

# MULTI-TIME SCALES IN ADAPTIVE DYNAMICS: MICROSCOPIC INTERPRETATION OF A TRAIT SUBSTITUTION TREE MODEL

ANTON BOVIER AND SHI-DONG WANG

ABSTRACT. We consider a fitness-structured population model with competition and migration between nearest neighbors. Under a combination of large population and rare migration limits we are particularly interested in the asymptotic behavior of the total population partition on supporting trait sites. For the population without mutation on a finite trait space we obtain the equilibrium configuration and characterize the right time scale for fixation. For the model with mutation on an infinite trait space a jump process-trait substitution tree model is established on a rarer mutation time scale against the rare migration constrained in terms of a large population limit. Due to a change of the fitness landscape provoked by a new mutant, some temporarily unfit types can be recovered from time to time. In the end we shed light to illustrate sexual reproduction in a diploid population on the genetic level.

## CONTENTS

|   |    |
|---|----|
| 1. Introduction   | 1  |
| 2. Microscopic model  | 5  |
| 3. Early time window on an finite trait space as $K \rightarrow \infty$ | 6  |
| 4. Late time window with mutations as $K \rightarrow \infty$            | 9  |
| 5. Application to a diploid population                                  | 12 |
| 6. Outline of proofs  | 15 |
| References  | 21 |

## 1. INTRODUCTION

In recent years a spatially structured population with migration (dispersion) and local regulation, proposed by Bolker and Pacala [1], Dieckmann and Law [16] (in short BPDFL process), has attracted particular interest both from biologists and mathematicians. It has several advantages over general branching processes, which make it more natural as population models: the quadratic competition term is used to prevent the population size from escaping to infinity and the migration term is used to transport the population mass

---

*Date:* June 20, 2012.

*2000 Mathematics Subject Classification.* 92D25, 60J85, 37N25, 92D15, 60J75.

*Key words and phrases.* large population, rare migration, rare mutation, time scales separation, trait substitution tree.

A. Bovier is supported in part by the German Research Foundation in the SFB 611 and the Hausdorff Center for Mathematics. S.-D. Wang was supported by a Hausdorff Scholarship while at the University of Bonn, and EPSRC Grant EP/I01361X/1 while at the University of Oxford.

from one colony to unoccupied colonies for survival, and to further get colonized. There are mainly two highlights of related papers. For instance, Etheridge [10], Fournier and Méléard [13], Hutzenthaler and Wakolbinger [15], have studied the extinction and survival problems. Champagnat [3], Champagnat and Lambert [4], Champagnat and Méléard [5], Méléard and Tran [17], Dawson and Greven [8] focus more on its long time behavior by multi-scale analysis methods.

The main ingredient behind this model is logistic branching random walks, that is, a combination of logistic branching populations with spatial random walks (or migration) on trait sites. Under a combination of a large population and rare mutation limits, a so-called trait substitution sequence model (in short TSS) is derived in [3]. The heuristics leading to the TSS model is based on the biological assumptions of large population and rare mutation, and on another assumption that no two different types of individuals can coexist on a long time scale: the selective competition eliminates one of them. On the one hand, coexistence and diversity after entering of new mutants are not allowed due to the deficient spatial structure. On the other hand, natural selection is not only limited to competition mechanism but also is often combined with a survival strategy-migration mechanism. In spite of this heuristics, this model is still lack of a rigorous mathematical basis.

The adaptive-dynamics approach is controversially debated since it was criticized only feasible in the context of phenotypic approach. However, the link with its corresponding genetic insight has rarely been developed (see Eshel [9]). As far as sexual reproduction is concerned, population genetic models have dominated for many years since they have been proved powerful to model diploid populations on the genetic level. For a finite gene pool of fixed size, main evolutionary mechanisms like mutation, selection, and reproduction are theoretically tractable though they can take a role in a very complicated way especially after sexual reproduction gets involved. The effect of sexual reproduction is more complicated to characterize mostly because random shuffling of genes may create many genotypes for natural selection to act on, which makes mathematical analysis more difficult (e.g., see [7] for the case with three genotypes combined by two alleles). In contrast, adaptive-dynamics approach is mainly concerned with the long-term evolutionary property but usually ignores the genetic complications. Is there a way to embody features such as sexual reproduction arising on the genetic level but at the same time in which one can study its long term behavior via the quantitative trait method, i.e., taking advantage of adaptive-dynamics approach on the phenotypic level? This is the biological motivation of this paper.

In this paper we propose a new model to justify the above arguments. We introduce a spatial migration mechanism among possible genotypes, which can be viewed as a result of fusion of any pair of alleles out of a fixed finite allele pool of a diploid population. After natural selection acting on a short-term evolution time scale, the population can attain an equilibrium configuration according to the fitness landscape. Each time there enters a new mutant gene (allele) into the gene pool, the genotype space is enlarged due to formation of new genotypes, and the spatial migration can be used to characterize the reshuffling procedure on the way to a new equilibrium configuration. Loosely speaking, the spatial movement is used to compensate the simplicity of genetic reproduction in

adaptive-dynamics approach. The critical point we need to take care of is to distinguish these different time scales after introducing fitness spatial structure in the model.

The novelty of this model differs from previous models in three key aspects. Firstly, no genetic information is lost on any time scale. Some genotype containing a specific deleterious allele may be invisible due to its temporary low fitness on the migration time scale, but it can recover on a longer mutation time scale due to the entering of a new mutant allele and the reshuffling of the genotype space. For example, some epidemic virus may become popular periodically because of a change of its mutated genetic structure or a genetic change of its potential carrier. Secondly, thanks to the fitness spatial structure endowed on finite genotype space, coexistence is allowed under the assumption of nearest-neighbor competition and migration. This distinguishes our model from the classical adaptive-dynamics, which often converges to a monomorphic equilibrium. What is more, we derive a well-defined branching tree structure in the limiting system, which is like a spatial version of the Galton-Watson branching process. Last but not least, the idea of introducing the spatial migration to interpret sexual reproduction can provide a link between adaptive-dynamics and its genetics counterpart. In particular, similar consideration can be done to quantify more complicated sexual reproduction model than our toy case from the genetics side by mapping it to spatial migration model (see Section 5).

As a reminder, we want to mention some recent progress in the interacting fields of adaptive dynamics and genetics. Champagnat and Méléard [5] relax the assumption of non-coexistence condition in [3] and obtain a polymorphic evolution sequence (PES) as a generalization of the TSS model, allowing coexistence of several traits in the population. However, still unfit allelic traits can be excluded from the evolutionary history and may never recover, depending on the Jacobian matrices of Lotka-Volterra systems. Recently, Collet, Méléard and Metz [7] consider a diploid population model with sexual reproduction, and obtain that population behaves on the mutation time scale as a jump process moving between homozygous states (genotypes comprising of a pair of identical alleles). Although their model puts a rigorous basis on Mendelian diploids, as mentioned in [7, Section 6], it is still under the restriction of an unstructured population and single locus genetics. Evans et al. [6, 12] study a continuous time evolving distribution of genotypes called *mutation-selection balance* model where recombination acts on a faster time scale than mutation and selection. The intuition behind their asymptotic result is that the mutation preserves the Poisson property whereas selection and recombination respectively drive the population distribution away from and toward Poisson. If all three processes are operating together, one expects that recombination mechanism disappears in the limiting system. This in some sense motivates us to specify migration in adaptive-dynamics to express one kind of genetic reshuffling like recombination in Evan's model. And migration should act on some well-defined fitness structure.

In this paper we are interested in the case when the migrant event is rare with respect to branching events but not that rare as in [3] (see Figure 1). In contrast, we assume that there are infinite migrants from a resident population on the natural time scale. Let a parameter  $\epsilon$  be the migration rate and  $K$  be proportional to the initial population size. We will impose the rare migration constraint  $1 \ll K\epsilon \ll K$  on the population (see parameter region II in Figure 1). As far as a finite-trait dynamic system is concerned, to find out the exact fixation time scale expressed in terms of the migration rate and population size is of

particular interest for us. Since the original model is not easy for us to study due to the complicated interactions, we present here a slightly modified model of the one in [3] but retaining the essential machinery founded in the original model. This paper is restricted with nearest-neighbor competitions and migrations along the monotone fitness landscape. What is more, in order to study the long time behavior, we introduce mutations to drive the population to move forwards to more fitter configuration on a rare mutation time scale, which is longer than the fixation time scale. Note that the limit theorem arising in [3] can be applied consistently in the model developed in this paper.

The purpose of this paper and the accompanying one [2] is to justify a trait substitution tree process (in short TST) to illustrate the coexistence phenomenon with spatial structure in evolution theory, which is a purely atomic finite measure-valued process. The present one is derived from the microscopic point of view while the other one [2] is from the macroscopic point of view. Combining these two papers together with [3, 5], the entire framework on (rare) migration against (large) population limit can be fully characterized and it generates different rescaling limits-TSS and TST respectively on different time scales. In summary, the entire framework is as follows:

- Take large population and rare migration simultaneously by  $K\epsilon \ll \frac{1}{\ln K}$ , it leads to a TSS limit in [3].
- Firstly let  $K \rightarrow \infty$ , then add rare mutation by  $\ln \frac{1}{\epsilon} \ll \frac{1}{\sigma}$  as  $\epsilon \rightarrow 0$ , it leads to a TST limit in [2].
- Take large population, rare migration and even rarer mutation all simultaneously constrained by  $1 \ll K\epsilon \ll K$ ,  $\ln \frac{1}{\epsilon} \ll \frac{1}{K\sigma}$ . That is our goal in this paper.

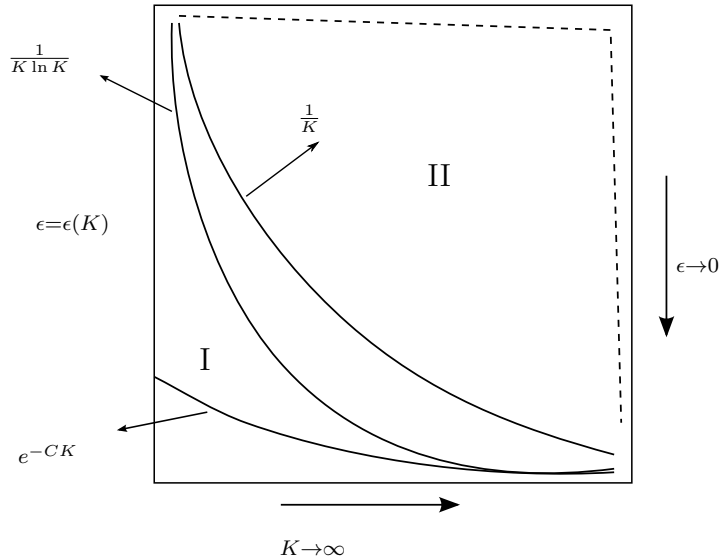


FIGURE 1. Parameter region separation: migration rate  $\epsilon$  against population size  $K$

The remainder of the paper is structured as follows. In Section 2, we present a description of the individual-based model. In Section 3, we consider the case without mutation but on a finite trait space, and characterize the rare migration limit against the

large population limit. In Section 4, concerning a modified population supported on an infinite trait space by introducing mutations, we justify a so-called trait substitution tree processes in the rare mutation limit, which also appeared in [2]. In section 5, we apply the previous results to a diploid population. In the last section, related proofs for results in previous sections are provided.

## 2. MICROSCOPIC MODEL

We begin with a description of an individual-based model. Assume that the population at time  $t$  is composed of a finite number  $I_t$  individuals characterized by their phenotypic traits  $x_1(t), \dots, x_{I_t}(t)$  belonging a compact subset  $\mathcal{X}$  of  $\mathbb{R}^d$ . We denote by  $\mathcal{M}_F(\mathcal{X})$  the set of non-negative finite measures on  $\mathcal{X}$ . Let  $\mathcal{M}_a(\mathcal{X}) \subset \mathcal{M}_F(\mathcal{X})$  be the set of counting measures on  $\mathcal{X}$ :

$$\mathcal{M}_a(\mathcal{X}) = \left\{ \sum_{i=1}^n \delta_{x_i} : x_1, \dots, x_n \in \mathcal{X}, n \in \mathbb{N} \right\}.$$

Then, the population process at time  $t$  can be represented as:

$$\nu_t = \sum_{i=1}^{I_t} \delta_{X_i(t)}.$$

Let  $B(\mathcal{X})$  denote the totality of functions on  $\mathcal{X}$  which are bounded and measurable. For any  $f \in B(\mathcal{X})$ ,  $\nu \in \mathcal{M}_F(\mathcal{X})$ , we use notation  $\langle \nu, f \rangle = \int f d\nu$ .

Let's specify the population process  $(\nu_t)_{t>0}$  by introducing a sequence of demographic parameters:

- $b(x)$  is the birth rate from an individual with trait  $x$ .
- $d(x)$  is the death rate of an individual with trait  $x$  because of "aging".
- $\alpha(x, y)$  is the competition kernel felt by some individual with trait  $x$  from another individual with trait  $y$ .
- $m(x, dy)$  is the migration law of an individual from trait site  $x$  to site  $y$ .
- $\mu(x)$  is the mutation rate of an individual with trait  $x$ .
- $p(x, dh)$  is the law of mutant variation  $h = y - x$  between a mutant  $y$  and its resident trait  $x$ . Since the mutant trait  $y = x + h$  should belong to  $\mathcal{X}$ , this law has its support in  $\mathcal{X} - x := \{y - x : y \in \mathcal{X}\} \subset \mathbb{R}^d$ .

To specify the model without mutation mechanism, the infinitesimal generator of the  $\mathcal{M}_a(\mathcal{X})$ -valued process is given as follows, for any  $F \in B(\mathcal{M}_a(\mathcal{X}))$ :

$$\begin{aligned} LF(\nu) &= \sum_{i=1}^I [F(\nu + \delta_{x_i}) - F(\nu)] b(x_i) \\ &\quad + \sum_{i=1}^I [F(\nu - \delta_{x_i}) - F(\nu)] \left( d(x_i) + \sum_{j \neq i}^I \alpha(x_i, x_j) \right) \\ &\quad + \sum_{i=1}^I \sum_{x_j \neq x_i} [F(\nu + \delta_{x_j} - \delta_{x_i}) - F(\nu)] m(x_i, x_j). \end{aligned} \tag{2.1}$$

The first term above describes the clonal reproduction at the mother's site. The second term describes death of an individual  $x_i$  either due to aging or competition from another individual  $x_j$ . And the last term describes the migration of an individual from trait site  $x_i$  to site  $x_j$ .

By introducing a parameter  $K \in \mathbb{N}$ , we rescale the population size and competition kernel by  $K$ . We will show later, as  $K$  tends to infinity, one can get different large population limits by various well-chosen rescaling procedures. Furthermore, the population process can be parameterized by another parameter  $\epsilon$  governing the rate of migration law  $m(x_i, x_j)$  in terms of population size scaling parameter  $K$ .

For any  $K \in \mathbb{N}$ , instead of studying the above process  $(\nu_t^K)_{t \geq 0}$ , it is more convenient to consider a sequence of rescaled measure-valued processes:

$$X_t^K := \frac{1}{K} \nu_t^K = \frac{1}{K} \sum_{i=1}^{I_t^K} \delta_{x_i} \quad (2.2)$$

where  $X^K$  is a  $\mathcal{M}_F(\mathcal{X})$ -valued process with the following infinitesimal generator:

$$\begin{aligned} L^K F(\nu) &= \int_{\mathcal{X}} \left[ F(\nu + \frac{\delta_x}{K}) - F(\nu) \right] b(x) K \nu(dx) \\ &\quad + \int_{\mathcal{X}} \left[ F(\nu - \frac{\delta_x}{K}) - F(\nu) \right] \left( d(x) + \int_{\mathcal{X}} \alpha(x, y) \nu(dy) \right) K \nu(dx) \\ &\quad + \epsilon \int_{\mathcal{X}} \int_{\mathcal{X}} \left[ F(\nu + \frac{\delta_y}{K} - \frac{\delta_x}{K}) - F(\nu) \right] m(x, dy) K \nu(dx). \end{aligned} \quad (2.3)$$

Notice that we rescale the competition kernel  $\alpha$  by  $K$  so that the system mathematically makes sense when we take a large population limit. From the biological point of view,  $K$  can be interpreted as scaling the resource or area available.

Let us denote by **(A)** the following assumptions.

**(A1):**  $\exists \bar{b}, \bar{d}, \underline{\alpha}, \bar{\alpha}$ , such that  $0 < b(x) \leq \bar{b}$ ,  $0 < d(x) \leq \bar{d}$ ,  $0 < \underline{\alpha} \leq \alpha(x, y) \leq \bar{\alpha}$ , and  $b(x) - d(x) > 0$ ,  $\forall x \in \mathcal{X}$ .

**(A2):**  $\forall x, y \in \mathcal{X}$ ,  $\bar{f}(x, y) \cdot \bar{f}(y, x) < 0$ , where the fitness functions

$$\begin{aligned} \bar{f}(y, x) &= b(y) - d(y) - \alpha(y, x) \bar{n}(x) \text{ and } \bar{n}(x) = \frac{b(x) - d(x)}{\alpha(x, x)}, \\ \bar{f}(x, y) &= b(x) - d(x) - \alpha(x, y) \bar{n}(y) \text{ and } \bar{n}(y) = \frac{b(y) - d(y)}{\alpha(y, y)}. \end{aligned}$$

Notice that assumption (A1) guarantees that the process with infinitesimal generator (2.3) is well defined (refer to [13]). Assumption (A2) gives the non-coexistence condition for any pair of competing populations.

### 3. EARLY TIME WINDOW ON AN FINITE TRAIT SPACE AS $K \rightarrow \infty$

We firstly review some already known results. Champagnat [3, Theorem 1] proved the following result by the time scales separation technique, which can be extended to a more general case in accelerated population dynamics [18].

**Theorem 3.1.** *Admit assumptions (A1) and (A2). Suppose that  $X_0^K = \frac{N_0^K}{K} \delta_x$  such that  $\frac{N_0^K}{K} \xrightarrow{\text{law}} n_0 > 0$  as  $K \rightarrow +\infty$ , and  $\forall C > 0$ ,*

$$\exp\{-CK\} \ll K\epsilon \ll \frac{1}{\ln K}. \quad (3.1)$$

*Then,  $(X_{t/K\epsilon}^K, t \geq 0)$  converges in the sense of f.d.d. to*

$$Y_t = \begin{cases} n_0 \delta_x, & t = 0 \\ \bar{n}(\eta_t) \delta_{\eta_t}, & t > 0 \end{cases}$$

*where the Markov jump process  $(\eta_t, t \geq 0)$  satisfies  $\eta_0 = x$  with an infinitesimal generator:*

$$A\varphi(x) = \int_{\mathcal{X}} (\varphi(y) - \varphi(x)) \bar{n}(x) \frac{[\bar{f}(y, x)]_+}{b(y)} m(x, dy). \quad (3.2)$$

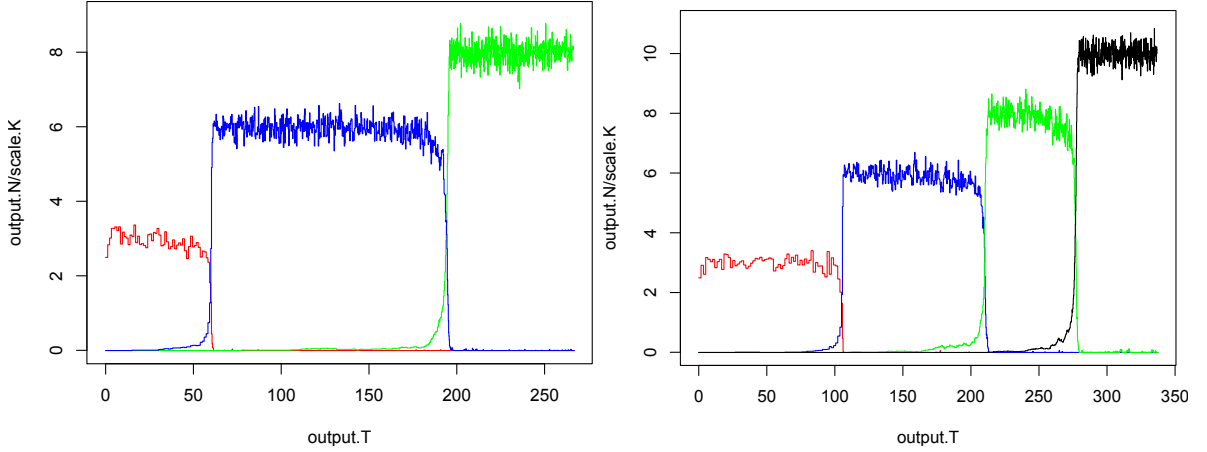


FIGURE 2. Simulations of the trait substitution sequence model arising in Theorem 3.1.

**Remark 3.2.**

- *The migration time scale is of order  $\frac{1}{K\epsilon}$  whereas the fixation time scale starting from one migrant is of order  $\ln K$ . The population is kept monomorphic on the rare migration time scale. The rare migration parameter region constrained by (3.1) is denoted by the region **I** in Figure 1.*
- *As showed in Figure 2, it simulates a TSS model with trait space comprising of three types in the left one while in the right one it simulates a four-type case. We mark the population density of trait  $x_0, x_1, x_2, x_3$  by red, blue, green and black colored curves respectively. Take  $b(x_0) = 3, b(x_1) = 6, b(x_2) = 8, b(x_3) = 10$  and death rates  $d(x_i) \equiv 0, i = 0, 1, 2, 3$ . Take competition kernel  $\alpha \equiv 1$ , migration kernel  $m \equiv 0.5$ , and migration parameter  $\epsilon = K^{-2}$ , where initial population size  $K = 100$ .*

In [2], we firstly let  $K$  tend to infinity in (2.3) and get a deterministic limit. Then, we consider the rescaling limit of the deterministic system supported on a finite trait space under a slow migration limit. That is actually an extreme case where it attains the limiting process (trait substitution tree) by taking sequence limits along the marginal

path (see dashed line in Figure 1). In terms of the individual-based population, it is of particular interest for us to give a microscopic interpretation of the TST process under some suitable conditions. Prior to the following theorem, we give assumption **(B)** as follows to attain our main result.

**(B1):** For any finite number of types  $L \in \mathbb{N}$ , it has a monotonously increasing fitness landscape:  $x_0 \prec x_1 \prec \dots \prec x_L$ , where  $x_0 \prec x_1$  denotes  $\bar{f}(x_0, x_1) < 0$ ,  $\bar{f}(x_1, x_0) > 0$ .

**(B2):** Nearest-neighbor migration and competition, i.e.  $m(x_i, x_j) = \alpha(x_i, x_j) = 0$  for any  $|i - j| > 1$ .

**(B3):** For any  $i \geq 2$ ,

$$\frac{i}{b(x_i) - d(x_i)} \geq \frac{1}{f(x_i, x_{i-1})} + \frac{1}{\bar{f}(x_{i-1}, x_{i-2})} + \dots + \frac{1}{\bar{f}(x_1, x_0)}. \quad (3.3)$$

Note that assumption (B3) is not necessary for us to obtain the following theorem. There actually exist a variety of different possible paths to converge to the equilibrium configuration determined up to the ordered sequence of traits as in assumption (B1). However, thanks to assumption (B3), it brings us a lot convenience to prove the theorem without losing intrinsic features.

We inherit some notations from [2], denote configurations by  $\Gamma^{(L)} := \sum_{i=0}^l \bar{n}(x_{2i})\delta_{x_{2i}}$  if  $L = 2l$  and  $\Gamma^{(L)} := \sum_{i=1}^{l+1} \bar{n}(x_{2i-1})\delta_{x_{2i-1}}$  if  $L = 2l + 1$  for any  $l \in \mathbb{N} \cup 0$ .

**Theorem 3.3.** *Admit assumptions **(A1)** and **B**. Consider the processes  $(X_t^K)_{t \geq 0}$  on the trait space  $\mathcal{X} = \{x_0, x_1, \dots, x_L\}$ . Suppose that  $X_0^K = \frac{N_0^K}{K}\delta_{x_0}$  such that  $\frac{N_0^K}{K} \xrightarrow{\text{law}} n_0 > 0$  as  $K \rightarrow +\infty$ , and*

$$1 \ll K\epsilon \ll K. \quad (3.4)$$

*Then there exists a constant  $\bar{t}_L > 0$ , such that for any  $t > \bar{t}_L$ ,  $\lim_{K \rightarrow \infty} X_{t \ln \frac{1}{\epsilon}}^K \stackrel{(d)}{=} \Gamma^{(L)}$  under the total variation norm.*

**Remark 3.4.**

- We illustrate the theorem by simulations (see Figure 3). We take all the same parameters as in Figure 2 except replacing  $\epsilon = K^{-\frac{4}{5}}$  in the three-type case and  $\epsilon = K^{-\frac{3}{4}}$  in the four-type case, initial population size  $K = 1000$ . Obviously, they both satisfy conditions (3.4).
- The parameter region for rare migration constrained by (3.4) is denoted by the upper right area **II** in Figure 1. As analyzed in Theorem 3.3, the fixation time scale is of order  $\ln \frac{1}{\epsilon}$ . The stable configuration for the three-type case is  $\Gamma^{(2)} = 3\delta_{x_0} + 8\delta_{x_2}$  and it is  $\Gamma^{(3)} = 6\delta_{x_1} + 10\delta_{x_3}$  for the four-type case. We will show in Theorem 4.2 that the TST process jumps from  $\Gamma^{(2)}$  to  $\Gamma^{(3)}$  on an even rarer mutation time scale of order  $\frac{1}{K\sigma}$  (see Figure 4).

#### 4. LATE TIME WINDOW WITH MUTATIONS AS $K \rightarrow \infty$

Following the framework we build up in [2], in order to study the asymptotic behavior on a much longer time scale, we introduce a mutation mechanism into the population



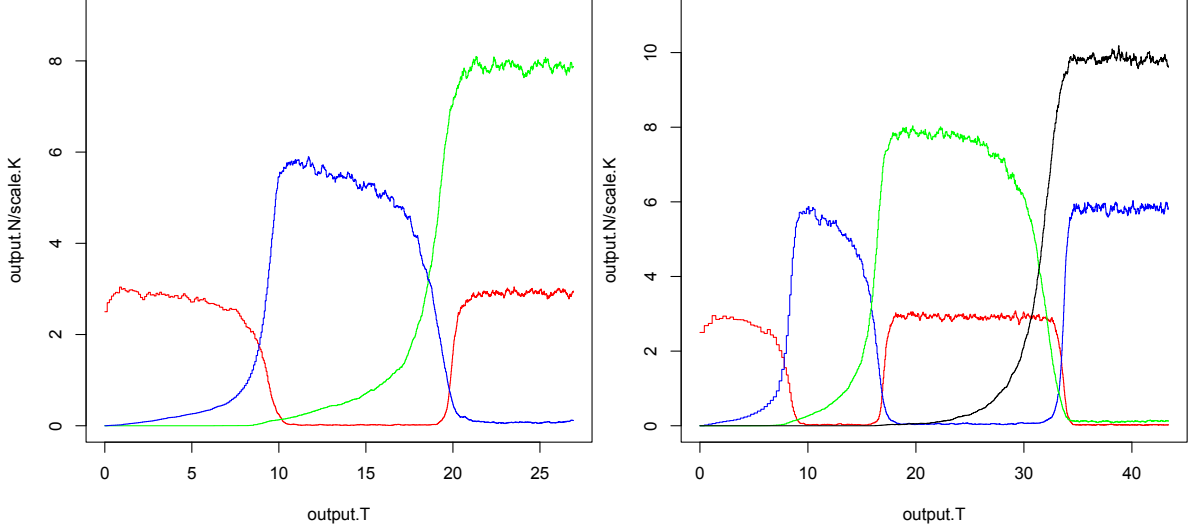


FIGURE 3. Simulations of a trait substitution tree model arising in Theorem 3.3 on a three- and four-type trait space.

generated by (2.3). We now study the model with mutations formulated by the following generator supported on a compact set  $\mathcal{X}$ :

$$\begin{aligned}
 L^{K,\epsilon,\sigma}F(\nu) &= \int_{\mathcal{X}} \left[ F\left(\nu + \frac{\delta_x}{K}\right) - F(\nu) \right] b(x)K\nu(dx) \\
 &+ \int_{\mathcal{X}} \left[ F\left(\nu - \frac{\delta_x}{K}\right) - F(\nu) \right] \left( d(x) + \int_{\mathcal{X}} \alpha(x,y)\nu(dy) \right) K\nu(dx) \\
 &+ \epsilon \int_{\mathcal{X}} \int_{\mathcal{X}} \left[ F\left(\nu + \frac{\delta_y}{K} - \frac{\delta_x}{K}\right) - F(\nu) \right] m(x,dy) 1_{\{y \in \text{supp}\{\nu\}\}} K\nu(dx) \\
 &+ \sigma \int_{\mathcal{X}} \int_{\mathbb{R}^d} \left[ F\left(\nu + \frac{\delta_{x+h}}{K}\right) - F(\nu) \right] \mu(x)p(x,dh)K\nu(dx).
 \end{aligned} \tag{4.1}$$

Here we denote the process by  $X^{K,\epsilon,\sigma}$  with one more superscript  $\sigma$ , distinguishing from the one without mutation in the previous section.

Notice that the mutation kernel  $p(x,dh)$  is used to introduce a new trait site to the previous finite trait space and enlarge the supporting trait space by one each time there enters a mutant, whereas the migration kernel only acts on current support sites of the population. Later we will see, under some rare mutation constraint (with respect to migration rate), the dominating power for fixation is mainly from exponential growth of initial migration particles. Before proceeding towards the main theorem, we now give some assumptions and the definition of the trait substitution tree, which appeared in [2].

Assumption (C).

**(C1):** For any given distinct traits  $\{x_0, x_1, \dots, x_n\} \subset \mathcal{X}, n \in \mathbb{N}$ , there exists a total order permutation

$$x_{n_0} \prec x_{n_1} \prec \dots \prec x_{n_{n-1}} \prec x_{n_n}, \tag{4.2}$$

where  $x \prec y$  means that the fitness functions satisfy  $\bar{f}(x, y) = b(x) - d(x) - \alpha(x, y)\bar{n}(y) < 0$ , and  $\bar{f}(y, x) = b(y) - d(y) - \alpha(y, x)\bar{n}(x) > 0$ .

For simplicity of notation, we always assume  $x_0^{(n)} \prec x_1^{(n)} \prec \dots \prec x_n^{(n)}$  with  $x_i^{(n)} = x_{n_i}$  for  $0 \leq i \leq n$ . By adding a new trait  $x$  whose fitness is between  $x_j^{(n)}$  and  $x_{j+1}^{(n)}$  for some  $0 \leq j \leq n$ , we relabel new traits as following

$$x_0^{(n+1)} \prec x_1^{(n+1)} \prec \dots \prec x_n^{(n+1)} \prec x_{n+1}^{(n+1)}, \quad (4.3)$$

where  $x_i^{(n+1)} = x_i^{(n)}$  for  $0 \leq i \leq j$ ,  $x_{j+1}^{(n+1)} = x$  and  $x_i^{(n+1)} = x_{i-1}^{(n)}$  for  $j+2 \leq i \leq n+1$ . **(C2):** Competition and migration only occurs between nearest neighbors, i.e. for totally ordered traits in (C1), we have  $m(x_i^{(n)}, x_j^{(n)}) = \alpha(x_i^{(n)}, x_j^{(n)}) \equiv 0$  for  $|i-j| > 1$ .

Under above assumptions we can rewrite the generator (4.1) as following

$$\begin{aligned} L^{K, \epsilon, \sigma} F(\nu) &= \int_{\mathcal{X}} \left[ F\left(\nu + \frac{\delta_x}{K}\right) - F(\nu) \right] b(x) K \nu(dx) \\ &+ \int_{\mathcal{X}} \left[ F\left(\nu - \frac{\delta_x}{K}\right) - F(\nu) \right] \left( d(x) + \int_{\mathcal{X}} \alpha(x, y) 1_{\{x^-, x, x^+\}} \nu(dy) \right) K \nu(dx) \\ &+ \epsilon \int_{\mathcal{X}} \int_{\mathcal{X}} \left[ F\left(\nu + \frac{\delta_y}{K} - \frac{\delta_x}{K}\right) - F(\nu) \right] 1_{\{x^-, x^+\}} m(x, dy) K \nu(dx) \\ &+ \sigma \int_{\mathcal{X}} \int_{\mathbb{R}^d} \left[ F\left(\nu + \frac{\delta_{x+h}}{K}\right) - F(\nu) \right] \mu(x) p(x, dh) K \nu(dx) \end{aligned} \quad (4.4)$$

where  $x^-$  and  $x^+$ , specified by the total order relation in assumption (C1), are elements in  $\text{supp}\{\nu\} \subset \mathcal{X}$  satisfying

$$x^- = \sup\{y \in \text{supp}\{\nu\} : \bar{f}(y, x) < 0\}$$

and

$$x^+ = \inf\{y \in \text{supp}\{\nu\} : \bar{f}(y, x) > 0\}.$$

On the migration time scale, there are a variety of different paths to approach the equilibrium configuration by specifying different coefficients. However, the equilibrium configuration of a finite trait system is always the same up to the ordered sequence determined as in assumption (C1) and the time scale for convergence is always of order  $O(\ln \frac{1}{\epsilon})$  as showed in Theorem 3.3.

**Definition 4.1.** A Markov jump process  $\{\Gamma_t : t \geq 0\}$  characterized as following is called a trait substitution tree (in short TST) with the ancestor  $\Gamma_0 = \bar{n}(x_0)\delta_{x_0}$ .

- (i): For any nonnegative integer  $l$ , it jumps from  $\Gamma^{(2l)} := \sum_{i=0}^l \bar{n}(x_{2i}^{(2l)})\delta_{x_{2i}^{(2l)}}$  to  $\Gamma^{(2l+1)}$  with transition rate  $\bar{n}(x_{2k}^{(2l)})\mu(x_{2k}^{(2l)})p(x_{2k}^{(2l)}, dh)$  for any  $0 \leq k \leq l$ , where
- $\Gamma^{(2l+1)} = \sum_{i=1}^j \bar{n}(x_{2i-1}^{(2l)})\delta_{x_{2i-1}^{(2l)}} + \bar{n}(x_{2k}^{(2l)} + h)\delta_{x_{2k}^{(2l)} + h} + \sum_{i=j+1}^l \bar{n}(x_{2i}^{(2l)})\delta_{x_{2i}^{(2l)}}$   
if there exists  $0 \leq j \leq l$  s.t.  $x_{2j}^{(2l)} \prec x_{2k}^{(2l)} + h \prec x_{2j+1}^{(2l)}$ ,

- $\Gamma^{(2l+1)} = \sum_{i=1}^j \bar{n}(x_{2i-1}^{(2l)}) \delta_{x_{2i-1}^{(2l)}} + \sum_{i=j}^l \bar{n}(x_{2i}^{(2l)}) \delta_{x_{2i}^{(2l)}}$   
 if there exists  $0 \leq j \leq l$  s.t.  $x_{2j-1}^{(2l)} \prec x_{2k}^{(2l)} + h \prec x_{2j}^{(2l)}$ .

Then, we relabel the traits according to the total order relation as in (C1):

$$x_0^{(2l+1)} \prec x_1^{(2l+1)} \prec \dots \prec x_{2l}^{(2l+1)} \prec x_{2l+1}^{(2l+1)}, \quad (4.5)$$

where in associate with the first case

$$\begin{aligned} x_i^{(2l+1)} &:= x_i^{(2l)} \text{ for } 0 \leq i \leq 2j, & x_{2j+1}^{(2l+1)} &:= x_{2k}^{(2l)} + h, \\ x_i^{(2l+1)} &:= x_{i-1}^{(2l)} \text{ for } 2j+2 \leq i \leq 2l+1, \end{aligned}$$

and in associate with the second case

$$\begin{aligned} x_i^{(2l+1)} &:= x_i^{(2l)} \text{ for } 0 \leq i \leq 2j-1, & x_{2j}^{(2l+1)} &:= x_{2k}^{(2l)} + h, \\ x_i^{(2l+1)} &:= x_{i-1}^{(2l)} \text{ for } 2j+1 \leq i \leq 2l+1. \end{aligned}$$

(ii): For nonnegative integer  $l$ , it jumps from  $\Gamma^{(2l+1)} := \sum_{i=1}^{l+1} \bar{n}(x_{2i-1}^{(2l+1)}) \delta_{x_{2i-1}^{(2l+1)}}$  to  $\Gamma^{(2l+2)}$

with transition rate  $\bar{n}(x_{2k-1}^{(2l+1)}) \mu(x_{2k-1}^{(2l+1)}) p(x_{2k-1}^{(2l+1)}, dh)$  for any  $1 \leq k \leq l+1$ , where

- $\Gamma^{(2l+2)} = \sum_{i=1}^j \bar{n}(x_{2(i-1)}^{(2l+1)}) \delta_{x_{2(i-1)}^{(2l+1)}} + \bar{n}(x_{2k-1}^{(2l+1)} + h) \delta_{x_{2k-1}^{(2l+1)} + h} + \sum_{i=j+1}^{l+1} \bar{n}(x_{2i-1}^{(2l+1)}) \delta_{x_{2i-1}^{(2l+1)}}$   
 if there exists  $1 \leq j \leq l+1$  s.t.  $x_{2j-1}^{(2l+1)} \prec x_{2k-1}^{(2l+1)} + h \prec x_{2j}^{(2l+1)}$ ,
- $\Gamma^{(2l+1)} = \sum_{i=1}^j \bar{n}(x_{2(i-1)}^{(2l+1)}) \delta_{x_{2(i-1)}^{(2l+1)}} + \sum_{i=j}^{l+1} \bar{n}(x_{2i-1}^{(2l+1)}) \delta_{x_{2i-1}^{(2l+1)}}$   
 if there exists  $1 \leq j \leq l+1$  s.t.  $x_{2j-2}^{(2l+1)} \prec x_{2k-1}^{(2l+1)} + h \prec x_{2j-1}^{(2l+1)}$ .

Then, we relabel the traits according to the total order relation as in (C1):

$$x_0^{(2l+2)} \prec x_1^{(2l+2)} \prec \dots \prec x_{2l+1}^{(2l+2)} \prec x_{2l+2}^{(2l+2)}, \quad (4.6)$$

where in associate with the first case

$$\begin{aligned} x_i^{(2l+2)} &:= x_i^{(2l+1)} \text{ for } 0 \leq i \leq 2j-1, & x_{2j}^{(2l+2)} &:= x_{2k-1}^{(2l+1)} + h, \\ x_i^{(2l+2)} &:= x_{i-1}^{(2l+1)} \text{ for } 2j+1 \leq i \leq 2l+2, \end{aligned}$$

and in associate with the second case

$$\begin{aligned} x_i^{(2l+2)} &:= x_i^{(2l+1)} \text{ for } 0 \leq i \leq 2j-2, & x_{2j-1}^{(2l+2)} &:= x_{2k-1}^{(2l+1)} + h, \\ x_i^{(2l+2)} &:= x_{i-1}^{(2l+1)} \text{ for } 2j \leq i \leq 2l+2. \end{aligned}$$

**Theorem 4.2.** Admit assumption (A1) and (C). Consider the process  $\{X_t^{K,\epsilon,\sigma}, t \geq 0\}$  described by the generator (4.4). Suppose that  $X_0^{K,\epsilon,\sigma} = \frac{N_0^K}{K} \delta_{x_0}$  and  $\frac{N_0^K}{K} \rightarrow \bar{n}(x_0)$  in law as  $K \rightarrow \infty$ . In addition to the condition (3.4), suppose it also holds that

$$\ln \frac{1}{\epsilon} \ll \frac{1}{K\sigma} \ll e^{KC} \quad \text{for any } C > 0. \quad (4.7)$$

Then  $(X_{t/K\sigma}^{K,\epsilon,\sigma})_{t \geq 0}$  converges as  $K \rightarrow \infty$  to the trait substitution tree  $(\Gamma_t)_{t \geq 0}$  defined in Definition 4.1 in the sense of f.d.d. on  $\mathcal{M}_F(\mathcal{X})$  equipped with the topology induced by mappings  $\nu \mapsto \langle \nu, f \rangle$  with  $f$  a bounded measurable function on  $\mathcal{X}$ .

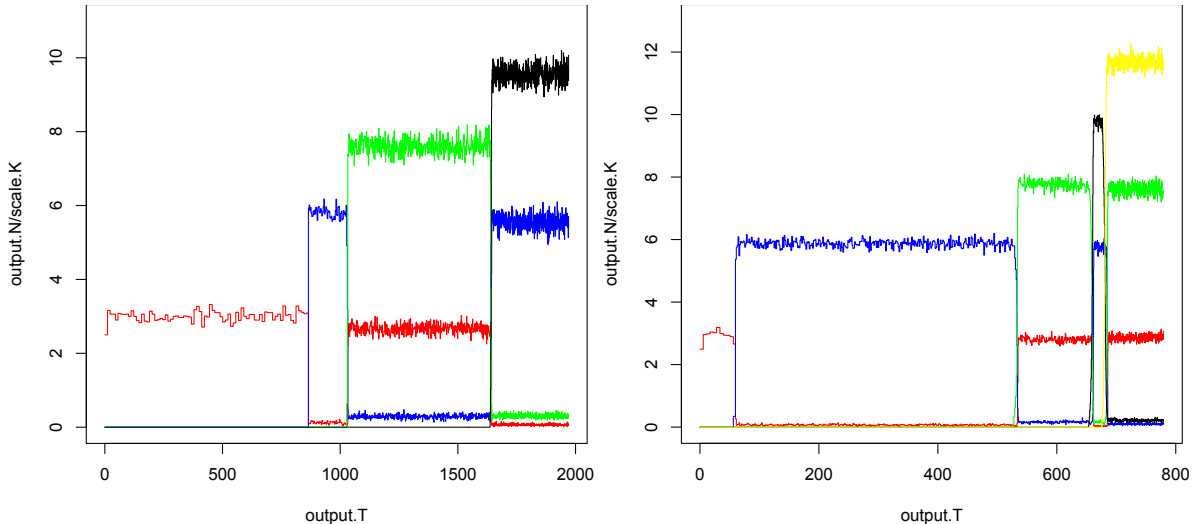


FIGURE 4. Simulations of a trait substitution tree on the mutation time scale arising in Theorem 4.2 on four- and five-type trait space.

- Remark 4.3.**
- There are two time scales for the individual-based population, which can be observed from Theorem 3.3 and the generator (4.4). One is the fixation time scale of order  $\ln \frac{1}{\epsilon}$  while the other one is the mutation time scale of order  $\frac{1}{K\sigma}$ , which are constrained on LHS of the inequality (4.7). By adopting the time scales separation technique used in [3], we can get a nice limiting structure-TST in the large population limit. The RHS of the inequality (4.7) is used to guarantee that system can not drift away from the TST equilibrium configuration on the mutation time scale (see Freidlin and Wentzell [14]).
  - As showed in Figure 4, we simulate the trait substitution tree processes by introducing a mutation mechanism. Note that the simulation shows a special case where the population always reproduces a mutant which is more fitter than any of already existing traits. The birth rate of red-colored population is 3, while the blue one, the green one, the black one and the yellow one have birth rates 6, 8, 10, 12 resp.. Their death rates are constant 0. We take  $\epsilon = K^{-0.8}$  and  $\sigma = K^{-1.5}$ , where initial scaling parameter  $K = 400$ . On a longer mutation time scale, the fixation process due to migration is not visible any more. However, if we zoom into the infinitesimal fixation period, we get the picture as in Figure 3.

## 5. APPLICATION TO A DIPLOID POPULATION

We begin by introducing some terminology from population genetics in the same setting as [11, Chapter 10.1].

In this section we restrict our attention on one-locus diploid populations, where the chromosomes occur in the form of homologous pairs. More precisely, an individual's

genetic makeup with respect to a particular locus, as indicated by the unordered pair of alleles situated there (one on each chromosome), is referred to be its genotype. Therefore, if there are  $n$  possible alleles, i.e. finite allele space  $\mathcal{A}^n := \{A_1, \dots, A_n\}$ , at a given locus, they can fuse  $g(n) := C_n^2 + n = n(n+1)/2$  possible genotypes. We denote the genotype space by  $G^n := \{(A_i, A_j), 1 \leq i \leq j \leq n\}$ . Without loss of generality one can take the total allele space  $\mathcal{A}$  to be a compact subset of  $\mathbb{R}^d$ , or simply a continuum interval  $[0, 1]$ . Denote the total genotype space by  $G$ . To keep consistent with notations in Section 2, we can endow quantitative trait value for every different genotype by the following measurable mapping  $\Phi : G^n \mapsto \mathcal{X}$  such that

$$\Phi((A_i, A_j)) = x_l \quad \text{for } 1 \leq l \leq g(n), 1 \leq i \leq j \leq n \quad (5.1)$$

where  $\Phi$  is symmetric for  $i, j$ , and traits  $(x_l)_{1 \leq l \leq g(n)}$  are ordered according to their relative fitness as in assumption (C1). In accordance with notations in Section 2, we list the following parameters for diploid populations. For any  $g \in G$ ,

- denote by  $B(g) := b(\Phi(g))$  the birth rate from an individual of genotype  $g$ .
- denote by  $D(g) := d(\Phi(g))$  the death rate of an individual of genotype  $g$  because of “aging”.
- denote by  $\hat{\alpha}(g_1, g_2) := \alpha(\Phi(g_1), \Phi(g_2))$  the competition kernel felt by an individual of genotype  $g_1$  from another of genotype  $g_2$ . We restrict the competition acting on individuals of the same genotype and within the nearest-neighbors (ordered in terms of their fitness values).
- denote by  $\hat{m}(g_1, dg_2)$  the replacement law of an individual of genotype  $g_1$  by another  $g_2$  when one of allele pair of the mother individual  $g_1$  undergoes fusion with another gamete allele from any father to form a new type  $g_2$ . Thus,  $g_2$  is chosen to be any genotype fused in a way that one allele is from  $g_1$  and the other one can be anyone from the whole supporting allele space.
- denote by  $\hat{\mu}(g) := \mu(\Phi(g))$  the mutation rate of an individual of genotype  $g$ .
- denote by  $\hat{p}(A, da)$  the law of mutant variation between a mutant and its resident allele type  $A$ .

We also limit our discussion to monoecious populations, those in which each individual can act as either a male or a female parent. The reproductive process can be briefly described as follows.

**Sexual reproduction.** Sexual reproduction is a creation of a new organism by combining genetic gametes from two parental individuals. Suppose each individual can produce a large amount of germ cells, cells of the same genotype. Some germ cells split into two gametes (each carries one chromosome information from the homologous pair in the original cell), and any two gametes from different individuals form another genotype which is different from the parents. The above procedure is called meiosis and fusion of gametes. Note that sexual reproduction can be realized in a manner of gamete replacement of one of the mother individual’s gamete. For instance, the effect of reproduction from a (mother) individual of genotype  $(A_i, A_j)$  by fusing with any gamete  $A_k$  generated from any other (father) individual is a replacement of the individual  $(A_i, A_j)$  by a new individual of genotype either  $(A_i, A_k)$  or  $(A_k, A_j)$ . We assume that each father individual can generate enough amount of germ cells to provide gametes for the replacement procedure of mother individuals. The size of offspring individuals resulted from sexual reproduction

only depends on the size of replaced mother individuals. Based on the above simplicity, one can think of sexual reproduction as a spatial migration of an individual from a genotypic site  $(A_i, A_j)$  (say  $g_1$ ) to another genotypic site  $(A_i, A_k)$  or  $(A_k, A_j)$  (say  $g_2$ ) with migration kernel  $\hat{m}(g_1, dg_2)$ , and the migration rate only depends on the size density of the (mother) genotype  $(A_i, A_j)$ . Then we can employ the results on the spatial migration model in Section 3 and Section 4, which follows later. Here the sexual reproduction is based on allelic level but is represented on genotypic level.

**Clonal birth and death.** Besides the sexual reproduction, meanwhile, there are also reproductive birth which does not apply fusion of gametes from two parents. Instead, an offspring carries a clonal copy of the parent's genotype  $g$  with birth rate  $B(g)$ . This kind of local birth is carried out as defined in Section 2. Similarly as before, the death of an individual is governed by a quadratic form of its own size density and the density of its nearest-neighbors. These two events are based on genotypic level.

**Mutation.** Mutation occurs due to a change of one of alleles in a genotype, and the allele space is enlarged by one. Suppose there are  $n$  different allele types before a mutation event. After mutation, the size of the genotype space is enlarged to be of size  $C_{n+1}^2 + n + 1$  from  $C_n^2 + n$ . The mutation event is governed by a mutation kernel  $\hat{p}(A, da)$  on the allele space  $\mathcal{A}$ . The mutation event is based on allelic level.

Notice that all the above events are density dependent, which means their transition rates are proportional to the local density of the population size. We consider it as finite measure-valued processes  $(\mathbf{g}_t)_{t \geq 0}$  on the genotype space  $G$ . Denote by  $\mathcal{M}_F(G)$  the totality of finite measures on  $G$ . Now we write down the infinitesimal generator of the diploid model with  $1/K$ -scaled weight for a proper testing function  $F$

$$\begin{aligned}
 \mathcal{L}^K F(\mathbf{g}) &= \int_G \left[ F\left(\mathbf{g} + \frac{\delta_g}{K}\right) - F(\mathbf{g}) \right] B(g) K \mathbf{g}(dg) \\
 &+ \int_G \left[ F\left(\mathbf{g} - \frac{\delta_g}{K}\right) - F(\mathbf{g}) \right] \left( D(g) + \int_G \hat{\alpha}(g, \tilde{g}) \mathbf{g}(d\tilde{g}) \right) K \mathbf{g}(dg) \\
 &+ \epsilon \int_G \int_G \left[ F\left(\mathbf{g} + \frac{\delta_{g_2}}{K} - \frac{\delta_{g_1}}{K}\right) - F(\mathbf{g}) \right] \hat{m}(g_1, dg_2) 1_{\{g_2: g_1 \cap g_2 \neq \emptyset; g_2 \in \text{supp}\{\mathbf{g}\}\}} K \mathbf{g}(dg_1) \\
 &+ \sigma \int_G \int_{\mathcal{A}} \left[ F\left(\mathbf{g} + \frac{\delta_{(A_1+a, A_2)}}{K}\right) - F(\mathbf{g}) \right] \delta_g((A_1, A_2)) \hat{\mu}(g) \hat{p}(A_1, da) K \mathbf{g}(dg)
 \end{aligned} \tag{5.2}$$

where  $g_1 \cap g_2 \neq \emptyset$  means that genotype  $g_1$  and  $g_2$  share at least one same allelic type in their allele pairs.

**Definition 5.1.** A Markov jump process  $(\Delta_t)_{t \geq 0}$  with initial  $\Delta_0 = n_0 \delta_{(A_1, A_1)}$  for some  $A_1 \in \mathcal{A}$  is called a genotype substitution tree if it satisfies the follows. For any given  $n \in \mathbb{N}$  and  $\mathcal{A}^n = \{A_1, A_2, \dots, A_n\}$ , let  $\Gamma^{(g^{(n)})}$  (defined in Definition 4.1) be the equilibrium configuration for traits distribution determined by genotypes in  $G^n = \{(A_i, A_j) : 1 \leq i \leq j \leq n\}$  and (5.1). Let  $\Delta^{(n)}$  be  $\Gamma^{(g^{(n)})}$ 's corresponding equilibrium on the genotype space. Suppose that the number  $g^{(n)}$  is even, the transition rate for the Markov jump process  $(\Delta_t)$  from  $\Delta^{(n)}$  to  $\Delta^{(n+1)}$  due to mutation variation of magnitude  $a$  on an allele  $A_j$  ( $1 \leq j \leq n$ )

is

$$\sum_{i=1}^{g(n)/2} \bar{n}(x_{2i}) \hat{\mu}(x_{2i}) (1 + \delta_j(k)) 1_{\{x_{2i} = \Phi((A_j, A_k)), \exists 1 \leq k \leq n\}} \hat{p}(A_j, da) \quad (5.3)$$

where  $(x_i)_{1 \leq i \leq g(n)}$  are indexed according to an increasing order of fitness values of  $\{\Phi((A_i, A_j)) : 1 \leq i \leq j \leq n\}$ .

Note that  $(1 + \delta_j(k))$  appeared in above rate is used to distinguish cases whether the mutant allele is from a homozygous genotype or not.

**Theorem 5.2.** Consider a sequence of processes  $\{(\mathbf{g}_t^K)_{t \geq 0}, K \in \mathbb{N}\}$  defined by above generator with initial condition  $\mathbf{g}_0^K = n_K \delta_{(A_1, A_1)}$ , where  $n_K$  converges to  $n_0$  as  $K \rightarrow \infty$ . If mutation and migration rates satisfy constraints (3.4) and (4.7), as  $K \rightarrow \infty$ ,  $(\mathbf{g}_{\frac{t}{K\sigma}}^K)$  converges to a genotype substitution tree process  $(\Delta_t)$  in the sense of f.d.d. on  $\mathcal{M}_F(G)$  equipped with the topology induced by mappings  $\mathbf{g} \mapsto \langle \mathbf{g}, f(\Phi(\cdot)) \rangle$  with  $f$  a bounded measurable function on  $\mathcal{X}$ .

The above theorem can be obtained as a corollary of Theorem 4.2.

## 6. OUTLINE OF PROOFS

In order to illustrate the basis idea of proofs, we start with a three-trait toy model. But notice that our analysis is not reduced only to the three-trait case. All the machinery is still available for any finite-trait space, which will be shown later. However, the explicit proofs are more difficult to write down without some restrictive conditions. That is why we impose assumption **(B3)** in Theorem 3.3.

**Proposition 6.1.** Admit the same condition as in Theorem 3.3. Consider a sequence of processes on a trait space  $\mathcal{X} = \{x_0, x_1, x_2\}$ . Then, there exists a constant  $\bar{t}_2 > 0$ , such that for any  $t > \bar{t}_2$

$$\lim_{K \rightarrow \infty} X_{t \ln \frac{1}{\epsilon}}^K \stackrel{d}{=} \Gamma^{(2)} \quad (6.1)$$

under the total variation norm.

*Proof.* (see Figure 5).

Let  $\xi_t^K(x_0) := \frac{N_t^K}{K}$  and  $\xi_t^K(x_i) := \frac{N_t^{K,i}}{K} = \langle X_t^K, 1_{\{x_i\}} \rangle$  for  $i = 1, 2$ .

**Step 1.** Firstly, consider the emergence and growth of population at trait site  $x_1$ . Set  $S_1^\epsilon = \inf\{t > 0 : \xi_t^K(x_1) \geq \epsilon\}$ . Thanks to  $\frac{N_0^K}{K} \rightarrow n_0 > 0$  in law as  $K \rightarrow \infty$  and by applying the law of large numbers of random processes (see Chap.11, Ethier and Kurtz 1986), one obtains from the last term in generator (2.3) that, for any  $\delta > 0$ ,  $T > 0$ ,

$$\lim_{K \rightarrow \infty} \mathbb{P} \left( \sup_{0 \leq t \leq T} \left| \frac{\xi_t^K(x_1)}{\epsilon} - n_t(x_1) \right| < \delta \right) = 1$$

where  $n_t(x_1)$  is governed by equation  $\dot{n}(x_1) = m(x_0, x_1)n_0$  with initial  $n_0(x_1) = 0$ . Therefore,

$$\lim_{K \rightarrow \infty} \mathbb{P} \left( \frac{1}{m(x_0, x_1)n_0} - \delta < S_1^\epsilon < \frac{1}{m(x_0, x_1)n_0} + \delta \right) = 1, \quad (6.2)$$

that is,  $S_1^\epsilon$  is of order 1.

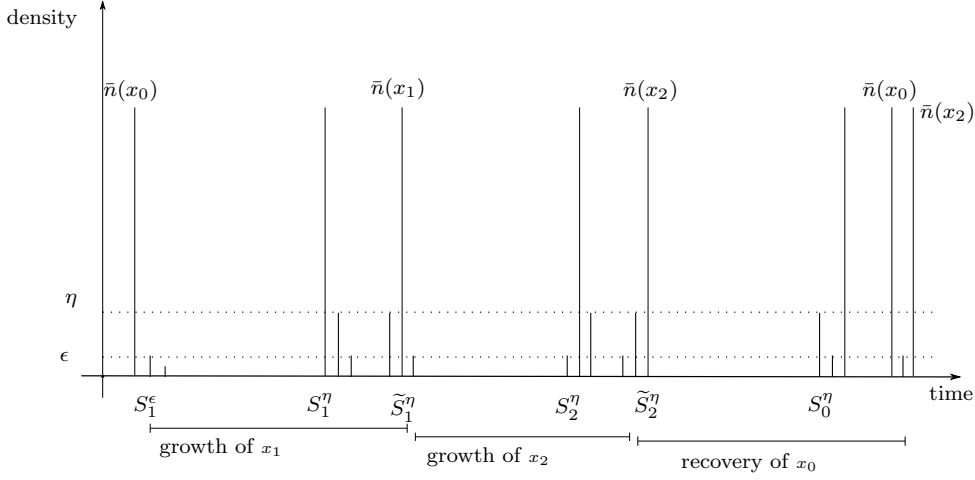


FIGURE 5. Phase evolution of mass bars in early time window on the three-trait site space

For any  $\eta > 0$ , set  $S_1^\eta = \inf\{t : t > S_1^\epsilon, \xi_t^K(x_1) \geq \eta\}$ . Consider a sequence of rescaled processes  $\left(\frac{N_t^{K,1}}{K\epsilon}\right)_{t \geq S_1^\epsilon}$  with  $\frac{N_{S_1^\epsilon}^{K,1}}{K\epsilon} = \frac{\xi_{S_1^\epsilon}^K(x_1)}{\epsilon} \rightarrow 1$  as  $K \rightarrow \infty$ . As before, by law of large numbers of random processes (see Chap.11 Ethier and Kurtz 1986), one obtains, for any  $\delta > 0, T > 0$ ,

$$\lim_{K \rightarrow \infty} \mathbb{P} \left( \sup_{0 \leq t \leq T} \left| \frac{N_t^{K,1}}{K\epsilon} - m_t \right| < \delta \right) = 1, \quad (6.3)$$

where  $m_t$  is governed by equation  $\dot{m} = \bar{f}(x_1, x_0)m = (b(x_1) - d(x_1) - \alpha(x_1, x_0)\bar{n}(x_0))m$  with  $m_0 = 1$ .

Set  $T_1^{\eta/\epsilon} = \inf\{t - S_1^\epsilon : t > S_1^\epsilon, \frac{N_t^{K,1}}{K\epsilon} \geq \eta/\epsilon\}$ , and  $t_1^{\eta/\epsilon} = \inf\{t > 0 : m_t \geq \eta/\epsilon\}$ . Then, for any  $\delta > 0$ , there exists  $\delta' > 0$  such that

$$\begin{aligned} & \lim_{K \rightarrow \infty} \mathbb{P} \left( \left( \frac{1}{\bar{f}(x_1, x_0)} - \delta \right) \ln \frac{1}{\epsilon} < S_1^\eta - S_1^\epsilon < \left( \frac{1}{\bar{f}(x_1, x_0)} + \delta \right) \ln \frac{1}{\epsilon} \right) \\ &= \lim_{K \rightarrow \infty} \mathbb{P} \left( \left( \frac{1}{\bar{f}(x_1, x_0)} - \delta \right) \ln \frac{1}{\epsilon} < T_1^{\eta/\epsilon} < \left( \frac{1}{\bar{f}(x_1, x_0)} + \delta \right) \ln \frac{1}{\epsilon} \right) \\ &= \lim_{K \rightarrow \infty} \mathbb{P} \left( \left( \frac{1}{\bar{f}(x_1, x_0)} - \frac{\delta}{2} \right) \ln \frac{1}{\epsilon} < t_1^{\eta/\epsilon} < \left( \frac{1}{\bar{f}(x_1, x_0)} + \frac{\delta}{2} \right) \ln \frac{1}{\epsilon} \right), \quad (6.4) \\ & \sup_{0 \leq t \leq t_1^{\eta/\epsilon}} \left| \frac{N_t^{K,1}}{K\epsilon} - m_t \right| < \delta' \\ &= 1 \end{aligned}$$

where the last equal sign is due to (6.3).

After population of trait  $x_1$  reaches some  $\eta$  threshold, the dynamics  $(\xi_t^K(x_0), \xi_t^K(x_1))$  can be approximated by the solution of a two-dimensional Lotka-Volterra equations.



Then, it takes time of order 1 (mark this time coordinator by  $\tilde{S}_1^\eta$ ) for the two subpopulations switching their mass distribution and gets attracted into  $\eta$ -neighborhood of the stable equilibrium  $(0, \bar{n}(x_1))$ .

**Step 2.** Now consider the emerging and growth of population  $\xi_t^K(x_2) := \langle X_t^K, 1_{\{x_2\}} \rangle$  at trait site  $x_2$ . Set  $S_2^\epsilon = \inf\{t : t > \tilde{S}_1^\eta, \xi_t^K(x_2) \geq \epsilon\}$ . Similarly as is done for  $S_1^\epsilon$  in (6.2), one can get that  $\lim_{K \rightarrow \infty} \mathbb{P}(S_2^\epsilon - \tilde{S}_1^\eta = O(1)) = 1$ . On a longer time scale, we will not distinguish  $S_2^\epsilon$  from  $\tilde{S}_1^\eta$ .

Set  $S_2^\eta = \inf\{t : t > S_2^\epsilon, \xi_t^K(x_2) \geq \eta\}$ . One follows the same procedure to derive (6.4) and asserts that for any  $\delta > 0$ ,

$$\lim_{K \rightarrow \infty} \mathbb{P} \left( \left( \frac{1}{\bar{f}(x_2, x_1)} - \delta \right) \ln \frac{1}{\epsilon} < S_2^\eta - \tilde{S}_1^\eta < \left( \frac{1}{\bar{f}(x_2, x_1)} + \delta \right) \ln \frac{1}{\epsilon} \right) = 1. \quad (6.5)$$

Note that assumption **(B3)**  $\frac{2}{b(x_2) - d(x_2)} \geq \frac{1}{f(x_1, x_0)} + \frac{1}{f(x_2, x_1)}$  guarantees that  $\xi_t^K(x_2)$  can not grow so fast in exponential rate  $b(x_2) - d(x_2)$  such that it reaches some  $\eta$ -level before  $S_2^\eta$ .

During time period  $(\tilde{S}_1^\eta, S_2^\eta)$ , population at site  $x_0$ , on one hand, decreases due to the competition from more fitter trait  $x_1$ . On the other hand, it can not go below  $\epsilon$  level due to the successive migration in a portion of  $\epsilon$  from site  $x_1$ . More precisely, by neglecting migrant contribution,  $\xi_t^K(x_0)$  converges  $n_t(x_0)$  in probability as  $K$  tends to  $\infty$ , where

$$\dot{n}_t(x_0) = (b(x_0) - d(x_0) - \alpha(x_0, x_1)\bar{n}(x_1))n_t(x_0) = \bar{f}(x_0, x_1)n_t(x_0) \quad (6.6)$$

with  $n_0(x_0) = \eta$ . Let  $\Delta S_2^\eta = S_2^\eta - \tilde{S}_1^\eta$ . Then, for any  $\delta > 0$ ,

$$\begin{aligned} & \lim_{K \rightarrow \infty} \mathbb{P} \left( \xi_{S_2^\eta}^K(x_0) \in (n_{\Delta S_2^\eta}(x_0) - \delta, n_{\Delta S_2^\eta}(x_0) + \delta) \right) \\ &= \lim_{K \rightarrow \infty} \mathbb{P} \left( \eta e^{\bar{f}(x_0, x_1)\Delta S_2^\eta} - \delta < \xi_{S_2^\eta}^K(x_0) < \eta e^{\bar{f}(x_0, x_1)\Delta S_2^\eta} + \delta \right) \\ &= \lim_{K \rightarrow \infty} \mathbb{P} \left( \eta e^{|\bar{f}(x_0, x_1)|/\bar{f}(x_2, x_1)} - \delta < \xi_{S_2^\eta}^K(x_0) < \eta e^{|\bar{f}(x_0, x_1)|/\bar{f}(x_2, x_1)} + \delta \right) \\ &= 1 \end{aligned} \quad (6.7)$$

where the second equality is due to (6.5). Taking the migration from site  $x_1$  into account, we thus have

$$\lim_{K \rightarrow \infty} \mathbb{P} \left( \xi_{S_2^\eta}^K(x_0) = O(\epsilon^{|\bar{f}(x_0, x_1)|/\bar{f}(x_2, x_1)} \vee \epsilon) \right) = 1. \quad (6.8)$$

We proceed as before for  $\tilde{S}_1^\eta$  in step 1. After time  $S_2^\eta$ , the mass bars on dimorphic system  $(\xi_t^K(x_1), \xi_t^K(x_2))$  can be approximated by ODEs and will be switched again in time of order 1 (marked by  $\tilde{S}_2^\eta$  as in Figure 5), and they are attracted into  $\eta$ -neighborhood of  $(0, \bar{n}(x_2))$ . As for the population density on site  $x_0$ , one obtains from (6.8)

$$\lim_{K \rightarrow \infty} \left( \xi_{S_2^\eta}^K(x_0) = O(\epsilon^{c_1}) \right) = 1 \quad (6.9)$$

where  $c_1 = \frac{|\bar{f}(x_0, x_1)|}{\bar{f}(x_2, x_1)} \wedge 1 \leq 1$ .

**Step 3.** We now consider the recovery of subpopulation at trait site  $x_0$ . Recovery arises because of the lack of effective competitions from its neighbor site  $x_1$ , or under negligible competitions since the local population density on  $x_1$  is very low under the

control of its fitter neighbor  $x_2$ . Without loss of generality, we suppose  $c_1 := \frac{|\bar{f}(x_0, x_1)|}{f(x_2, x_1)} < 1$  in (6.8).

Set  $S_0^\eta = \inf\{t : t > \tilde{S}_2^\eta, \xi_t^K(x_0) \geq \eta\}$ . We proceed as before in step 1. From (6.8),  $\frac{\xi_{\tilde{S}_2^\eta}^K(x_0)}{\epsilon^{c_1}}$  converges to some positive constant (say  $m_0$ ) in probability as  $K \rightarrow \infty$ . Thus, by applying law of large numbers to the sequence of processes  $\frac{N_t^K}{K\epsilon^{c_1}}$ , for any  $\delta > 0, T > 0$ ,

$$\lim_{K \rightarrow \infty} \mathbb{P} \left( \sup_{0 \leq t \leq T} \left| \frac{\xi_t^K(x_0)}{\epsilon^{c_1}} - m_t \right| < \delta \right) = 1 \quad (6.10)$$

where  $m_t$  is governed by logistic equation  $\dot{m} = (b(x_0) - d(x_0))m$  starting with a positive initial  $m_0$ .

Following the same way to obtain (6.4), time length  $S_0^\eta - \tilde{S}_2^\eta$  can be approximated by time needed for dynamics  $m$  to approach  $\eta/\epsilon^{c_1}$  level, which is of order  $\frac{c_1}{(b(x_0) - d(x_0))} \ln \frac{1}{\epsilon}$ , i.e. for any  $\delta > 0$ ,

$$\lim_{K \rightarrow \infty} \mathbb{P} \left( \left( \frac{c_1}{b(x_0) - d(x_0)} - \delta \right) \ln \frac{1}{\epsilon} < S_0^\eta - \tilde{S}_2^\eta < \left( \frac{c_1}{b(x_0) - d(x_0)} + \delta \right) \ln \frac{1}{\epsilon} \right) = 1. \quad (6.11)$$

At the same time,  $\xi_t^K(x_1)$  converges in probability to  $\psi_t$  which satisfies equation  $\dot{\psi} = \bar{f}(x_1, x_2)\psi$  with  $\psi_{\tilde{S}_2^\eta} = \eta$ . Then, we can justify the following estimate for population density at site  $x_1$ ,

$$\lim_{K \rightarrow \infty} \mathbb{P} \left( \xi_{S_0^\eta}^K(x_1) = O(\epsilon^{c_2} \vee \epsilon) \right) = 1 \quad (6.12)$$

where  $c_2 = \frac{c_1 |\bar{f}(x_1, x_2)|}{b(x_0) - d(x_0)}$ .

We now combine all these estimates (6.4), (6.5), (6.11) together, and conclude that

$$\lim_{K \rightarrow \infty} \mathbb{P} \left( \|X_{t \ln \frac{1}{\epsilon}}^K - \Gamma^{(2)}\| < \delta \right) = 1 \quad (6.13)$$

for  $t > \bar{t}_2 := \frac{1}{f(x_1, x_0)} + \frac{1}{f(x_2, x_1)} + \frac{c_1}{b(x_0) - d(x_0)}$  under the total variation norm  $\|\cdot\|$  on  $\mathcal{M}_F(\mathcal{X})$ .  $\square$

*Proof of Theorem 3.3.* We proceed the proof by the induction method over the superscript  $L \in \mathbb{N}$  of trait space  $\mathcal{X}^{(L)} = \{x_0, x_1, \dots, x_L\}$ .

(1). When  $L = 2$ , it is already proved in Proposition 6.1 that there exists a constant  $\bar{t}_2 > 0$  such that for any  $t > \bar{t}_2$

$$\lim_{K \rightarrow \infty} X_{t \ln \frac{1}{\epsilon}}^K \stackrel{(d)}{=} \Gamma^{(2)} \quad (6.14)$$

under the total variation norm.

(2). Without loss of generality, suppose it holds that for any  $L = 2l$  there exists a constant  $\bar{t}_{2l}$  such that for any  $t > \bar{t}_{2l}$

$$\lim_{K \rightarrow \infty} X_{t \ln \frac{1}{\epsilon}}^K \stackrel{(d)}{=} \Gamma^{(L)}. \quad (6.15)$$

We need to prove the same relation also holds for the case  $L = 2l + 1$ .

We firstly consider the invasion time scale of population at site  $x_{2l+1}$ .

Denote by  $\xi_t^K(x_{2l+1}) := \langle X_t^K, 1_{\{x_{2l+1}\}} \rangle$ . If  $K\epsilon^{2l+1} \ll 1$ , it follows a similar proof as in Proposition 6.1. So, now we only need to consider the case when  $K\epsilon^{2l+1} \gg 1$ , that

is, the mass at site  $x_{2l+1}$  is large in the very beginning. In fact, since  $\frac{N_0^K}{K} \rightarrow n_0$  in law as  $K \rightarrow \infty$  and the nearest-neighbor mass migrates from site  $x_0$  to site  $x_{2l+1}$  by passing through  $x_1, \dots, x_{2l}$ , one applies the law of large numbers for random processes and obtains that

$$\lim_{K \rightarrow \infty} \mathbb{P} \left( \sup_{0 \leq t \leq T} |\xi_t^K(x_{2l+1}) - n_t(x_{2l+1})| < \delta \right) = 1 \quad (6.16)$$

where  $n_t(x_{2l+1})$  satisfies the equation  $\dot{n}_t(x_{2l+1}) = \epsilon^{2l+1} \prod_{j=1}^{2l+1} m(x_j, x_{j-1}) n_0$ . So, it takes time of order 1 for  $\xi_t^K(x_{2l+1})$  to reach  $\epsilon^{2l+1}$  level (mark the time coordinator by  $S_{2l+1}^\epsilon$ ).

Set  $S_{2l+1}^\eta = \inf\{t : t > S_{2l+1}^\epsilon, \xi_t^K(x_{2l+1}) \geq \eta\}$ . For  $t \in (S_{2l+1}^\epsilon, S_{2l+1}^\eta)$ , again by law of large numbers,  $\frac{\xi_t^K(x_{2l+1})}{\epsilon^{2l+1}}$  converges to  $\phi_t$  which satisfies  $\phi_0 = 1$  and

$$\dot{\phi} = (b(x_{2l+1}) - d(x_{2l+1}))\phi. \quad (6.17)$$

Thus,  $\Delta S_{2l+1}^\eta := S_{2l+1}^\eta - S_{2l+1}^\epsilon$  can be approximated by the time length (say  $\Delta t_{2l+1}$ ) needed for dynamics  $\phi$  to reach  $\eta/\epsilon^{2l+1}$  level, i.e.

$$\begin{aligned} \lim_{K \rightarrow \infty} \mathbb{P} \left( \left( \frac{2l+1}{b(x_{2l+1}) - d(x_{2l+1})} - \delta \right) \ln \frac{1}{\epsilon} < \Delta S_{2l+1}^\eta < \left( \frac{2l+1}{b(x_{2l+1}) - d(x_{2l+1})} + \delta \right) \ln \frac{1}{\epsilon} \right) \\ \lim_{K \rightarrow \infty} \mathbb{P} \left( \left( \frac{2l+1}{b(x_{2l+1}) - d(x_{2l+1})} - \delta \right) \ln \frac{1}{\epsilon} < \Delta t_{2l+1} < \left( \frac{2l+1}{b(x_{2l+1}) - d(x_{2l+1})} + \delta \right) \ln \frac{1}{\epsilon} \right) \\ = 1. \end{aligned} \quad (6.18)$$

We inherit the notation  $S_{2l}^\eta$  as the hitting time of  $\eta$ -level for the population at site  $x_{2l}$ . Due to the hypothesis for  $L = 2l$  case, we know that  $S_{2l}^\eta$  is of order

$$\left( [\bar{f}(x_{2l}, x_{2l-1})]^{-1} + \dots + [\bar{f}(x_1, x_0)]^{-1} \right) \ln \frac{1}{\epsilon}. \quad (6.19)$$

Thanks to assumption **(B3)**, i.e.

$$\frac{2l+1}{b(x_{2l+1}) - d(x_{2l+1})} > \frac{1}{\bar{f}(x_{2l}, x_{2l-1})} + \dots + \frac{1}{\bar{f}(x_1, x_0)}, \quad (6.20)$$

it implies that before time  $S_{2l}^\eta$ , population at site  $x_{2l+1}$  is still under negligible level (of order  $\epsilon^c$  for some positive constant  $c$ ) and can not influence the invasion process up to  $x_{2l}$ .

Following a similar procedure as deriving (6.19) (see Figure 5) to analyze the colonization of population at site  $x_{2l+1}$  due to migration from site  $x_{2l}$  with exponential rate  $\bar{f}(x_{2l+1}, x_{2l})$ , one obtains that  $S_{2l+1}^\eta$  should be of order

$$\left( [\bar{f}(x_{2l+1}, x_{2l})]^{-1} + \dots + [\bar{f}(x_1, x_0)]^{-1} \right) \ln \frac{1}{\epsilon}. \quad (6.21)$$

Comparing two time scale estimates (6.18) and (6.21) for  $S_{2l+1}^\eta$  under assumption **(B3)**, one gets (6.21) is the right one for the fixation of population at site  $x_{2l+1}$ .

Now we consider the total recovery time by summing up recovery time of all subpopulation on every second site backwards from  $x_{2l+1}$  to  $x_0$ , one can do calculations repeatedly as in Step 3 of the proof for Proposition 6.1. More precisely, for  $1 \leq i \leq l$ , the initial population  $\xi^K(x_{2i-1})$  on site  $x_{2i-1}$  which is prepared for recovering is no less than  $\epsilon$ -level due to

the consistent migration from its fitter neighbor site  $x_{2i}$ . On the other hand, it grows exponentially at least with a rate  $\bar{f}(x_{2i-1}, x_{2i-2}) = b(x_{2i-1}) - d(x_{2i-1}) - \alpha(x_{2i-1}, x_{2i-2})\bar{n}(x_{2i-2})$  due to the possibly strongest competition from its unfit neighbor  $x_{2i-2}$ . In all, the recovery time (mark by  $S_{2i-1}^{\eta, 2}$ ) for the population  $\xi^K(x_{2i-1})$  to reach  $\eta$ -level can be bounded from above

$$\lim_{K \rightarrow \infty} \mathbb{P} \left( S_{2i-1}^{\eta, 2} < \left( \frac{1}{\bar{f}(x_{2i-1}, x_{2i-2})} + \delta \right) \ln \frac{1}{\epsilon} \right) = 1. \quad (6.22)$$

We now combine both time estimates (6.21) and (6.22). Let

$$\bar{t}_{2l+1} := 2 \left( [\bar{f}(x_{2l+1}, x_{2l})]^{-1} + \dots + [\bar{f}(x_1, x_0)]^{-1} \right). \quad (6.23)$$

Then, one can conclude that for any  $t > \bar{t}_{2l+1}$ , for any  $\delta > 0$  and  $0 \leq i \leq l$ ,

$$\begin{aligned} \lim_{K \rightarrow \infty} \mathbb{P} \left( \left| \langle X_{t \ln \frac{1}{\epsilon}}^K, 1_{\{x_{2i+1}\}} \rangle - \bar{n}(x_{2i+1}) \right| < \delta \right) &= 1, \\ \lim_{K \rightarrow \infty} \mathbb{P} \left( \langle X_{t \ln \frac{1}{\epsilon}}^K, 1_{\{x_{2i}\}} \rangle < \delta \right) &= 1. \end{aligned} \quad (6.24)$$

It follows the conclusion for any  $t > \bar{t}_{2l+1}$ ,

$$\lim_{K \rightarrow \infty} X_{t \ln \frac{1}{\epsilon}}^K \stackrel{(d)}{=} \Gamma^{(2l+1)}. \quad (6.25)$$

□

*Proof of Theorem 4.2.* The proof of this result is similar to the proof of [3, Theorem 1]. We will not repeat all the details and only focus more on supporting lemmas which are cornerstones of the proof.

For any  $\varepsilon > 0$ ,  $t > 0$ ,  $L \in \mathbb{N}$ ,  $B \subset \mathcal{X}$  measurable, take the integer part  $L_1 := \lfloor \frac{L+2}{2} \rfloor$  and denote by

$$\begin{aligned} A^{K, \varepsilon, \sigma}(\varepsilon, t, L, B) := & \left\{ \text{Supp}(X_{\frac{t}{K\sigma}}^{K, \varepsilon, \sigma}) \text{ has } L+1 \text{ elements, and } L_1 \text{ out of them, say } \{x_1, \dots, \right. \\ & x_{L_1}\} \subset B, \text{ satisfy } \left| \langle X_{\frac{t}{K\sigma}}^{K, \varepsilon, \sigma}, 1_{\{x_i\}} \rangle - \bar{n}(x_i) \right| < \varepsilon, 1 \leq i \leq L_1, \\ & \text{and the other } L+1-L_1 \text{ traits, say } y_1, \dots, y_{L-L_1}, \text{ satisfy} \\ & \left. \langle X_{\frac{t}{K\sigma}}^{K, \varepsilon, \sigma}, 1_{\{y_j\}} \rangle < \varepsilon, 1 \leq j \leq L+1-L_1 \right\}. \end{aligned} \quad (6.26)$$

To the end, it is enough to establish that

$$\lim_{K \rightarrow \infty} \mathbb{P} (A^{K, \varepsilon, \sigma}(\varepsilon, t, L, B)) = \mathbb{P} (\text{Supp}(\Gamma_t) \subset B \text{ and has } L_1 \text{ elements} ) \quad (6.27)$$

where  $(\Gamma_t)_{t \geq 0}$  is defined in Definition 4.1. □

The first key ingredient of the proof is the characterization of exponentially distributed waiting time of each mutation event. It can be proved from the expression of the generator (4.4) as done in [3, Lemma 2 (c)]. We will not show the details here.

**Lemma 6.2.** *Assume that  $X_0^{K,\epsilon,\sigma} = \Gamma^{(L)}$ , w.o.l., take  $L = 2l$ . Let  $\tau$  be the first mutation time after 0. Then,*

$$\lim_{K \rightarrow \infty} \mathbb{P} \left( \tau > \frac{t}{K\sigma} \right) = \exp \left( -t \sum_{i=0}^l \bar{n}(x_{2i}^{(2l)}) \mu(x_{2i}^{(2l)}) \right). \quad (6.28)$$

$$\lim_{K \rightarrow \infty} \mathbb{P} \left( \text{at time } \tau, \text{ mutant comes from trait } x_{2k}^{(2l)} \right) = \frac{\bar{n}(x_{2k}^{(2l)}) \mu(x_{2k}^{(2l)})}{\sum_{i=0}^l \bar{n}(x_{2i}^{(2l)}) \mu(x_{2i}^{(2l)})}. \quad (6.29)$$

The second ingredient can be seen as a corollary of Theorem 3.3. It demonstrates that fixation of new configuration takes time of order  $\ln \frac{1}{\epsilon}$ , which is invisible on the mutation time scale.

**Lemma 6.3.** *Assume that  $X_0^{K,\epsilon,\sigma} = \Gamma^{(2l)} + \frac{1}{K} \delta_{x_{2k}^{(2l)} + h}$  for some  $0 \leq k \leq l$ . Then there exists a constant  $C > 0$ , for any  $\delta > 0$ , such that*

$$\lim_{K \rightarrow \infty} \mathbb{P} \left( \tau > C \ln \frac{1}{\epsilon}, \quad \sup_{t \in (C \ln \frac{1}{\epsilon}, \tau)} \|X_t^{K,\epsilon,\sigma} - \Gamma^{(2l+1)}\| < \delta \right) = 1 \quad (6.30)$$

where  $\Gamma^{(2l+1)}$  is defined as in Definition 4.2 (i) and  $\|\cdot\|$  is the total variation distance.

*Proof.* From Lemma 6.2, one concludes that, for any  $C > 0$ ,

$$\lim_{\epsilon \rightarrow 0} \mathbb{P} \left( \tau^\epsilon > C \ln \frac{1}{\epsilon} \right) = 1.$$

According to the fitness landscape, there will be one and only one ordered position for the new arising trait  $x_{2k}^{(2l)} + h$  in  $\Gamma^{(2l)}$ . Suppose there exists  $x_{2j}^{(2l)}$  such that  $x_{2k}^{(2l)} + h$  fits between  $x_{2j}^{(2l)}$  and  $x_{2j+1}^{(2l)}$ . Then, one has the local fitness order

$$x_{2j-1}^{(2l)} \prec x_{2j}^{(2l)} \prec x_{2k}^{(2l)} + h \prec x_{2j+1}^{(2l)}. \quad (6.31)$$

Since it is unpopulated for both traits  $x_{2j-1}^{(2l)}$  and  $x_{2j+1}^{(2l)}$  in  $\Gamma^{(2l)}$ , we consider  $(x_{2j}^{(2l)}, x_{2k}^{(2l)} + h)$  as an isolated pair without competition from others. As the same analysis as being done in Proposition 6.1, the two-type system will converge to  $(0, \bar{n}(x_{2k}^{(2l)} + h) \delta_{x_{2k}^{(2l)} + h})$  in time of order  $O(\ln \frac{1}{\epsilon})$ . On the right hand side of the isolated pair, nothing changes due to their isolation. Whereas on the left hand side of the pair, trait  $x_{2j-1}^{(2l)}$  increases exponentially due to the decay of its fitter neighbor  $x_{2j}^{(2l)}$ . So on and so forth, the mass occupation flips on the left hand side of  $x_{2j}^{(2l)}$ . As the same arguments in the finite trait space case (see *Proof of Theorem 3.3*), the entire rearrangement process can be completed in time of order  $O(\ln \frac{1}{\epsilon})$ .

In a similar method, we can prove the other case when the fitness location of  $x_{2k}^{(2l)} + h$  is on the left hand side of  $x_{2j}^{(2l)}$ , that is,

$$x_{2j-1}^{(2l)} \prec x_{2k}^{(2l)} + h \prec x_{2j}^{(2l)} \prec x_{2j+1}^{(2l)}.$$

In all, we conclude the new configuration  $\Gamma^{(2l+1)}$  by relabeling the traits as done in Definition 4.1 (i).  $\square$

Thus we conclude the proof of the Theorem 4.2.

## REFERENCES

- [1] B. Bolker and S. Pacala. Using moment equations to understand stochastically driven spatial pattern formation in ecological systems. *Theor. Popul. Biol.*, 52:179–197, 1997.
- [2] A. Bovier and S. D. Wang. Trait substitution tree on two timescales analysis. 2011. Preprint.
- [3] N. Champagnat. A microscopic interpretation for adaptive dynamics trait substitution sequence models. *Stoch. Proc. Appl.*, 116:1127–1160, 2006.
- [4] N. Champagnat and A. Lambert. Evolution of discrete populations and the canonical diffusion of adaptive dynamics. *Ann. Appl. Probab.*, 17:102–155, 2007.
- [5] N. Champagnat and S. Méléard. Polymorphic evolution sequence and evolutionary branching. *Probab. Theor. and Relat. Field.*, 148, 2010.
- [6] A. Clayton and S. N. Evans. Mutation-selection balance with recombination: convergence to equilibrium for polynomial selection costs. *SIAM J. Appl. Math.*, 69:1772–1792, 2009.
- [7] P. Collet, S. Méléard, and J.A.J. Metz. Random modeling of adaptive dynamics for sexual populations. *Preprint*, 2011.
- [8] D. A. Dawson and A. Greven. Multiscale analysis: Fisher-wright diffusions with rare mutations and selection, logistic branching system. 2010.
- [9] I. Eshel. On the changing concept of evolutionary population stability as a reflection of a changing point of view in the quantitative theory of evolution. *J. Math. Biol.*, 34:485–510, 1996.
- [10] A. M. Etheridge. Survival and extinction in a locally regulated population. *Ann. Appl. Probab.*, 14:188–214, 2004.
- [11] S. N. Ethier and T. G. Kurtz. *Markov processes: characterization and convergence*. John Wiley and Sons, New York, 1986.
- [12] S. N. Evans, D. Steinsaltz, and K. W. Wachter. A mutation-selection model for general genotypes with recombination. 2007.
- [13] N. Fournier and S. Méléard. A microscopic probabilistic description of a locally regulated population and macroscopic approximation. *Ann. Appl. Probab.*, 14:1880–1919, 2004.
- [14] M. I. Freidlin and A. D. Wentzell. *Random perturbations of dynamical systems*. Springer, New York, 1984.
- [15] M. Hutzenthaler and A. Wakolbinger. Ergodic behavior of locally regulated branching populations. *Ann. Appl. Probab.*, 17:474–501, 2007.
- [16] R. Law and U. Dieckmann. Moment approximations of individual-based models. *The Geometry of Ecological Interactions: Simplifying Spatial Complexity*, pages 252–270, 2002.
- [17] S. Méléard and V. C. Tran. Trait substitution sequence process and canonical equation for age-structured populations. *Journal of Math. Biol.*, 58:881–921, 2009.
- [18] S. D. Wang. Fixation and substitution of nearly neutral mutants in a locally regulated population. 2011. Preprint.

A. BOVIER, INSTITUT FÜR ANGEWANDTE MATHEMATIK, RHEINISCHE FRIEDRICH-WILHELMS-UNIVERSITÄT BONN, ENDENICHER ALLEE 60, 53115 BONN, GERMANY  
*E-mail address:* bovier@uni-bonn.de

S.-D. WANG, DEPARTMENT OF STATISTICS, UNIVERSITY OF OXFORD, 1 SOUTH PARKS ROAD, OXFORD, OX1 3TG, UK  
*E-mail address:* shidong.wang@stats.ox.ac.uk