Theoretical Population Biology 81 (2012) 168–178

Contents lists available at SciVerse ScienceDirect

Theoretical Population Biology

journal homepage: www.elsevier.com/locate/tpb

Estimation of the rate and effect of new beneficial mutations in asexual populations

Wei Zhang [a], Vasudha Sehgal [a], Duy M. Dinh [b], Ricardo B.R. Azevedo [b], Tim F. Cooper [b], Robert Azencott [a,*]

[a] Department of Mathematics, University of Houston, Houston, TX 77204-3008, USA

[b] Department of Biology and Biochemistry, University of Houston, Houston, TX 77204-5001, USA

ARTICLE INFO

Article history:
Received 16 December 2010
Available online 3 December 2011

Keywords:
Mutation rate
Adaptive mutations
Experimental evolution
Parameter estimation
Evolutionary simulations
Large deviation theory

ABSTRACT

The rate and effect of available beneficial mutations are key parameters in determining how a population adapts to a new environment. However, these parameters are poorly known, in large part because of the difficulty of designing and interpreting experiments to examine the rare and intrinsically stochastic process of mutation occurrence. We present a new approach to estimate the rate and selective advantage of beneficial mutations that underlie the adaptation of asexual populations. We base our approach on the analysis of experiments that track the effect of newly arising beneficial mutations on the dynamics of a neutral marker in evolving bacterial populations and develop efficient estimators of mutation rate and selective advantage. Using extensive simulations, we evaluate the accuracy of our estimators and conclude that they are quite robust to the use of relatively low experimental replication. To validate the predictions of our model, we compare theoretical and experimentally determined estimates of the selective advantage of the first beneficial mutation to fix in a series of ten replicate populations. We find that our theoretical predictions are not significantly different from experimentally determined selection coefficients. Application of our method to suitably designed experiments will allow estimation of how population evolvability depends on demographic and initial fitness parameters.

1. Introduction

Adaptive evolution is driven by the emergence of beneficial mutations and their subsequent spread due to natural selection. Models of this process have been developed and have demonstrated, for example, the importance of stochastic sampling events in the establishment of beneficial mutations in evolving populations (e.g., Haldane, 1927; Fisher, 1930; Kimura, 1962). A key difficulty in the application of these models is that many of the relevant population genetic parameters – for example, the underlying rate and distribution of the effects of beneficial mutations – are poorly known, limiting the application of models to predict general features of evolutionary dynamics in real populations.

Recent experiments have begun to address this problem. For example, experimental and theoretical analyses of evolving viral populations have been able to test key predictions of models of adaptation (Rokyta et al., 2005, 2008). This work took advantage of the small genome size, high mutation rate and large mutation effect sizes typical of viruses to isolate genotypes that were shown to have single adaptive mutations and that experienced minimal selection through competition between co-occurring lineages. These genotypes were used to directly estimate evolutionary parameters, including the underlying distribution of beneficial mutation effect sizes. Several experiments have been carried out to estimate the effect of beneficial mutations in bacterial populations (Rozen et al., 2002; Barrett et al., 2006). These experiments have led to many insights, but have some limitations. Notably, it is technically difficult to ensure that observed fitness changes are due to single mutations, and to evaluate the effect that interference between competing mutations will have on biasing the fitness effect of the mutations that fix (Rozen et al., 2002). Some experimental designs reduce these issues, but they typically consider adaptation caused by a subset of all available beneficial mutations (Kassen and Bataillon, 2006; MacLean and Buckling, 2009; McDonald et al., 2011). Alternatively, approaches have been developed to infer evolutionary parameters from the dynamics of a marker trait ('marker divergence' experiments: Imhof and Schlotterer, 2001; Rozen et al., 2002; Hegreness et al., 2006; Perfeito et al., 2007; Barrick et al., 2010). The details of these experiments differ, but most have in common the application of some kind of model to infer underlying evolutionary parameters from changes in the frequency of a neutral marker that is linked to newly arising beneficial mutations (Fig. 1).

* Correspondence to: Department of Mathematics, University of Houston, 651 PGH, 4800 Calhoun, Houston, TX 77204-3008, USA.

E-mail addresses: weizhang@math.uh.edu (W. Zhang), vasudha@math.uh.edu (V. Sehgal), ddinh4@uh.edu (D.M. Dinh), razevedo@uh.edu (R.B.R. Azevedo), tfcooper@uh.edu (T.F. Cooper), razencot@math.uh.edu (R. Azencott).

0040-5809/$ – see front matter © 2011 Elsevier Inc. All rights reserved.
doi:10.1016/j.tpb.2011.11.005

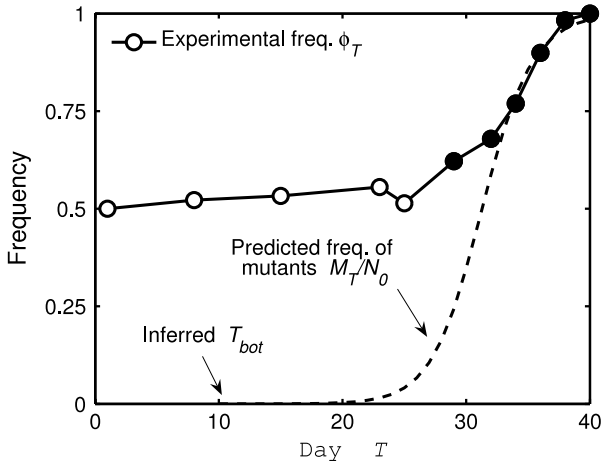

Fig. 1. Marker divergence experiment: empirical observations and inferred parameters. Points indicate the observed frequency of a neutral marker in an *E. coli* population. Changes in frequency are due to the origin and rise of a newly occurring beneficial mutation. To infer the dynamics of this mutation, we determine the timing and rate of deviation in marker frequency (solid points) from its initial baseline (hollow points) and use this information to fit an evolutionary model that estimates the time at which the beneficial mutation first becomes established in the population ($T_{bot} = 10$) and its selective advantage ($s = 0.967 \pm 0.03$, estimate $\pm$ standard deviation). The predicted mutant frequency based on a deterministic approximation using the estimated values of T_{bot} and s is shown as a dashed line. Note that the rate of change in marker frequency is slower than that of the underlying beneficial mutation.

Analyses of marker divergence experiments have revealed several important aspects of the evolutionary dynamics of asexual populations, for example the influential role of interference between competing mutations in determining the distribution of mutations that contribute to adaptation (Rozen et al., 2002; Hegreness et al., 2006; Perfeito et al., 2007; Kao and Sherlock, 2008). Indeed, because marker dynamics can be followed in evolving populations at a relatively high temporal resolution, it is possible to observe early beneficial mutations being outcompeted by later arising mutations of larger effect (Hegreness et al., 2006). In this case, it is possible to infer underlying parameters separately for both mutations. Moreover, marker divergence experiments allow estimation of an effective beneficial mutation rate, which is not usually possible through any direct measure of evolved genotypes.

Here we extend previous work by developing efficient, unbiased estimators of mutational parameters for marker divergence experiments. We apply our method to experimental data collected from populations of *Escherichia coli* and consider the influence of parameters controlled by the experimenter on the accuracy of our estimates of the rate and effect size of newly arising beneficial mutations. In the sections below we first set out the experimental design and develop a stochastic model to account for mutation dynamics in the evolving populations. We then use this model to develop new estimators of key evolutionary parameters. Finally, we evaluate the accuracy of these estimators, both computationally and experimentally.

2. Experimental design

2.1. Strains

The bacterial evolution experiment we focus on in this paper was designed to estimate the rate and selective advantage of newly arising beneficial mutations by matching experimental and simulation dynamics of a linked neutral marker (Hegreness et al., 2006; Barrick et al., 2010). Here, we extend this approach by developing an analytical model that estimates the same parameters directly from observed data. We apply this model to an experiment where we evolved $\mathcal{N} = 11$ replicate *E. coli* populations of effective size $N_e \approx 10^5$. Populations were started from the ancestor of an ongoing evolution experiment, except that half of the cells had an Ara$^+$ marker and half had an Ara$^-$ marker, causing cells to grow as white and red colonies, respectively, on indicator medium (Lenski et al., 1991).

Table 1
Notation.

Parameter	Description (value in our experiment)
A_T	Number of ancestors at the beginning of day T
a_t	Number of ancestors at time t within a day
D	Dilution factor (200)
M_T	Number of mutants at the beginning of day T
m_t	Number of mutants at time t within a day
μ	Mutation rate
$\mathcal{N}$	Number of replicate populations (11)
N_0	Initial population size (5×10^4)
N_{sat}	Saturation population size (10^7)
ν	Logarithm of the mutation rate
P_{bot}	Probability of bottleneck crossing by mutants
ϕ_T	Frequency of winning marker on day T
ρ	Rate of cell division
s	Selective advantage
$\hat{s}$	Estimator of s
$\widetilde{M}_T$	Number of mutants at the end of day T
T	Time among days
T_{bot}	Time of first bottleneck crossing
T_{fix}	Time when ϕ_T reaches 95%
t	Time within a day during the growth phase
t_{sat}	Saturation time within a day
Z_T	Number of new mutants at the end of day T

2.2. Marker divergence experiments

Population growth stopped when nutrients were depleted after ~10 h. Every 24 h, a subpopulation of approximate size N_0 was randomly sampled and transferred to a new culture containing fresh growth medium ($N_0 = N_{sat}/D$, where D is the daily dilution factor and N_{sat} is the size of the population at the time of sampling; Table 1). This transfer step was repeated daily for all $\mathcal{N}$ populations. No cell death occurs between the depletion of nutrients and the transfer time in the experimental conditions we used (Vasi et al., 1994). The frequency of the two cell marker types was recorded periodically by plating cells on indicator media. The impact of measurement error on recorded marker frequencies is neglected in the present paper to facilitate asymptotic theoretical analysis of the mutational parameter estimators.

During population growth a mutant genotype with a selective advantage $s > 0$ can occur with probability μ at each cell division. All individuals are asexual and, therefore, when a fitter genotype fixes in a population, it will drive the ancestral genotype extinct and cause the fixation of the marker type on which it occurred. There is no simple way to directly identify new beneficial mutations in fitter genotypes, making it difficult to identify and follow their dynamics directly. For this reason, we infer the dynamics of beneficial mutations through their effect on the frequency of the observable Ara markers (Fig. 1) The data are available at http://datadryad.org/.

2.3. Fitness assays

Fitness estimates were obtained from four evolved clones of the "winning" marker type isolated from each of the 11 replicate experimental populations. The fitness of each clone was measured relative to the ancestor as described previously except using a GFP-expressing derivative of the ancestor as a reference strain (Lenski