiment as in a), but across a range of test population sizes. Each circle is the median of 1000 replicates (100 genotypes x 10 replicates), error bars as in (a).

for 10^4 generations (i.e., an order of magnitude longer) in order to rule out the possibility that small populations only take a greater number of generations to decrease in fitness. Again, we found that small-population genotypes decreased in fitness (median = 2%) less than large-population genotypes (median of 10%; Fig. S2). Finally, we performed this test for drift robustness across a range of small population sizes (10^1 - 10^2 individuals) for 10^3 generations and found similar outcomes (Fig. 1b). These data demonstrate that small-population genotypes are able to withstand the deleterious effects of genetic drift significantly better than large-population genotypes: they are drift-robust.

Next, we explored *why* small-population genotypes were more robust to drift than large-population genotypes. We first looked at differences in the distribution of fitness effects (DFE) between small-population genotypes and large-population genotypes (Fig 2a). Both show the typical DFE found in biological organisms: most mutations are either lethal or have little effect [1]. However, there are some differences. Small-population genotypes have an excess of neutral, beneficial, and strongly deleterious mutations (defined as deleterious mutations with a fitness effect greater than or equal to 5%; Fig. 2b), while large-population genotypes have an excess of small-effect deleterious mutations (defined as deleterious mutations with a fitness effect less than 5%; Fig. 2c). Additionally, the average mutation is more deleterious for small-population genotypes than large-population genotypes (median $s = -0.38$ vs. median $s = -0.33$, Mann Whitney U=3048.0, $n = 100$, $p < 9.30 \times 10^{-5}$ one-tailed; Fig 2d). We confirmed that these trends hold for the average fraction of mutation-type per genotype (rather than averaging over genotypes) as well. Small populations had more neutral mutations (median of 0.40 vs. median 0.30, Mann Whitney U=630.5, $n = 100$, $p < 6.64 \times 10^{-27}$ one-tailed), more large-effect deleterious mutations, including lethal mutations (median of 0.44 vs. median 0.37, Mann Whitney U=2373.0, $n = 100$, $p < 6.91 \times 10^{-11}$

one-tailed), and fewer small-effect deleterious mutations (median of 0.12 vs. median 0.32, Mann Whitney U=33.0, $n = 100$, $p < 3.44 \times 10^{-34}$ one-tailed; Fig. 3a).

The difference in the fraction of small-effect deleterious mutations is striking, as drift-caused fitness decreases depend on the fixation of deleterious mutations with a small effect size. Therefore, differences in small-effect deleterious mutations may cause differences in drift robustness. To test this relationship, we investigated the correlation between the fraction of small-effect deleterious mutations (Fig. 3a) and the decrease in fitness in a strong drift environment (Fig. 1a). These two variables are strongly anti-correlated (Pearson's $r = -0.92$; $p < 4 \times 10^{-84}$), demonstrating that a genotype's drift robustness is determined by the fraction of small-effect mutational neighbors (Fig. 3b).

The experiments described above used organisms with a fixed genome size (50 instructions), one mode of reproduction (asexual), and one genomic mutation rate (0.1 mutations/genome/generation). To test whether the evolution of drift robustness is a robust phenomenon, we repeated the above experiments while relaxing each of these conditions. We also performed experiments where the small populations had initially evolved for the same number of generations as the large populations. These additional experiments showed that small-population genotypes are still more robust to genetic drift than large-population genotypes when the small population initial adaptation time was 10^4 generations (Fig. S3), when genome size was variable (Fig. S4), when reproduction was sexual (Fig. S5), and when the genomic mutation rate was 0.01 mutations/genomes/generation (Fig. S6). These results confirm that the evolution of drift robustness in small populations holds for a broad set of conditions.

We next explored the evolutionary pressure driving the evolution of drift robustness. There are two possible explanations for the evolution of drift robustness: 1) selection drives the evolution of drift robustness, or 2) drift robustness evolves not due to direct selection, but due to some other process. To determine which of these two hypotheses likely explains the evolution of drift robustness, we first performed competition experiments to test if drift robustness is under selection in small populations. For these competitions, we used 3 small-population genotypes and 12 large-population genotypes that evolved the exact same growth rate (fitness) in the original experiments. By using genotypes with equal growth rates, we were able to eliminate growth-rate differences as a factor in the competition outcomes, and focus on the role of other traits.

We performed 100 competition experiments between each combination of small- and large-population genotype in populations of 100 individuals. Based on these competitions, we concluded that drift robustness was not under selection. While we confirmed a relationship between the frequency of small-population genotype success and the ratio of small-effect deleterious mutation fractions (Fig. S7, Pearson's $r = 0.34$, $p < 0.05$), it is the opposite trend one would expect if a lower fraction of small-effect deleterious mutations was under selection (the correlation would have been negative). In addition, we found no significant relationship between the frequency of small-population genotype success and the ratio of small-effect deleterious mutation fractions when we analyzed the data for each small-population genotype on its own (Table S1; Fig. S7).

The results from these competitions suggest that drift robustness is not under positive selection. Previous work has shown that small populations evolve towards a dynamic equilibrium where the fixation of slightly-deleterious mutations is balanced by the fixation

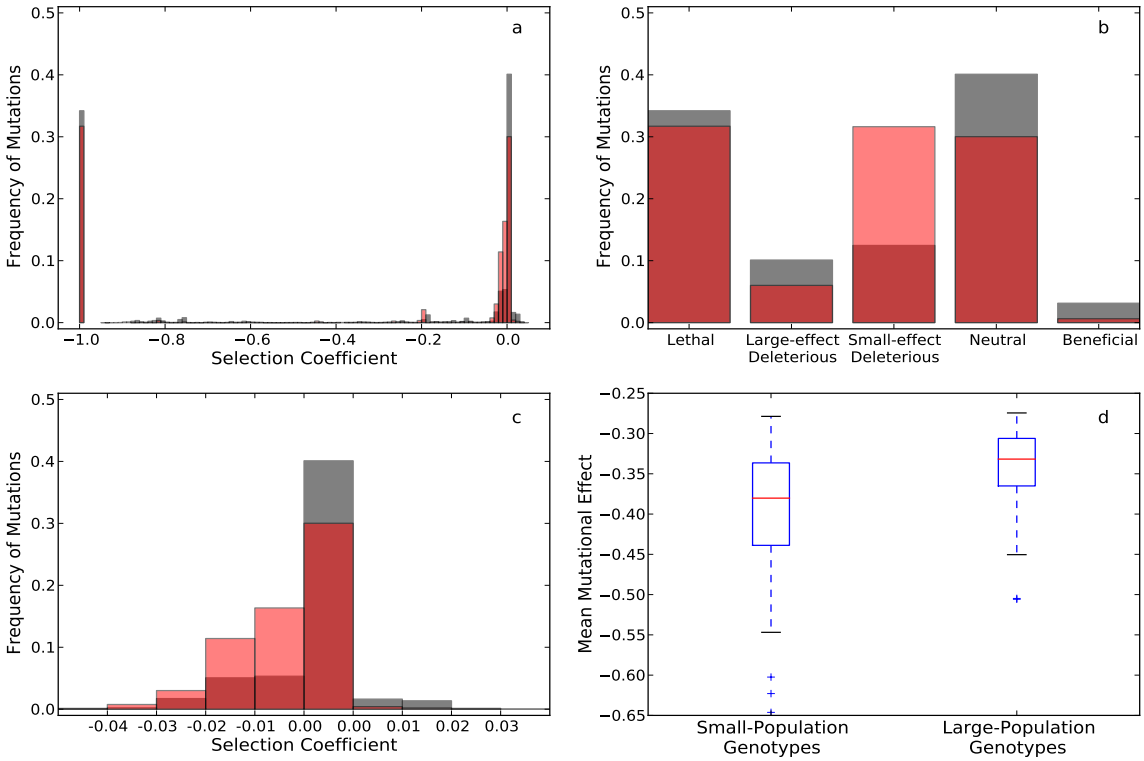

Figure 2: Differences in mutational effects between small-population genotypes and large-population genotypes. Colors as in Figure 1. For boxplots, red lines are medians, edges of the box are first and third quartile, whiskers are at most 1.5 times the relevant quartile, and the plus signs are outliers. a) The combined distribution of fitness effects (DFE) across all 100 small-population genotypes and 100 large-population genotypes. b) Same data as in panel a, but grouped into different classes of mutations. See main text for descriptions of small-effect deleterious mutations and large-effect deleterious mutations. c) The frequency of mutations with small selection coefficients across all 100 small-population genotypes and 100 large-population genotypes. Full DFE shown in panel a. d) The mean mutational effect (selection coefficient) of every possible point mutation (1250 mutations) for each genotype.

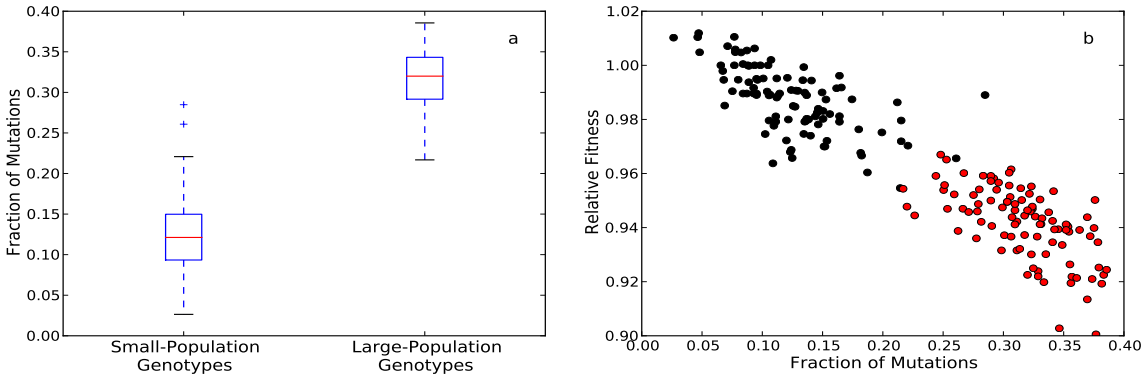

Figure 3: Relationship between small-effect deleterious mutations and drift robustness. Colors as in Figure 1. For boxplots, red lines are medians, edges of the box are first and third quartile, whiskers are at most 1.5 times the relevant quartile, and the plus signs are outliers. a) The fraction of point mutations that had a deleterious fitness effect less than 5% for each genotype (i.e., small-effect deleterious mutations). b) Relationship between fraction of small-effect deleterious mutations (panel a) and the median relative fitness from the drift robustness test (Fig. 1a).

of beneficial mutations [28, 29]. We wondered whether similar dynamics occurred in these populations, and hypothesized that drift-selection balance limits the fraction of small-effect deleterious mutations (Fig. 4a), that is, if a "drift-barrier" restricts the fraction of small-effect deleterious mutations [33]. For a genotype with a fraction of small-effect deleterious mutations above this line, drift will cause the fixation of small-effect deleterious mutations and decrease the fraction of small-effect deleterious mutations. For a genotype with a fraction of small-effect deleterious mutations below the drift-barrier, selection will cause the fixation of small-effect beneficial mutations and increase the fraction of small-effect deleterious mutations. Small populations can still adapt to novel fitness peaks, but must do so under this constraint. If this hypothesis is correct, small populations can then only adapt to drift-robust fitness peaks (those with few small-effect deleterious mutations), as they cannot persist on drift-fragile fitness peaks (those with many small-effect deleterious mutations). This would explain why small-population and large-population genotypes possess different DFEs, even if they are of similar fitness.

To test if population size limited the fraction of small-effect deleterious mutations in a genotype's DFE, we took the 3 small-population and 12 large-population genotypes used in the competition experiments above and evolved them at the "opposite" population size from which they originally evolved. As expected, small-population genotypes increased their fraction of small-effect deleterious mutations after evolution in a large population (Fig. 4b). Likewise, large-population genotypes decreased their fraction of small-effect deleterious mutations after evolution in a small population (Fig. 4b). These results suggest that drift-selection balance does limit the accumulation of small-effect deleterious mutations in the DFEs of small populations.

Finally, to further test drift-selection balance as the evolutionary cause of drift robustness,

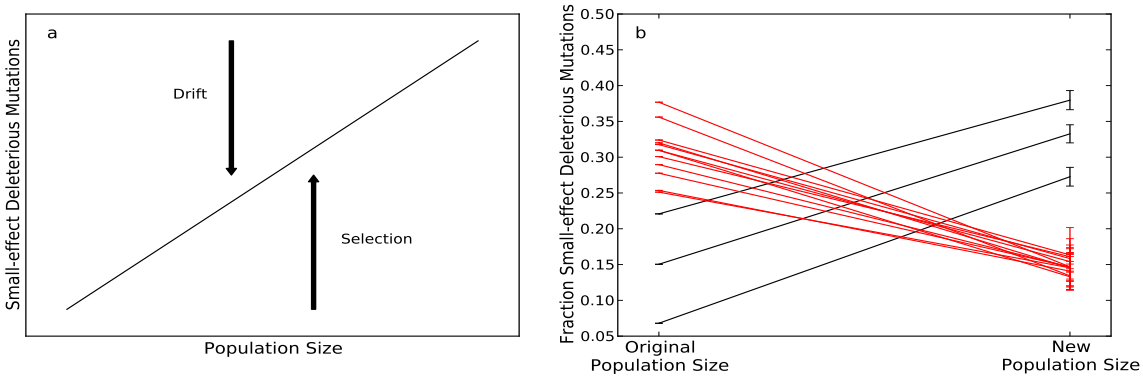

Figure 4: Drift-selection balance limits the fraction of small-effect deleterious mutations in small populations. a) Conceptual diagram of the balance between selection and drift and how it limits the number of small-effect deleterious mutations as a function of population size. b) The change in the fraction of small-effect deleterious mutations after evolution at the "opposite" population size. Small-population genotypes (black) evolved in a large population and large-population genotypes (red) evolved in a small population. Error bars are 95% confidence intervals of the mean.

we repeated the initial adaptation phase of these experiments, except we reverted deleterious mutants to their parent genotype when they entered the population (as in [34]). This setup prevents deleterious (but not lethal) mutations from entering the population and thus prevents these mutations from fixing due to genetic drift. Under these conditions, genotypes from small populations are not expected to evolve drift robustness, as drift cannot cause small populations to fall off drift-fragile peaks. In this treatment, small-population genotypes evolved on average greater fitness, compared to small-population genotypes that evolved in the presence of deleterious mutations (Fig. 5a; median of 1.94 vs. median of 2.25, Mann Whitney U=2432.0, $n = 100$, $p < 1.76 \times 10^{-10}$ one-tailed). Furthermore, small-population genotypes that evolved in the reversion treatment were more drift fragile than the original small-population genotypes (Fig. 5b; median fitness decrease of 5% vs. median fitness decrease of 1%). As expected from this result, the DFE of small-population genotypes from the reversion treatment differs from the DFE of the original small-population genotypes (Fig. 5c). These genotypes had fewer neutral mutations and more small-effect deleterious mutations (Fig. 5d), more large-effect deleterious mutations, and more lethal mutations. These results demonstrate that the evolution of drift robustness requires prior adaptation under the hypothesized drift-selection balance dynamics.

Discussion

Our results suggest the following explanation for the evolution of drift robustness in small populations. Genetic drift limits the number of small-effect beneficial mutations that can be fixed by selection and as a consequence small populations must adapt to peaks with a limited fraction of small-effect deleterious mutations. Genotypes in large populations are not under

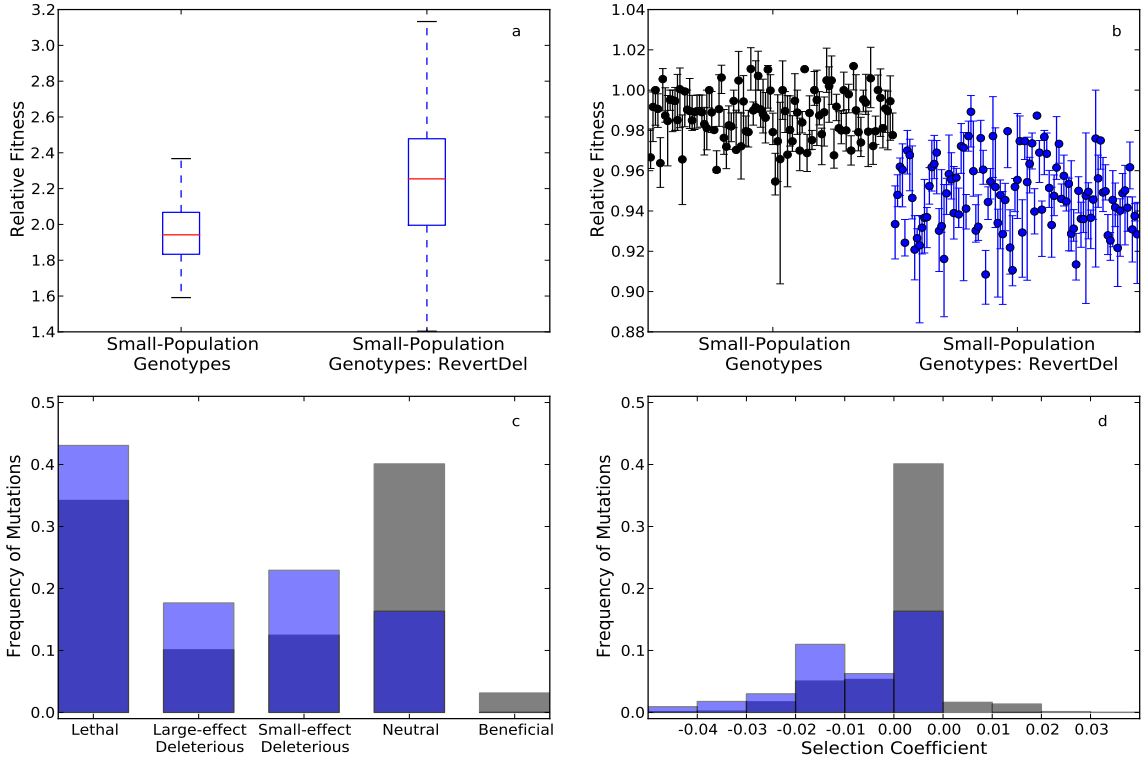

Figure 5: The evolution of drift robustness in small populations with or without deleterious mutations in the initial adaptation experiments. Gray refers to genotypes from the original experiments and blue refers to genotypes that evolved with deleterious mutations reverted a) Relative fitness to the ancestral genotype after 10^6 generations of adaptation. Box plots as in Fig. 2. b) Relative fitness before and after 10^3 generations of evolution with test population size of 50 individuals. Circles represent the median value of 10 replicates for one genotype; error bars are 1st and 3rd quartile. c) Distribution of fitness effects, separated into different mutational classes. d) Distribution of fitness effects for mutations with small selection coefficients.

this adaptive constraint and can adapt to fitness peaks with an abundance of small-effect deleterious mutations. However, this drift-selection constraint does not ultimately limit the adaptive potential of small populations (although it does decrease the speed of adaptation). Small populations can evolve towards fitness peaks of heights similar to those reached by large populations, but they must do so while maintaining drift robustness (or they will fall from these peaks). Therefore, small populations will evolve to drift-robust fitness peaks, while large populations climb drift-fragile fitness peaks. How do small populations adapt to drift-robust fitness peaks? It is likely they do so by fixing beneficial mutations that do not subsequently increase the fraction of small-effect deleterious mutations, but instead those that increase the fraction of neutral, large-effect deleterious, and lethal mutations.

Our data also suggest that the evolution of drift robustness in small populations is a different phenomenon than the evolution of mutational robustness in small populations. Previous studies provided two predictions for the evolution of mutational robustness in small populations. First, small populations should preferentially evolve to lower fitness peaks with more "redundancy," defined as a decreased average deleterious mutational effect [9,14]. Second, small populations should evolve "global" robustness mechanisms, such as error-correction mechanisms, that affect many loci [10, 11]. Neither of these predictions were verified in our experiments. Small populations evolved to fitness peaks with an increased average deleterious mutational effect (Fig. 2d). Furthermore, no global mechanism was required for the evolution of drift robustness.

It is likely that the evolution of drift robustness has many consequences for the evolutionary dynamics of small populations. The trend of small populations evolving to drift-robust fitness peaks provides another explanation for why small asexual populations can persist in the presence of Muller's ratchet, in addition to evidence that the ratchet can be balanced by beneficial mutations [28, 29]. The difference in DFEs between small-population genotypes and large-population genotypes also suggests that genetic drift may cause small populations to evolve different genomic architectures than large populations, as previously proposed [35,36]. However, we leave the study of the role of population size and genetic drift in the evolution of genome architecture to future work.

Candidates for organisms with drift-robust genomes include bacterial endosymbionts [24], genome-containing eukaryotic organelles (such as mitochondria) [37], and RNA viruses [38], all of which go through bottlenecks at some point in their life cycle, and multicellular eukaryotes with small effective population sizes [39]. There is evidence that both bacterial endosymbionts [17, 40–42] and RNA viruses [43, 44] have evolved alternate genome architectures in response to their population-genetic environment. Additionally, genomic-architecture traits found in eukaryotes may be the result of genetic drift [36]. A recent study has suggested that humans have a greater likelihood of strongly-deleterious and slightly-beneficial mutations than *Drosophila melanogaster* and this is related to humans' smaller effective population size [45]. This result is similar to the relationship between the distribution of fitness effects and population size that we found here. However, there has been no systematic study of how different organisms respond to strong genetic drift. Future work with biological organisms should establish the circumstances that cause organisms to vary in their robustness to genetic drift. Furthermore, experimental evolution may be able to produce organisms with drift-robust genomes whose architecture can be studied directly.

Online Methods

Avida

Experimental evolution was carried out using the digital evolution system Avida version 2.14. In Avida, a population of self-replicating computer programs is subject to Darwinian evolution as follows. Each of the programs ("avidians") consists of a genome of sequential computer instructions, drawn from an alphabet of twenty-six possible instructions. Together, these loci encode the ability for an avidian to create a new daughter avidian, copy its genome into the new avidian, and divide off the offspring. During this process, mutations can be introduced into the offspring's genome at a controlled rate, introducing genetic variation into the population. When a new offspring is placed into the population (and the population is at carrying capacity), a random individual is replaced by the new avidian, a process that introduces genetic drift into Avida populations. Avidians differ in their replication speed due to different genomic sequences, so that avidians that can replicate faster will out-compete slower-replicating types. Therefore, because variation is heritable, and because this variation leads to differential reproduction, an Avida population undergoes Darwinian evolution by natural selection. Avida has previously been used to study many concepts that are difficult to test with biological systems [46–52].

The Avida world consists of a toroidal grid of N cells, where N is the maximum population size. Each cell can be occupied by at most one avidian, although a cell may be empty. Upon reproduction, the offspring avidian is placed into an empty cell (if the population is below capacity) or into a random cell, where it replaces the already-present avidian. Although the default Avida setting places offspring into one of nine neighboring cells (including the parent) so as to emulate growth on a surface, in the present experiments any cell may be selected for replacement, to simulate a well-mixed environment.

Time in Avida is set according to "updates" (the time it takes for an avidian population to execute a give number of instructions). During each update, $30N$ instructions are executed across the population, where N is again the population size. In order to be able to execute its code, and avidian must have a resource, measured as "Single Instruction Processing" units (SIPs). At the beginning of each update, SIPs are distributed to programs in the population in proportion to a quantity called "merit", which is related to a genotype's ability to exploit the environment (see [30] for details).

In most of the experiments performed here, merit was held constant across all individuals, so on average 30 SIPs were distributed to each individual every update. The one treatment were merit could vary between individuals was the treatment where genome size was not fixed and could evolve. Merit is proportional to genome size in order to offset the decreased replication speed that comes with a larger genome; individuals with larger genomes thus have greater merits.

It should be noted that in most Avida experiments, populations can evolve the ability to perform certain Boolean logic calculations that can improve their merit and hence their fitness [48]. In the experiments performed here, the evolution of these logic calculations was set to be neutral and not under positive selection. Instead, the route for an avidian to improve its fitness was solely by reducing the number of instruction executions needed to copy its genome. A population will typically evolve a faster replication speed by increasing

the number of instructions that copy instructions from the parent genome to the offspring genome. When this copy number increase occurs, more instructions are copied per update, resulting in faster replication and greater fitness.

Reproduction is asexual in all but one of the treatments performed here. We also tested a treatment where avidians could reproduce sexually (see [53] for a more in-depth overview of sexual reproduction in Avida). In this treatment, when an avidian successfully copies its genome and executes a divide instruction, the new offspring's genome is not immediately placed into the population. Instead, the program waits until another avidian has executed a divide instruction. Then, the program selects a contiguous section of the genome from each offspring and swaps them, after which both offspring are placed into the population as usual.

Although Avida uses the update as its unit of time, experiments such as those performed here are often run for a given number of generations (the time it takes for the entire population to be replaced). The experiment ends when the average generation across all of the individuals in the population reaches a pre-specified number. Each individual's generation counter is equal to its parent's generation plus one. Therefore, while Avida experiments occur for a set number of generations, the population does not evolve with discrete generations. If fitness differs between individuals and lineages in the population, there can be variation in the individuals' generations in the population.

Experimental Design

The experiments performed for this study can be broken up into five sections. First, initial adaptation experiments were performed to generate genotypes adapted to small and large population size environments. For most treatments, the small populations (10^2 individuals) evolved for 10^6 generations and the large populations (10^4 individuals) evolved for 10^4 generations. This choice kept the total number of mutations during the course of the experiment constant on average. We also performed a set of experiments where we evolved the small populations for 10^4 generations only (i.e., the same amount of time as the large populations) to show that the unequal time of initial adaptation does not alter the results. All of these treatments had a genomic mutation rate of 10^{-1} mutations/generation/genome. We also performed a low mutation rate treatment with a genomic mutation rate of 10^{-2} mutations/generation/genome. For this treatment, the small populations evolved for 10^7 generations and the large populations evolved for 10^5 generations to compensate for the lower mutation rate. For every treatment, the ancestor organism for the initial adaptation treatments was the default Avida ancestor of 100 instructions, but with an altered genome length of 50 instructions. This alteration was performed by removing 50 nop-C instructions from the default genome (these instructions are inert). For each treatment and population size we performed 100 replicate experiments.

The second experimental step was to perform a test to measure the drift robustness of individuals evolved at a small population size versus individuals evolved at a large population size. From each small and large population, we used the most abundant individual to form a set of 100 small-population genotypes and 100 large-population genotypes per treatment. For each of these genotypes, we evolved 10 populations (2000 replicates in total) at various small population sizes for 10^3 generations (although the genotypes that evolved at a lower mutation rate were evolved during this test for 10^4 generations). For all treatments, this test

was performed at a population size of 50 individuals. For the main treatment, the test was also performed at population sizes ranging from 10 individuals to 100 individuals. We also evolved 10 populations of each genotype (2000 replicates in total) at a population size of 50 individuals for 10^4 generations for the main treatment. All treatment parameters were kept constant between the initial adaptation experiments and these tests, except for population size.

The third set of experiments tested whether drift robust small-population genotypes had a competitive advantage over large-population genotypes. We took three small-population genotypes and 12 large-population genotypes from the main treatment that had evolved the same reproduction speed. Then, for every combination of small-population genotype and large-population genotype ($3 \times 12 = 36$ combinations) we competed the two genotypes at an initial 1:1 ratio in a population of 100 individuals for 10^3 generations. We subsequently determined which of the two genotypes was the ancestor of the entire current population (in all competitions, one genotype eventually out-competes the other).

The fourth set of experiments tested whether population size determined the fraction of small-effect deleterious mutations in a genotype's DFE. To determine this, we evolved 10 replicates of the small-population genotypes from the competition experiments above for a further 10^4 generations in populations of 10^4 individuals. We also evolved ten replicates of twelve large-population genotypes for a further 10^4 generations in populations of 10^2 individuals, then measured the change in fraction of small-effect deleterious mutations.

The final set of experiments tested whether deleterious mutations were responsible for the evolution of drift robustness in small populations. We repeated the initial adaptation experiment and the drift robustness test with 50 individuals under the same parameter settings as the experiments featured in the main text. However, during the initial adaptation experiment, we reverted any deleterious mutations that appeared in the population [34]. In this setup, the Avida world examines the fitness cost of every new point mutation. If this new mutant has decreased fitness relative to its parent, the mutant is removed from the population and an exact copy of its parent is placed into the population instead.

Data Analysis

We calculated statistics for the evolved avidians using Avida's Analyze Mode [30]. In Analyze Mode, the experimenter can run an avidian through its life-cycle (until reproduction) and calculate several genotype characteristics. Fitness was calculated as the ratio between the number of instructions in the genome (the sequence length) to the number of instruction executions needed to copy the genome and reproduce (this is an unbiased predictor of the actual number of offspring).

In order to calculate the distribution of fitness effects for each genotype and other related mutational measures, each point mutation was generated for each genotype ($25 \times L$ mutations, where L is the number of instructions in the genome). The fitness effect of each mutation was calculated as $s = \frac{w_m}{w_0} - 1$, where w_m is the fitness of the mutant and w_0 was the fitness of the genotype. The average mutational effect of each genotype is the arithmetic mean of these fitness effects. The fraction of mutations of a given fitness effect was calculated as the number of mutations with that fitness effect divided by $25L$.

Statistical analyses were performed using the NumPy [54], SciPy [55], and Pandas [56] Python modules. Figures were created with the Matplotlib [57] Python module.

Code Availability

The Avida software is available for free use (https://github.com/devosoft/avida). Avida configuration scripts and data from Avida experiments is available at the Dryad data repository ().

Acknowledgements

T.L. acknowledges a Michigan State University Distinguished Fellowship, a BEACON fellowship, and the Russell B. Duvall award for support. This work was supported in part by Michigan State University through computational resources provided by the Institute for Cyber-Enabled Research. This material is based in part upon work supported by the National Science Foundation under Cooperative Agreement No. DBI-0939454. 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.

References

[1] Eyre-Walker, A. & Keightley, P. D. The distribution of fitness effects of new mutations. *Nature Reviews Genetics* **8**, 610–618 (2007).

[2] Crow, J. F. Some possibilities for measuring selection intensities in man. *Human Biology* **30**, 1–13 (1958).

[3] Kimura, M., Maruyama, T. & Crow, J. F. The mutation load in small populations. *Genetics* **48**, 1303 (1963).

[4] Kimura, M. & Maruyama, T. The mutational load with epistatic gene interactions in fitness. *Genetics* **54**, 1337 (1966).

[5] Agrawal, A. F. & Whitlock, M. C. Mutation load: the fitness of individuals in populations where deleterious alleles are abundant. *Annual Review of Ecology, Evolution, and Systematics* **43**, 115–135 (2012).

[6] Visser, J. *et al.* Perspective: evolution and detection of genetic robustness. *Evolution* **57**, 1959–1972 (2003).

[7] Wilke, C. O. & Adami, C. Evolution of mutational robustness. *Mutation Research* **522**, 3–11 (2003).

[8] van Nimwegen, E., Crutchfield, J. P. & Huynen, M. Neutral evolution of mutational robustness. *Proceedings of the National Academy of Sciences* **96**, 9716–9720 (1999).

[9] Krakauer, D. C. & Plotkin, J. B. Redundancy, antiredundancy, and the robustness of genomes. *Proceedings of the National Academy of Sciences* **99**, 1405–1409 (2002).

[10] Gros, P.-A. & Tenaillon, O. Selection for chaperone-like mediated genetic robustness at low mutation rate: impact of drift, epistasis and complexity. *Genetics* **182**, 555–564 (2009).

[11] Rajon, E. & Masel, J. Evolution of molecular error rates and the consequences for evolvability. *Proceedings of the National Academy of Sciences* **108**, 1082–1087 (2011).

[12] Wilke, C. O., Wang, J. L., Ofria, C., Lenski, R. E. & Adami, C. Evolution of digital organisms at high mutation rates leads to survival of the flattest. *Nature* **412**, 331–333 (2001).

[13] Edlund, J. A. & Adami, C. Evolution of robustness in digital organisms. *Artificial life* **10**, 167–179 (2004).

[14] Elena, S. F., Wilke, C. O., Ofria, C. & Lenski, R. E. Effects of population size and mutation rate on the evolution of mutational robustness. *Evolution* **61**, 666–674 (2007).

[15] Montville, R., Froissart, R., Remold, S. K., Tenaillon, O. & Turner, P. E. Evolution of mutational robustness in an RNA virus. *PLoS Biol* **3**, e381 (2005).

[16] Sanjuán, R., Cuevas, J. M., Furió, V., Holmes, E. C. & Moya, A. Selection for robustness in mutagenized RNA viruses. *PLoS Genet* **3**, e93 (2007).

[17] Sabater-Muñoz, B. *et al.* Fitness trade-offs determine the role of the molecular chaperonin GroEL in buffering mutations. *Molecular Biology and Evolution* **32**, 2681–2693 (2015).

[18] Poon, A. & Otto, S. P. Compensating for our load of mutations: freezing the meltdown of small populations. *Evolution* **54**, 1467–1479 (2000).

[19] Whitlock, M. C. Fixation of new alleles and the extinction of small populations: drift load, beneficial alleles, and sexual selection. *Evolution* **54**, 1855–1861 (2000).

[20] Haigh, J. The accumulation of deleterious genes in a population–Muller's ratchet. *Theoretical Population Biology* **14**, 251–267 (1978).

[21] Lynch, M., Bürger, R., Butcher, D. & Gabriel, W. The mutational meltdown in asexual populations. *Journal of Heredity* **84**, 339–344 (1993).

[22] Muller, H. J. The relation of recombination to mutational advance. *Mutation Research* **1**, 2–9 (1964).

[23] Lynch, M. Mutation accumulation in transfer RNAs: Molecular evidence for Muller's ratchet in mitochondrial genomes. *Molecular Biology and Evolution* **13**, 209–220 (1996).

[24] Moran, N. A. Accelerated evolution and Muller's rachet in endosymbiotic bacteria. *Proceedings of the National Academy of Sciences* **93**, 2873–2878 (1996).

[25] Gordo, I. & Charlesworth, B. The speed of Muller's ratchet with background selection, and the degeneration of Y chromosomes. *Genetical Research* **78**, 149–161 (2001).

[26] Felsenstein, J. The evolutionary advantage of recombination. *Genetics* **78**, 737–756 (1974).

[27] Lande, R. Risk of population extinction from fixation of deleterious and reverse mutations. *Genetica* **102**, 21–27 (1998).

[28] Silander, O. K., Tenaillon, O. & Chao, L. Understanding the evolutionary fate of finite populations: the dynamics of mutational effects. *PLoS Biol* **5**, e94 (2007).

[29] Goyal, S. *et al.* Dynamic mutation–selection balance as an evolutionary attractor. *Genetics* **191**, 1309–1319 (2012).

[30] Ofria, C., Bryson, D. M. & Wilke, C. O. Avida: A software platform for research in computational evolutionary biology. In Maciej Komosinski, A. A. (ed.) *Artificial Life Models in Software*, 3–35 (Springer London, 2009).

[31] Kawecki, T. J. *et al.* Experimental evolution. *Trends in Ecology & Evolution* **27**, 547–560 (2012).

[32] LaBar, T. & Adami, C. Different evolutionary paths to complexity for small and large populations of digital organisms. *PLoS Computational Biology* **12** (2016).

[33] Sung, W., Ackerman, M. S., Miller, S. F., Doak, T. G. & Lynch, M. Drift-barrier hypothesis and mutation-rate evolution. *Proceedings of the National Academy of Sciences* **109**, 18488–18492 (2012).

[34] Covert, A. W., Lenski, R. E., Wilke, C. O. & Ofria, C. Experiments on the role of deleterious mutations as stepping stones in adaptive evolution. *Proceedings of the National Academy of Sciences* **110**, E3171–E3178 (2013).

[35] Lynch, M. *The Origins of Genome Architecture* (Sinauer Associates, Sunderland, MA, 2007).

[36] Lynch, M., Bobay, L.-M., Catania, F., Gout, J.-F. & Rho, M. The repatterning of eukaryotic genomes by random genetic drift. *Annual Review of Genomics and Human Genetics* **12**, 347 (2011).

[37] Neiman, M. & Taylor, D. R. The causes of mutation accumulation in mitochondrial genomes. *Proceedings of the Royal Society of London B* **276**, 1201–1209 (2009).

[38] Zwart, M. P. & Elena, S. F. Matters of size: Genetic bottlenecks in virus infection and their potential impact on evolution. *Annual Review of Virology* **2**, 161–179 (2015).

[39] Lynch, M. & Conery, J. S. The origins of genome complexity. *Science* **302**, 1401–1404 (2003).

[40] Fares, M. A., Ruiz-González, M. X., Moya, A., Elena, S. F. & Barrio, E. Endosymbiotic bacteria: groEL buffers against deleterious mutations. *Nature* **417**, 398–398 (2002).

[41] Kuo, C.-H., Moran, N. A. & Ochman, H. The consequences of genetic drift for bacterial genome complexity. *Genome Research* **19**, 1450–1454 (2009).

[42] Kelkar, Y. D. & Ochman, H. Genome reduction promotes increase in protein functional complexity in bacteria. *Genetics* **193**, 303–307 (2013).

[43] Elena, S. F., Carrasco, P., Daròs, J.-A. & Sanjuán, R. Mechanisms of genetic robustness in RNA viruses. *EMBO Reports* **7**, 168–173 (2006).

[44] Holmes, E. C. *The evolution and emergence of RNA viruses* (Oxford University Press, 2009).

[45] Huber, C. D., Kim, B. Y., Marsden, C. D. & Lohmueller, K. E. Determining the factors driving selective effects of new nonsynonymous mutations. BioRxiv http://dx.doi.org/10.1101/071209 (2016).

[46] Lenski, R. E., Ofria, C., Collier, T. C. & Adami, C. Genome complexity, robustness and genetic interactions in digital organisms. *Nature* **400**, 661–664 (1999).

[47] Adami, C., Ofria, C. & Collier, T. C. Evolution of biological complexity. *Proceedings of the National Academy of Sciences* **97**, 4463–4468 (2000).

[48] Lenski, R. E., Ofria, C., Pennock, R. T. & Adami, C. The evolutionary origin of complex features. *Nature* **423**, 139–144 (2003).

[49] Adami, C. Digital genetics: unravelling the genetic basis of evolution. *Nature Reviews Genetics* **7**, 109–118 (2006).

[50] Goldsby, H. J., 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* **109**, 13686–13691 (2012).

[51] Goldsby, H. J., Knoester, D. B., Ofria, C. & Kerr, B. The evolutionary origin of somatic cells under the dirty work hypothesis. *PLoS Biol* **12**, e1001858 (2014).

[52] Zaman, L. *et al.* Coevolution drives the emergence of complex traits and promotes evolvability. *PLoS Biol* **12**, e1002023 (2014).

[53] Misevic, D., Ofria, C. & Lenski, R. E. Sexual reproduction reshapes the genetic architecture of digital organisms. *Proceedings of the Royal Society of London B: Biological Sciences* **273**, 457–464 (2006).

[54] Van Der Walt, S., Colbert, S. C. & Varoquaux, G. The NumPy array: a structure for efficient numerical computation. *Computing in Science & Engineering* **13**, 22–30 (2011).

[55] Jones, E., Oliphant, T., Peterson, P. *et al.* SciPy: Open source scientific tools for Python, 2001– (2015). URL `http://www.scipy.org`.

[56] McKinney, W. *Python for data analysis: Data wrangling with Pandas, NumPy, and iPython* (O'Reilly Media, Inc., 2012).

[57] Hunter, J. D. *et al.* Matplotlib: A 2D graphics environment. *Computing in Science and Engineering* **9**, 90–95 (2007).