Saumil Shah
Guide - Sutirth Dey

Adaptation to the same environment can result in loss or gain of a trait based on different population sizes

End Semester Report
Fall 2018

Introduction

Fitness is a central concept in evolutionary biology. The classical description of the fitness of an individual is related to its reproductive output. Fitness was considered to be an attribute of an individual, but it can also be used to elucidate other higher or lower biological units such as the population or the gene. In the context of altruistic behavior, group selection ensues, and a set of individuals with the highest group fitness prevail. Allelic or genotypic fitness describes the contribution of an allele or a genotype to the next generation.

However, the fitness of a particular allele or genotype depends on several other factors. There are interactions in the genome between different loci which lead to epistasis, meaning fitness of a gene depends on the available genetic background of the organism. Epistasis can be classified into various categories based on the description of the interaction of the genes. One such class of epistasis is called sign epistasis, where the sign of the fitness of a gene is under epistatic control. This indicates that an allele is beneficial in some genetic backgrounds while deleterious in other.

Another such controlling factor is the environment. The fitness of a genotype depends on the surrounding environment. For example, the fittest genotype during an ice age might not be the fittest genotype once the ice age is over. A population must alter its genotype distribution according to the change in environment, and this takes place via the interaction of biological traits of an individual with the surroundings. The heritable genetic basis of any biological trait that gives it a fitness advantage over other traits becomes more common in the population. Consequently, a population will always adapt and improve their fitness in a given environment.

If the gain of a trait translates into higher fitness in an environment, then the population will get selected to retain the trait in that environment. The opposite is true for the case where the loss of a trait translates into higher fitness. Unwanted traits in an environment with costs or adverse effects will be lost over generations. Hence the trait in question can either be useful or unnecessary but not both hence it will either be enhanced or lost but never both in a population. Therefore utility of a trait in a given environment determines the evolutionary trajectory taken by the population in that environment.

Contrary to above-developed understanding, recent experiments have revealed that adaptation to the same environment can result in the decay or enhancement of the same character based on differences in population size. *Escherichia coli* populations were cultured in an environment with a cocktail of three antibiotics in the experiments. While adapting to this environment, *E. coli* cultures with small population size enhanced their efflux activity compared to the ancestors, whereas large population lost the efflux activity. Noting the tendency of the trait to undergo a retroverted selection based on numerical changes in population size, we are interested in understanding how quantitative changes can give rise to qualitative changes. Here we have developed a simple system using a toolkit of basic evolutionary mechanisms, such as epistasis and mutations, that reproduce similar results as the experiments.

Objectives

1. Simulate for the phenomenon proposed in the title.
2. Determine the minimal set of conditions which can lead to such incident.
3. Verify whether this phenomenon is deterministic and explore the possibility of mathematical modeling.

Methods

Setup

Asexual haploid individuals were chosen for this study to keep the complexity to a minimum. Agent-based computer simulations were carried out to reproduce the proposed phenomenon. The simulations included simple Wright-Fisher model, having random genetic drift along with mutations to mimic the dynamics of asexual populations. The simulations were subjected to discrete generations and were parameterized by population size. During the simulations, all the replicates maintained the specified population size after the random genetic drift being applied every generation (Wright-Fisher step). These simulations were carried out for a wide range of population sizes, spanning from 10^1 to 10^7 individuals. All the simulations were carried out for 1000 generations.

Sign Epistasis

The genotype of all individuals was composed of two loci, A and B, each with three alleles L (low expression), O (intermediate expression), and H (high expression). This gave rise to nine genotypes. The fitness of genes was such that

higher expression of the gene resulted in higher fitness. Gene A exhibited sign epistasis on Gene B backgrounds such that high expression of gene A was beneficial in B_L or B_O backgrounds but deleterious in B_H background as illustrated in **figure 1**.

A_OB_O was chosen as the wild-type for the simulations. All the simulations started with clonal populations of wild-type individuals. Fitness values for each genotype used in the simulation represented the number of offsprings produced per individual per generation, shown in **figure 1**. These fitness values were kept constant throughout the simulations to account for one single environment, except non-epistatic controls, where the A_LB_H and A_HB_H values were interchanged.

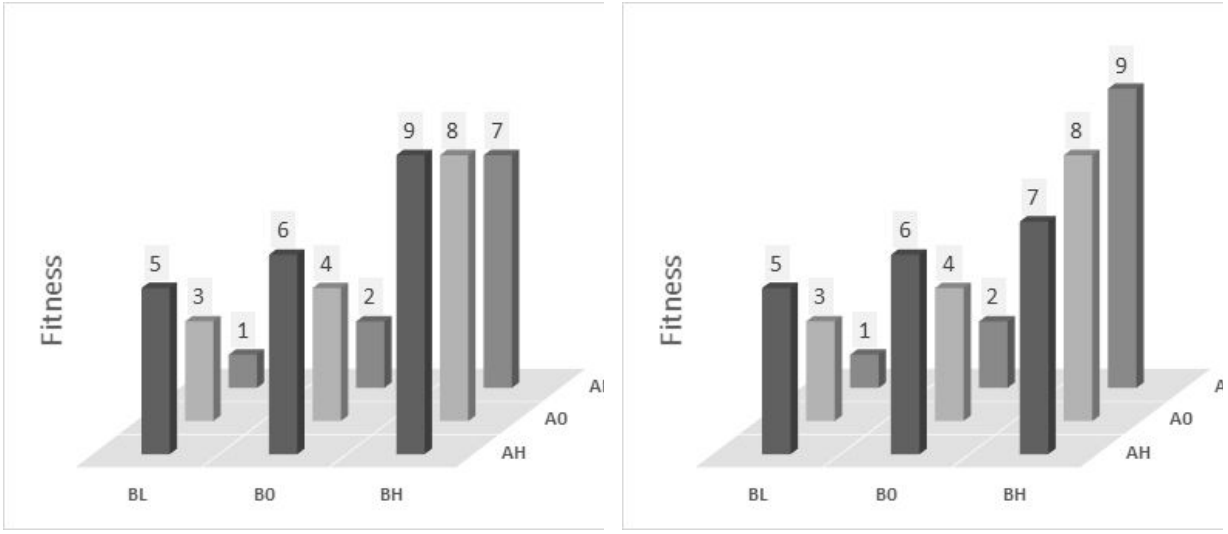

Figure 1: Non-epistatic (left) and Epistatic (right) fitness landscapes.

Mutation Rates

Locus A and Locus B had different mutation rates (mutation probability per generation per individual). Mutations on locus A were more common than mutations on locus B. Deleterious mutations, which lowered the expression of a gene, i.e., A_O to A_L , were more common than beneficial mutations, which enhanced its expression, i.e., A_O to A_H. Mutation rates of locus A and locus B are shown in **figure 2**. Mutation rates were kept such that multiple mutations cannot

occur at the same time. The exact values used in the simulations were as shown in **table 1.**

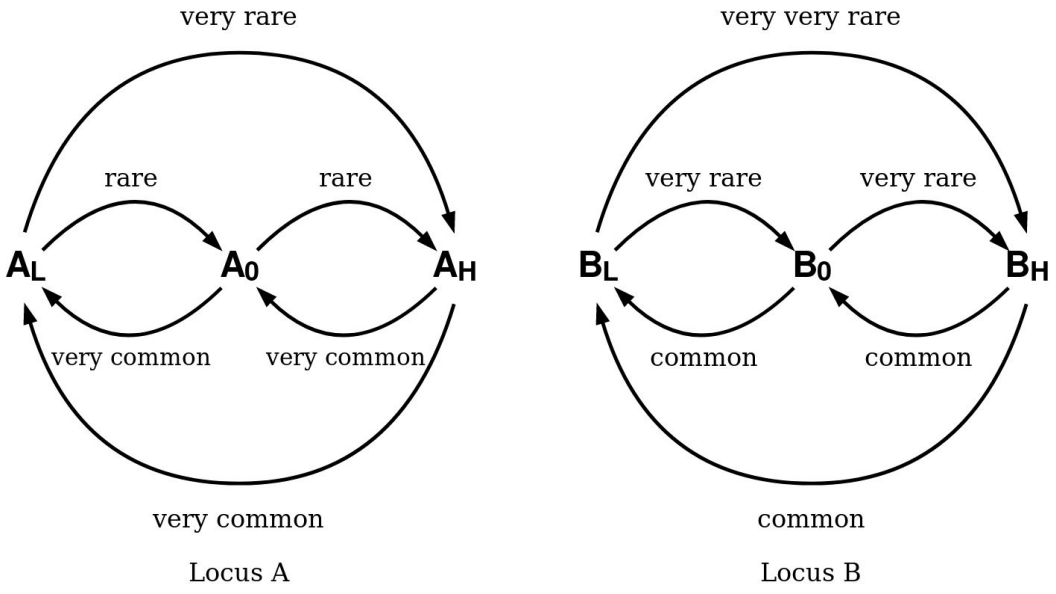

Figure 2: Mutation rates on locus A and locus B.

Mutation rate	Value
very common	$1*10^{-3}$
common	$5*10^{-4}$
rare	$5*10^{-5}$
very rare	$5*10^{-8}$
very very rare	$1*10^{-9}$

Table 1: Mutation rates per generation per individual.

Controls

To ensure that a system with epistasis and differential mutation rates was the minimal setup required to reproduce a similar phenomenon we relaxed them one by one, i.e. we simulated non-epistatic case, both locus with A-like mutation rate case, both locus with B-like mutation case in addition to non-epistatic A-like mutation rates and non-epistatic B-like mutation rates. All the cases mentioned above did not show the proposed phenomenon.

The Law of Rare Events

Imitating a rare event is very costly regarding computations. To determine which individuals will acquire the mutation, one has to sample from a Binomial distribution $\sim B(N,p)$ for each individual every generation, where N is the number of individuals and p is the mutation rate. But when N is large, and p is very small the expected number of total Binomial success can be approximated by a Poisson random variable $\sim Pois(Np)$. This means that one has to sample only once from a Poisson distribution $\sim Pois(Np)$ every generation and pick that many individuals and give them the mutation. Since individuals with the same genotype are indistinguishable one can pick at random.

Results

The results obtained from simulations are summarized **figure 3**. All populations were observed to increase their mean fitness. The genotype A_HB_O got fixed in populations with sizes smaller than 10^5 individuals. Whereas in the population sizes larger than or equal to 10^5 individuals the genotype A_LB_H got fixed. Adapting to the same environment led to two different evolutionary outcomes, i.e., the trait A got enhanced in the small population and in large population it got decayed, based on different population sizes. Small populations achieved fitness value of 6 and large populations achieved fitness value of 9. At population sizes close to 10^5 not all the replicates behave the same.

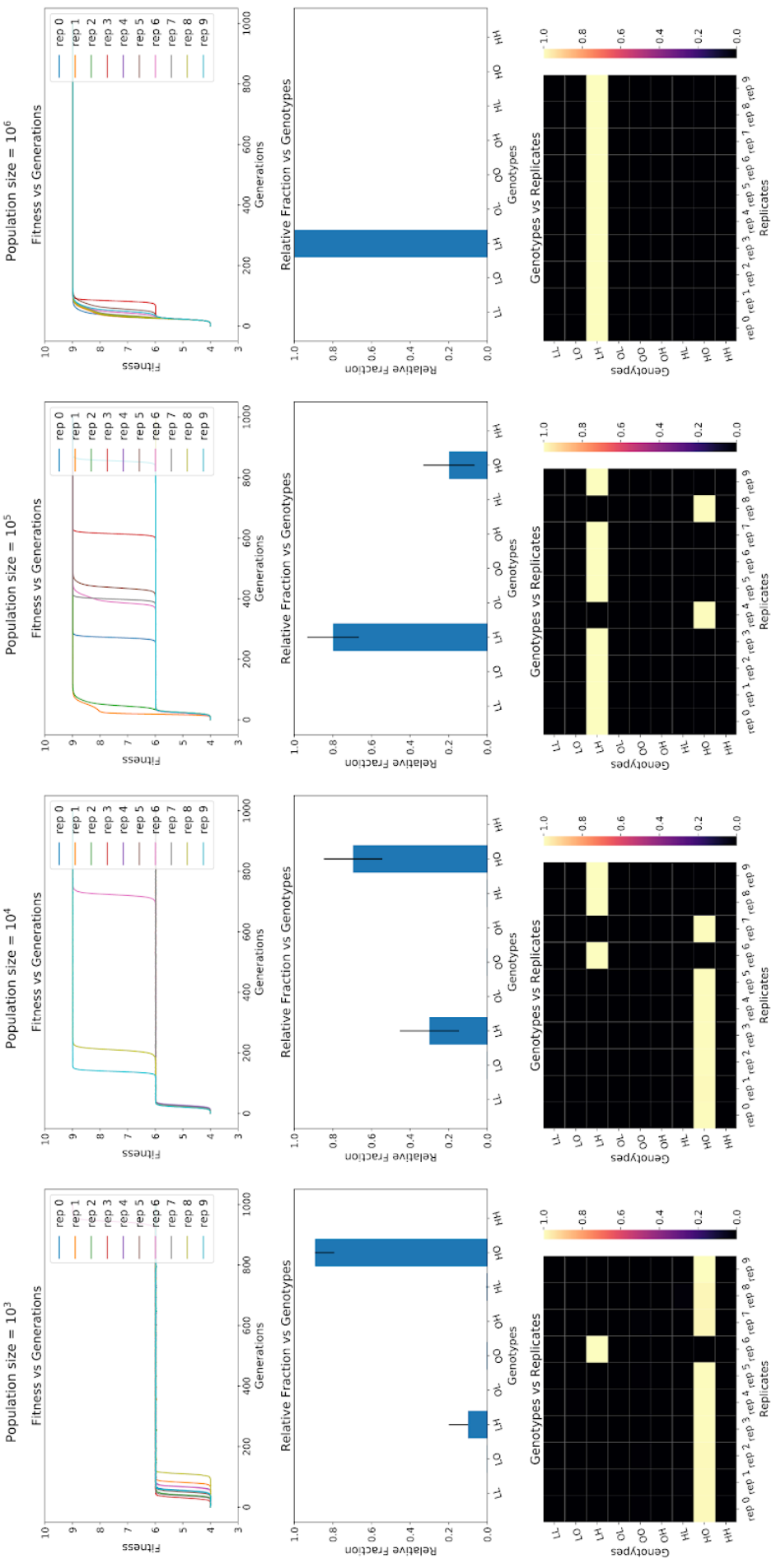

Figure 3

Figure 3: Comparison between different population sizes. Four columns are different population sizes 10^3, 10^4, 10^5 and 10^6, respectively. First row is fitness vs generation of all replicates. Second row is mean relative fraction (±SE) of all the genotypes. Third panel is replicate wise relative fraction of all the genotypes.

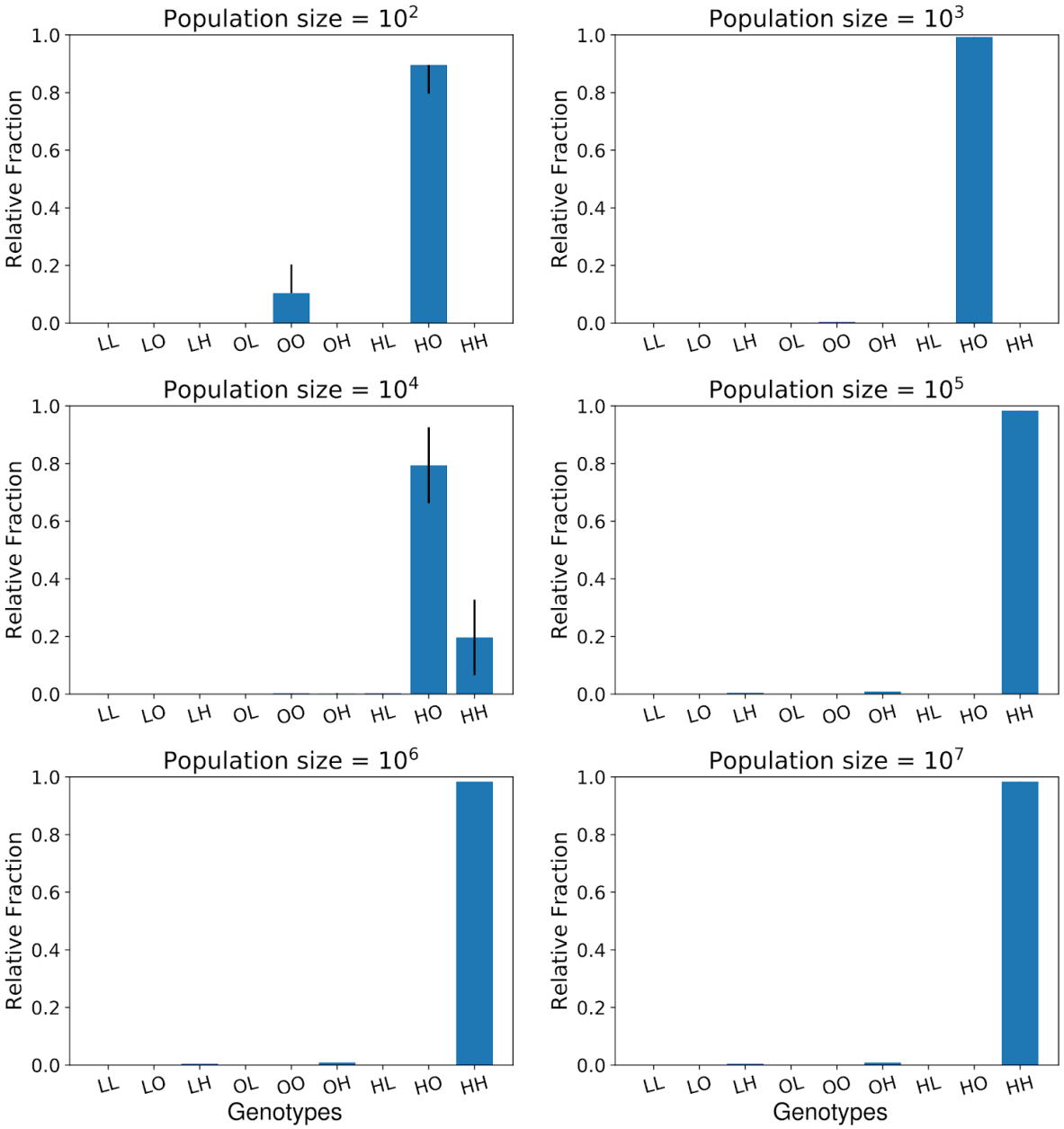

Figure 4: No epistasis control. Relative fractions (±SE) vs genotypes for different population size.

The fitness values of A_HB_H (7) and A_LB_H (9) were switched to collapse epistasis, which resulted in all the population of different sizes having only the A_H allele and the A_L allele was absent from all the populations.

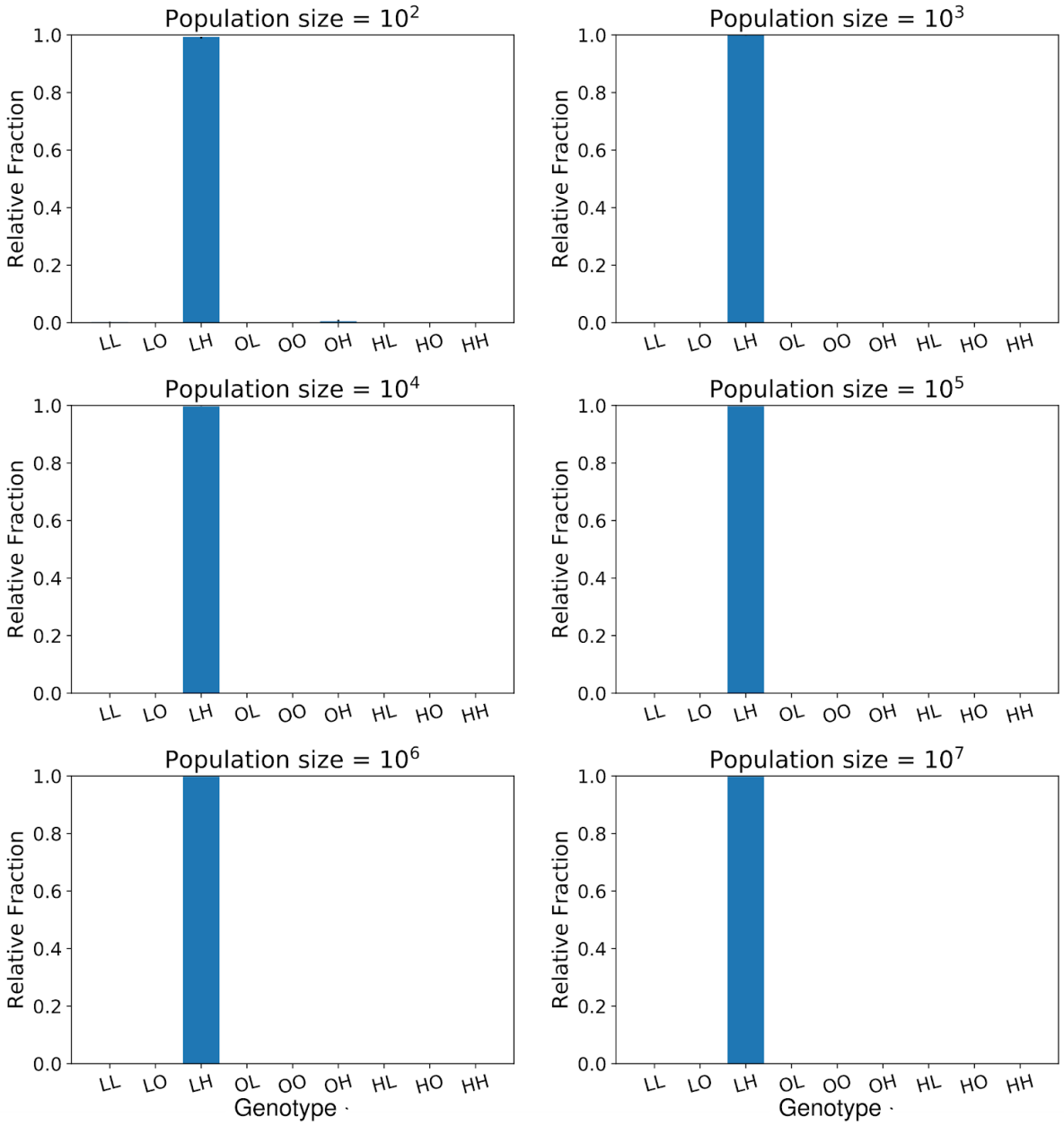

Figure 5: Locus A like mutation rate control. Relative fractions (±SE) vs genotypes for different population size.

The mutation rates at the locus B were changed to have mutation rates exactly like locus A, and this resulted into the B_H allele getting fixed in the population and under the effect of sign epistasis all population got fixed having only the A_L genotype.

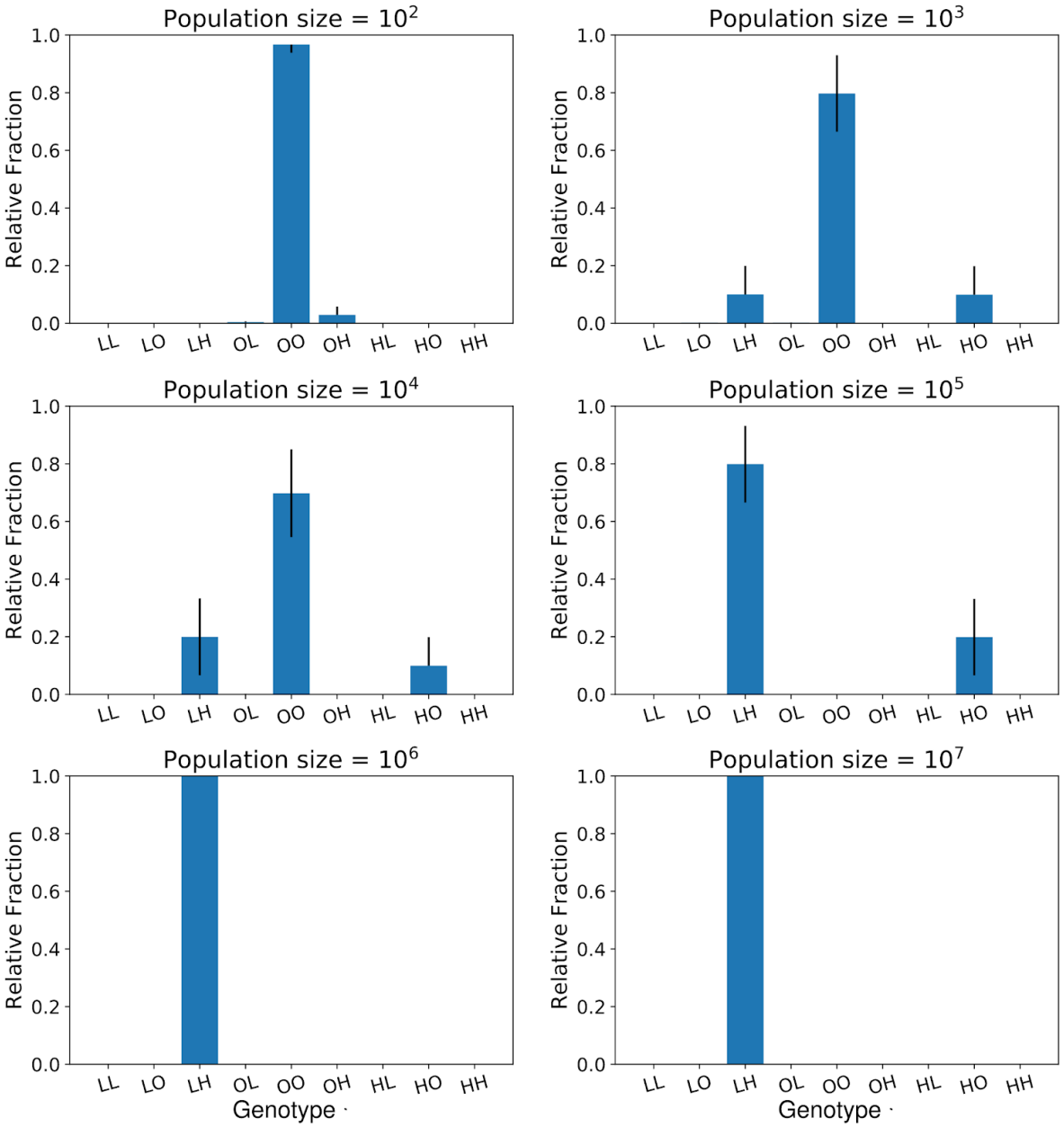

Figure 6: Locus B like mutation rate control. Relative fractions (±SE) vs genotypes for different population size.

The mutation rates at the locus A were changed to have mutation rates exactly like locus B, and this resulted into the B_H allele getting fixed only at very large population sizes, greater than 10^5 individuals. Hence these populations had only the A_L genotype. The smaller populations of this control could not obtain The A_H because the A_O to A_H mutation was made very infrequent. Hence the wild-type genotype was persistent.

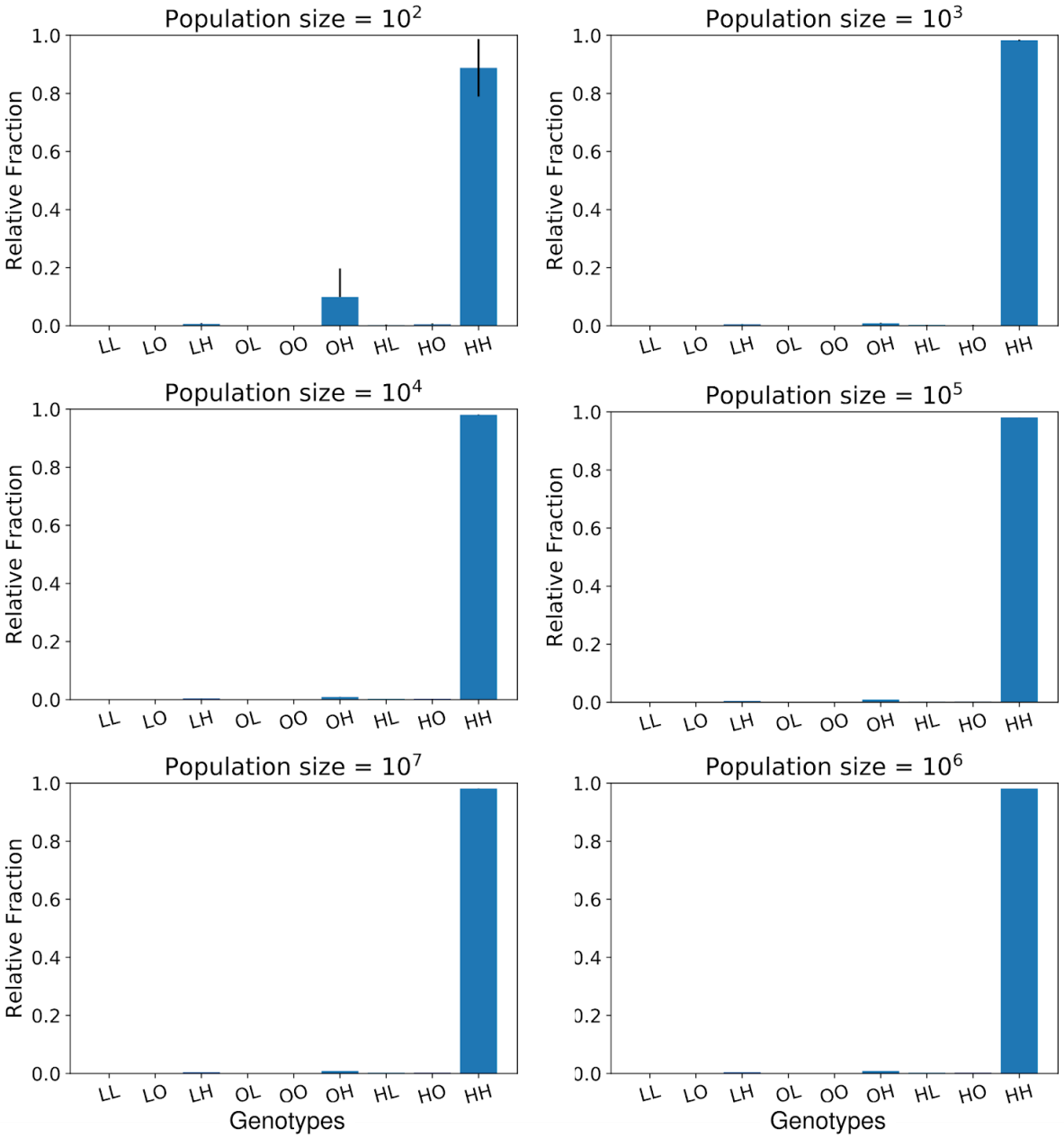

Figure 7: No epistasis with locus A like mutation rate control. Relative fractions (±SE) vs genotypes for different population size.

The absence of epistasis and frequent mutations allowed these population to access the beneficial mutations at lower population sizes, which resulted in best-fit genotype A_HB_H getting fixed in all the populations.

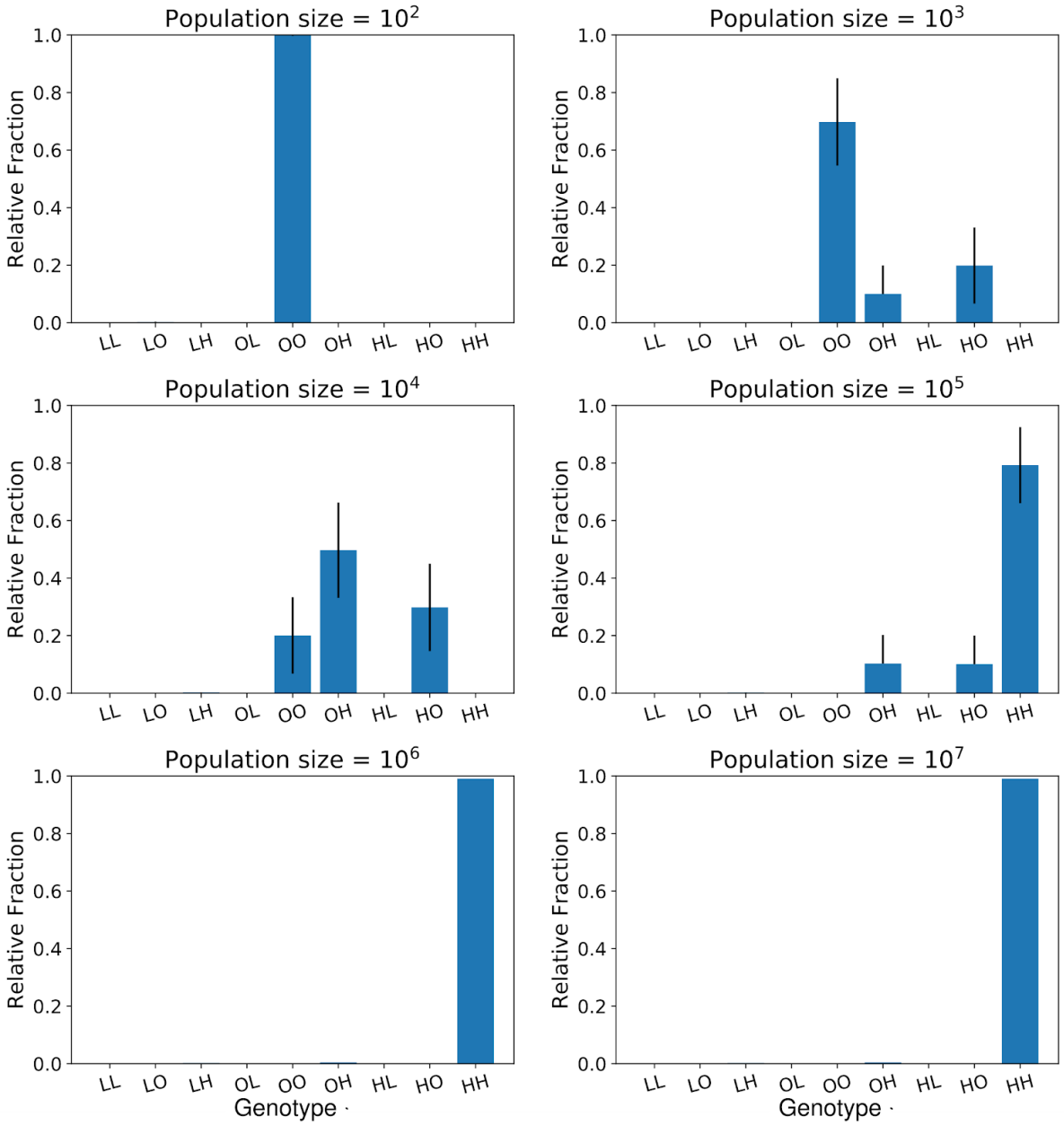

Figure 8: No epistasis with locus B like mutation rate control. Relative fractions (±SE) vs genotypes for different population size.

The absence of epistasis and rare mutations did not allow smaller populations to access the beneficial mutations, which resulted in wild-type being the dominant genotype. Whereas the populations larger than 10^5 individuals could access the beneficial mutations resulting in best-fit genotype A_HB_H getting fixed in the populations.

Discussion

Results obtained by simulation show that it is strikingly possible to gain or lose a trait in the same environment just by varying population size. Given a fitness landscape with sign epistasis and some mutation constraints, this is an apparent outcome but what could have made this possible in case of *E. coli* populations adapting to an environment of an antibiotic cocktail.

One could think of locus A as something that requires significant maintenance cost for its expression and is large in length. On the other hand, locus B can be thought of as a structural protein that does not require much maintenance and is shorter in length. If an Efflux gene was at locus A and some structural protein which is involved in the action of the antibiotic was at locus B, then this would fit the description required for the phenomenon of retroverted selection. Having more efflux is beneficial in antibiotic environments as it would pump the antibiotic out of the cell and hence the phenotype with higher efflux will have higher fitness. Unless some protein at locus B acquires a rare mutation which changes its structure without significant loss in function rendering the antibiotic useless. Now efflux is not required anymore, and cells are paying a cost to maintain this gene which will reduce the fitness of high efflux individuals. This is sign epistasis of locus A in locus B backgrounds, which translates to a fitness landscape just like in **figure 1**.

But this beneficial mutation at locus B is only accessible by a large population because of its rarity. A population without this mutation will take a trajectory which leads to A_H and a population with the beneficial mutation at locus B will take a trajectory which leads to A_L. This is equivalent to saying a small *E. coli* population will gain efflux and a large *E coli.* population will lose efflux.

There was an interesting observation that the small and large both experimental *E. coli* cultures had similar fitness values. This can't be achieved with a two-locus setting, with adding a third locus, similar to locus A, equal fitness for the small and large population can be achieved. Three-locus simulation results are provided in **figure 9** in the supplementary section. Two-locus genotypes were represented by a 2-digit number and the fitness landscape was 2-dimensional as described in the **pseudocode** in the supplementary section. Using a 3-digit number, and a 3-dimensional fitness array this can easily be extended to simulate three-locus cases.

Supplementary Material

Pseudocode

```
****************************************

function pick is
      input       population        P
                  sub_sample_size   K

      output      sub_sample        S

      set S to 1xK array of K random elements selected from P

      return S

****************************************

function count is
      input       population        P

      output      frequencies       F

      set F to 1x9 array of counts of p in P
      // p belongs to ordered array [0,1,2,10,11,12,20,21,22]

      return F

****************************************

function reproduce is
      input       population        P
                  fitness_landscape W

      output      population        P

      replace each element p of P with W(p) many copies in P
      // w(p) is fitness of individual p
      // p belongs to array [0,1,2,10,11,12,20,21,22]

      return P

****************************************
```

```
****************************************

function mutate is
      input       population        P
                  mutation_rates    A, B

      output      population        P

      initialize  number_of_mutations     E, F
      // E and F are 3x3 matrices were matrix element Eij/Fij is the count
of mutation event i -> j on locus A/B

      set Eij to ~Pois(count(ij)*Aij)
      set Fij to ~Pois(count(ij)*Bij)

      //P(1,i) are individuals with allele i at locus A
      //P(2,j) are individuals with allele j at locus B
      pick(P(1,i), Eij) and replace 1st digit i with j
      pick(P(2,i), Fij) and replace 2nd digit i with j

      return P

****************************************

algorithm wf_sims_with_mut is
      input       mutation_rates    A, B
                  fitness_landscape W
                  population_size   N
                  generations       G
                  replicates        R

// alleles [L,O,H] are represented by integers [0,1,2]
// A and B are 3x3 matrices where matrix element Aij/Bij is the rate of
mutation event i -> j on locus A/B
// W is a 3x3 matrix where matrix element Wij is the fitness of genotype ij

      output      frequencies       F

// F is a Rx9 vector where component Fij is the frequency of jth genotype
at the end of ith replicate; j belongs to ordered array
[LL,LO,LH,OL,OO,OH,HL,HO,HH]
      initialize  frequencies       F
      set F to Rx9 array of zeros
```

```
      for r := 1 to r := R step 1 do

            initialize    population      P
            set P to 1xN array of 11

            // each individuals are elements of P
            //   genotypes   [LL,LO,LH,OL,OO,OH,HL,HO,HH]   represented   by
integers [0,1,2,10,11,12,20,21,22])

            for g := 1 to g := G step 1 do

                  P := reproduce(P, W)
                  // updates P for reproduction
                  // size of P changed to sum of fitness of all individuals

                  P := mutate(P, A, B)
                  // updates P for mutation events
                  // size of P stays unchanged

                  P := pick(P,N)
                  //  updates  P  for  random  genetic  drift  (Wright-Fisher
step)
                  // size of P changed to N

                  // end of a generation
            end

            F(r):= count(P)
            // F(r) is rth row of matrix F

            // end of a replicate
      end

      return F

      // end of algorithm
end
```

Three-locus

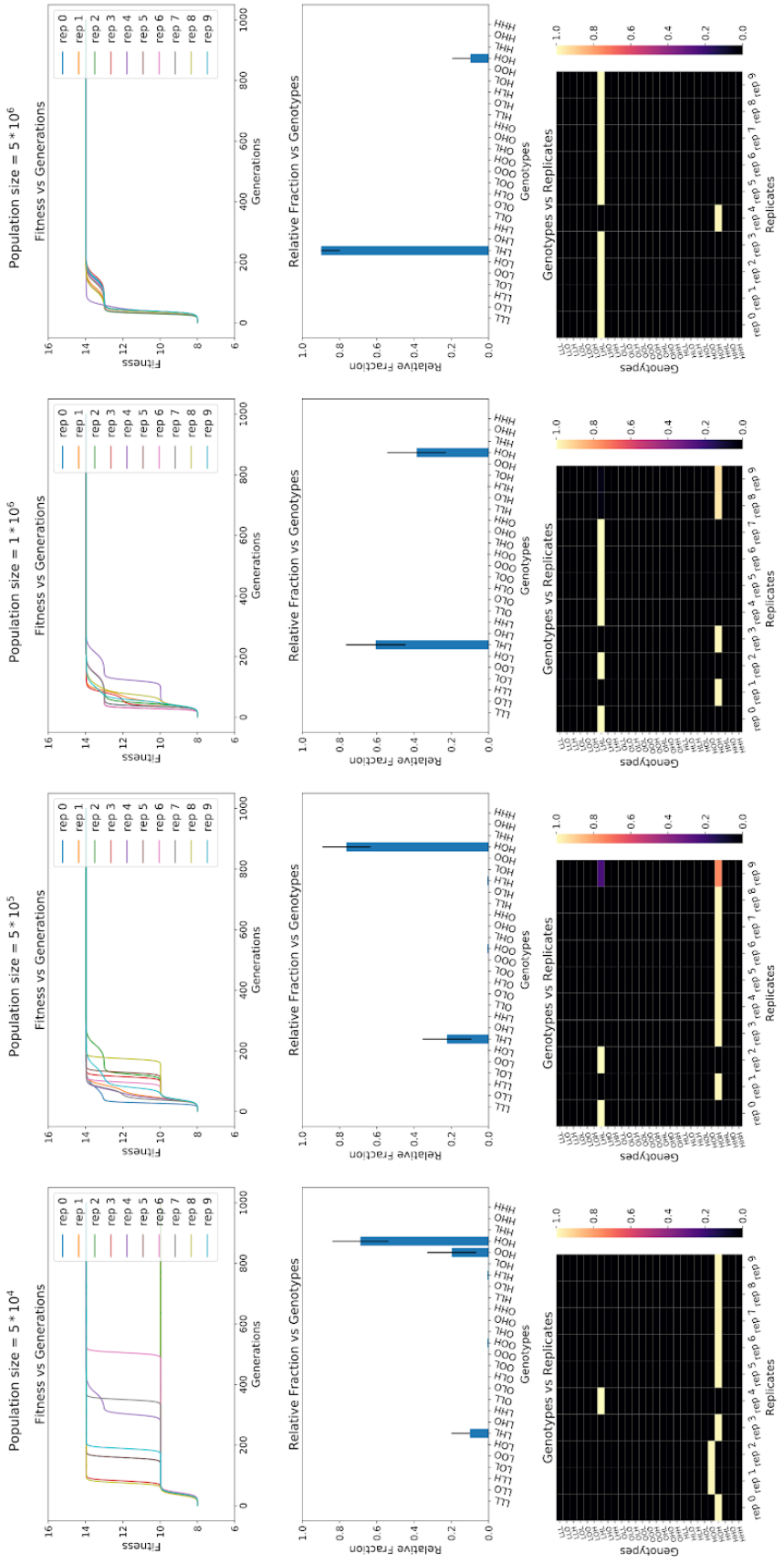

Figure 9

Figure 9: Comparison between different population sizes with three loci. Four columns are different population sizes 10^3, 10^4, 10^5 and 10^6, respectively. First row is fitness vs generation of all replicates. Second row is mean relative fraction (±SE) of all the genotypes. Third panel is replicate wise relative fraction of all the genotypes.

Reference

1. Patten, M. (2010). Levels of Selection. Encyclopedia of Animal Behavior, 272-276. doi:10.1016/B978-0-08-045337-8.00112-1.

2. Raynes, Y., Gazzara, M. R., & Sniegowski, P. D. (2012, 03). Contrasting Dynamics Of A Mutator Allele In Asexual Populations Of Differing Size. *Evolution, 66*(7), 2329-2334. doi:10.1111/j.1558-5646.2011.01577.x

3. Weinreich, D. M., Watson, R. A., & Chao, L. (2005, 06). Perspective: Sign Epistasis And Genetic Constraint On Evolutionary Trajectories. *Evolution, 59*(6), 1165-1174. doi:10.1111/j.0014-3820.2005.tb01768.x

4. Chavhan, Y., Karve, S., & Dey, S. (Under review). Adapting in larger numbers can make populations more vulnerable to environmental changes.