

A STOCHASTIC ADAPTIVE DYNAMICS MODEL FOR BACTERIAL POPULATIONS WITH MUTATION, DORMANCY AND TRANSFER

JOCHEN BLATH¹, TOBIAS PAUL² AND ANDRÁS TÓBIÁS³

(25 January 2023)

Abstract: This paper introduces a stochastic adaptive dynamics model for the interplay of several crucial traits and mechanisms in bacterial evolution, namely dormancy, horizontal gene transfer (HGT), mutation and competition. In particular, it combines the recent model of Champagnat, Méléard and Tran (2021) involving HGT with the model for competition-induced dormancy of Blath and Tóbiás (2020).

Our main result is a convergence theorem which describes the evolution of the different traits in the population on a ‘doubly logarithmic scale’ as piece-wise affine functions. Interestingly, even for a relatively small trait space, the limiting process exhibits a non-monotone dependence of the success of the dormancy trait on the dormancy initiation probability. Further, the model establishes a new ‘approximate coexistence regime’ for multiple traits that has not been observed in previous literature.

MSC 2010. 60J85, 92D25.

Keywords and phrases. Dormancy, seed bank, competition, horizontal gene transfer, mutation, stochastic population model, large population limit, multitype branching process with immigration, multitype logistic branching process, invasion fitness, individual-based model, coexistence.

1. INTRODUCTION AND BIOLOGICAL MOTIVATION

1.1. Motivation and Previous Work. The stochastic individual based modelling and analysis of the dynamics and evolution of bacterial populations has attracted significant interest in recent years (see e.g. [Cha06, FM04, BCF⁺16, BCF⁺18, LFL17, BB18]). This can on the one hand be motivated externally by the relevance of bacterial population dynamics in biology, medicine and industry, and on the other hand internally by the presence of interesting and distinctive features which invite new modelling approaches and lead to new patterns and results. Two of these distinct features, which have only rather recently been incorporated in population genetic/dynamic models in a systematic way, are *horizontal gene transfer* and *dormancy*.

The first feature, horizontal gene transfer (HGT), can in an abstract sense be understood as the ability of individuals to transfer parts of their genome (resp. the corresponding traits) to other living individuals, for example via exchange of plasmids during bacterial conjugation [LT46]. This is in contrast to the hereditary ‘vertical transfer’, where genes are copied from parent to daughter cell during binary fission. Essentially, HGT may thus be interpreted as an evolutionary strategy to increase the production of (one’s own) favourable traits. HGT comes in several different forms, but for the assumptions of this paper, we will only consider a mechanism that can be motivated from transfer via

¹Goethe-Universität Frankfurt, Robert-Mayer-Straße 10, 60325 Frankfurt am Main, blath@math.uni-frankfurt.de

²HU Berlin, Rudower Chaussee 25, 12489 Berlin, t.paul@math.hu-berlin.de

³Budapest University of Technology and Economics, Műegyetem rkp. 3., H-1111 Budapest and Alfréd Rényi Institute of Mathematics, Reáltanoda utca 13-15., 1053 Budapest, Hungary, tobias@cs.bme.hu

conjugation. However, it is known that carrying a large quantity of plasmids slows down cell division and as such reduces the reproduction rate (cf. [Bal13]). Such a trade-off leads to interesting questions about the optimality of HGT strategies. HGT has received increasing attention from the modelling side in the last decades, and is now considered as an additional and relatively novel major evolutionary force in bacterial populations (see e.g. [BP14, KW12, SL77]).

A second common feature in microbial population dynamics is the wide-spread ability of individuals to enter a reversible state of low/vanishing metabolic activity. Such a dormancy trait comes in many guises, but the general feature seems to be that it allows individuals to survive (e.g. in the form of an endospore or cyst) during adverse conditions. It can be triggered by environmental cues (responsive switching), but may also happen spontaneously (stochastic bet hedging) see [LJ11, LdHWBB21] for recent overviews. Again, as for HGT, such a trait comes with a significant reproductive trade-off, since the maintenance of a dormancy trait requires a substantial machinery, and thus consumes resources which are unavailable for reproduction.

Interestingly, both mechanisms (HGT and dormancy) also play a crucial role in the context of antibiotic resistance, though in very different ways. While the exchange of resistance genes via horizontal transfer can lead to multi-resistant microbial populations (see e.g. [Ben08]), dormancy in the form of *persister cells* can be the cause of chronic infections, since these dormant cells with their vanishing metabolism seem to be protected from antibiotic treatment ([Lew10]).

However, HGT and dormancy are of course not the only features of bacterial population dynamics, and interact with classical mechanisms such as reproduction (and hereditary effects), mutation, selection, and competition. Only recently, the joint effects resulting from these mechanisms seem to have moved into the focus of mathematical modellers. However, given the complexity of bacterial dynamics and the underlying mechanisms, and in view of the sheer number of different evolutionary forces involved in such communities, it is clear that mathematical modelling has to start with simple, idealized scenarios in order to begin to understand basic patterns emerging from such complex interactions. This process has been initiated in the last decade.

Indeed, the papers of Billiard et al [BCF⁺16, BCF⁺18] have investigated the consequences of a simple directional HGT mechanism in stochastic individual based models with a focus on its interplay with competition, mutation, and the maintenance of polymorphic variability. In [CMT21], the approach is transferred and extended into an adaptive dynamics setting with moderately large mutation rates (as previously considered in [DM11], see also [CKS21]), providing a rather new and sophisticated mathematical machinery that leads to interesting scaling limits and emergent behaviour on a ‘doubly logarithmic scale’. It is shown that HGT can have major consequences for the long-term behaviour of the affected systems, including coexistence, evolutionary suicide and evolutionary cyclic behaviour, depending on the strength of the transfer rate.

Regarding dormancy (and the resulting seed banks), this feature has now been well established as an evolutionary force in population genetics, starting with [KKL01], and become a topic of investigation in coalescence theory (cf. [BGCKWB16, BEGC⁺15, BGCKWB20]). In ecology, dormancy and seed banks have been investigated for several decades, starting with Cohen [Coh66], and this lead to a rich (mostly deterministic) theory, see e.g. [LdHWBB21] for many further references. Traditional seed bank theory is complemented by quantitative research on phenotypic switches in microbial communities, cf. e.g. [KL05]. However, the mathematical analysis of dormancy in stochastic individual based models, in particular in the framework of adaptive dynamics, seems to be still in its infancy. Yet, several building blocks are already available. The interplay with competition has been investigated in [BT20], where it is shown that dormancy traits responding to competitive pressure can invade and fixate in a resident population despite a substantial reproductive trade-off. One step further, the interplay of dormancy with competition and directional HGT has been investigated in [BT21], where coexistence regimes of HGT and dormancy traits are being established.

1.2. Overview of the Present Paper. In the present paper, we are attempting to combine the evolutionary forces of mutation, selection, competition, HGT and dormancy within the adaptive dynamics framework of [CMT21]. In particular, we aim to obtain an analogue of their key convergence result, and to investigate the resulting macroscopic behaviour in dependence of the strength of a ‘dormancy initiation parameter’.

Let us briefly sketch some of the aspects of our model. We will consider a finite set of possible traits \mathcal{X} where each trait reproduces randomly. The trait space is the intersection of a constant multiple of the integer grid \mathbb{Z}^2 with the square $[0, 4]^2$. The first coordinate x of the trait (x, y) expresses the strength of dormancy (increasing with x), and the second coordinate y corresponds to the strength of HGT (increasing with y), as we will explain below. To incorporate reproductive trade-offs, the birth rate of an individual of trait (x, y) is strictly decreasing both in x and in y . Further we consider natural death at a fixed rate 1 for any active individual, which may be thought of as death by age. We also involve ‘death by competition’ for active individuals. This gives the death rates a dependence on the current population size. Now, traits (x, y) can become dormant instead of dying by competition with probability proportional to x . The dormant individuals are not competing for resources and hence do not contribute towards nor are affected by death by competition. Dormant individuals will also not take part in reproduction nor horizontal transfer. The dormant individuals will switch back to their active state at a fixed rate and have only a natural death rate, which usually is less than the one for active individuals. For horizontal transfer, we will assume that at a population size dependent rate, any given two active individuals meet. In this event, the individual with the ‘stronger’ HGT trait, ie. with the higher y -coordinate, transfers its trait to the other individual. Lastly, mutations occur randomly at birth with a power law with respect to the carrying capacity K . More precisely, the probability of a mutation at birth is $K^{-\alpha}$ for some $\alpha \in (0, 1)$. The mutations will either increase the x -coordinate or the y -coordinate, to the next possible value. In particular, we assume that it is not possible for both the ability to become dormant and the ability to perform horizontal transfer to be improved by one mutation.

We are interested in the dynamics of our model on the $\log K$ time-scale as $K \rightarrow \infty$. Our main result Theorem 2.2 describes convergence properties as in [CMT21, Theorem 2.1] or [CKS21, Theorem 2.2]. However, in its proof the auxiliary processes that we have to consider are now mostly bi-type (with one component representing the active individuals of a trait and the other component representing the dormant ones), which goes beyond their frameworks. Regarding our bi-type setting, some invasion properties have been studied in [BT20], where the form of HGT is slightly different.

Here, the mutation rate scales like $K^{-\alpha}$ for some power $\alpha \in (0, 1)$. Consequently, mutants relevant for the evolution of the population are not separated from each other in time. This is a major difference from the classical ‘Champagnat scaling’ discussed in [Cha06], where mutations are less frequent and cannot influence each other. In the polynomial mutation regime, under suitable assumptions, the logarithm of the size of any trait (with base K) converges to a piecewise linear function on the $\log K$ time scale as $K \rightarrow \infty$, as we will discuss below. In a population genetic framework, such a mutation regime was studied in [DM11] in a model with clonal interference. In the adaptive dynamics literature, this scaling of mutations occurred before in [Sma17, BCS19]. From a mathematical point of view, the main novelty of the paper [CMT21] is the systematic study of logistic birth-and-death processes with non-constant immigration, as it was also noted in [CKS21].

In our analysis, we will assume that the population is always of the same order as the carrying capacity, which already poses significant technical challenges, as the length of the present manuscript indicates. In particular, behaviours such as evolutionary suicide are not included in our analysis. In Section 3, we will explore the limiting dynamics for a couple of fixed parameters. We are able to recover some cyclic behaviour already observed in [CMT21]. In addition, the introduction of dormancy seems to allow for the system to be driven towards a state of coexistence in the following sense: At no point

in time there are more than two traits with size of order K , but on the $\log K$ timescale, there exists a finite time $T_1 < \infty$ such that for all $\varepsilon > 0$ there exists a time $T_0 < T_1$ such that on the time interval $[T_0, T_1]$ at least three traits are of order at least $K^{1-\varepsilon}$, which means that at least three traits are simultaneously macroscopic on a suitable interval. This behaviour has been found previously by [CKS21] in the case of asymmetric competition without HGT. In the model studied in [DM11], the set of points where the limiting piecewise linear process changes slopes may also have a finite accumulation point, see Lemma 1 therein.

The remainder of this paper is organized as follows. In Section 2 below, we present our model and our main result. Section 3 contains numerical results regarding some fixed choices of parameters for our model. The proof of our main convergence result, Theorem 2.2 will be carried out in Section 4.

In preparation of proving the convergence properties for our model, we analyse bi-type branching processes in Appendix A. We will see that similar properties hold for bi-type processes as they have been shown in [CMT21, Appendix B] for one-type processes. However, the addition of a second component to the considered processes is sufficient to only allow the ideas of the proof to carry over. The details of the proofs, in particular Theorem A.10, are more involved and require significant amounts of preparation.

In Appendix B, we consider several properties of logistic branching processes. Here, we can also make use of the ideas from [BT20], since we are interested in showing that after some time an initially resident trait is driven towards a small population size, while an invasive species becomes resident. As there are many cases of this competition to be distinguished, we also make use of the ideas in [BT21] in the case of competition between a bi-type process and a single-type process.

2. PRESENTATION OF THE MODEL AND MAIN RESULT

We construct a continuous time Markov jump process as follows: Let $K \in \mathbb{N}$ be a number, which controls the population size and is referred to as the *carrying capacity*. Further we consider the *trait space* $\mathcal{X} := \{0, \delta, \dots, L\delta\}^2 = ([0, 4] \cap \delta\mathbb{N})^2$, where $\delta > 0$ is a fixed real number and $L := \lfloor \frac{4}{\delta} \rfloor$. Here, the choice of the number 4 is arbitrary, it follows the paper [CMT21]. As already anticipated, the first coordinate x of the trait (x, y) of an individual expresses the strength of dormancy of the individual, and the second coordinate y of its trait expresses its strength of HGT. For each trait (x, y) we may have *active* or *dormant* individuals (in fact, if $x = 0$, then individuals cannot be dormant). We use the notation $N_{m,n}^{K,a}(t)$ and $N_{m,n}^{K,d}(t)$ to refer to the active and dormant population size respectively of trait $(m\delta, n\delta)$ at time $t \geq 0$.

- Active individuals of trait (x, y) give birth to another individual at rate

$$b(x, y) = 4 - \frac{x + y}{2}.$$

Fixing $\alpha \in (0, 1)$, the child carries the trait $(x + \delta, y)$ with probability $\frac{K^{-\alpha}}{2}$, and with the same probability it carries the trait $(x, y + \delta)$. Otherwise the offspring has trait (x, y) . Also, if a mutated trait would not belong to \mathcal{X} anymore, the offspring does not mutate and carries the parental trait (x, y) . The decreasing birth rate as x and y increase reflects the trade-off between high reproduction and other survival mechanisms.

- There is competition over resources between active individuals, which we incorporate into the death rate. Let $C > 0$ and $p \in (0, \frac{1}{4})$ be fixed. Active individuals of trait $(x, y) \in \mathcal{X}$ die at rate

$$d((x, y), N^{K,a}) = 1 + \frac{C(1 - px)N^{K,a}}{K},$$

where $N^{K,a}$ denotes the entire active population size $N^{K,a} = \sum_{m,n=0}^L N_{m,n}^{K,a}$.

- Active individuals of trait (x, y) can become dormant at rate

$$c((x, y), N^{K,a}) = \frac{CpxN^{K,a}}{K}.$$

In particular, we are interested in 'competition induced switching', where due to competition from other individuals a part of the population becomes dormant. Individuals with a high value in the first trait component x are thus able to efficiently avoid death in favour of dormancy.

- Dormant individuals of any trait die at a natural rate $\kappa \geq 0$ and become active again at rate $\sigma > 0$. Usually κ will be a small rate, significantly less than 1, so that dormant individuals are less likely to die than active individuals. This reflects the immunity of dormant individuals to external pressures.
- An active individual of trait (x, y) can transfer its trait to a given active individual with trait (\tilde{x}, \tilde{y}) at rate

$$\tau((x, y), (\tilde{x}, \tilde{y}), N^{K,a}) = \frac{\tau}{N^{K,a}} \mathbb{1}_{y > \tilde{y}}.$$

Note that dormant individuals are neither affected by nor are able to perform transfer. Here, traits with a large second component y are advantageous.

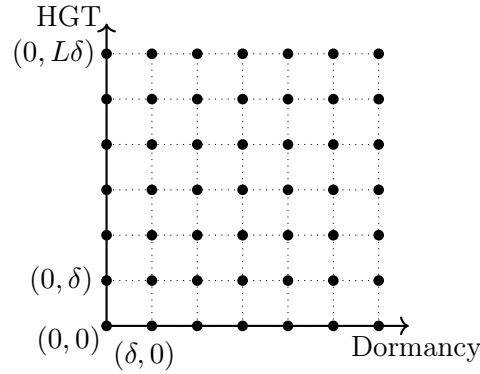


FIGURE 1. A visualization of the trait space \mathcal{X} . The strength of dormancy in a trait increases as the first component increases and the strength of HGT increases with the second component.

Comparing with Theorem A.3, we are concerned with the total size of each trait $(m\delta, n\delta) \in \mathcal{X}$, which we will denote by $N_{m,n}^K(t) := N_{m,n}^{K,a}(t) + N_{m,n}^{K,d}(t)$, and the corresponding exponents

$$N_{m,n}^K(t \log K) = K^{\beta_{m,n}^K(t)} - 1 \iff \beta_{m,n}^K(t) := \frac{\log(1 + N_{m,n}^K(t \log K))}{\log K}. \quad (2.1)$$

We are interested in the behaviour of $\beta_{m,n}^K$ as $K \rightarrow \infty$, that is, we want to understand the evolution of the population sizes on the $\log K$ timescale. Since our death and dormancy rate are dependent on the population size, there may be two cases: Either there is a single trait (x, y) , which has a population size of order K , in which case we refer to the trait (x, y) as *resident*; or the entire population is of size $o(K)$, in which case we refer to the trait with the largest population size as *dominant*. For our purposes, we will only consider the case where there is always one resident trait.

Now, assume that the trait $(x, y) = (m\delta, n\delta)$ is resident. Then, for large K , we can approximate the dynamics of $(N_{m,n}^{K,a}(t), N_{m,n}^{K,d}(t))$ as $K(z^a(t), z^d(t))$, where $(z^a(t), z^d(t))$ solves the ordinary differential

equation

$$\begin{aligned}\dot{z}^a(t) &= \left(3 - \frac{x+y}{2} - Cz^a(t)\right) z^a(t) + \sigma z^d(t) \\ \dot{z}^d(t) &= Cpx(z^a(t))^2 - (\kappa + \sigma)z^d(t).\end{aligned}\tag{2.2}$$

Indeed, this approximation follows from [EK86, Theorem 11.2.1]. We want to calculate a stable equilibrium of this system, which has already been done in [BT20, Section 2.2]. There it is shown that the only coordinate-wise non-negative asymptotically stable equilibrium of the system (2.2) is given for $3 - \frac{x+y}{2} > 0$ as $(\bar{z}_{m,n}^a, \bar{z}_{m,n}^d)$, where

$$\bar{z}_{m,n}^a = \frac{(3 - \frac{x+y}{2})(\kappa + \sigma)}{C(\kappa + (1 - px)\sigma)} \quad \text{and} \quad \bar{z}_{m,n}^d = \frac{px(3 - \frac{x+y}{2})^2(\kappa + \sigma)}{C(\kappa + (1 - px)\sigma)^2}.\tag{2.3}$$

Observe that this also holds true in the case where $x = 0$, in which case the equilibrium size of the dormant population is 0, and the active population size is $\frac{1}{C}(3 - \frac{x+y}{2})$, which corresponds to the equilibrium of the differential equation

$$\dot{z}(t) = \left(3 - \frac{x+y}{2} - Cz(t)\right) z(t).$$

If $3 - \frac{x+y}{2} < 0$, then there is no positive equilibrium and the fixed point $(0, 0)$ becomes asymptotically stable. This can be seen from linearizing the system (2.2), which yields the Jacobian

$$A(0, 0) = \begin{pmatrix} 3 - \frac{x+y}{2} & \sigma \\ 0 & -(\kappa + \sigma) \end{pmatrix},$$

whose determinant is positive and trace is negative. Hence both eigenvalues must be negative, showing that in this case $(0, 0)$ indeed is a stable equilibrium.

In order to have a well-defined process, we also need to introduce a starting condition. Initially, we assume the trait $(0, 0)$ to be close to its equilibrium, which is of size

$$N_{0,0}^{K,a}(0) = \left\lfloor \frac{3K}{C} \right\rfloor.\tag{2.4}$$

Since the effective mutation rate in a population of order K^c is $K^{c-\alpha}$, we choose all other starting conditions to be

$$N_{0,n}^{K,a}(0) = \lfloor K^{1-n\alpha} \rfloor \quad \text{and} \quad (N_{m,n}^{K,a}(0), N_{m,n}^{K,d}(0)) = \lfloor (K^{1-(m+n)\alpha}, K^{1-(m+n)\alpha}) \rfloor\tag{2.5}$$

if $n\alpha < 1$ and $(m+n)\alpha < 1$ respectively and 0 otherwise. Indeed, this choice is consistent with Lemma A.12, which would suggest that on the $\log K$ timescale we otherwise would immediately obtain a population of our chosen initial size. In addition, this choice shows that

$$\beta_{m,n}^K(0) \xrightarrow{K \rightarrow \infty} (1 - (m+n)\alpha) \vee 0.$$

Our next goal is to define the *invasion fitness* – also known as the initial rate of growth – $S((\tilde{x}, \tilde{y}), (x, y))$ of a single individual of trait (\tilde{x}, \tilde{y}) in a population, where the trait (x, y) is resident, i.e. at its equilibrium size. Hence, we consider the active population given by $K\bar{z}^a$ from (2.3). In particular we assume $\frac{x+y}{2} < 3$. We distinguish two cases:

Case 1: $\tilde{x} = 0$: In this case, the population size of trait $(0, \tilde{y})$ follows the dynamics of a usual one-dimensional birth and death process. Hence, we define the initial growth rate S as the asymptotic difference of birth and death rate, where we need to take into account the horizontal

transfer as additional births or deaths as follows

$$\begin{aligned} S((0, \tilde{y}), (x, y)) &:= \lim_{K \rightarrow \infty} b(0, \tilde{y}) - d((0, \tilde{y}), K\bar{z}^a) + \frac{K\bar{z}^a}{K\bar{z}^a + 1} \tau \mathbb{1}_{y < \tilde{y}} - \frac{K\bar{z}^a}{K\bar{z}^a + 1} \tau \mathbb{1}_{y > \tilde{y}} \\ &= 3 - \frac{\tilde{y}}{2} - \frac{(3 - \frac{x+y}{2})(\kappa + \sigma)}{\kappa + (1 - px)\sigma} + \tau \operatorname{sign}(\tilde{y} - y). \end{aligned}$$

Case 2: $\tilde{x} > 0$: Here we have transfer between the active and dormant populations. Hence the growth rate corresponds to that of a bi-type branching process. Being consistent with the definition thereof in Appendix A, we define the components of (A.3) asymptotically in accordance with Notation A.4. We set

$$\begin{aligned} r_1 &:= \lim_{K \rightarrow \infty} b(\tilde{x}, \tilde{y}) - d(K\bar{z}^a) + \frac{K\bar{z}^a}{K\bar{z}^a + 1} \tau \mathbb{1}_{y < \tilde{y}} - \frac{K\bar{z}^a}{K\bar{z}^a + 1} \tau \mathbb{1}_{\tilde{y} < y} \\ &= 3 - \frac{\tilde{x} + \tilde{y}}{2} - \frac{(3 - \frac{x+y}{2})(\kappa + \sigma)}{\kappa + (1 - px)\sigma} + \tau \operatorname{sign}(\tilde{y} - y), \\ r_2 &:= 0 - \kappa - \sigma = -(\kappa + \sigma), \\ \sigma_1 &:= \lim_{K \rightarrow \infty} \frac{Cp\tilde{x}(K\bar{z}^a + 1)}{K} = \frac{p\tilde{x}(3 - \frac{x+y}{2})(\kappa + \sigma)}{\kappa + (1 - px)\sigma}, \quad \sigma_2 := \sigma. \end{aligned}$$

Then the invasion fitness is defined by

$$S((\tilde{x}, \tilde{y}), (x, y)) := \frac{r_1 + r_2 + \sqrt{(r_1 - r_2)^2 + 4\sigma_1\sigma_2}}{2}.$$

This number is the largest eigenvalue of the mean matrix of the corresponding approximating bi-type branching process, which is given by

$$\begin{pmatrix} r_1 & \sigma_1 \\ \sigma_2 & r_2 \end{pmatrix}$$

We refer to Appendix A for details on the derivation of the initial growth rate of bi-type branching processes.

Note that distinguishing these two cases is necessary: If we were to model the behaviour of individuals of traits $(0, \tilde{y})$ as bi-type branching processes without switching into the dormant state, we would have – using the definition from the second case with $\sigma_1 = 0$ – that

$$S((\tilde{x}, \tilde{y}), (x, y)) = \max\{r_1, r_2\} \geq -(\kappa + \sigma).$$

In particular, for bi-type processes the invasion fitness is bounded from below by the total rate at which individuals exit the dormancy component. This lower bound is not reasonable for individuals which cannot become dormant.

Example 2.1. We are not able to exclude the possibility of long-term coexistence in the sense that $\operatorname{sign}(S((\tilde{x}, \tilde{y}), (x, y))) = -\operatorname{sign}(S((x, y), (\tilde{x}, \tilde{y})))$. As an example we may choose $C = 1$, $\tau = 1.3$, $\delta = 0.9$, $\kappa = 0$, $\sigma = 1$ and $p = 0.23$. Then an explicit computation shows

$$S((2\delta, 4\delta), (0, 2\delta)) \approx 0.22 \quad \text{and} \quad S((0, 2\delta), (2\delta, 4\delta)) \approx 0.29.$$

In these cases, an invasion would lead to coexistence, which we will exclude from our main theorem.

Using the above definitions of the invasion fitness, we can state our convergence result, which is very similar to [CMT21, Theorem 2.1].

Theorem 2.2. *Let $\alpha \in (0, 1)$, $\delta \in (0, 4)$, $\tau \geq 0$, $p \in (0, \frac{1}{4})$, $\kappa \geq 0$ and $\sigma > 0$ such that $S((\tilde{x}, \tilde{y}), (x, y)) \neq 0$ for all $(x, y), (\tilde{x}, \tilde{y}) \in \mathcal{X}$ with $(x, y) \neq (\tilde{x}, \tilde{y})$. Further assume that the transitions are as in the beginning of this section and that the initial conditions (2.4) and (2.5) are satisfied.*

- (i) Then there exists a time $T_0 > 0$ such that the sequences $\beta_{m,n}^K(t)$ from (2.1) converge as $K \rightarrow \infty$ in probability in $L^\infty([0, T])$ for all $T < T_0$ towards a deterministic piecewise affine continuous function $t \mapsto \beta_{m,n}(t)$ such that $\beta_{m,n}(0) = (1 - (m+n)\alpha) \vee 0$, which is characterized as follows.
- (ii) We define the sequence $s_k \geq 0$ and $(m_k^*, n_k^*) \in \{0, \dots, L\}^2$ inductively: Set $s_0 = 0$ and $(m_1^*, n_1^*) = (0, 0)$. Assume that for $k \geq 1$ we have constructed $s_{k-1} < T_0$ and (m_k^*, n_k^*) and assume that $\beta_{m,n}(s_{k-1}) \neq 0$ for some $(m\delta, n\delta) \in \mathcal{X}$. Then we define

$$s_k := \inf \left\{ t > s_{k-1} \mid \exists (m, n) \neq (m_k^*, n_k^*), \beta_{m,n}(t) = \beta_{m_k^*, n_k^*}(t) \right\}$$

Using this definition, we can distinguish three cases:

- (a) If $\beta_{m_k^*, n_k^*}(s_k) > 0$ define

$$(m_{k+1}^*, n_{k+1}^*) = \arg \max_{(m,n) \neq (m_k^*, n_k^*)} \beta_{m,n}(s_k)$$

if the argmax is unique. Otherwise we stop the induction and set $T_0 = s_k$.

- (b) If in case (a) we have

$$S((m_k^*\delta, n_k^*\delta), (m_{k+1}^*\delta, n_{k+1}^*\delta)) < 0 \quad \text{and} \quad S((m_{k+1}^*\delta, n_{k+1}^*\delta), (m_k^*\delta, n_k^*\delta)) > 0,$$

and $(m_{k+1}^* + n_{k+1}^*)\delta < 6$, then we continue our induction. Otherwise set $T_0 = s_k$ and stop the induction.

- (c) If there exists some $(m, n) \in \{0, \dots, L\}^2 \setminus \{(m_k^*, n_k^*)\}$ such that $\beta_{m,n}(s_k) = 0$ and $\beta_{m,n}(s_k - \varepsilon) > 0$ for all $\varepsilon > 0$ sufficiently small, then we also stop the induction and set $T_0 = s_k$.

- (iii) The function $\beta_{m,n}$ is defined for $t \in [s_{k-1}, s_k]$ as

$$\beta_{0,0}(t) = \left[\mathbb{1}_{\beta_{0,0}(s_{k-1}) > 0} \left(\beta_{0,0}(s_{k-1}) + \int_{s_{k-1}}^t S((0,0), (m_k^*\delta, n_k^*\delta)) \, ds \right) \right] \vee 0$$

and for $m \neq 0$ or $n \neq 0$ we have

$$\begin{aligned} \beta_{m,n}(t) &= \left(\beta_{m,n}(s_{k-1}) + \int_{t_{(m,n),k} \wedge t}^t S((m\delta, n\delta), (m_k^*\delta, n_k^*\delta)) \, ds \right) \\ &\vee (\beta_{m-1,n}(t) - \alpha) \vee (\beta_{m,n-1}(t) - \alpha) \vee 0, \end{aligned}$$

where $\beta_{-1,n} \equiv \beta_{m,-1} \equiv 0$ and the time $t_{(m,n),k}$ is defined by

$$t_{(m,n),k} := \begin{cases} \inf \{ t \geq s_{k-1} \mid \beta_{m-1,n}(t) = \alpha \text{ or } \beta_{m,n-1}(t) = \alpha \}, & \text{if } \beta_{m,n}(s_{k-1}) = 0 \\ s_{k-1} & \text{otherwise.} \end{cases}$$

The proof of this theorem will be discussed in Section 4. In light of the convergence theorems derived in Appendix A, this result is not very surprising. The defined fitness function determines the rate of growth of the corresponding branching process in the same way that the largest eigenvalue of the mean matrix of a single or bi-type branching process does.

Also, note that the fitness functions are constant on each time interval, so we may replace the integral by multiplying with the length of the integrated interval. We have chosen this representation to allow for a more direct comparison with [CMT21, Theorem 2.1].

Remark 2.3. We will shortly discuss the conditions, listed in part (ii) in our theorem, which lead to an end of the induction.

- (a) At the time s_k at least one new trait, other than the previously resident trait, becomes of order K in the population. Hence, we want to ensure that the resulting competition between the different traits only occurs between two traits, so that we can apply our results from Appendix B. This condition requires at most two traits to be of size of order K simultaneously.

- (b) As we have seen in Example 2.1, there is not necessarily competitive exclusion. The first requirement in this case ensures that the invading trait becomes resident, while the initially resident trait declines in size, so there is no coexistence. The second condition $(m_{k+1}^* + n_{k+1}^*)\delta < 6$ is needed for the invading trait to have a positive equilibrium size.
- (c) If there is a trait, which is almost, but not fully, extinct at the time at which there is a change in the resident trait, we are not able to determine against which of the two traits of size K there is competition. Therefore, we want to ensure each trait with small population size to be fully extinct at the time when a change in the resident trait occurs.

3. EXAMPLES

3.1. Examples for limiting Functions in Theorem 2.2. In this section we will consider specific, arbitrary choices of parameters for our model to find some range of resulting behaviours for the limiting functions $\beta_{m,n}$ established in Theorem 2.2. As we will see, the dynamics are already quite complicated in the case of very few traits. In particular, a full analysis in the case of $2\delta < 4 < 3\delta$ as in [CMT21, Section 3] is not feasible. The main problem from our simulation appears to be the non-periodicity of our systems.

For all of the upcoming examples we choose $\delta = 1.51$, $\tau = 1.3$, $\kappa = 0$, $\sigma = 1$ and $\alpha = 0.5$. We will vary the dormancy parameter p which will show us plenty of qualitatively different results.

Example 3.1 ($p = 0.21$). For now we let $p = 0.21$. Then we can plot the limiting function and obtain the following graphics.

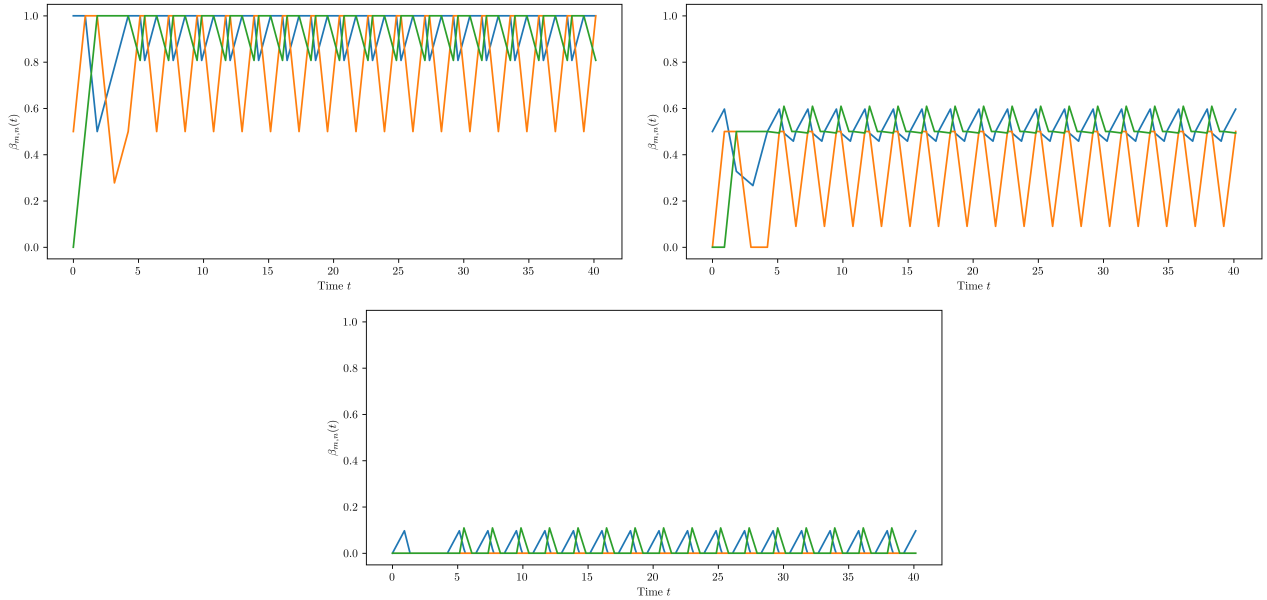


FIGURE 2. ($p = 0.21$) Top left: The dynamics of $\beta_{0,0}$ (blue), $\beta_{0,1}$ (orange) and $\beta_{0,2}$ (green). Top right: The dynamics of $\beta_{1,0}$ (blue), $\beta_{1,1}$ (orange) and $\beta_{1,2}$ (green). Bottom: The dynamics of $\beta_{2,0}$ (blue), $\beta_{2,1}$ (orange) and $\beta_{2,2}$ (green).

In this case we see that a similar behaviour as in [CMT21] is recovered: all traits exhibit (almost) periodicity. This stems from the fact that the traits with a dormancy component are not sufficiently fit. While the trait $(\delta, 0)$ is fit against the trait $(0, 0)$ and the trait $(\delta, 2\delta)$ is fit against the trait $(0, \delta)$, especially during the times when the trait $(0, 2\delta)$ is resident, all dormancy traits have a negative fitness and are only kept alive through the incoming migration. Hence, the essential components of the dynamics can be reduced to the case without dormancy.

Example 3.2 ($p = 0.22$). In this case, the resulting dynamics are given in the figure below.

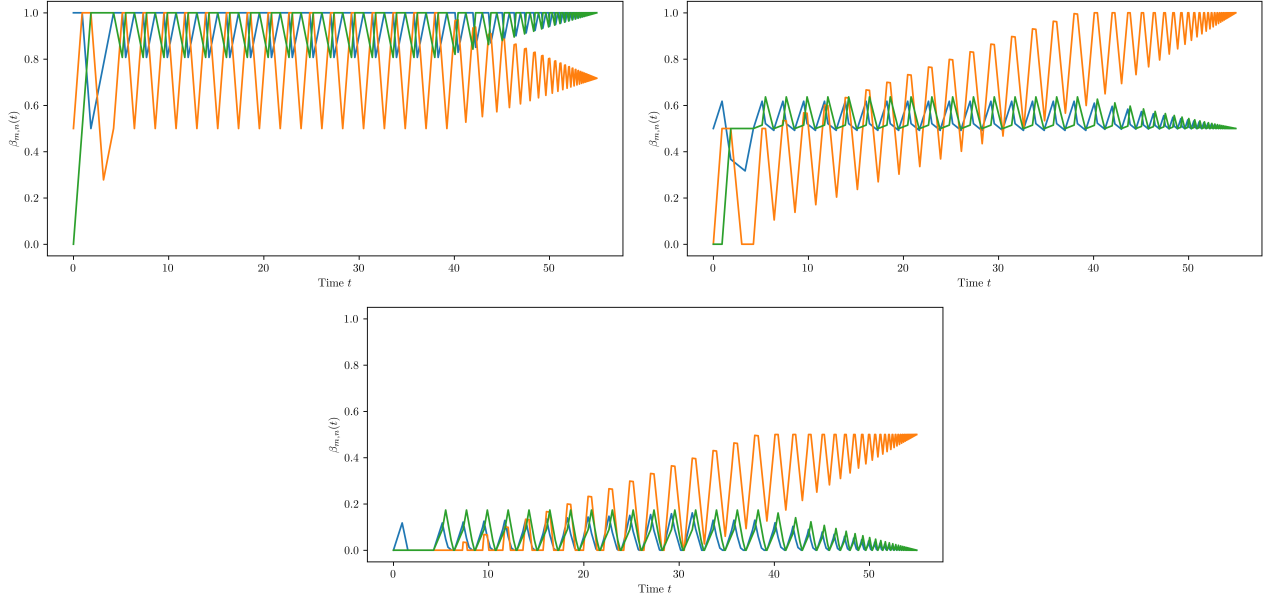


FIGURE 3. ($p = 0.22$) Top left: The dynamics of $\beta_{0,0}$ (blue), $\beta_{0,1}$ (orange) and $\beta_{0,2}$ (green). Top right: The dynamics of $\beta_{1,0}$ (blue), $\beta_{1,1}$ (orange) and $\beta_{1,2}$ (green). Bottom: The dynamics of $\beta_{2,0}$ (blue), $\beta_{2,1}$ (orange) and $\beta_{2,2}$ (green).

Here, two phases are to be distinguished: At first, we observe a very similar behaviour as in the case $p = 0.21$. In fact, for the traits $(0, \ell\delta)$, $\ell \in \{0, 1, 2\}$, the functions are at first identical to the previous case. However, the trait (δ, δ) is now sufficiently fit that its population size overall increases with each cycle until at one point it becomes resident. From this point onwards, we see that the functions are approaching a coexistence limit in the sense that for all k we have $s_{k+1} > s_k$ with $\lim_{k \rightarrow \infty} s_k < \infty$ and

$$\lim_{k \rightarrow \infty} \beta_{0,0}(s_k) = \lim_{k \rightarrow \infty} \beta_{0,2}(s_k) = \lim_{k \rightarrow \infty} \beta_{1,1}(s_k) = 1.$$

We will prove this claim below. Thus, although we have excluded the possibility of coexistence of any traits in the formulation of our Theorem 2.2 by demanding that the fitness functions need to have opposite signs, the system converges to an equilibrium. The reason behind this is the fact that we have demanded opposite signs, but the absolute values of the relative fitnesses of two traits are not necessarily, and often will not be, the same. This allows traits with dormancy to experience a large growth while they are not resident and fit against the dominant trait, but only a slow decline in population size when they are unfit against the dominant trait. In [CMT21] the fitness functions are antisymmetric functions in the sense that $S(x, y) = -S(y, x)$ for traits x, y and therefore such behaviour cannot be observed. The traits $(2\delta, \ell\delta)$ are again only driven by immigration through mutations.

We will now show inductively that the sequence $(s_k)_{k \in \mathbb{N}}$ converges by considering the system where there are only the traits $(0, 0)$, (δ, δ) and $(0, 2\delta)$. This reduction is justified by our simulations above, since all other traits become of order $o(K)$ after time 40. Further we assume the initial condition of our reduced system to be

$$\beta_{0,0}(0) = 1, \quad \beta_{1,1}(0) = 1, \quad \text{and} \quad \beta_{0,2}(0) = x_0 \in (0, 1),$$

that is, we assume that at the starting point of the system, the trait (δ, δ) has just become resident in the population which is only possible, if the trait $(0, 0)$ has been previously resident. In particular, the trait $(0, 2\delta)$ is unfit against the trait $(0, 0)$ and therefore must be of order $o(K)$. We will now construct a sequence of intermediate times until a similar configuration with $\beta_{0,0}(t) = \beta_{1,1}(t) = 1$

and $\beta_{0,2}(t) = x_1 > x_0$ is reached as is displayed in Figure 4. We calculate the individual fitnesses as determined by the fitness function. We obtain

$$\begin{aligned} S((\delta, \delta), (0, 0)) &= \frac{-\delta + \tau - \sigma + \sqrt{(\tau - \delta + \sigma)^2 + 12p\delta}}{2} & S((0, 2\delta), (0, 0)) &= \tau - \delta \\ S((0, 0), (\delta, \delta)) &= 3 - \frac{3 - \delta}{1 - p\delta} - \tau & S((0, 2\delta), (\delta, \delta)) &= 3 - \delta - \frac{3 - \delta}{1 - p\delta} + \tau \\ S((\delta, \delta), (0, 2\delta)) &= \frac{-\tau - \sigma + \sqrt{(\sigma - \tau)^2 + 4p\delta(3 - \delta)}}{2} & S((0, 0), (0, 2\delta)) &= \delta - \tau. \end{aligned}$$

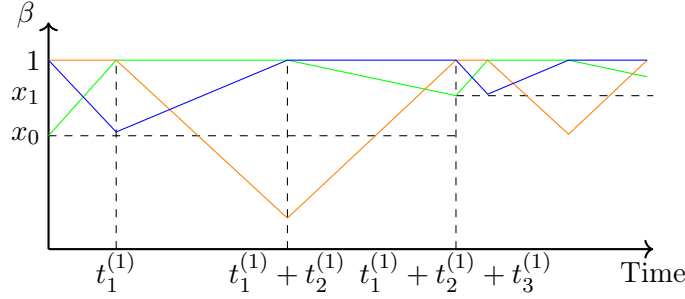


FIGURE 4. Illustration for the successive construction of the times $t_i^{(1)}$. Blue represents $\beta_{0,0}$, orange is $\beta_{1,1}$ and green is $\beta_{0,2}$.

We can therefore explicitly calculate that in this system the trait $(0, 2\delta)$ becomes resident after time

$$t_1^{(1)} := \frac{1 - x_0}{S((0, 2\delta), (\delta, \delta))} = C_1(1 - x_0).$$

At this time, we have the sizes

$$\beta_{0,0}(t_1^{(1)}) = 1 + t_1^{(1)} \cdot S((0, 0), (\delta, \delta)), \quad \beta_{1,1}(t_1^{(1)}) = 1 \quad \text{and} \quad \beta_{0,2}(t_1^{(1)}) = 1.$$

In the next step, the traits are competing with $(0, 2\delta)$. Therefore the trait $(0, 0)$ becomes resident after time

$$t_2^{(1)} := \frac{1 - (1 + t_1^{(1)} \cdot S((0, 0), (\delta, \delta)))}{S((0, 0), (0, 2\delta))} = -\frac{S((0, 0), (\delta, \delta))}{S((0, 0), (0, 2\delta))} \cdot t_1^{(1)} = C_2(1 - x_0).$$

We obtain

$$\beta_{0,0}(t_1^{(1)} + t_2^{(1)}) = 1, \quad \beta_{1,1}(t_1^{(1)} + t_2^{(1)}) = 1 + t_2^{(1)} \cdot S((\delta, \delta), (0, 2\delta)) \quad \text{and} \quad \beta_{0,2}(t_1^{(1)} + t_2^{(1)}) = 1.$$

The third phase of this system consists of competition of the other traits with $(0, 0)$. In this case, the trait (δ, δ) becomes resident again after time

$$t_3^{(1)} := \frac{1 - (1 + t_2^{(1)} \cdot S((\delta, \delta), (0, 2\delta)))}{S((\delta, \delta), (0, 0))} = -\frac{S((\delta, \delta), (0, 2\delta))}{S((\delta, \delta), (0, 0))} \cdot t_2^{(1)} = C_3(1 - x_0).$$

We can calculate that

$$\beta_{0,0}(t_1^{(1)} + t_2^{(1)} + t_3^{(1)}) = 1, \quad \beta_{1,1}(t_1^{(1)} + t_2^{(1)} + t_3^{(1)}) = 1 \quad \text{and} \quad \beta_{0,2}(t_1^{(1)} + t_2^{(1)} + t_3^{(1)}) = 1 + t_3^{(1)} \cdot S((0, 2\delta), (0, 0)).$$

In particular, we recover our starting condition after time $t_1^{(1)} + t_2^{(1)} + t_3^{(1)}$ where $S((0, 2\delta), (0, 0)) < 0$ implies

$$\beta_{0,2}(t_1^{(1)} + t_2^{(1)} + t_3^{(1)}) = 1 - C(1 - x_0) =: x_1,$$

with $C \in (0, 1)$. Repeating this process inductively shows that after the n -th such cycle, we obtain the condition

$$\beta_{0,0} \left(\sum_{k=1}^n t_1^{(k)} + t_2^{(k)} + t_3^{(k)} \right) = 1, \quad \beta_{1,1} \left(\sum_{k=1}^n t_1^{(k)} + t_2^{(k)} + t_3^{(k)} \right) = 1$$

and

$$\beta_{0,2} \left(\sum_{k=1}^n t_1^{(k)} + t_2^{(k)} + t_3^{(k)} \right) = 1 - (1 - x_0)C^n =: x_n.$$

Thus, as $n \rightarrow \infty$, the functions converge to 1 at the endpoint of each cycle. It remains to show, that the time steps are summable. Indeed, we find

$$\begin{aligned} \sum_{k=1}^n t_1^{(k)} + t_2^{(k)} + t_3^{(k)} &= \sum_{k=1}^n C_1 \cdot (1 - x_k) + C_2 \cdot (1 - x_k) + C_3 \cdot (1 - x_k) \\ &= (C_1 + C_2 + C_3) \sum_{k=1}^n (1 - x_0)C^k, \end{aligned}$$

which converges as $n \rightarrow \infty$. Therefore, we have proven that our choice of parameters leads to coexistence after finite time.

Example 3.3 ($p = 0.23$). We obtain the following functions.

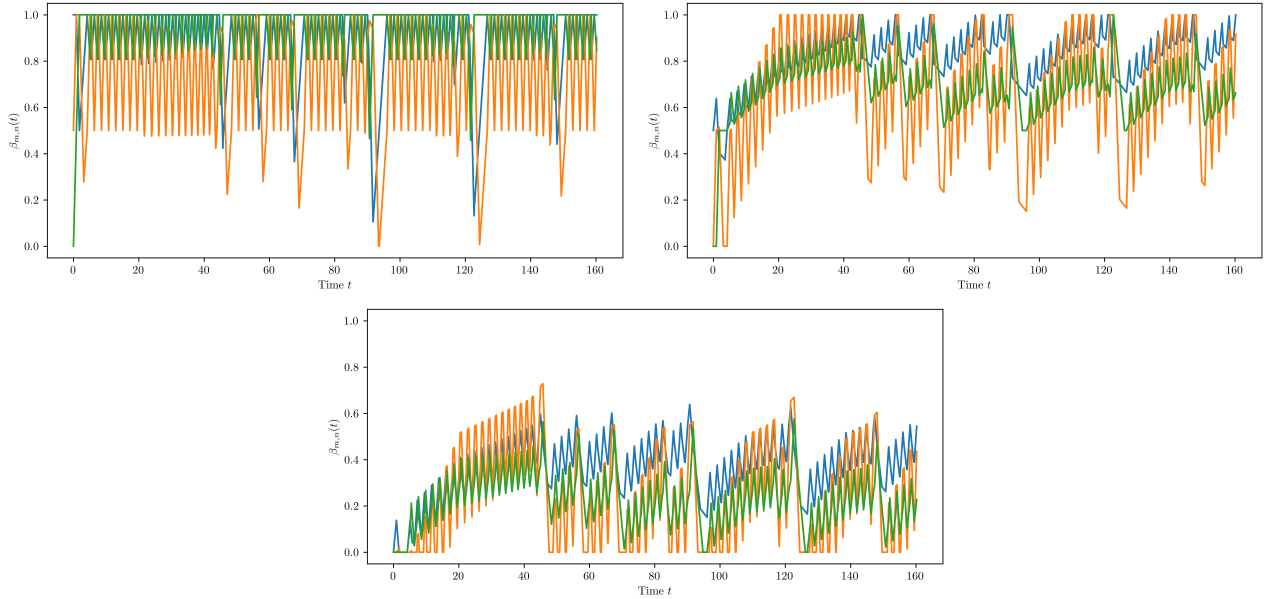


FIGURE 5. ($p = 0.23$) Top left: The dynamics of $\beta_{0,0}$ (blue), $\beta_{0,1}$ (orange) and $\beta_{0,2}$ (green). Top right: The dynamics of $\beta_{1,0}$ (blue), $\beta_{1,1}$ (orange) and $\beta_{1,2}$ (green). Bottom: The dynamics of $\beta_{2,0}$ (blue), $\beta_{2,1}$ (orange) and $\beta_{2,2}$ (green).

Here, there are even more phases to distinguish: At first we have the growth phase of the trait (δ, δ) until it becomes resident for the first time shortly after time 20. Note that, due to the increased value of p , the traits $(\delta, 0)$ and $(\delta, 2\delta)$ have an increased fitness as well and are slightly increasing in size. Now the larger value of p increases the equilibrium population size of trait (δ, δ) , which in turn implies that while (δ, δ) is resident, the traits $(0, 0)$ and $(0, 2\delta)$ have a lower fitness. Thus the times for which the traits $(0, 0)$ and $(0, 2\delta)$ are resident will be prolonged slightly. This is sufficient for the trait $(\delta, 0)$ (which only has a positive fitness while the trait $(0, 0)$ is resident) to become resident for the first time around time 43. Since the advantage from dormancy is not large enough to give the traits $(0, \ell\delta)$ an

overall negative fitness, we then have alternating times during which the traits $(\delta, \ell\delta)$ are growing and the traits $(0, \ell\delta)$ are cyclically resident followed by a short phase where one or more of the traits $(\delta, \ell\delta)$ become resident and the traits $(0, \ell\delta)$ experience a short but sharp decline in size. We do not know if the functions become periodic eventually, however from simulations we conjecture that this is not necessarily the case.

Example 3.4 ($p = 0.234$). Here, we observe an interesting change in the dynamics. Due to the increased fitness of the dormant individuals, it takes a shorter amount of time until one of the traits with dormancy becomes resident. In addition, they are able to stay resident for longer periods. Since the trait $(0, 0)$ is unfit against both $(\delta, 0)$ and (δ, δ) , it has an overall lower fitness. Other than in the case $p = 0.23$, the trait $(\delta, 2\delta)$ does not become resident fast enough to prevent $(0, 0)$ from becoming extinct. Once $(0, 0)$ is extinct, it cannot be resurrected since there are no incoming mutations. Hence we see a significant change. The dormant traits are not yet strong enough to prevent the trait $(0, 2\delta)$ from becoming resident. After $(0, 2\delta)$ is resident, all traits with dormancy become extinct or are only kept alive due to incoming mutations.

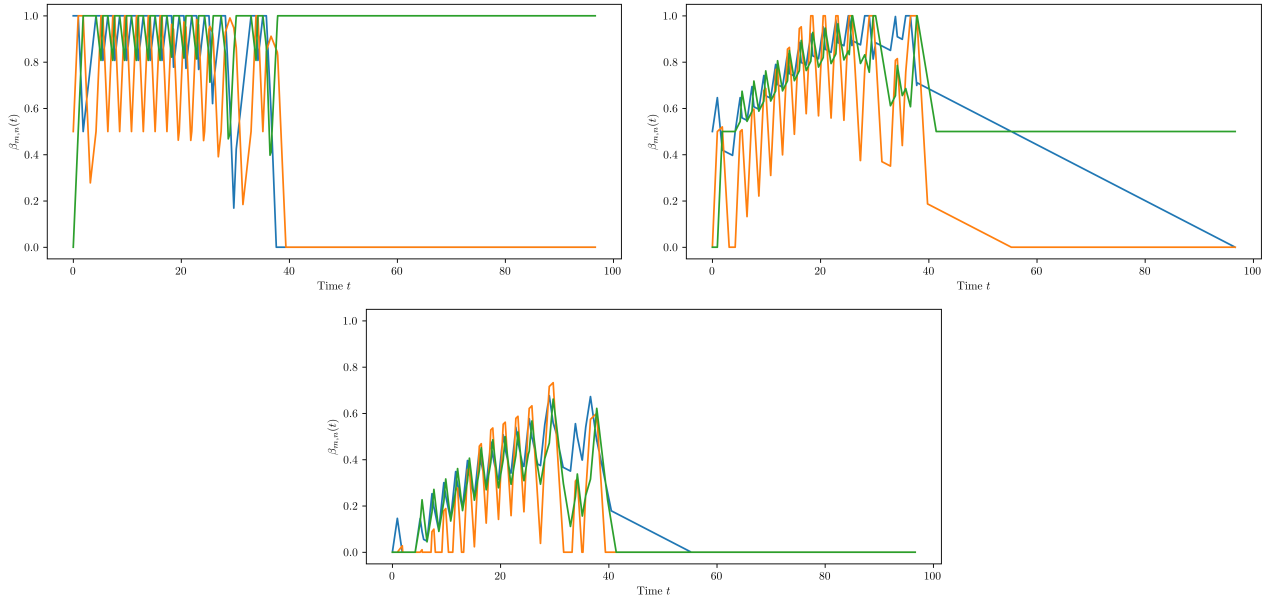


FIGURE 6. ($p = 0.234$) Top left: The dynamics of $\beta_{0,0}$ (blue), $\beta_{0,1}$ (orange) and $\beta_{0,2}$ (green). Top right: The dynamics of $\beta_{1,0}$ (blue), $\beta_{1,1}$ (orange) and $\beta_{1,2}$ (green). Bottom: The dynamics of $\beta_{2,0}$ (blue), $\beta_{2,1}$ (orange) and $\beta_{2,2}$ (green).

Example 3.5 ($p = 0.24$). In this case, the dormancy is sufficiently strong such that the overall fitness of the non-dormant traits is negative when the traits $(\delta, \ell\delta)$ become resident. Thus we are getting again coexistence as in Example 3.2, but now between the three traits $(\delta, \ell\delta)$.

3.2. Extending Theorem 2.2. One may ask the question of how the dynamics change as we alter the remaining parameters. Note that Theorem 2.2 does not cover the case where a trait $(x, y) \in \mathcal{X}$ with $\frac{x+y}{2} > 3$ becomes dominant. Although we have not treated this case formally, the convergence claimed in Theorem 2.2 should extend in a natural way: If the trait $(x, y) \in \mathcal{X}$ becomes dominant but is unfit on its own, then the entire population size drops to $o(K)$ immediately on the $\log K$ timescale. Therefore, we can define the fitness functions in these cases as before, but omit all factors which are scaled by K , that is we set the death rate to 1 and the switching rate from active to dormant to 0.

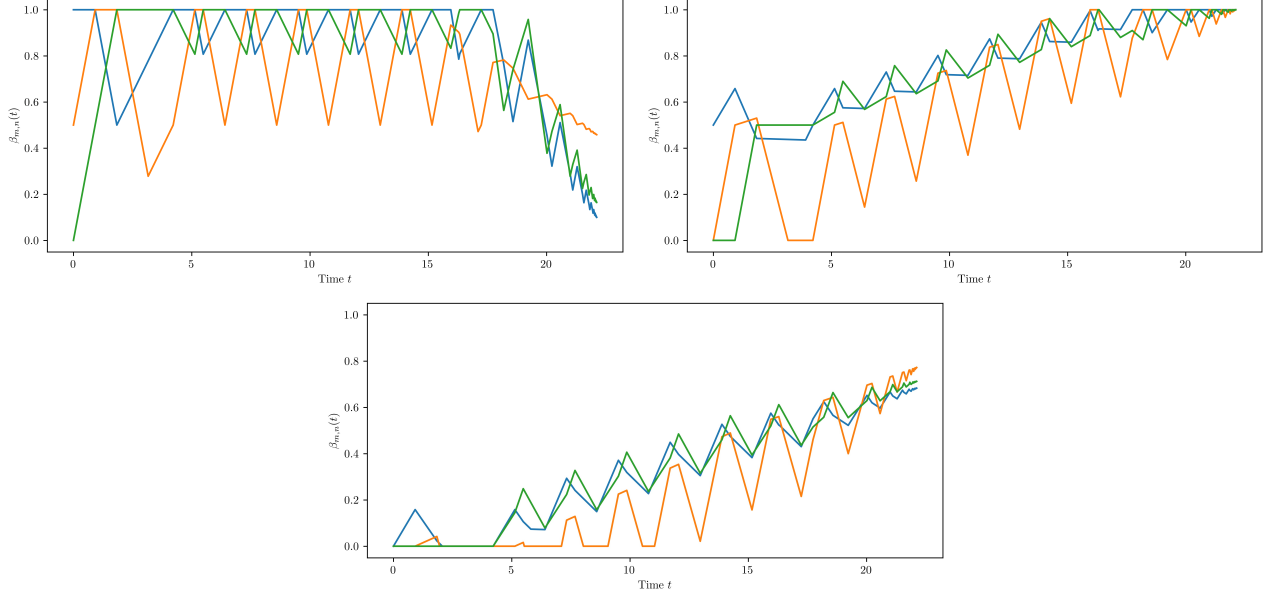


FIGURE 7. ($p = 0.24$) Top left: The dynamics of $\beta_{0,0}$ (blue), $\beta_{0,1}$ (orange) and $\beta_{0,2}$ (green). Top right: The dynamics of $\beta_{1,0}$ (blue), $\beta_{1,1}$ (orange) and $\beta_{1,2}$ (green). Bottom: The dynamics of $\beta_{2,0}$ (blue), $\beta_{2,1}$ (orange) and $\beta_{2,2}$ (green).

Then, we obtain for $\tilde{x} > 0$ the fitness

$$S((\tilde{x}, \tilde{y}), (x, y)) := \frac{r_1 + r_2 + \sqrt{(r_1 - r_2)^2 + 4\sigma_1\sigma_2}}{2} = \max\{r_1, r_2\}$$

where we use the notation $r_1 := 3 - \frac{\tilde{x} + \tilde{y}}{2} + \tau \text{sign}(\tilde{y} - y)$, $r_2 := -(\kappa + \sigma)$, $\sigma_1 := 0$ and $\sigma_2 := \sigma$. For $\tilde{x} = 0$ we set

$$S((0, \tilde{y}), (x, y)) := r_1 = 3 - \frac{\tilde{y}}{2} + \tau \text{sign}(\tilde{y} - y).$$

Also, we need to use this definition of the fitness function when the population size is of the order $o(K)$ but the dominant trait has not reached a size of order K . With these extensions to the fitness function, the limiting functions $\beta_{m,n}$ should satisfy the formula stated in Theorem 2.2 (iii). Note that the fitness of the traits with dormancy is bounded from below by $r_2 = -(\kappa + \sigma)$. Hence it may happen that at a point where there are exactly two dominant traits and normally a change in the dominant trait would occur both traits have the same negative slope. In these cases it is not obvious how to continue. Indeed, using the definition of the times s_k we would obtain $s_{k+1} = s_k$ and we cannot proceed any further.

Example 3.6. Let $\delta = 1.85$, $\tau = 1.3$, $p = 0.248$, $\kappa = 0$, $\sigma = 1$ and $\alpha = 0.5$. In this case we get the functions displayed in Figure 8.

Interestingly, here we have a finite time horizon T_0 and convergence of $\beta_{2,\ell}(s_k) \rightarrow 1$ as $k \rightarrow \infty$ for $\ell = 0, 1, 2$ although there are repeatedly short periods of macroscopic extinction where the entire population size is of order $o(K)$.

Example 3.7. If we set $\delta = 1.92$, $\tau = 1.3$, $p = 0.248$, $\kappa = 0$, $\sigma = 1$ and $\alpha = 0.5$, that is all parameters the same as in the previous example but for δ , then a similar but simultaneously new behaviour emerges.

The new aspect here is the finite time horizon T_0 where we have again convergence of $\beta_{2,\ell}(s_k)$ as $k \rightarrow \infty$ for $\ell = 0, 1, 2$, but now $\lim_{k \rightarrow \infty} \beta_{2,\ell}(s_k) < 1$. We may say that in this case the system of individuals is generally unfit, since it is not able to remain of order K at least periodically.

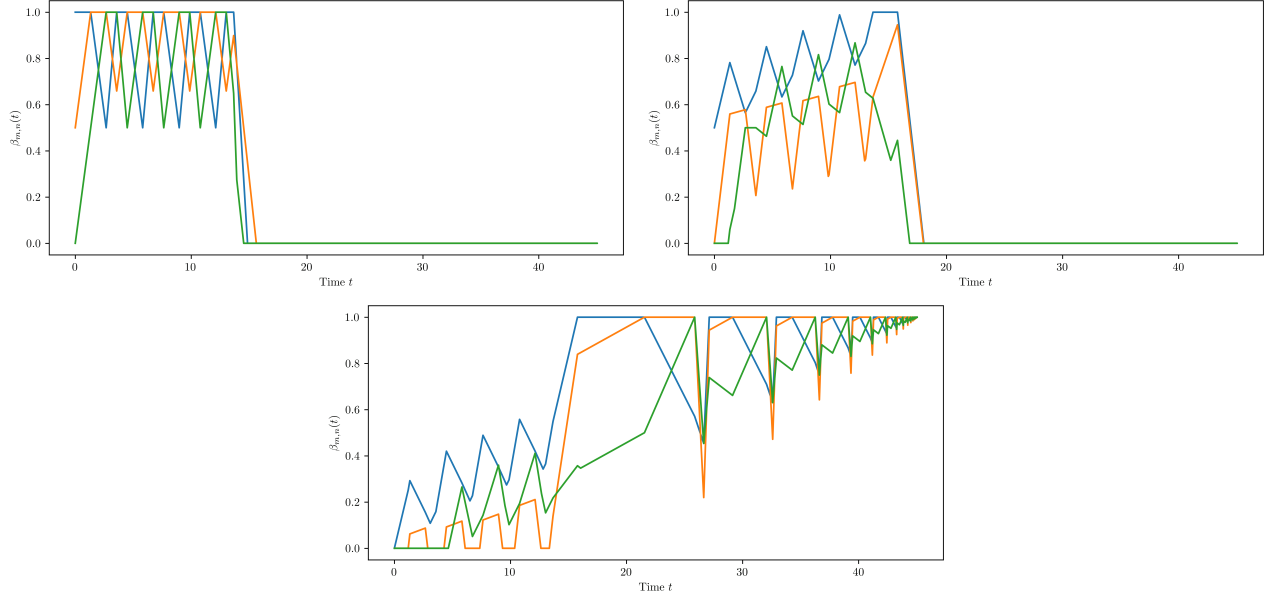


FIGURE 8. Functions for Example 3.6. Top left: The dynamics of $\beta_{0,0}$ (blue), $\beta_{0,1}$ (orange) and $\beta_{0,2}$ (green). Top right: The dynamics of $\beta_{1,0}$ (blue), $\beta_{1,1}$ (orange) and $\beta_{1,2}$ (green). Bottom: The dynamics of $\beta_{2,0}$ (blue), $\beta_{2,1}$ (orange) and $\beta_{2,2}$ (green).

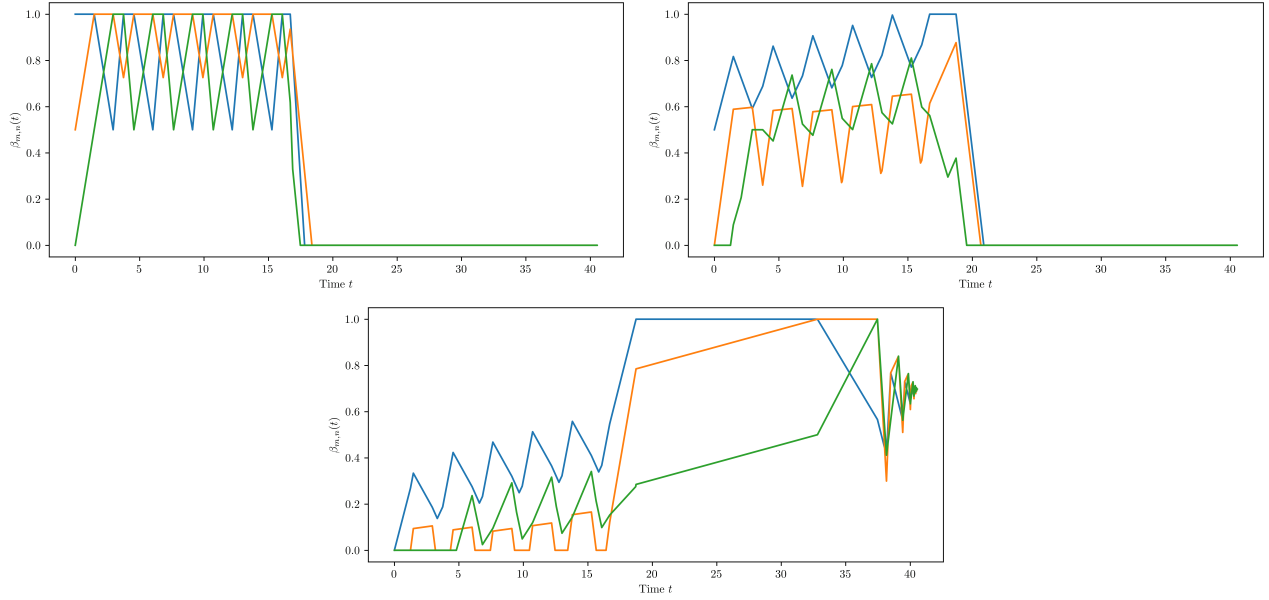


FIGURE 9. Functions for Example 3.7. Top left: The dynamics of $\beta_{0,0}$ (blue), $\beta_{0,1}$ (orange) and $\beta_{0,2}$ (green). Top right: The dynamics of $\beta_{1,0}$ (blue), $\beta_{1,1}$ (orange) and $\beta_{1,2}$ (green). Bottom: The dynamics of $\beta_{2,0}$ (blue), $\beta_{2,1}$ (orange) and $\beta_{2,2}$ (green).

In all of our simulations where an unfit trait becomes dominant, we have observed either one of the mentioned convergences or two traits with the same negative slope. In particular, we have not been able to observe evolutionary suicide and conjecture that due to the introduction of dormancy, evolutionary suicide is not possible. The reason for our conjecture lies in our fundamental modelling assumptions: only traits which can become dormant can also be unfit. Furthermore, assuming $\delta, C, \alpha, \tau, \kappa$ and σ fixed, due to the continuity of the functions $\beta_{m,n}$, we conjecture that the qualitative behaviours observed

(cyclic, driving towards coexistence, alternating but not periodic patterns) can be categorized into values of p coming from open intervals $I \subseteq (0, \frac{1}{4})$ and as such it would be interesting to explicitly calculate these threshold values.

3.3. Simulations. Another point of interest is the size of the carrying capacity K . We know from Theorem 2.2 that as $K \rightarrow \infty$ the exponents of the stochastic system converge under suitable rescaling of time towards the functions $\beta_{m,n}$. However, in reality the carrying capacity will be finite and thus we may ask how large K needs to be, such that the limiting functions $\beta_{m,n}$ give a good description of the stochastic system, more precisely $\beta_{m,n}^K$. For this we conducted simulations but came to the conclusion, that explicitly simulating the Markov process is not feasible for $K > 10^6$. The reason is twofold: On the one side, we need to increase the time horizon for the simulations as K increases (since we are working on the $\log K$ time scale) and on the other side, the time steps between events become smaller as the population size increases.

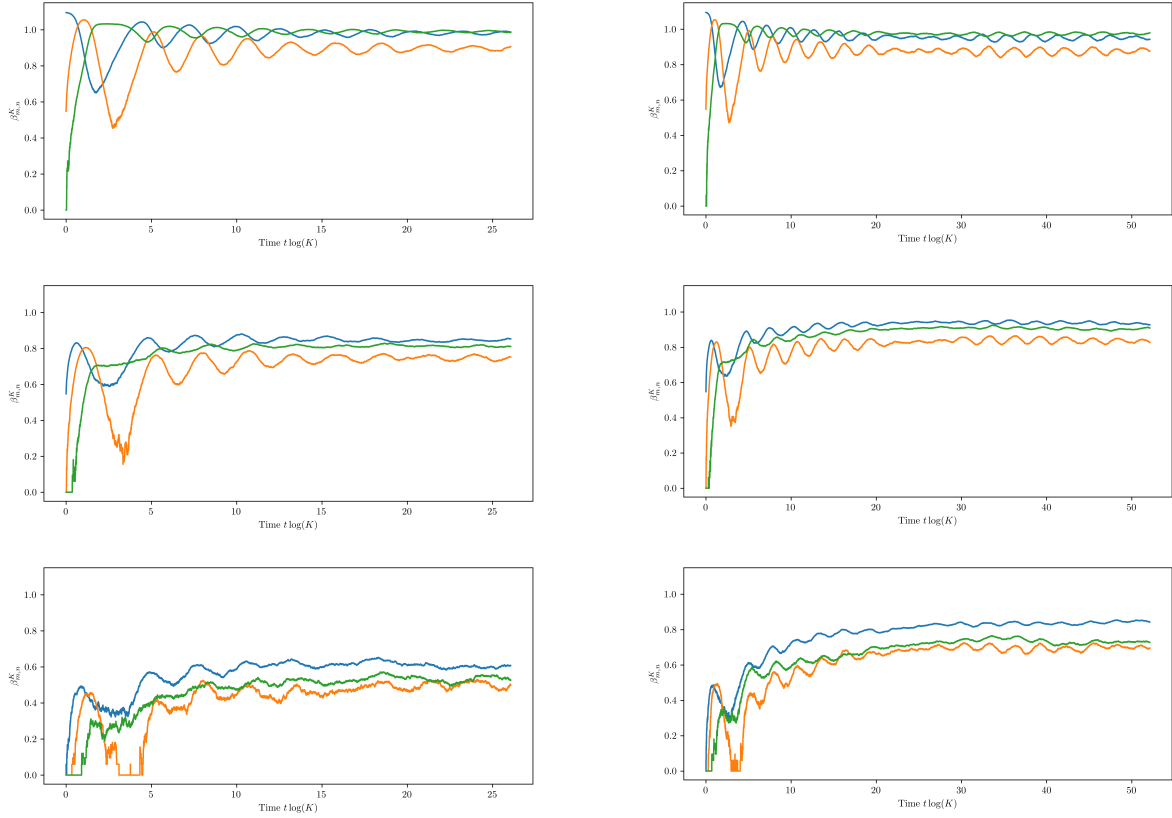


FIGURE 10. Left: Simulations with the parameters as in Example 3.1 and $C = 1$, $K = 10^5$. Right: Simulations with parameters as in Example 3.2 and $C = 1$, $K = 10^5$. From top to bottom we are increasing the index m of $\beta_{m,n}^K$ by 1 and in each plot $n = 0$ is blue, $n = 1$ is orange and $n = 2$ is green.

From our simulations with $K = 10^5$ in Figure 10 we are able to see, that the stochastic process resembles very little spontaneous jumps when the population size is large. Note that the images on the bottom of Figure 10 appear to be filled with jumps visible to the eye, which is due to the fact that $\sqrt{K} \approx 316$, so having an exponent of size $\frac{1}{2}$ means in terms of the population that around 316 individuals are alive. Therefore, a single event causes a relatively large change in the population. Otherwise, the curves appear to be smooth, which leads us to a more efficient way of simulating the

dynamics. We know, that on compact intervals the dynamics of $(\frac{N_{m,n}^{K,a}}{K}, \frac{N_{m,n}^{K,d}}{K})$ without migration can be approximated by the solution of the differential equation

$$\begin{aligned} \dot{x}_{m,n}^a(t) &= \sigma x_{m,n}^d(t) \\ &+ x_{m,n}^a(t) \left[3 - \frac{(m+n)\delta}{2} - C \sum_{m',n'=0}^L x_{m',n'}^a(t) + \tau \frac{\sum_{m'=0}^L \left(\sum_{n'=0}^{n-1} x_{m',n'}^a(t) - \sum_{n'=n+1}^L x_{m',n'}^a(t) \right)}{\sum_{m',n'=0}^L x_{m',n'}^a(t)} \right] \\ \dot{x}_{m,n}^d(t) &= pm\delta \cdot C x_{m,n}^a(t) \sum_{m',n'=0}^L x_{m',n'}^a(t) - (\sigma + \kappa) x_{m,n}^d(t). \end{aligned}$$

We also need to take into account the mutations which occur at birth with probability $K^{-\alpha}$. Since this probability tends to 0 as $K \rightarrow \infty$, we do not have a mutation term in the differential equation on its own. However, as we are more interested in simulating the dynamics for some fixed K , we alter the derivative of the active component to be

$$\dot{x}_{m,n}^a(t) \leftarrow \dot{x}_{m,n}^a(t) + \left(4 - \frac{(m+n-1)\delta}{2} \right) K^{-\alpha} (x_{m-1,n}(t) + x_{m,n-1}(t)),$$

which leads to a mixed approximation of the stochastic system. Now, choosing K fixed, we have on one side the usual approximation via an ODE and on the other side we have a non-zero mutation probability which is in accordance with the model. Determining the solution to these systems is numerically very efficient compared to a direct simulation and allows us to simulate the behaviour for large K . We refer to the exponents of the population sizes determined by solving the system as $\gamma_{m,n}^K$. However, we need to choose time steps Δt for solving the ODE, which leads to complications: The process $N_{m,n}^K$ is only taking integer values, so in particular, if the rescaled process satisfies $\frac{N_{m,n}^K}{K} < \frac{1}{K}$, then the population should be extinct. Now, if Δt is too small compared with $1/K$, then it may happen that the immigration during a time step of length Δt is not sufficiently strong to start the population. Another numerical issue is the time horizon, on which we need to solve the differential equation. After rescaling, we need to solve until time $T \log K$, which in our cases would usually have $T \in [50, 200]$ and thus may lead to some numerical instabilities. In particular, systems such as in Example 3.4 are sensitive to small deviations.

Comparing Figure 11 with the stochastic simulations, the ODE approach gives us a similar behaviour. Hence, we are confident that the solution to the differential equation will be similar to the stochastic system if we increase K . Obviously, these plots (stochastic simulation and ODE solution) have very little in common with the limits which we have discussed in the corresponding examples. However, when thinking of bacterial populations, $K = 10^5$ is still very small.

In Figure 12 with $K = 10^{15}$, the limiting functions are a much better approximation of the exponents, although Δt is still too small in relation to K for us to see any mutations arriving in $\gamma_{2,n}^K$ when $p = 0.21$. Also, note that the coexistence, which we observed for $p = 0.22$ in the limit, is in fact a normal cycle of residency between $(0, 0)$, (δ, δ) and $(0, 2\delta)$. Only when letting $K \rightarrow \infty$, these cycles become ever shorter and lead to coexistence. Another interesting effect of finite populations is the prolonged duration it takes for the trait (δ, δ) to become resident in the population. As K increases, this duration becomes shorter on the $\log K$ time scale. Although we cannot be certain about the reason for this mechanism, we think that it may be due to the competition phases which vanish on the $\log K$ time scale as $K \rightarrow \infty$ but take up a non-negligible amount of time for fixed K . In particular, the competition against traits with dormancy takes longer due to the dormancy component and hence the convergence is slower in K compared with systems with only HGT and no dormancy.

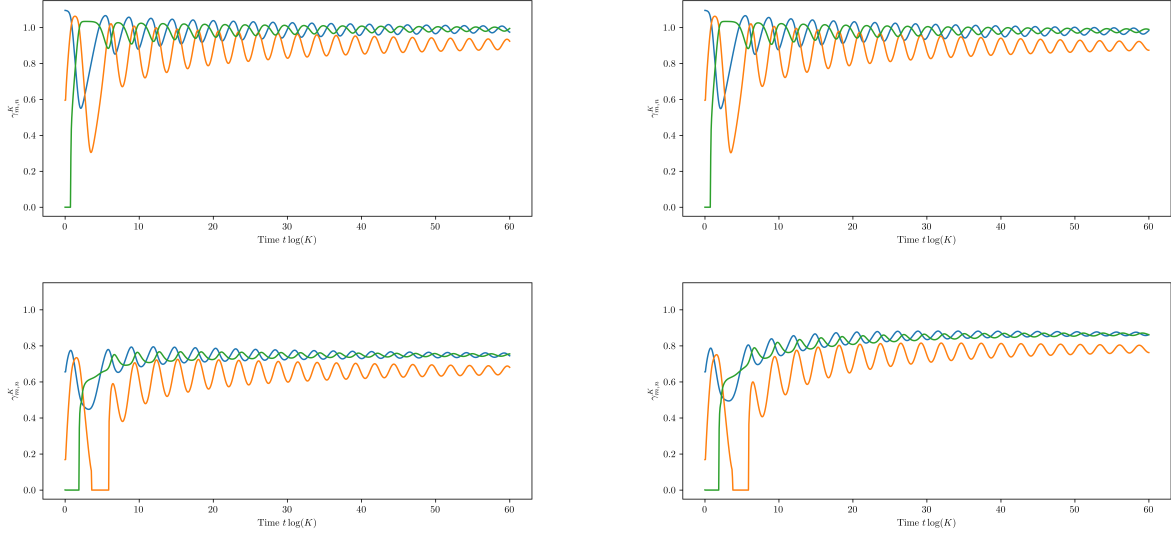


FIGURE 11. Left: Solving the ODE with Euler scheme, $K = 10^5$ and $\Delta t = T \log(K) K^{-1}$ with parameters as in Example 3.1 and $C = 1$. Right: Solving the ODE with Euler scheme, $K = 10^5$ and $\Delta t = T \log(K) K^{-1}$ with parameters as in Example 3.2 and $C = 1$. From top to bottom we have the usual arrangement of the Exponents $\gamma_{m,n}^K$. We have omitted the plots $\gamma_{2,n}^K$ due to the lack of incoming mutations.

4. PROOF OF THEOREM 2.2

We will give a short sketch of the proof, which is very similar to [CMT21, Theorem 2.1]. The idea is to decompose the time scale into two different kinds of phases: First there are long phases $[\sigma_k^K \log K, \theta_k^K \log K]$ which then are followed by short intermediate phases $[\theta_k^K \log K, \sigma_{k+1}^K \log K]$. During the long phases, there is exactly one trait, whose population size is close to its equilibrium and all other traits are of size $o(K)$. During the short phases, another trait emerges and becomes significant for competitive events and due to competition the initially resident trait is replaced by the emerging trait. We will show that

$$\lim_{K \rightarrow \infty} \sigma_{k+1}^K = \lim_{K \rightarrow \infty} \theta_k^K = s_k$$

with probability converging to 1 and hence on the $\log K$ timescale the intermediate phases vanish.

Since we only want to show this theorem in the case where only fit individuals (with a positive active equilibrium size) can become resident, we do not need to distinguish these cases, so our proof is simplified in this aspect compared to [CMT21]. However, during the intermediate phases we need to observe whether none, one or both of the involved traits can become dormant and in which way the horizontal transfer is acting, if at all.

Thus the proof will be performed by induction on k . During the long phases, we will make heavy use of coupling arguments to show the convergence $\beta_{m,n}^K \rightarrow \beta_{m,n}$. This will again be done by induction on the traits, where we need a nested induction, since the horizontal transfer can be exerted onto all traits with a lower second component. For these phases, we will make extensive use of Theorem A.3 and Theorem A.1, so we refer to Appendix A. During the intermediate phases, we need the corresponding competition results, which can be found in Appendix B.

To make the structure of the induction more obvious, we give the general idea here: The trait space \mathcal{X} can be visualized as the δ -grid on $[0, 4]^2$ and we first show the convergence on a time interval for the trait $(0, 0)$ as the base case. Then we advance our induction in the direction of dormancy to the trait $(\delta, 0)$, where we make another base case in order to highlight the differences in the bi-type case. This

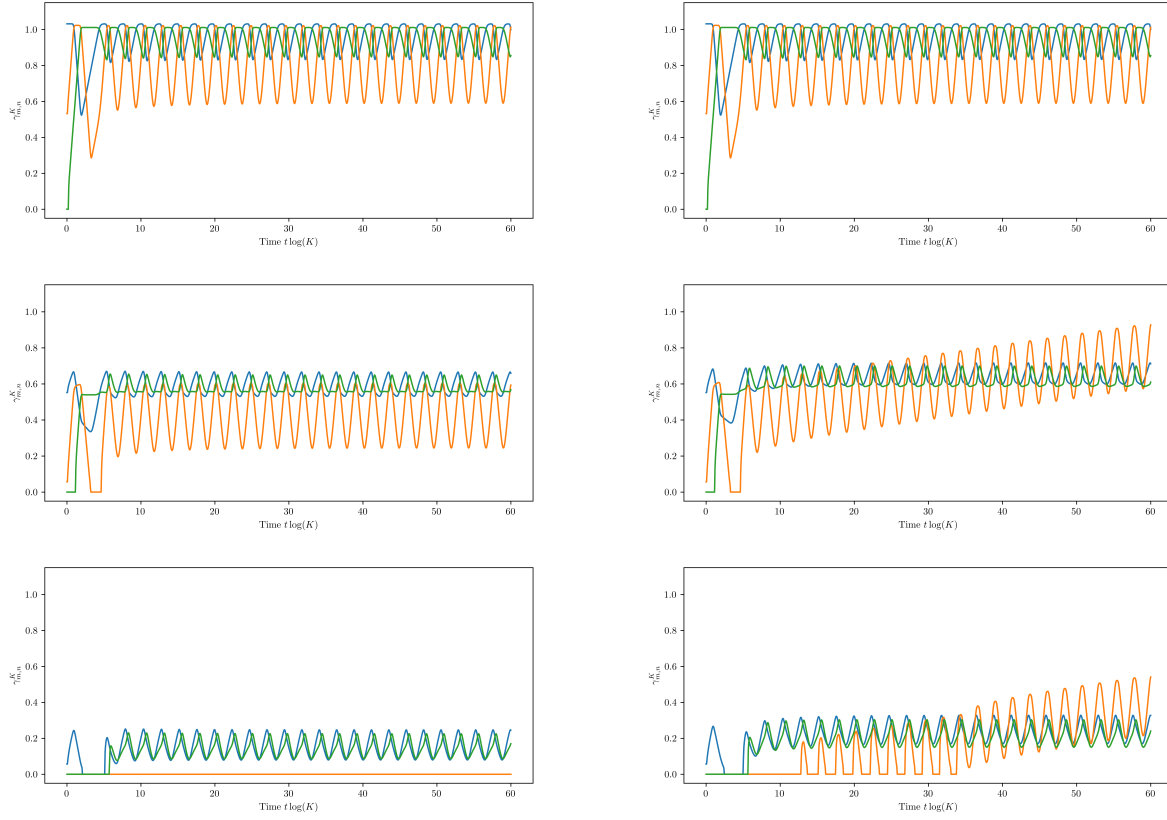


FIGURE 12. Left: Solving the ODE with Euler scheme, $K = 10^{15}$ and $\Delta t = T \log(K) 10^{-5}$ with parameters as in Example 3.1 and $C = 1$. Right: Solving the ODE with Euler scheme, $K = 10^{15}$ and $\Delta t = T \log(K) 10^{-5}$ with parameters as in Example 3.2 and $C = 1$. From top to bottom we have the usual arrangement of the Exponents $\gamma_{m,n}^K$.

is then followed by the induction step for traits $(m\delta, 0)$. In this fashion, we can then assume the result to hold for all traits $(\tilde{m}\delta, \tilde{n}\delta)$ with $\tilde{n} \leq n$ and $\tilde{m} \in \{0, \dots, L\}$ for some fixed $n \in \{0, \dots, L-1\}$. Then we can show the result for traits $(m\delta, (n+1)\delta)$ via an induction on m as for the case of $(m\delta, 0)$.

Throughout the proof we will use various kinds of branching processes. We denote by $BP_K(b, d, \beta)$ a one-dimensional branching process with birth rate b , death rate d and initial condition $\lfloor K^\beta - 1 \rfloor$. Also, we denote by $BPI_K(b, d, a, c, \beta)$ a one-dimensional branching process with birth rate b , death rate d , immigration at rate $K^c e^{at}$ and initial condition $\lfloor K^\beta - 1 \rfloor$. We refer to [CMT21, Appendix A,B] for results concerning these processes. With $BBPI_K(b_1, b_2, d_1, d_2, \sigma_1, \sigma_2, a, c, \beta, \gamma)$ we denote a two-dimensional branching process with birth rates b_1, b_2 , death rates d_1, d_2 , switching rates σ_1, σ_2 , immigration into the first coordinate at rate $K^c e^{at}$ and initial condition $\lfloor (K^\beta - 1, K^\gamma - 1) \rfloor$. We refer to Appendix A.

Further, we denote by $LBDI_K(b, d, C, \gamma)$ a one-dimensional logistic birth and death process with birth rate b , death rate $d + \frac{CN}{K}$, where N denotes the population size, and immigration at a predictable rate $\gamma(t)$ at time $t \geq 0$. We refer to [CMT21, Appendix C]. Also $LBDI_K(b_1, d_1, d_2, \sigma_2, p, C, \gamma_1)$ denotes the distribution of a two-dimensional logistic birth and death process with birth rates $b_1, 0$, death rates $d_1 + \frac{(1-p)CN}{K}, d_2$, where N denotes the population size of the first component, switching rates $\frac{pCN}{K}, \sigma_2$ and immigration into the first component at a predictable rate $\gamma_1(t)$ at time $t \geq 0$. We refer to Appendix B.

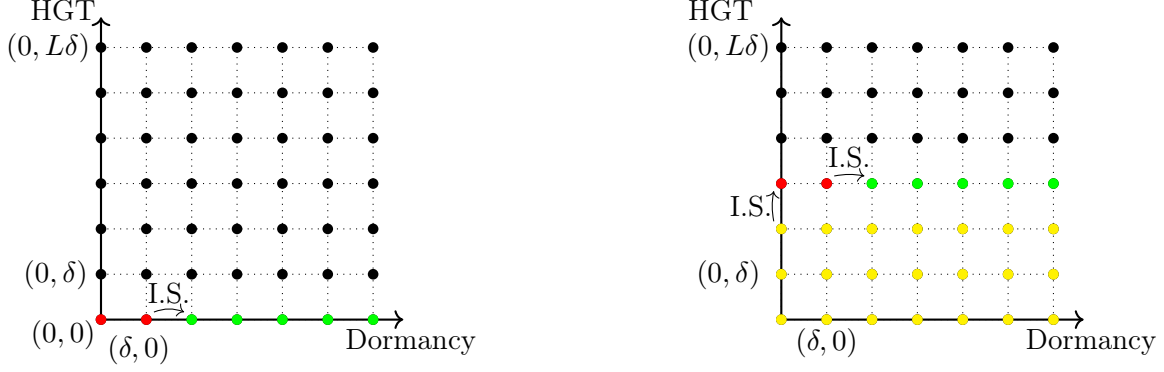


FIGURE 13. The schematic induction: In the left picture, red are base cases, and the convergence for the remaining green traits are proven via the induction step. This is Step 1a) in the proof. Then in Step 1b) we make the induction step in the direction of HGT which is shown in the second image. For this we assume the result to be shown for all yellow traits. Then we have the base cases in red and subsequently another induction step again in green.

Proof of Theorem 2.2. We distinguish two cases: Either there is only one phase (that is, all traits are unfit against the initially resident trait $(0,0)$) or there are at least two phases. In either case we now consider a fixed time $T > 0$.

4.1. Proof of Theorem 2.2, Case 1.

Case 1: $S((x,y), (0,0)) \leq 0$ for all $(x,y) \in \mathcal{X}$: As in [CMT21], we define a time θ_1^K , during which trait $(0,0)$ is resident. Let $\varepsilon_1 > 0$ and $\rho > 0$ and define the time

$$\theta_1^K := \inf \left\{ t \geq 0 \mid N_{0,0}^K(t \log K) \notin \left[\left(\frac{3}{C} - 3\varepsilon_1 \right) K, \left(\frac{3}{C} + 3\varepsilon_1 \right) K \right] \right. \\ \left. \text{or } \sum_{(m,n) \neq (0,0)} N_{m,n}^K(t \log K) \geq \rho \varepsilon_1 K \right\}.$$

Then we easily calculate that $\beta_{m,n}(t) = (1 - (n+m)\alpha) \vee 0$ for $t \leq \theta_1^K \wedge T$.

Step 0: Deriving bounds on the rates. Since all of the following couplings in Step 1 will need some bounds on the rates, we derive them here for our process up to time θ_1^K . In particular, the arrivals in the active component of trait (m,n) due to reproduction or transfer at time $t \leq \theta_1^K \wedge T$ occur at rate

$$N_{m,n}^{K,a}(t \log K) \left(\left(4 - \frac{(m+n)\delta}{2} \right) (1 - K^{-\alpha}) + \tau \frac{\sum_{m'=0, n' < n}^L N_{m',n'}^{K,a}(t \log K)}{\sum_{m',n'=0}^L N_{m',n'}^{K,a}(t \log K)} \right),$$

which satisfies for K large enough and $n > 0$ the inequality

$$\begin{aligned} & N_{m,n}^{K,a}(t \log K) \left(4 - \frac{(m+n)\delta}{2} - \varepsilon_1 + \tau \frac{3 - 3C\varepsilon_1}{3 + C(3 + \rho)\varepsilon_1} \right) \\ & \leq N_{m,n}^{K,a}(t \log K) \left(\left(4 - \frac{(m+n)\delta}{2} \right) (1 - K^{-\alpha}) + \tau \frac{\sum_{m'=0, n' < n}^L N_{m',n'}^{K,a}(t \log K)}{\sum_{m',n'=0}^L N_{m',n'}^{K,a}(t \log K)} \right) \\ & \leq N_{m,n}^{K,a}(t \log K) \left(4 - \frac{(m+n)\delta}{2} + \tau \right). \end{aligned}$$

These bounds are true for $n = 0$ when we remove the terms involving τ . The arrivals due to incoming mutations occur at rate

$$(N_{m-1,n}^{K,a}(t \log K) + N_{m,n-1}^{K,a}(t \log K)) \left(4 - \frac{(m+n-1)\delta}{2} \right) \frac{K^{-\alpha}}{2}.$$

Further, the departures from the active population due to transfer or death occur at rate

$$N_{m,n}^{K,a}(t \log K) \left(1 + \frac{C(1-pm\delta)}{K} \sum_{m',n'=0}^L N_{m',n'}^{K,a}(t \log K) + \tau \frac{\sum_{m'=0,n'>n}^L N_{m',n'}^{K,a}(t \log K)}{\sum_{m',n'=0}^L N_{m',n'}^{K,a}(t \log K)} \right),$$

which can be bounded for ρ small enough by

$$\begin{aligned} & N_{m,n}^{K,a}(t \log K) (4 - 3pm\delta - 3C(1-pm\delta)\varepsilon_1) \\ & \leq N_{m,n}^{K,a}(t \log K) \left(1 + \frac{C(1-pm\delta)}{K} \sum_{m',n'=0}^L N_{m',n'}^{K,a}(t \log K) + \tau \frac{\sum_{m'=0,n'>n}^L N_{m',n'}^{K,a}(t \log K)}{\sum_{m',n'=0}^L N_{m',n'}^{K,a}(t \log K)} \right) \\ & \leq N_{m,n}^{K,a}(t \log K) \left(4 - 3pm\delta + (3C + \rho)(1-pm\delta)\varepsilon_1 + \tau \frac{C\rho\varepsilon_1}{3-3\varepsilon_1} \right). \end{aligned}$$

Similarly, the active to dormant transfer rate is given by

$$N_{m,n}^{K,a}(t \log K) \frac{Cpm\delta}{K} \sum_{m',n'=0}^L N_{m',n'}^{K,a}(t \log K),$$

which satisfies the bounds

$$\begin{aligned} & N_{m,n}^{K,a}(t \log K) (3pm\delta - 3Cpm\delta\varepsilon_1) \\ & \leq N_{m,n}^{K,a}(t \log K) \frac{Cpm\delta}{K} \sum_{m',n'=0}^L N_{m',n'}^{K,a}(t \log K) \\ & \leq N_{m,n}^{K,a}(t \log K) (3pm\delta + Cpm\delta(3 + \rho)\varepsilon_1). \end{aligned}$$

Step 1: Induction on the traits. We will now show by induction on m and n that the bounds

$$K^{\beta_{m,n}(t)-(m+n+1)\varepsilon_1} - 1 \leq N_{m,n}^K(t \log K) \leq K^{\beta_{m,n}(t)+(m+n+1)\varepsilon_1} - 1 \quad (4.1)$$

hold true for $t \leq \theta_1^K \wedge T$. In this situation, the condition reads as

$$K^{((1-(m+n)\alpha) \vee 0) - (m+n+1)\varepsilon_1} - 1 \leq N_{m,n}^K(t \log K) \leq K^{((1-(m+n)\alpha) \vee 0) + (m+n+1)\varepsilon_1} - 1$$

Step 1a): Traits $(0, y)$. For $m = n = 0$ this is obviously satisfied by definition of θ_1^K .

Base case: $n = 0$.

Base case: $m = 1$. For $m = 1$ and $n = 0$ we couple the process $N_{1,0}^K$ with processes $\tilde{Z}_{1,0}^K$ and $\hat{Z}_{1,0}^K$, such that component-wise

$$\tilde{Z}_{1,0}^K(t \log K) \leq N_{1,0}^K(t \log K) \leq \hat{Z}_{1,0}^K(t \log K),$$

where $\tilde{Z}_{1,0}^K$ is a $BBPI_K(4 - \frac{\delta}{2} - \bar{C}\varepsilon_1, 0, 4 - 3p\delta + 2\bar{C}\varepsilon_1, \kappa, 3p\delta - \bar{C}\varepsilon_1, \sigma, 0, 1 - \alpha - \varepsilon_1, 1 - \alpha - \varepsilon_1)$ and $\hat{Z}_{1,0}^K$ is a $BBPI_K(4 - \frac{\delta}{2} + \bar{C}\varepsilon_1, 0, 4 - 3p\delta - 2\bar{C}\varepsilon_1, \kappa, 3p\delta + \bar{C}\varepsilon_1, \sigma, 0, 1 - \alpha + \varepsilon_1, 1 - \alpha + \varepsilon_1)$, where

$$\bar{C} := 1 + (1 \vee \tau)(1 + C)(6 + \rho).$$

Indeed, this coupling is justified by the bounds on the rates derived in Step 0 of this proof. Hence, applying Theorem A.3 (i), we see that

$$\begin{aligned} & \lim_{K \rightarrow \infty} \frac{\log(1 + \tilde{Z}_{1,0}^K(t \log K))}{\log K} \\ &= (1 - \alpha - \varepsilon_1) + \left(0 \vee \frac{-\frac{\delta}{2} - 2\bar{C}\varepsilon_1 - (\kappa + \sigma) + \sqrt{(-\frac{\delta}{2} - 2\bar{C}\varepsilon_1 + \kappa + \sigma)^2 + 4(3p\delta - \bar{C}\varepsilon_1)\sigma}}{2} \cdot t \right) \\ &\geq 1 - \alpha - \varepsilon_1 \end{aligned}$$

and similarly, since $S((\delta, 0), (0, 0)) < 0$, we have for ε_1 sufficiently small that

$$\begin{aligned} & \lim_{K \rightarrow \infty} \frac{\log(1 + \hat{Z}_{1,0}^K(t \log K))}{\log K} \\ &= (1 - \alpha + \varepsilon_1) + \left(0 \vee \frac{-\frac{\delta}{2} + 2\bar{C}\varepsilon_1 - (\kappa + \sigma) + \sqrt{(-\frac{\delta}{2} + 2\bar{C}\varepsilon_1 + \kappa + \sigma)^2 + 4(3p\delta + \bar{C}\varepsilon_1)\sigma}}{2} \cdot t \right) \\ &\leq 1 - \alpha + \varepsilon_1. \end{aligned}$$

Thus, the claim is shown for $N_{1,0}^K$. For the remainder of the proof, we will use the shorthand notation $S((\tilde{x}, \tilde{y}), (x, y)) \pm C_*\varepsilon_1$ to indicate the rate of growth of a bi-type branching process whose birth, death, switching and transfer rates are modified by some factor of ε_1 and otherwise coincide with those of a bi-type branching process whose growth rate is given by $S((\tilde{x}, \tilde{y}), (x, y))$.

Induction step for $m - 1 \rightarrow m$, $n = 0$: Now assume that it has been shown that

$$K^{((1-(m-1)\alpha)\vee 0) - m\varepsilon_1} - 1 \leq N_{m-1,0}^K(t \log K) \leq K^{((1-(m-1)\alpha)\vee 0) + m\varepsilon_1} - 1. \quad (4.2)$$

Then, we can couple the process $N_{m,0}^K(t \log K)$ with different processes $\tilde{Z}_{m,0}^K$ and $\hat{Z}_{m,0}^K$ such that

$$\tilde{Z}_{m,0}^K(t \log K) \leq N_{m,0}^K(t \log K) \leq \hat{Z}_{m,0}^K(t \log K),$$

where the distribution of $\tilde{Z}_{m,0}^K$ is determined by $BBPI_K(4 - \frac{m\delta}{2} - \bar{C}\varepsilon_1, 0, 4 - 3pm\delta + 2\bar{C}\varepsilon_1, \kappa, 3pm\delta - \bar{C}\varepsilon_1, \sigma, 0, (1 - (m-1)\alpha)_+ - \alpha - m\varepsilon_1, (1 - m\alpha - m\varepsilon_1)_+)$ and the distribution of $\hat{Z}_{m,0}^K$ is $BBPI_K(4 - \frac{m\delta}{2} + \bar{C}\varepsilon_1, 0, 4 - 3pm\delta - 2\bar{C}\varepsilon_1, \kappa, 3pm\delta + \bar{C}\varepsilon_1, \sigma, 0, (1 - (m-1)\alpha)_+ - \alpha + m\varepsilon_1, (1 - m\alpha + m\varepsilon_1)_+)$.

Then, we see from Theorem A.3 (i) or (iii) applied accordingly that for $\varepsilon_1 > 0$ small enough

$$\begin{aligned} & \lim_{K \rightarrow \infty} \frac{\log(1 + \tilde{Z}_{m,0}^K(t \log K))}{\log K} \\ &= \begin{cases} ((1 - (m-1)\alpha)_+ - \alpha - m\varepsilon_1) \vee (1 - m\alpha - m\varepsilon_1)_+ + (S((m\delta, 0), (0, 0)) - C_*\varepsilon_1)t, & \text{if } m\alpha < 1 \\ 0, & \text{otherwise} \end{cases} \\ &\geq \beta_{m,0}(t) - m\varepsilon_1 \end{aligned}$$

and similarly

$$\begin{aligned} & \lim_{K \rightarrow \infty} \frac{\log(1 + \widehat{Z}_{m,0}^K(t \log K))}{\log K} \\ &= \begin{cases} ((1 - (m-1)\alpha)_+ - \alpha + m\varepsilon_1) \vee ((1 - m\alpha + m\varepsilon_1)_+ + (S((m\delta, 0), (0, 0)) + C_*\varepsilon_1)t), & \text{if } m\alpha < 1 \\ 0, & \text{otherwise,} \end{cases} \\ &\leq \beta_{m,0}(t) + m\varepsilon_1. \end{aligned}$$

Step 1b): Traits $(m\delta, n\delta)$. Induction step for $n-1 \rightarrow n$: Assume now that for all $m \in \{0, \dots, L\}$ and all $n' \leq n-1$ the bounds

$$K^{((1-(m+n')\alpha)\vee 0) - (m+n'+1)\varepsilon_1} - 1 \leq N_{m,n'}^K(t \log K) \leq K^{((1-(m+n')\alpha)\vee 0) + (m+n'+1)\varepsilon_1} - 1$$

hold.

Base case $m = 0$: For the process $N_{0,n}^K$ we only have incoming migration from $N_{0,n-1}^K$. Hence, we can couple

$$\widetilde{Z}_{0,n}^K(t \log K) \leq N_{0,n}^K(t \log K) \leq \widehat{Z}_{0,n}^K(t \log K),$$

where $\widetilde{Z}_{0,n}^K$ is a $BPI_K(4 - \frac{n\delta}{2} + \tau - \bar{C}\varepsilon_1, 4 + \bar{C}\varepsilon_1, 0, (1 - (n-1)\alpha)_+ - \alpha - n\varepsilon_1, (1 - n\alpha - n\varepsilon_1)_+)$ and $\widehat{Z}_{0,n}^K$ is a $BPI_K(4 - \frac{n\delta}{2} + \tau + \bar{C}\varepsilon_1, 4 - \bar{C}\varepsilon_1, 0, (1 - (n-1)\alpha)_+ - \alpha + \varepsilon_1, (1 - n\alpha + n\varepsilon_1)_+)$, which as before gives the sought bounds from (4.1) by applying Theorem A.1 (i) or (iii).

Base case $m = 1$: Here, we have incoming mutations as mentioned in the beginning of the proof from two different populations, for which we already have suitable bounds. Thus, we can couple the process $N_{1,n}^K$ as usual with

$$\widetilde{Z}_{1,n}^K(t \log K) \leq N_{1,n}^K(t \log K) \leq \widehat{Z}_{1,n}^K(t \log K),$$

where the distribution of $\widetilde{Z}_{1,n}^K$ is $BBPI_K(4 - \frac{(n+1)\delta}{2} + \tau - \bar{C}\varepsilon_1, 0, 4 - 3p\delta + 2\bar{C}\varepsilon_1, \kappa, 3p\delta - \bar{C}\varepsilon_1, \sigma, 0, (1 - n\alpha)_+ - \alpha - (n+1)\varepsilon_1, (1 - (n+1)\alpha - (n+1)\varepsilon_1)_+)$ and the law of $\widehat{Z}_{1,n}^K$ is determined by a $BBPI_K(4 - \frac{(n+1)\delta}{2} + \tau + \bar{C}\varepsilon_1, 0, 4 - 3p\delta - 2\bar{C}\varepsilon_1, \kappa, 3p\delta + \bar{C}\varepsilon_1, \sigma, 0, (1 - n\alpha)_+ - \alpha + (n+1)\varepsilon_1, (1 - (n+1)\alpha + (n+1)\varepsilon_1)_+)$. Applying Theorem A.3 (i) or (iii) as in the case $m = 1, n = 0$ yields the claim (4.1).

Induction step $m-1 \rightarrow m$: Assume that we have shown the bounds

$$K^{((1-(m'+n)\alpha)\vee 0) - (m'+n+1)\varepsilon_1} - 1 \leq N_{m',n}^K(t \log K) \leq K^{((1-(m'+n)\alpha)\vee 0) + (m'+n+1)\varepsilon_1} - 1$$

for all $m' \leq m-1$. Then, we couple the process $N_{m,n}^K$ as usual with

$$\widetilde{Z}_{m,n}^K(t \log K) \leq N_{m,n}^K(t \log K) \leq \widehat{Z}_{m,n}^K(t \log K),$$

where the distribution of $\widetilde{Z}_{m,n}^K$ is $BBPI_K(4 - \frac{(m+n)\delta}{2} + \tau - \bar{C}\varepsilon_1, 0, 4 - 3pm\delta + 2\bar{C}\varepsilon_1, \kappa, 3pm\delta - \bar{C}\varepsilon_1, \sigma, 0, (1 - (m+n-1)\alpha)_+ - \alpha - (m+n)\varepsilon_1, (1 - (m+n)\alpha - (m+n)\varepsilon_1)_+)$ and the law of $\widehat{Z}_{m,n}^K$ is $BBPI_K(4 - \frac{(m+n)\delta}{2} + \tau + \bar{C}\varepsilon_1, 0, 4 - 3pm\delta - 2\bar{C}\varepsilon_1, \kappa, 3pm\delta + \bar{C}\varepsilon_1, \sigma, 0, (1 - (m+n-1)\alpha)_+ - \alpha + (m+n)\varepsilon_1, (1 - (m+n)\alpha + (m+n)\varepsilon_1)_+)$. Again, applying Theorem A.3 (i) or (iii) as before gives the claim (4.1).

Step 2: Showing $\theta_1^K \geq T$. In Step 1 we have shown that the process $\frac{\log(1 + N_{m,n}^K(t \log K))}{\log K}$ converges in probability in the space $L^\infty([0, \theta_1^K \wedge T])$ towards $\beta_{n,m}(t) = (1 - (n+m)\alpha)_+$. It now suffices to show

that $\theta_1^K \geq T$, which can be done in the same manner as in [CMT21]. As we have computed above, for all $t \leq \theta_1^K \wedge T$ we have with high probability

$$\sum_{(m,n) \neq (0,0)} N_{m,n}^K(t \log K) \leq K^{\max_{(m,n) \neq (0,0)} \beta_{m,n}(t) + \frac{\alpha}{2}} = K^{1 - \frac{\alpha}{2}}.$$

In particular at time $t = \theta_1^K \wedge T$, we have $\sum_{(m,n) \neq (0,0)} N_{m,n}^K(t \log K) < \rho \varepsilon_1 K$ with high probability. Hence, for K large enough, we see that up to time $\theta_1^K \wedge 2T$ we can couple

$$Z_{0,0,1}^K(t \log K) \leq N_{0,0}^K(t \log K) \leq Z_{0,0,2}^K(t \log K),$$

where $Z_{0,0,1}^K$ is a $LBDI_K(4(1-\varepsilon), 1+C\varepsilon, C, 0)$ and $Z_{0,0,2}^K$ is a $LBDI_K(4, 1, C, 0)$. Applying [CMT21, Lemma C.1 (i)] to both processes shows that at time $t = \theta_1^K \wedge T$ the process $N_{0,0}^K(t \log K)$ is still close to its equilibrium size with high probability. Thus $\theta_1^K > T$ with probability converging to 1 as $K \rightarrow \infty$ and the proof is completed in this case.

4.2. Proof of Theorem 2.2, Case 2.

Case 2: In the second case, we consider $S((x, y), (0, 0)) > 0$ for some $(x, y) \in \mathcal{X}$.

4.2.1. Phase 1. Note that the bounds on the arrival, departure and migration rates derived in Step 0 of Case 1 remain true. Unfortunately, we are not able to give a closed form for the limiting function β as in [CMT21, Section 4.2.2] without a significant number of cases to be distinguished. However, in this case there exists some time $s_1 \leq T_0$, at which for the first time for some $m_2^*, n_2^* \in \{0, \dots, L\}$ we have $\beta_{m_2^*, n_2^*}(s_1) = \beta_{0,0}(s_1) = 1$. Due to our assumptions in the Theorem, m_2^*, n_2^* are unique. Now, we split the time interval $[0, s_1]$ into subintervals, on which all $\beta_{m,n}$ are affine functions. That is, there exists a finite number of times $0 = t_0 < t_1 < \dots < t_\ell \leq s_1$ such that on the interval $[t_{i-1}, t_i]$ all functions $\beta_{m,n}$ are of the form

$$\beta_{m,n}(t) = \beta_{m,n}(t_{i-1}) + a_{m,n}(t - t_{i-1}), \quad t \in [t_{i-1}, t_i],$$

for some constants $a_{m,n}$ which may depend on the time interval. This representation as an affine linear function can be seen from Theorem 2.2 (iii). We will now show by induction, first on n and then on m , that $\beta_{m,n}^K \rightarrow \beta_{m,n}$ as $K \rightarrow \infty$ on the interval $[0, t_1 \wedge \theta_1^K \wedge T]$. Showing the convergence on the other intervals $[t_{i-1}, t_i \wedge \theta_1^K \wedge T]$ can be done in the same way.

Step 1: The induction on the traits. Recalling the time $t_{(m,n),1}$ from Theorem 2.2 (iii), we want to show that for all $t \in [0, t_1 \wedge \theta_1^K \wedge T]$ the bounds

$$\beta_{m,n}(t) - C_* \varepsilon_1 \leq \frac{\log(1 + N_{m,n}^K(t \log K))}{\log K} \leq \beta_{m,n}(t) + C_* \varepsilon_1 \quad (4.3)$$

hold with probability converging to 1 as $K \rightarrow \infty$ for some constant C_* , which may depend on m and n . In particular, the notation C_* does not necessarily refer to any particular constant, but more to a suitable constant, which is sufficiently large. For $m = n = 0$, the bounds hold trivially by definition of θ_1^K .

In the following, we use the notation

$$(\beta_{m,n}(t) + C_* \varepsilon_1)_\times := \begin{cases} \beta_{m,n}(t) + C_* \varepsilon_1, & \text{if } \beta_{m,n}(t) > 0 \\ 0, & \text{otherwise.} \end{cases}$$

Step 1a): Traits $(0, y)$. Base case: $n = 0$.

Base case: $m = 1$. We can couple as in Case 1 the process $N_{1,0}^K$ with processes $\tilde{Z}_{1,0}^K$ and $\hat{Z}_{1,0}^K$, such that component-wise

$$\tilde{Z}_{1,0}^K(t \log K) \leq N_{1,0}^K(t \log K) \leq \hat{Z}_{1,0}^K(t \log K),$$

where the distribution of $\tilde{Z}_{1,0}^K$ is determined by $BBPI_K(4 - \frac{\delta}{2} - \bar{C}\varepsilon_1, 0, 4 - 3p\delta + 2\bar{C}\varepsilon_1, \kappa, 3p\delta - \bar{C}\varepsilon_1, \sigma, 0, \beta_{0,0}(0) - \alpha - C_*\varepsilon_1, \beta_{1,0}(0) - C_*\varepsilon_1)$ and $\hat{Z}_{1,0}^K$ is a $BBPI_K(4 - \frac{\delta}{2} + \bar{C}\varepsilon_1, 0, 4 - 3p\delta - 2\bar{C}\varepsilon_1, \kappa, 3p\delta + \bar{C}\varepsilon_1, \sigma, 0, \beta_{0,0}(0) - \alpha + C_*\varepsilon_1, \beta_{1,0}(0) + C_*\varepsilon_1)$, where again

$$\bar{C} := 1 + (1 \vee \tau)(1 + C)(6 + \rho).$$

Obviously, we obtain the same convergence as before, but the inequalities derived may not apply anymore. By Theorem A.3 (i), we have the convergence

$$\begin{aligned} \lim_{K \rightarrow \infty} \frac{\log(1 + \tilde{Z}_{1,0}^K(t \log K))}{\log K} &= (\beta_{0,0}(0) - \alpha - C_*\varepsilon_1) \vee (\beta_{1,0}(0) - C_*\varepsilon_1 + S((\delta, 0), (0, 0))t) \\ &= (\beta_{1,0}(0) + S((\delta, 0), (0, 0))t) \vee (\beta_{0,0}(t) - \alpha) - C_*\varepsilon_1 \end{aligned}$$

and

$$\begin{aligned} \lim_{K \rightarrow \infty} \frac{\log(1 + \hat{Z}_{1,0}^K(t \log K))}{\log K} &= (\beta_{0,0}(0) - \alpha + C_*\varepsilon_1) \vee (\beta_{1,0}(0) + C_*\varepsilon_1 + S((\delta, 0), (0, 0))t) \\ &= (\beta_{1,0}(0) + S((\delta, 0), (0, 0))t) \vee (\beta_{0,0}(t) - \alpha) + C_*\varepsilon_1, \end{aligned}$$

by using $\beta_{1,0}(0) = \beta_{0,0}(0) - \alpha$.

Induction step for $m-1 \rightarrow m$, $n=0$: Now, assume that (4.3) has been shown for all $m' \leq m-1$ and $n=0$. Our goal is to show that (4.3) also holds for m and $n=0$. For this purpose, we couple with $\tilde{Z}_{m,0}^K$ and $\hat{Z}_{m,0}^K$ such that

$$\tilde{Z}_{m,0}^K(t \log K) \leq N_{m,0}^K(t \log K) \leq \hat{Z}_{m,0}^K(t \log K),$$

where the distribution of $\tilde{Z}_{m,0}^K$ is determined by $BBPI_K(4 - \frac{m\delta}{2} - \bar{C}\varepsilon_1, 0, 4 - 3pm\delta + 2\bar{C}\varepsilon_1, \kappa, 3pm\delta - \bar{C}\varepsilon_1, \sigma, a_{m-1,0}, \beta_{m-1,0}(0) - \alpha - C_*\varepsilon_1, (\beta_{m,0}(0) - C_*\varepsilon_1)_+)$ and the law of $\hat{Z}_{m,0}^K$ is $BBPI_K(4 - \frac{m\delta}{2} + \bar{C}\varepsilon_1, 0, 4 - 3pm\delta - 2\bar{C}\varepsilon_1, \kappa, 3pm\delta + \bar{C}\varepsilon_1, \sigma, a_{m-1,0}, \beta_{m-1,0}(0) - \alpha + C_*\varepsilon_1, (\beta_{m,0}(0) + C_*\varepsilon_1)_\times)$.

We distinguish the cases where $\beta_{m,0}(0) - C_*\varepsilon_1$ is strictly positive (to apply Theorem A.3 (i)) or $(\beta_{m,0}(0) - C_*\varepsilon_1)_+ = 0$ to apply Theorem A.3 (ii) or (iii), depending on $a_{m-1,0}$ being strictly positive or non-positive, which yields the convergence

$$\begin{aligned} &\lim_{K \rightarrow \infty} \frac{\log(1 + \tilde{Z}_{m,0}^K(t \log K))}{\log K} \\ &= \begin{cases} (\beta_{m-1,0}(0) - \alpha - C_*\varepsilon_1 + a_{m-1,0}t) \vee (\beta_{m,0}(0) - C_*\varepsilon_1 + S((m\delta, 0), (0, 0))t) \vee 0, & \text{if } \beta_{m,0}(0) - C_*\varepsilon_1 > 0 \\ (S((m\delta, 0), (0, 0)) \vee a_{m-1,0}) \left(t - \frac{|\beta_{m-1,0}(0) - \alpha - C_*\varepsilon_1|}{a_{m-1,0}} \right) \vee 0, & \text{if } \beta_{m,0}(0) - C_*\varepsilon_1 \leq 0 < a_{m-1,0} \\ 0, & \text{if } \beta_{m,0}(0) - C_*\varepsilon_1, a_{m-1,0} \leq 0. \end{cases} \end{aligned}$$

Note that in the second case it holds $\frac{|\beta_{m-1,0}(0) - \alpha - C_*\varepsilon_1|}{a_{m-1,0}} = t_{(m,0),1} + C_*\varepsilon_1$. Even though the functions $\beta_{m,n}$ do not have a change in slope by our choice of t_1 , this limit of the coupled process may. Also, if the maximum in the second case is attained by $a_{m-1,0}$, then the second case reads as

$$a_{m-1,0}t + \beta_{m-1,0}(0) - \alpha - C_*\varepsilon_1$$

by using that $\beta_{m-1,0}(0) - \alpha - C_*\varepsilon_1 \leq 0$. In particular, using $\beta_{m-1,0}(t) = \beta_{m-1,0}(0) + a_{m-1,0}t$, we obtain in each case

$$\lim_{K \rightarrow \infty} \frac{\log\left(1 + \tilde{Z}_{m,0}^K(t \log K)\right)}{\log K} \geq [(\beta_{m,0}(0) + S((m\delta, 0), (0, 0))(t - (t \wedge t_{(m,0),1}))) \vee (\beta_{m-1,0}(t) - \alpha) \vee 0] - C_*\varepsilon_1.$$

A similar application of Theorem A.3 for $\hat{Z}_{m,0}^K$ entails

$$\lim_{K \rightarrow \infty} \frac{\log\left(1 + \hat{Z}_{m,0}^K(t \log K)\right)}{\log K} \geq [(\beta_{m,0}(0) + S((m\delta, 0), (0, 0))(t - (t \wedge t_{(m,0),1}))) \vee (\beta_{m-1,0}(t) - \alpha) \vee 0] + C_*\varepsilon_1,$$

which finishes the induction for $n = 0$.

Step 1b): Traits $(m\delta, n\delta)$. Induction step: $n - 1 \rightarrow n$. We assume the bounds in (4.3) have been shown for all $m' \in \{0, \dots, L\}$ and $n' \leq n - 1$.

Base case: $m = 0$. Here, the immigration is only coming from $N_{0,n-1}^K$ and hence we can couple with processes $\tilde{Z}_{0,n}^K$ and $\hat{Z}_{0,n}^K$ such that

$$\tilde{Z}_{0,n}^K(t \log K) \leq N_{0,n}^K(t \log K) \leq \hat{Z}_{0,n}^K(t \log K),$$

where $\tilde{Z}_{0,n}^K$ is a $BPI_K(4 - \frac{n\delta}{2} + \tau - \bar{C}\varepsilon_1, 4 + 2\bar{C}\varepsilon_1, a_{0,n-1}, \beta_{0,n-1}(0) - \alpha - C_*\varepsilon_1, (\beta_{0,n}(0) - C_*\varepsilon_1)_+)$ and the law of $\hat{Z}_{0,n}^K$ is $BPI_K(4 - \frac{n\delta}{2} + \tau + \bar{C}\varepsilon_1, 4 - 2\bar{C}\varepsilon_1, a_{0,n-1}, \beta_{0,n-1}(0) - \alpha + C_*\varepsilon_1, (\beta_{0,n}(0) + C_*\varepsilon_1)_\times)$. Now applying Theorem A.1 shows (4.3) in this case.

Base case: $m = 1$. This case can be treated as the induction step below.

Induction step: $m - 1 \rightarrow m$. Now, we assume that for all $m' \leq m - 1$ we have shown the inequality (4.3). Then, it also holds for m since we can again distinguish the immigration from outside to be dominated either from $N_{m-1,n}^K$ or from $N_{m,n-1}^K$ and then we can couple as usual (in the case that the immigration is dominated by $N_{m-1,n}^K$) with processes $\tilde{Z}_{m,n}^K$ and $\hat{Z}_{m,n}^K$ such that

$$\tilde{Z}_{m,n}^K(t \log K) \leq N_{m,n}^K(t \log K) \leq \hat{Z}_{m,n}^K(t \log K),$$

where $\tilde{Z}_{m,n}^K$ is given by a $BBPI_K(4 - \frac{(m+n)\delta}{2} + \tau - \bar{C}\varepsilon_1, 0, 4 - 3pm\delta + 2\bar{C}\varepsilon_1, \kappa, 3pm\delta - \bar{C}\varepsilon_1, \sigma, a_{m-1,n}, \beta_{m-1,n}(0) - \alpha - C_*\varepsilon_1, (\beta_{m,n}(0) - C_*\varepsilon_1)_+)$ and the law of $\hat{Z}_{m,n}^K$ is determined by $BBPI_K(4 - \frac{(m+n)\delta}{2} + \tau + \bar{C}\varepsilon_1, 0, 4 - 3pm\delta - 2\bar{C}\varepsilon_1, \kappa, 3pm\delta + \bar{C}\varepsilon_1, \sigma, a_{m-1,n}, \beta_{m-1,n}(0) - \alpha + C_*\varepsilon_1, (\beta_{m,n}(0) + C_*\varepsilon_1)_\times)$. As before, applying Theorem A.3 in each case shows the claimed inequality (4.3) with probability converging to 1 as $K \rightarrow \infty$. This finishes the induction for the first phase.

Performing the induction in n and m as above also for the remaining intervals $[t_{i-1}, t_i \wedge \theta_1^K \wedge T]$ shows the bounds (4.3) on the entire interval $[0, s_1 \wedge \theta_1^K \wedge T]$ with probability converging to 1 due to the Markov property. The only changes that need to be made are in the starting conditions of the coupled processes, where we replace $\beta_{m,n}(0)$ by $\beta_{m,n}(t_{i-1})$.

Step 2: Deriving a lower bound for θ_1^K . Next, we will show that $(s_1 - \eta) \wedge T < \theta_1^K$ for any $\eta > 0$ with high probability. Assume for now $T > s_1 - \eta$. By definition of s_1 , all functions $\beta_{m,n}$ are bounded away

from 1 on the interval $[0, s_1 - \eta]$ except for $\beta_{0,0}$. Hence for all $t \leq \theta_1^K \wedge (s_1 - \eta)$ we have

$$\sum_{(m,n) \neq (0,0)} N_{m,n}^K(t \log K) \leq K^{\max_{(m,n) \neq (0,0)} \beta_{m,n}(t) + \tilde{\varepsilon}} \leq K^{1-\tilde{\varepsilon}}$$

for $\tilde{\varepsilon} > 0$ sufficiently small with probability converging to 1. Hence, to show $s_1 - \eta < \theta_1^K$, we also need to exclude the possibility of $N_{0,0}^K$ exiting a neighbourhood of its equilibrium size. Indeed, we can couple the process $N_{0,0}^K$ with processes

$$Z_{0,0,1}^K(t \log K) \leq N_{0,0}^K(t \log K) \leq Z_{0,0,2}^K(t \log K),$$

where $Z_{0,0,1}^K$ is a $LBDI_K(4(1-\varepsilon), 1 + C\varepsilon, C, 0)$ and $Z_{0,0,2}^K$ is a $LBDI_K(4, 1, C, 0)$ and $\varepsilon > 0$. As in case 1, applying [CMT21, Lemma C.1 (i)] to both processes shows that at time $t = \theta_1^K \wedge (s_1 - \eta)$ the process $N_{0,0}^K(t \log K)$ is close to its equilibrium size with high probability. Thus, recalling $T > s_1 - \eta$, $\theta_1^K > (s_1 - \eta) \wedge T$ with probability converging to 1 as $K \rightarrow \infty$. In particular, for $T < s_1 - \eta$ it holds $\theta_1^K > (s_1 - \eta) \wedge T$ with high probability.

Therefore, we can conclude by letting $\varepsilon_1 \downarrow 0$ that the convergence $\beta_{m,n}^K \rightarrow \beta_{m,n}$ in probability as $K \rightarrow \infty$ on the interval $[0, (s_1 - \eta) \wedge T]$ holds true.

4.2.2. Intermediate Phase 1. In this intermediate phase, we will show that the resident trait $(0, 0)$ experiences competition with an invasive trait $(m_2^* \delta, n_2^* \delta)$, which we will show to be of order K at the end of the first phase $[0, \theta_1^K \log K]$. Hence our goal is twofold: Firstly we want to show that $\theta_1^K \rightarrow s_1$ as $K \rightarrow \infty$ in probability. Secondly, we want to show that at some time $\sigma_2^K \log K = \theta_1^K \log K + T(\varepsilon_1, \rho)$ the competition leads to the invasive trait becoming resident and its size being close to its equilibrium size. At the same time $N_{0,0}^K$ becomes smaller than $\rho \varepsilon_1 K$.

Step 1: Convergence of $\theta_1^K \rightarrow s_1$. We know from the end of the previous section where we proved $\beta_{m,n}^K \rightarrow \beta_{m,n}$ on the interval $[0, (s_1 - \eta) \wedge T]$ that $s_1 - \eta < \theta_1^K$ with high probability. Thus, to show $\theta_1^K \rightarrow s_1$, it suffices to show $\theta_1^K < s_1 + \eta$ for any $\eta > 0$ with probability converging to 1 as $K \rightarrow \infty$.

Towards a contradiction, assume that $\theta_1^K \geq s_1 + \eta$. Then, the couplings on the interval $[t_{\ell-1}, t_\ell \wedge \theta_1^K]$ with $t_\ell = s_1$ can be extended until time $t_\ell^* = s_1 + \eta$ since the couplings are valid as long as the time t satisfies $t \leq \theta_1^K$. In particular, for the coupling of $N_{m_2^*, n_2^*}^K$ we obtain the lower bound

$$\begin{aligned} & \lim_{K \rightarrow \infty} \frac{\log(1 + N_{m_2^*, n_2^*}^K(t \log K))}{\log K} \\ & \geq \left[(\beta_{m_2^*, n_2^*}(t_{\ell-1}) + S((m_2^* \delta, n_2^* \delta), (0, 0))((t - t_{\ell-1}) - ((t - t_{\ell-1}) \wedge t_{(m_2^*, n_2^*), 1}))) \right. \\ & \quad \vee (\beta_{m_2^*-1, n_2^*}(t_{\ell-1}) + a_{m_2^*-1, n_2^*}(t - t_{\ell-1}) - \alpha) \\ & \quad \left. \vee (\beta_{m_2^*, n_2^*-1}(t_{\ell-1}) + a_{m_2^*, n_2^*-1}(t - t_{\ell-1}) - \alpha) \vee 0 \right] - C_* \varepsilon_1. \end{aligned}$$

We know however that at time s_1 the last expression converges to 1 as $\varepsilon_1 \downarrow 0$ and by definition of s_1 the lower bound is strictly increasing on some interval $[s_1 - \eta, s_1]$. Since the lower bound is the maximum of different affine functions, it remains strictly increasing on the interval $[s_1, s_1 + \eta]$. In particular, for ε_1 small enough, at time $s_1 + \eta$ the lower bound becomes larger than 1, which is a contradiction since

$$\lim_{K \rightarrow \infty} \frac{\log(1 + N_{m_2^*, n_2^*}^K(t \log K))}{\log K} \leq 1$$

for all $t \geq 0$. Hence with probability converging to 1 we have $\theta_1^K < s_1 + \eta$. We conclude $\theta_1^K \rightarrow s_1$ in probability.

Step 2: Emergence of a new population. For our second goal, we need to show that $N_{0,0}^K$ does not exit a neighbourhood of its equilibrium, so that at time θ_1^K the population of trait $(m_2^*\delta, n_2^*\delta)$ emerges. This part of the proof is identical to [CMT21, Section 4.2.3], but is repeated here for the reader's convenience. Unfortunately, we cannot use the coupling from the previous phase anymore since $\theta_1^K \rightarrow s_1$ and therefore $\beta_{m,n}(\theta_1^K)$ are not bounded away from 1. However, we do know that for K sufficiently large, the emigration from trait $(0,0)$, which occurs at rate $4K^{-\alpha}$, can be bounded by $C\rho\varepsilon_1$. Then, on the time interval $[0, \theta_1^K \wedge T]$, we can couple

$$Z_{0,0,1}^K(t \log K) \leq N_{0,0}^K(t \log K) \leq Z_{0,0,2}^K(t \log K),$$

where $Z_{0,0,1}^K$ is a $LBDI_K(4 - C\rho\varepsilon_1, 1 + \frac{\tau\rho\varepsilon_1}{3/C - 3\varepsilon_1} + C\rho\varepsilon_1, C, 0)$ and $Z_{0,0,2}^K$ is a $LBDI_K(4, 1, C, 0)$. We easily identify the equilibria

$$\bar{z}_{0,0,1} = \frac{3}{C} - \varepsilon_1 \left(2\rho + \frac{\tau\rho}{3 - 3C\varepsilon_1} \right) \quad \text{and} \quad \bar{z}_{0,0,2} = \frac{3}{C}.$$

Now, we choose ρ sufficiently small such that $\bar{z}_{0,0,1}$ is contained in the chosen domain around the equilibrium of $N_{0,0}^K$, that is $\bar{z}_{0,0,1} \in [\frac{3}{C} - 3\varepsilon_1, \frac{3}{C} + 3\varepsilon_1]$, which holds as soon as

$$2\rho + \frac{\tau\rho}{3 - 3C\varepsilon_1} < 3.$$

Applying [CMT21, Lemma C.1.] to $Z_{0,0,1}^K$ and $Z_{0,0,2}^K$ shows that

$$\lim_{K \rightarrow \infty} \mathbb{P} \left(\forall t \in [0, s_1 + \eta]: \frac{Z_{0,0,1}^K(t \log K)}{K} \geq \frac{3}{C} - 3\varepsilon_1 \right) = 1$$

and similarly

$$\lim_{K \rightarrow \infty} \mathbb{P} \left(\forall t \in [0, s_1 + \eta]: \frac{Z_{0,0,2}^K(t \log K)}{K} \leq \frac{3}{C} + 3\varepsilon_1 \right) = 1.$$

Note that the coupling above is only true until time $\theta_1^K < s_1 + \eta$, but the bounds for the processes $Z_{0,0,*}^K$ with $* \in \{1, 2\}$ are still true for any later times. In particular, we obtain for the time θ_1^K that

$$\lim_{K \rightarrow \infty} \mathbb{P} \left(\frac{Z_{0,0,1}^K(\theta_1^K \log K)}{\log K} \geq \frac{3}{C} - 3\varepsilon_1 \right) = 1 = \lim_{K \rightarrow \infty} \mathbb{P} \left(\frac{Z_{0,0,2}^K(\theta_1^K \log K)}{\log K} \leq \frac{3}{C} + 3\varepsilon_1 \right).$$

Since at time θ_1^K the coupling still holds, we see

$$\lim_{K \rightarrow \infty} \mathbb{P} \left(\frac{N_{0,0}^K(\theta_1^K \log K)}{\log K} \in \left[\frac{3}{C} - 3\varepsilon_1, \frac{3}{C} + 3\varepsilon_1 \right] \right) = 1.$$

Hence, by definition of θ_1^K we must have

$$\sum_{(m,n) \neq (0,0)} N_{m,n}^K(\theta_1^K \log K) \geq \rho\varepsilon_1 K$$

with probability converging to 1 as $K \rightarrow \infty$. Since we have assumed that at any given time at most two of the limiting exponents $\beta_{m,n}$ may be 1 and we already know from above that $\beta_{m_2^*, n_2^*}(s_1) = 1$, it must hold for some $\tilde{\varepsilon} > 0$ that

$$\max_{(m,n) \notin \{(0,0), (m_2^*, n_2^*)\}} \beta_{m,n}(s_1) \leq 1 - \tilde{\varepsilon}.$$

Since we have shown the convergences $\beta_{m,n}^K \rightarrow \beta_{m,n}$ on $[0, s_1 - \eta]$, by the continuity of the exponents (see Lemma A.15) and the convergence $\theta_1^K \rightarrow s_1$ we conclude

$$\sum_{(m,n) \notin \{(0,0), (m_2^*, n_2^*)\}} N_{m,n}^K(\theta_1^K \log K) \leq K^{1-\tilde{\varepsilon}}$$

with high probability for $\tilde{\varepsilon} > 0$ sufficiently small. Hence, it must hold $N_{m_2^*, n_2^*}^K(\theta_1^K \log K) \geq \rho \varepsilon_1 K/2$ with probability converging to 1. It is important to note that by definition of θ_1^K , this is equivalent to demanding

$$N_{m_2^*, n_2^*}^K(\theta_1^K \log K) \in \left[\frac{\rho \varepsilon_1 K}{2}, \rho \varepsilon_1 K \right]$$

which enables us to apply the Propositions from Appendix B in combination with Remark B.16.

Step 3: Competition. Now that we have established the emergence of the invasive trait $(m_2^* \delta, n_2^* \delta)$, we need to distinguish the two cases $m_2^* = 0$ and $m_2^* > 0$.

Case(a): $m_2^* = 0$. In this case we can proceed as in [CMT21], as the invading trait is a one-dimensional process which necessarily performs horizontal transfer. Firstly, we note again due to continuity of the exponent that

$$\sum_{(m,n) \notin \{(0,0), (m_2^*, n_2^*)\}} N_{m,n}^K(t \log K) \leq K^{1-\frac{\tilde{\varepsilon}}{4}}$$

for all $t \in [\theta_1^K, \theta_1^K + s]$ for $s > 0$ sufficiently small with probability converging to 1. Being consistent with the notation in [CMT21, Section C.2.2], we define for any time t the functions

$$\begin{aligned} b_1^K(t) &= 4(1 - K^{-\alpha}), \quad b_2^K(t) = \left(4 - \frac{(m_2^* + n_2^*)\delta}{2} \right) (1 - K^{-\alpha}), \\ d_1^K(t) &= d_2^K(t) = 1 + \left(\frac{C}{K} + \frac{\tau}{\sum_{m=0, n>0}^L N_{m,n}^{K,a}(t)} \right) \sum_{(m,n) \notin \{(0,0), (m_2^*, n_2^*)\}} N_{m,n}^{K,a}(t), \\ \tau^K(t) &= \tau \cdot \frac{N_{0,0}^{K,a}(t) + N_{m_2^*, n_2^*}^{K,a}(t)}{\sum_{m,n=0}^L N_{m,n}^{K,a}(t)}, \quad \gamma_1^K(t) = 0, \quad \gamma_2^K(t) \leq 4K^{-\alpha} N_{0,0}^{K,a}(t). \end{aligned}$$

Note that for the immigration rate γ_2^K we would need to consider the incoming immigration from the neighbouring traits. However, if $(0,0)$ is not one of them, those traits are of size of order strictly less than K , so the upper bound for $\gamma_2^K(t)$ is justified. The above functions except for γ_2^K converge on the interval $[\theta_1^K \log K, (\theta_1^K + s) \log K]$ to $b_1 = 4$, $b_2 = 4 - \frac{(m_2^* + n_2^*)\delta}{2}$, $d_1 = d_2 = 1$, τ and 0 respectively in order of appearance. For $\gamma_2^K(t)$ we obtain the convergence $\frac{\gamma_2^K(t)}{K} \rightarrow 0$ as $K \rightarrow \infty$.

Now, we can apply the Markov property at time θ_1^K and subsequently Lemma C.3 from [CMT21], which gives us the existence of a finite time $T(\rho, \varepsilon_1)$ such that with probability converging to 1 we have

$$N_{0,0}^K(\theta_1^K \log K + T(\rho, \varepsilon_1)) \leq \rho \varepsilon_1 K$$

and

$$\frac{N_{m_2^*, n_2^*}^{K,a}(\theta_1^K \log K + T(\rho, \varepsilon_1))}{K} \in \left[\bar{z}_{m_2^*, n_2^*}^a - \varepsilon_2, \bar{z}_{m_2^*, n_2^*}^a + \varepsilon_2 \right],$$

where $\bar{z}_{m_2^*, n_2^*}^a$ denotes the active equilibrium population size (which in this case coincides with the total equilibrium population size) of the rescaled process $\frac{N_{m_2^*, n_2^*}^K}{K}$.

Hence, we can define the end of the first intermediate phase as

$$\sigma_2^K \log K = \theta_1^K \log K + T(\rho, \varepsilon_1).$$

In particular, we have $\sigma_2^K \rightarrow s_1$ in probability as $K \rightarrow \infty$. At time $\sigma_2^K \log K$ we can use the continuity of the exponent and are left with the following bounds on our populations

$$N_{0,0}^K(\sigma_2^K \log K) \in [K^{1-\varepsilon_1}, \rho \varepsilon_1 K], \quad \frac{N_{m_2^*, n_2^*}^K(\theta_1^K \log K + T(\rho, \varepsilon_1))}{K} \in \left[\bar{z}_{m_2^*, n_2^*}^a - \varepsilon_2, \bar{z}_{m_2^*, n_2^*}^a + \varepsilon_2 \right]$$

and for all $(m, n) \notin \{(0, 0), (m_2^*, n_2^*)\}$ we have, again using the continuity argument from Lemma A.15,

$$\frac{\log(1 + N_{m,n}^K(\sigma_2^K \log K))}{\log K} \in [\beta_{m,n}(s_1) - \varepsilon_2, \beta_{m,n}(s_1) + \varepsilon_2].$$

Note that populations for which $\beta_{m,n}(s_1) = 0$ are actually extinct at time σ_2^K . This is due to our assumption that in this case we must have $\beta_{m,n}(t) = 0$ on an interval $[s_1 - \varepsilon, s_1]$, which due to our starting condition implies negative fitness and weak immigration and hence by Lemma A.16 extinction of the population. Then, applying Lemma A.13 shows for K sufficiently large that there is no immediate resurrection of the population after time s_1 .

Case(b): $m_2^* > 0$. Now, the individuals of the invading trait are able to become dormant. Hence, we have competition between a resident one-dimensional process and an invading two-dimensional process. Note that we may or may not have horizontal transfer exhibited from the invading trait. As in Case(a) we define a number of functions and apply the corresponding result on competition. The functions to be defined are

$$\begin{aligned} a_1^K(t) &= 4(1 - K^{-\alpha}), \quad b_1^K(t) = \left(4 - \frac{(m_2^* + n_2^*)\delta}{2}\right)(1 - K^{-\alpha}), \\ d_1^K(t) &= 1 + \left(\frac{C}{K} + \frac{\tau}{\sum_{m=0, n>0}^L N_{m,n}^{K,a}(t)}\right) \sum_{(m,n) \notin \{(0,0), (m_2^*, n_2^*)\}} N_{m,n}^{K,a}(t), \quad d_2^K(t) \equiv \kappa, \\ \gamma_1^K(t) &= 0, \quad \gamma_2^K(t) \leq 4K^{-\alpha} N_{0,0}^{K,a}(t). \end{aligned}$$

If there is no horizontal transfer (that is $n_2^* = 0$), we set $\tau^K(t) \equiv 0$. Otherwise we set

$$\tau^K(t) = \tau \cdot \frac{N_{0,0}^{K,a}(t) + N_{m_2^*, n_2^*}^{K,a}(t)}{\sum_{m,n=0}^L N_{m,n}^{K,a}(t)}.$$

Since we have non-negative horizontal transfer exerted from the invading trait onto $(0, 0)$ and we have dormancy for the invading but not for the initially resident trait, we can apply Proposition B.23 together with Remark B.16 due to the same convergence arguments made in Case(a). Hence, there exists some finite time $T(\rho, \varepsilon_1)$ such that with probability larger than $1 - o_{\varepsilon_1}(1)$ we have

$$\begin{aligned} N_{0,0}^K(\sigma_2^K \log K) &\in [K^{1-\varepsilon_1}, \rho \varepsilon_1 K] \\ \frac{N_{m_2^*, n_2^*}^{K,a}(\theta_1^K \log K + T(\rho, \varepsilon_1))}{K} &\in [\bar{z}_{m_2^*, n_2^*}^a - \varepsilon_2, \bar{z}_{m_2^*, n_2^*}^a + \varepsilon_2], \end{aligned}$$

and

$$\frac{N_{m_2^*, n_2^*}^{K,d}(\theta_1^K \log K + T(\rho, \varepsilon_1))}{K} \in [\bar{z}_{m_2^*, n_2^*}^d - \varepsilon_2, \bar{z}_{m_2^*, n_2^*}^d + \varepsilon_2]$$

as $K \rightarrow \infty$, where $\bar{z}_{m_2^*, n_2^*}^d$ is the equilibrium size of the dormant component of the rescaled process. Then, at time

$$\sigma_2^K \log K = \theta_1^K \log K + T(\rho, \varepsilon_1)$$

we have the same bounds as in Case(a) with the only difference in the equilibrium size of the process $N_{m_2^*, n_2^*}^K$.

4.2.3. Phase k . We will now consider a time interval $[\sigma_k^K \log K, \theta_k^K \log K]$, where $\sigma_k^K \rightarrow s_{k-1}$ and $\theta_k^K \rightarrow s_k$ in probability. Thus, we consider $k \geq 2$ and assume that we have already completed step $k-1$. In particular, we assume that we have defined a stopping time σ_k^K with the convergence property

mentioned above such that for the resident population of trait $(m_k^* \delta, n_k^* \delta)$ the bounds

$$\frac{N_{m_k^*, n_k^*}^{K,a}(\sigma_k^K \log K)}{K} \in [\bar{z}_{m_k^*, n_k^*}^a - \varepsilon_k, \bar{z}_{m_k^*, n_k^*}^a + \varepsilon_k]$$

and

$$\frac{N_{m_k^*, n_k^*}^{K,d}(\sigma_k^K \log K)}{K} \in [\bar{z}_{m_k^*, n_k^*}^d - \varepsilon_k, \bar{z}_{m_k^*, n_k^*}^d + \varepsilon_k]$$

hold. Furthermore, we assume that for the previously resident trait we have

$$K^{1-\varepsilon_k} \leq N_{m_{k-1}^*, n_{k-1}^*}^K(\sigma_k^K \log K) \leq \rho \varepsilon_k K.$$

For all remaining traits $(m\delta, n\delta) \notin \{(m_{k-1}^* \delta, n_{k-1}^* \delta), (m_k^* \delta, n_k^* \delta)\}$, we assume $N_{m,n}^K(\sigma_k^K \log K) = 0$ if $\beta_{m,n}(s_{k-1}) = 0$ and otherwise we assume

$$\frac{\log(1 + N_{m,n}^K(\sigma_k^K \log K))}{\log K} \in [\beta_{m,n}(s_{k-1}) - \varepsilon_k, \beta_{m,n}(s_{k-1}) + \varepsilon_k].$$

As in the base case, we introduce the time θ_k^K , which is the time until the active part of the resident trait leaves a neighbourhood of its equilibrium or a new trait emerges, that is

$$\theta_k^K := \inf \left\{ t \geq \sigma_k^K \mid N_{m_k^*, n_k^*}^{K,a}(t \log K) \notin \left[(\bar{z}_{m_k^*, n_k^*}^a - 3\varepsilon_k)K, (\bar{z}_{m_k^*, n_k^*}^a + 3\varepsilon_k)K \right] \right. \\ \left. \text{or } \sum_{(m,n) \neq (m_k^*, n_k^*)} N_{m,n}^K(t \log K) \geq \rho \varepsilon_k K \right\}.$$

Step 0: Deriving bounds on the rates. Similarly to Step 0 in Case 1 of the proof, we can derive similar bounds on the birth, death and migration rates on the time interval $[\sigma_k^K, \theta_k^K]$. The bounds for the birth and arrival due to horizontal transfer rates are

$$4 - \frac{(m+n)\delta}{2} + \tau \mathbb{1}_{n > n_k^*} \pm C_* \varepsilon_k,$$

and for the death and emigration due to horizontal transfer we obtain the bounds

$$1 + \bar{z}_{m_k^*, n_k^*}^a(1 - pm\delta) + \tau \mathbb{1}_{n_k^* > n} \pm C_* \varepsilon_k.$$

The immigration rates stay the same as in the base case, since they do not depend on the resident trait population size. The active to dormant switching rate then satisfy the bounds

$$pm\delta \bar{z}_{m_k^*, n_k^*}^a \pm C_* \varepsilon_k.$$

Step 1: Induction on the traits. As in the base case, we want to use the bounds given above to couple our processes accordingly and show by induction on the traits the upper and lower bounds on $\beta_{m,n}^K$. For this, we may again decompose the time interval $[s_{k-1}, s_k]$ into sections on which all $\beta_{m,n}$ are affine. On the first such subinterval which is of the form $t \in [s_{k-1}, t_1 \wedge \theta_k^K \wedge T]$, we can write

$$\beta_{m,n} = \beta_{m,n}(s_{k-1}) + a_{m,n}(t - s_{k-1}).$$

for some constants $a_{m,n} \in \mathbb{R}$. We will not fully carry out the induction, but give a broad idea, since it is very similar to the base case. If $(m_k^*, n_k^*) = (0, 0)$, we are in the same situation as in the base case, so we can use the Markov property at time $\sigma_k^K \log K$ and obtain the same results where in the couplings we need to replace $\beta_{m,n}(0)$ with $\beta_{m,n}(s_{k-1})$.

In the case where $(m_k^*, n_k^*) \neq (0, 0)$ and $\beta_{0,0}(s_{k-1}) > 0$, there is no incoming immigration into the trait $(0, 0)$ and hence we can use the coupling

$$\tilde{Z}_{0,0}^K(t \log K) \leq N_{0,0}^K(t \log K) \leq \hat{Z}_{0,0}^K(t \log K),$$

where $\tilde{Z}_{0,0}^K$ is a $BP_K(4 - C_*\varepsilon_k, 1 + \bar{z}_{m_k^*, n_k^*}^a + \tau \mathbb{1}_{n_k^* > 0} + C_*\varepsilon_k, \beta_{0,0}(s_{k-1}) - C_*\varepsilon_k)$ and $\hat{Z}_{0,0}^K$ is given as $BP_K(4 + C_*\varepsilon_k, 1 + \bar{z}_{m_k^*, n_k^*}^a + \tau \mathbb{1}_{n_k^* > 0} - C_*\varepsilon_k, \beta_{0,0}(s_{k-1}) + C_*\varepsilon_k)$. For our coupled processes, the convergence theorem [CMT21, Lemma A.1] implies the bounds

$$\begin{aligned} & \beta_{0,0}(s_{k-1}) + S((0, 0), (m_k^* \delta, n_k^* \delta))(t - s_{k-1}) - C_*\varepsilon_k \\ & \leq \frac{\log(1 + N_{0,0}^K(t \log K))}{\log K} \\ & \leq \beta_{0,0}(s_{k-1}) + S((0, 0), (m_k^* \delta, n_k^* \delta))(t - s_{k-1}) + C_*\varepsilon_k. \end{aligned}$$

If $\beta_{0,0}(s_{k-1}) = 0$, then due to the lack of immigration and our observation that populations with $\beta_{m,n}(s_{k-1}) = 0$ are actually extinct we have $N_{0,0}^K(t \log K) = 0$ for all $t \geq \sigma_k^K$.

As mentioned, we abbreviate the induction and assume that the bounds

$$\beta_{m,n}(t) - C_*\varepsilon_k \leq \frac{\log(1 + N_{m,n}^K(t \log K))}{\log K} \leq \beta_{m,n}(t) + C_*\varepsilon_k \quad (4.4)$$

have been shown up to the neighbouring traits of $(m\delta, n\delta)$ for all $t \in [s_{k-1}, t_1 \wedge \theta_k^K \wedge T]$. Then, we need to distinguish the cases where $m = 0$ and $m > 0$ as well as $n \geq n_k^*$ and $n < n_k^*$. The first distinction corresponds to the question of the ability to become dormant, whereas the second one dictates the way that horizontal transfer influences the dynamics. Furthermore, we need to distinguish whether $N_{m-1,n}^K$ or $N_{m,n-1}^K$ is larger (in terms of orders of powers of K) to determine which population is responsible for the immigration rate. Also, we need to separate the cases where $\beta_{m,n}(s_{k-1}) = 0$ or strictly larger than 0. In the first case, we need to couple with processes whose initial population size is also 0. Without loss of generality we assume $N_{m,n-1}^K$ to be of larger order than $N_{m-1,n}^K$ - the other case can be done by switching the corresponding indices. Then, we can couple

$$\tilde{Z}_{m,n}^K(t \log K) \leq N_{m,n}^K(t \log K) \leq \hat{Z}_{m,n}^K(t \log K)$$

where $\tilde{Z}_{m,n}^K$ and $\hat{Z}_{m,n}^K$ are $BPI_K(4 - \frac{(m+n)\delta}{2} + \tau \mathbb{1}_{n > n_k^*} \mp C_*\varepsilon_k, 1 + \bar{z}_{m_k^*, n_k^*}^a + \tau \mathbb{1}_{n < n_k^*} \pm 2C_*\varepsilon_k, a_{m,n-1}, \beta_{m,n-1}(s_{k-1}) - \alpha \mp C_*\varepsilon_k, (\beta_{m,n}(s_{k-1}) \mp C_*\varepsilon_k)_\times)$ in the case where $m = 0$ and otherwise they are determined by $BBPI_K(4 - \frac{(m+n)\delta}{2} + \tau \mathbb{1}_{n > n_k^*} \mp C_*\varepsilon_k, 0, 1 + (1 - pm\delta)\bar{z}_{m_k^*, n_k^*}^a + \tau \mathbb{1}_{n < n_k^*} \pm 2C_*\varepsilon_k, \kappa, pm\delta \mp C_*\varepsilon_k, \sigma, a_{m,n-1}, \beta_{m,n-1}(s_{k-1}) - \alpha \mp C_*\varepsilon_k, (\beta_{m,n}(s_{k-1}) \mp C_*\varepsilon_k)_\times)$.

Applying Theorem A.1 or A.3 accordingly shows the bounds (4.4) by definition of our fitness function. Continuing this process for all time intervals on which all $\beta_{m,n}$ are affine shows the bounds (4.4) on the entire interval $[s_{k-1}, s_k \wedge \theta_k^K \wedge T]$ with probability converging to 1.

Step 2: Deriving a lower bound for θ_k^K . As in Step 2 of the base case, we want to show that $(s_k - \eta) \wedge T < \theta_k^K$ with probability converging to 1. Again due to our assumption, we know that all functions $\beta_{m,n}$ except for $\beta_{m_k^*, n_k^*}$ are bounded away from 1 on the interval $[s_{k-1} + \eta, s_k - \eta]$. Therefore, it again suffices for showing $s_1 - \eta < \theta_k^K$ that $N_{m_k^*, n_k^*}^{K,a}$ does not exit a neighbourhood of its equilibrium until time $s_k - \eta$. For this purpose, we can couple with processes

$$Z_{m_k^*, n_k^*, 1}^K(t \log K) \leq N_{m_k^*, n_k^*}^K(t \log K) \leq Z_{m_k^*, n_k^*, 2}^K(t \log K)$$

up to time θ_k^K . Again we need to distinguish between the possibility of becoming dormant or not. If $m_k^* = 0$, we can choose $Z_{m_k^*, n_k^*, 1}^K$ as a $LBDI_K((4 - \frac{(m_k^* + n_k^*)\delta}{2})(1 - \varepsilon), 1 + C\varepsilon, C, K^{-\alpha} N_{m_k^*, n_k^* - 1}^*)$ and

$Z_{m_k^*, n_k^*, 1}$ as a $LBDI_K((4 - \frac{(m_k^* + n_k^*)\delta}{2}), 1, C, K^{-\alpha}N_{m_k^*, n_k^* - 1})$. If, on the other hand, we have $m_k^* > 0$, we need to distinguish where the immigration is coming from and can choose the process $Z_{m_k^*, n_k^*, 1}$ to be determined by a $LBDI_K((4 - \frac{(m_k^* + n_k^*)\delta}{2})(1 - \varepsilon), 1 + C\varepsilon, \kappa, \sigma, p, C, K^{-\alpha}N_{m_k^*, n_k^* - 1})$ if we assume the immigration to be dominated by $N_{m_k^*, n_k^* - 1}$. Then we can choose $Z_{m_k^*, n_k^*, 2}$ as a $LBDI_K((4 - \frac{(m_k^* + n_k^*)\delta}{2}), 1, \kappa, \sigma, p, C, K^{-\alpha}N_{m_k^*, n_k^* - 1})$. Now, applying [CMT21, Lemma C.1] to the first case and Corollary B.8 in the case of bi-type processes, we see that at time $s_k - \eta$ the process $N_{m_k^*, n_k^*}^K$ has not exited a neighbourhood of its equilibrium size with probability converging to 1. In particular, we must have $s_k - \eta < \theta_k^K$ with high probability.

4.2.4. *Intermediate Phase k.* The structure of this intermediate phase remains the same as in Section 4.2.2.

Step 1: Convergence of $\theta_k^K \rightarrow s_k$. This part of the proof can be taken from Step 1 in Intermediate Phase 1 with minor changes in the times and the resident trait and is not repeated here.

Step 2: Emergence of a new population. This part is also very similar. However, we may need to couple with logistic bi-type branching processes instead of single type. Since this is a straightforward adaptation similar to Step 2 of Phase k , we do not carry it out here. We do obtain however that

$$N_{m_{k+1}^*, n_{k+1}^*}^K(\theta_k^K \log K) \in \left[\frac{\rho \varepsilon_k K}{2}, \rho \varepsilon_k K \right]$$

and

$$\sum_{(m,n) \notin \{(m_k^*, n_k^*), (m_{k+1}^*, n_{k+1}^*)\}} N_{m,n}^K(\theta_k^K \log K) \leq K^{1-\frac{\varepsilon_k}{2}}.$$

Step 3: Competition. By assumption of the theorem, there is competition between the resident and the emerging trait. Distinguishing the cases, we can proceed as in Intermediate Phase 1 and define the corresponding birth, death, migration, switching and horizontal transfer rates which then allow us to apply one of the Propositions from B.15, B.17, B.23, B.25, B.26 and B.28 in conjunction with Remark B.16 or [CMT21, Lemma C.3], which in each case give us a finite time $T(\rho, \varepsilon_k)$ such that with probability larger than $1 - o_{\varepsilon_k}(1)$ we have, as $K \rightarrow \infty$, the bounds

$$N_{m_k^*, n_k^*}^K(\theta_k^K \log K + T(\rho, \varepsilon_k)) \in [K^{1-\varepsilon_k}, \rho \varepsilon_k K],$$

$$\frac{N_{m_{k+1}^*, n_{k+1}^*}^{K,a}(\theta_k^K \log K + T(\rho, \varepsilon_k))}{K} \in [\bar{z}_{m_{k+1}^*, n_{k+1}^*}^a - \varepsilon_k, \bar{z}_{m_{k+1}^*, n_{k+1}^*}^a + \varepsilon_k],$$

and

$$\frac{N_{m_{k+1}^*, n_{k+1}^*}^{K,d}(\theta_k^K \log K + T(\rho, \varepsilon_k))}{K} \in [\bar{z}_{m_{k+1}^*, n_{k+1}^*}^d - \varepsilon_k, \bar{z}_{m_{k+1}^*, n_{k+1}^*}^d + \varepsilon_k].$$

Thus, we can define the time $\sigma_{k+1}^K \log K := \theta_k^K \log K + T(\rho, \varepsilon_k)$, at which time the stated properties in the beginning of Step k are satisfied with high probability. That is, for $(m, n) \notin \{(0, 0), (m_2^*, n_2^*)\}$ we have again using the continuity argument from Lemma A.15

$$\frac{\log(1 + N_{m,n}^K(\sigma_{k+1}^K \log K))}{\log K} \in [\beta_{m,n}(s_k) - \varepsilon_{k+1}, \beta_{m,n}(s_k) + \varepsilon_{k+1}],$$

if $\beta_{m,n}(s_k) > 0$ and $N_{m,n}^K(\sigma_{k+1}^K \log K) = 0$ otherwise. To see the latter part, the argument from the end of Case(a) in Step 3 of Section 4.2.2 still applies. Thus, we have proven Theorem 2.2. \square

APPENDIX A. RESULTS ON BI-TYPE BRANCHING PROCESSES WITH IMMIGRATION

In this section, we derive a general convergence result for special bi-type branching processes. More specifically, we want to generalize the following theorem from [CMT21].

We denote the law of a one-dimensional branching process $(Z^K)_{t \geq 0}$ with birth rate $b \geq 0$, death rate $d \geq 0$ and time dependent immigration at rate $K^c e^{at}$ at time $t \geq 0$ with $a, c \in \mathbb{R}$ by $BPI_K(b, d, a, c, \beta)$, where $Z_0^K = \lfloor K^\beta - 1 \rfloor$.

Theorem A.1. *Let Z^K be a $BPI_K(b, d, a, c, \beta)$ with $c \leq \beta$ and assume either $\beta > 0$ or $c \neq 0$. Then the process $\frac{\log(1 + Z_{t \log K}^K)}{\log K}$ converges when K tends to infinity in probability in $L^\infty([0, T])$ for all $T > 0$ to the continuous, deterministic function $\bar{\beta}$ given by*

- (i) if $\beta > 0$, $\bar{\beta}: t \mapsto (\beta + rt) \vee (c + at) \vee 0$;
- (ii) if $\beta = 0$, $c < 0$ and $a > 0$, $\bar{\beta}: t \mapsto ((r \vee a)(t - \frac{|c|}{a})) \vee 0$;
- (iii) if $\beta = 0$, $c < 0$ and $a \leq 0$, $\bar{\beta}: t \mapsto 0$;

where $r = b - d$.

Proof. This is Theorem B.5 from [CMT21]. □

In the spirit of the above theorem, we consider the process $Z_t^K = (X_t^K, Y_t^K)$ with initial population $(X_0^K, Y_0^K) = (\lfloor K^\beta - 1 \rfloor, \lfloor K^\gamma - 1 \rfloor)$ and transition rates

$$(n, m) \mapsto \begin{cases} (n+1, m), & \text{at rate } b_1 n + K^c e^{at} \\ (n, m+1), & \text{at rate } b_2 m \\ (n-1, m+1), & \text{at rate } \sigma_1 n \\ (n+1, m-1), & \text{at rate } \sigma_2 m \\ (n-1, m), & \text{at rate } d_1 n \\ (n, m-1), & \text{at rate } d_2 m \end{cases}.$$

We refer to the rates $b_1, b_2 \geq 0$ as birth rates of X_t^K and Y_t^K respectively, $d_1, d_2 \geq 0$ as their respective death rates and $\sigma_1, \sigma_2 > 0$ are the switching rates. The additional $K^c e^{at}$ represents the immigration into the population from the outside, where $a, c \in \mathbb{R}$.

Notation A.2. *We denote the distribution of a bi-type branching process as introduced above by $BBPI_K(b_1, b_2, d_1, d_2, \sigma_1, \sigma_2, a, c, \beta, \gamma)$. If the initial condition satisfies $\beta = \gamma$, we use the shorthand notation $BBPI_K(b_1, b_2, d_1, d_2, \sigma_1, \sigma_2, a, c, \beta)$.*

We are now interested in finding some convergence results for the total population size $X_t^K + Y_t^K$ similar to those from Appendix B in [CMT21]. We will show the following theorem.

Theorem A.3. *Let $Z_t^K = (X_t^K, Y_t^K)$ be a $BBPI_K(b_1, b_2, d_1, d_2, \sigma_1, \sigma_2, a, c, \beta, \gamma)$ as introduced in Notation A.2. Further assume that $c \leq \beta \vee \gamma$ and $\beta \vee \gamma > 0$ or $c \neq 0$ and let λ as in (A.3). Then for all $T \geq 0$ the process*

$$s \mapsto \frac{\log(1 + X_{s \log K}^K + Y_{s \log K}^K)}{\log K}$$

converges in probability in $L^\infty([0, T])$ as $K \rightarrow \infty$ towards a deterministic function $\bar{\beta}: [0, T] \rightarrow \mathbb{R}$, which we describe in each case:

- (i) If $\beta \vee \gamma > 0$, then $\bar{\beta}(t) = ((\beta \vee \gamma) + \lambda t) \vee (c + at) \vee 0$.
- (ii) If $\beta \vee \gamma = 0$ and $c < 0$ and $a > 0$, then $\bar{\beta}(t) = (\lambda \vee a)(t - \frac{|c|}{a}) \vee 0$.

(iii) If $\beta \vee \gamma = 0$ and $c < 0$ and $a \leq 0$, then $\bar{\beta}(t) = 0$.

The proof of the theorem will rely partly on Markov's, Chebyshev's and Doob's inequalities, so we first need to derive some bounds for the expected value and variance of our process.

A.1. Bounds on the Expectation and Variance. Our first step is to find the semimartingale decomposition of X_t^K and Y_t^K . In order to do so, we introduce some notation.

Notation A.4. In the following we write $r_1 := b_1 - d_1 - \sigma_1$ and $r_2 := b_2 - d_2 - \sigma_2$.

Lemma A.5. Consider the process $Z_t^K = (X_t^K, Y_t^K)$ as introduced above. Then there exist càdlàg martingales M_t^K, N_t^K starting at 0, such that

$$\begin{pmatrix} X_t^K \\ Y_t^K \end{pmatrix} = \begin{pmatrix} X_0^K \\ Y_0^K \end{pmatrix} + \begin{pmatrix} M_t^K \\ N_t^K \end{pmatrix} + \int_0^t \begin{pmatrix} r_1 X_s^K + \sigma_2 Y_s^K + K^c e^{as} \\ r_2 Y_s^K + \sigma_1 X_s^K \end{pmatrix} ds.$$

Proof. This decomposition follows from Dynkin's formula. \square

Our next goal is to identify the rate of growth of our population, which is directly linked to determining the expected value of the population size. In order to do so, we calculate the expected value for our process up to some constants.

Lemma A.6. The expected value (x_t^K, y_t^K) of (X_t^K, Y_t^K) solves the ordinary differential equation

$$\begin{pmatrix} \dot{x}_t^K \\ \dot{y}_t^K \end{pmatrix} = \begin{pmatrix} r_1 & \sigma_2 \\ \sigma_1 & r_2 \end{pmatrix} \begin{pmatrix} x_t^K \\ y_t^K \end{pmatrix} + \begin{pmatrix} K^c e^{at} \\ 0 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} x_0^K \\ y_0^K \end{pmatrix} = \begin{pmatrix} K^\beta - 1 \\ K^\gamma - 1 \end{pmatrix}. \quad (\text{A.1})$$

Proof. This is a direct consequence of Lemma A.5, where we can apply the expected value on both sides. Interchanging the expected value and integral on the right hand side by Fubini shows that (x_t^K, y_t^K) is absolutely continuous. Differentiating both sides gives the differential equation (A.1). \square

Note that this differential equation can be solved easily: The matrix

$$\begin{pmatrix} r_1 & \sigma_2 \\ \sigma_1 & r_2 \end{pmatrix} = SDS^{-1} \quad (\text{A.2})$$

can be diagonalised, because its eigenvalues λ and $\tilde{\lambda}$ can be written as

$$\lambda = \frac{r_1 + r_2 + \Delta}{2} \quad \text{and} \quad \tilde{\lambda} = \frac{r_1 + r_2 - \Delta}{2}, \quad (\text{A.3})$$

where $\Delta = \sqrt{(r_1 - r_2)^2 + 4\sigma_1\sigma_2} \neq 0$ and hence $\lambda > \tilde{\lambda}$. In particular we are now able to give a characterization of the expected values for X_t^K and Y_t^K .

Lemma A.7. The expected values (x_t^K, y_t^K) of (X_t^K, Y_t^K) satisfy for $t > 0$ the asymptotic relation

$$(x_t^K, y_t^K) = \begin{cases} \Theta(K^c e^{at}), & \text{if } a > \lambda \\ \Theta((x_0^K + y_0^K + K^c)e^{\lambda t}), & \text{if } \lambda > a \\ \Theta((x_0^K + y_0^K + (1+t)K^c)e^{\lambda t}), & \text{if } \lambda = a, \end{cases}$$

where we use the notation $f^K = \Theta(g^K)$ for two families of functions $f^K, g^K: [0, \infty) \rightarrow \mathbb{R}$ if there exists some finite constant $C > 0$ such that for all $t \geq 0$ we have

$$\lim_{K \rightarrow \infty} \frac{f^K(t)}{g^K(t)} = C.$$

In fact, there exists a constant $\tilde{C} > 0$ sufficiently large such that for all $K \geq 0$ and all $t \geq 0$ we have

$$x_t^K, y_t^K \leq \tilde{C} \left([(x_0^K + y_0^K + (1+t)K^c)e^{\lambda t}] \vee [K^c e^{at}] \right).$$

Proof. The solution to the differential equation (A.1) is known to be

$$\begin{pmatrix} x_t^K \\ y_t^K \end{pmatrix} = Se^{Dt}S^{-1} \begin{pmatrix} x_0^K \\ y_0^K \end{pmatrix} + \int_0^t Se^{D(t-s)}S^{-1} \begin{pmatrix} K^ce^{as} \\ 0 \end{pmatrix} ds.$$

An explicit computation shows the claim. \square

In the following we will also need some bounds on the variation of X_t^K and Y_t^K . In order to derive them, we need some more preparation. In particular, we need to compute the quadratic variation. The purpose here is twofold: We need these variation terms once for finding an upper bound of the variance of X^K and Y^K . Secondly, we will later, in the proof of our convergence result, make use of Doob's inequality and hence need to calculate the expected value of some quadratic variation.

Lemma A.8. *The quadratic variation of the martingales M^K , N^K and $M^K + N^K$ as well as the quadratic covariation $[M^K, N^K]$ of M^K and N^K are given by*

$$\begin{aligned} [M^K]_t &= \int_0^t (b_1 + d_1 + \sigma_1)X_s^K + \sigma_2Y_s^K + K^ce^{as} ds, \\ [N^K]_t &= \int_0^t (b_2 + d_2 + \sigma_2)Y_s^K + \sigma_1X_s^K ds, \\ [M^K + N^K]_t &= 3 \int_0^t (b_1 + d_1)X_s^K + (b_2 + d_2)Y_s^K + K^ce^{as} ds, \\ [M^K, N^K]_t &= \int_0^t (r_1 + 2d_1)X_s^K + (r_2 + 2d_2)Y_s^K + K^ce^{as} ds. \end{aligned}$$

Proof. We only carry out the calculations for M^K . In an analogous fashion we can calculate the quadratic variation of N^K and of $M^K + N^K$. For the covariation $[M^K, N^K]$ we can use the polarization identity

$$[M^K, N^K]_t = \frac{1}{2} ([M^K + N^K]_t - [M^K]_t - [N^K]_t).$$

Applying Itô's formula to $(X_t^K)^2$ and Dynkin's formula with $f(x, y) = x^2$ shows that

$$\begin{aligned} \widehat{M}_t^K &= (X_t^K)^2 - (X_0^K)^2 - \int_0^t (2X_s^K + 1)(b_1X_s^K + \sigma_2Y_s^K + K^ce^{as}) + (\sigma_1 + d_1)(-2X_s^K + 1)X_s^K ds, \\ \widetilde{M}_t^K &= (X_t^K)^2 - (X_0^K)^2 - \int_0^t 2X_s^K((b_1 - d_1 - \sigma_1)X_s^K + \sigma_2Y_s^K + K^ce^{as}) ds - [M^K]_t, \end{aligned}$$

for some martingales \widehat{M}^K and \widetilde{M}^K starting at 0. By the uniqueness of the Doob-Meyer decomposition of $(X_t^K)^2 - (X_0^K)^2$ we see that $\widehat{M}_t^K = \widetilde{M}_t^K$ and hence

$$[M^K]_t = \int_0^t (b_1 + d_1 + \sigma_1)X_s^K + \sigma_2Y_s^K + K^ce^{as} ds.$$

\square

Now, we can make use of the quadratic variations to derive our bounds for the variance.

Lemma A.9. *There exists a constant $C_* \geq 0$ independent of K such that*

$$\mathbb{V}(X_t^K), \mathbb{V}(Y_t^K) \leq C_*(1 + t^2) \left((e^{2\lambda t} + e^{\lambda t})(x_0^K + y_0^K + K^c) + K^ce^{at} \right) \quad \text{for all } t \geq 0$$

Proof. We denote $u_t^K := \mathbb{V}(X_t^K)$, $v_t^K := \mathbb{V}(Y_t^K)$ and $w_t^K := \text{cov}(X_t^K, Y_t^K)$. Applying Itô's formula and Lemma A.8 to $(X_t^K - x_t^K)^2$ and $(Y_t^K - y_t^K)^2$ as well as using Integration by Parts for $(X_t^K - x_t^K)(Y_t^K - y_t^K)$

gives the differential equation

$$\begin{pmatrix} \dot{u}_t^K \\ \dot{v}_t^K \\ \dot{w}_t^K \end{pmatrix} = \begin{pmatrix} 2r_1 & 0 & 2\sigma_2 \\ 0 & 2r_2 & 2\sigma_1 \\ \sigma_1 & \sigma_2 & r_1 + r_2 \end{pmatrix} \begin{pmatrix} u_t^K \\ v_t^K \\ w_t^K \end{pmatrix} + \begin{pmatrix} (b_1 + d_1 + \sigma_1)x_t^K + \sigma_2 y_t^K + K^c e^{at} \\ (b_2 + d_2 + \sigma_2)y_t^K + \sigma_1 x_t^K \\ (r_1 + 2d_1)x_t^K + (r_2 + 2d_2)y_t^K + K^c e^{at} \end{pmatrix} \quad (\text{A.4})$$

with initial condition $(u_0^K, v_0^K, w_0^K) = (0, 0, 0)$. Now we can proceed as in Lemma A.7. The eigenvalues of the coefficient matrix are $2\tilde{\lambda} < r_1 + r_2 < 2\lambda$, so it is diagonalisable with matrices S, S^{-1} such that

$$\begin{pmatrix} 2r_1 & 0 & 2\sigma_2 \\ 0 & 2r_2 & 2\sigma_1 \\ \sigma_1 & \sigma_2 & r_1 + r_2 \end{pmatrix} = SDS^{-1},$$

where $D = \text{diag}(2\tilde{\lambda}, r_1 + r_2, 2\lambda)$. The solution to the differential equation (A.4) is given by

$$\begin{aligned} \begin{pmatrix} u_t^K \\ v_t^K \\ w_t^K \end{pmatrix} &= \int_0^t S e^{D(t-s)} S^{-1} \begin{pmatrix} (b_1 + d_1 + \sigma_1)x_s^K + \sigma_2 y_s^K + K^c e^{as} \\ (b_2 + d_2 + \sigma_2)y_s^K + \sigma_1 x_s^K \\ (r_1 + 2d_1)x_s^K + (r_2 + 2d_2)y_s^K + K^c e^{as} \end{pmatrix} ds \\ &\leq C_* \int_0^t e^{2\lambda(t-s)} \begin{pmatrix} x_s^K + y_s^K + K^c e^{as} \\ y_s^K + x_s^K \\ x_s^K + y_s^K + K^c e^{as} \end{pmatrix} ds, \end{aligned}$$

where $C_* \geq 0$ is a suitable constant independent of K and the inequality holds for each component. By Lemma A.7 we can further estimate the expected values x_s^K, y_s^K with the constant C_* changing from line to line by

$$\begin{aligned} u_t^K, v_t^K, w_t^K &\leq C_* e^{2\lambda t} \int_0^t e^{-\lambda s} (x_0^K + y_0^K + (1+s)K^c) + K^c e^{(a-2\lambda)s} ds \\ &\leq C_* e^{2\lambda t} (1+t+t^2) \left[e^{-\lambda t} (x_0^K + y_0^K + K^c) + K^c e^{(a-2\lambda)t} + (x_0^K + y_0^K + K^c) \right] \\ &\leq C_*(1+t^2) \left[(e^{2\lambda t} + e^{\lambda t})(x_0^K + y_0^K + K^c) + K^c e^{at} \right], \end{aligned}$$

which we have claimed. \square

A.2. A Special Case of Theorem A.3. With these preparations we are well situated to show a first convergence result for general bi-type branching processes, which is easily seen to be a special case of Theorem A.3 (i).

Theorem A.10. *Let $Z_t^K = (X_t^K, Y_t^K)$ be a bi-type branching process whose distribution is given by $BBPI_K(b_1, b_2, d_1, d_2, \sigma_1, \sigma_2, a, c, \beta, \gamma)$ with $\beta \vee \gamma > 0$. Assume that $c \leq \beta \vee \gamma$ and that $T > 0$ is such that*

$$\inf_{t \in [0, T]} ((\beta \vee \gamma) + \lambda t) \vee (c + at) > 0, \quad (\text{A.5})$$

where we recall λ from (A.3). Then the following convergence in probability holds in $L^\infty([0, T])$:

$$\left\{ s \mapsto \frac{\log(1 + X_{s \log K}^K + Y_{s \log K}^K)}{\log K} \right\} \xrightarrow{K \rightarrow \infty} \{s \mapsto ((\beta \vee \gamma) + \lambda s) \vee (c + as)\}.$$

Remark A.11. Note that due to the strictly positive switching rates, the same convergence also holds for the processes

$$\left\{ s \mapsto \frac{\log(1 + X_{s \log K}^K)}{\log K} \right\} \quad \text{and} \quad \left\{ s \mapsto \frac{\log(1 + Y_{s \log K}^K)}{\log K} \right\}$$

on the interval $(0, T]$ if $\beta \neq \gamma$ and on $[0, T]$ if $\beta = \gamma$. Intuitively, if these processes were of different sizes, the switching would immediately fill the difference. For a formal proof, a straightforward adaptation of the proof of Theorem A.10 is possible.

Proof. For the proof we make extensive use of ideas from [CMT21, Theorem B.1]. We define $\bar{\beta}_t := ((\beta \vee \gamma) + \lambda t) \vee (c + at)$.

Step 1: Semimartingale Arguments. For $\eta > 0$ to be determined, we define the set

$$\Omega_1^K := \left\{ \sup_{t \in [0, T \log K]} \left| e^{-\lambda t} (X_t^K + Y_t^K - (x_t^K + y_t^K)) \right| \leq K^\eta \right\}.$$

Our first goal is to identify a set of parameters η such that $\mathbb{P}(\Omega_1^K) \rightarrow 1$ as $K \rightarrow \infty$. For this, in [CMT21, Lemma B.3] it is shown, that the process of which the absolute value is taken in Ω_1^K is a martingale. Here however, due to the switching between X_t^K and Y_t^K we do not have a martingale. Instead we use Integration by Parts as well as Lemmata A.5 and A.6 to get the decomposition

$$\begin{aligned} & e^{-\lambda t} (X_t^K + Y_t^K - (x_t^K + y_t^K)) \\ &= \int_0^t -\lambda e^{-\lambda s} (X_s^K + Y_s^K - (x_s^K + y_s^K)) \, ds + \int_0^t e^{-\lambda s} \, d(X_s^K + Y_s^K) - \int_0^t e^{-\lambda s} \, d(x_s^K + y_s^K) \\ &= \int_0^t e^{-\lambda s} \, d(M_s^K + N_s^K) + \int_0^t e^{-\lambda s} ((r_1 + \sigma_1 - \lambda)(X_s^K - x_s^K) + (r_2 + \sigma_2 - \lambda)(Y_s^K - y_s^K)) \, ds. \end{aligned}$$

We denote the martingale $\int_0^t e^{-\lambda s} \, d(M_s^K + N_s^K)$ by \widetilde{M}_t^K . Hence, by Doob's inequality we have

$$\begin{aligned} & \mathbb{P} \left(\sup_{t \leq T \log K} \left| e^{-\lambda t} (X_t^K + Y_t^K - (x_t^K + y_t^K)) \right| \geq K^\eta \right) \\ & \leq \mathbb{P} \left(\sup_{t \leq T \log K} \left| \widetilde{M}_t^K \right| + \int_0^t |r_1 + \sigma_1 - \lambda| e^{-\lambda s} |X_s^K - x_s^K| + |r_2 + \sigma_2 - \lambda| e^{-\lambda s} |Y_s^K - y_s^K| \, ds \geq K^\eta \right) \\ & \leq C_* K^{-2\eta} \mathbb{E} \left[\left(\left| \widetilde{M}_{T \log K}^K \right| + \int_0^{T \log K} e^{-\lambda s} |X_s^K - x_s^K| + e^{-\lambda s} |Y_s^K - y_s^K| \, ds \right)^2 \right], \end{aligned}$$

where $C_* > 0$ is a suitable constant, which may change in the following from line to line. Now, successively using $(a + b + c)^2 \leq 3(a^2 + b^2 + c^2)$, Hölder's Inequality and Fubini's Theorem, we see that

$$\begin{aligned} & \mathbb{P} \left(\sup_{t \leq T \log K} \left| e^{-\lambda t} (X_t^K + Y_t^K - (x_t^K + y_t^K)) \right| \geq K^\eta \right) \\ & \leq C_* K^{-2\eta} \mathbb{E} \left[\left(\left| \widetilde{M}_{T \log K}^K \right| + \int_0^{T \log K} e^{-\lambda s} |X_s^K - x_s^K| + e^{-\lambda s} |Y_s^K - y_s^K| \, ds \right)^2 \right] \\ & \leq C_* K^{-2\eta} \left(\mathbb{E}[\widetilde{M}_{T \log K}^2] + \mathbb{E} \left[\left(\int_0^{T \log K} e^{-\lambda s} |X_s^K - x_s^K| \, ds \right)^2 + \left(\int_0^{T \log K} e^{-\lambda s} |Y_s^K - y_s^K| \, ds \right)^2 \right] \right) \\ & \leq C_* K^{-2\eta} \mathbb{E}[(\widetilde{M}_{T \log K}^K)^2] \\ & \quad + C_* K^{-2\eta} T \log K \mathbb{E} \left[\int_0^{T \log K} e^{-2\lambda s} |X_s^K - x_s^K|^2 \, ds + \int_0^{T \log K} e^{-2\lambda s} |Y_s^K - y_s^K|^2 \, ds \right] \\ & \leq C_* K^{-2\eta} \left(\mathbb{E}[(\widetilde{M}_{T \log K}^K)^2] + T \log K \int_0^{T \log K} e^{-2\lambda s} (\mathbb{V}(X_s^K) + \mathbb{V}(Y_s^K)) \, ds \right). \end{aligned} \tag{A.6}$$

We will now estimate the expectation and the integral separately. Firstly, using the definition of \widetilde{M}^K , we easily see using Itô's Isometry and Lemma A.8 that

$$\begin{aligned}\mathbb{E}[(\widetilde{M}_{T \log K}^K)^2] &= \mathbb{E} \left[\left(\int_0^{T \log K} e^{-\lambda s} d(M_s^K + N_s^K) \right)^2 \right] = \mathbb{E} \left[\int_0^{T \log K} e^{-2\lambda s} d[M^K + N^K]_s \right] \\ &= 3 \int_0^{T \log K} e^{-2\lambda s} ((b_1 + d_1)x_s^K + (b_2 + d_2)y_s^K + K^c e^{as}) ds.\end{aligned}$$

Furthermore, using the calculation of the expected values x_s^K and y_s^K up to constants from Lemma A.7, we can estimate

$$\mathbb{E}[(\widetilde{M}_{T \log K}^K)^2] \leq C_*(1 + T^2) \log^2(K) \cdot (K^\beta + K^\gamma + K^c + K^{\beta-\lambda T} + K^{\gamma-\lambda T} + K^{c-\lambda T} + K^{c+(a-2\lambda)T})$$

again for some suitable constant $C_* > 0$, which may change from line to line and can without loss of generality be chosen sufficiently large such that (A.6) holds as well. Since we assume $c \leq \beta \vee \gamma$, we obtain the estimate

$$\mathbb{E}[(\widetilde{M}_{T \log K}^K)^2] \leq C_*(1 + T^2) \log^2(K) \cdot K^{(\beta \vee \gamma) \vee ((\beta \vee \gamma) - \lambda T) \vee (c + (a-2\lambda)T)}. \quad (\text{A.7})$$

We now turn to the integral. Using Lemma A.9 and A.7 again, as well as $c \leq \beta \vee \gamma$, we see that

$$\begin{aligned}&\int_0^{T \log K} e^{-2\lambda s} (\mathbb{V}(X_s^K) + \mathbb{V}(Y_s^K)) ds \\ &\leq C_* K^{(\beta \vee \gamma) \vee ((\beta \vee \gamma) - \lambda T) \vee (c + (a-2\lambda)T)} (1 + T^3) \log^3(K)\end{aligned} \quad (\text{A.8})$$

for some constant $C_* \geq 0$ sufficiently large. Hence, plugging the estimates (A.7) and (A.8) into (A.6), we see that

$$\begin{aligned}1 - \mathbb{P}(\Omega_1^K) &\leq \mathbb{P} \left(\sup_{t \leq T \log K} \left| e^{-\lambda t} (X_t^K + Y_t^K - (x_t^K + y_t^K)) \right| \geq K^\eta \right) \\ &\leq C_* K^{-2\eta} K^{(\beta \vee \gamma) \vee ((\beta \vee \gamma) - \lambda T) \vee (c + (a-2\lambda)T)} (1 + T^3) \log^3(K).\end{aligned}$$

From now on, we will consider η such that

$$\frac{(\beta \vee \gamma) \vee ((\beta \vee \gamma) - \lambda T) \vee (c + (a-2\lambda)T)}{2} < \eta < \beta \vee \gamma. \quad (\text{A.9})$$

This condition ensures as shown above that $\lim_{K \rightarrow \infty} \mathbb{P}(\Omega_1^K) = 1$. On the set Ω_1^K , we can obtain

$$\begin{aligned}&\sup_{t \leq T} \left| \frac{\log(1 + X_{t \log K}^K + Y_{t \log K}^K)}{\log K} - \bar{\beta}_t \right| \\ &= \sup_{t \leq T} \frac{1}{\log K} \left| \log \left(\frac{1 + X_{t \log K}^K + Y_{t \log K}^K}{1 + x_{t \log K}^K + y_{t \log K}^K} \right) + \log \left(\frac{1 + x_{t \log K}^K + y_{t \log K}^K}{K^{\bar{\beta}_t}} \right) \right| \\ &\leq \sup_{t \leq T} \frac{1}{\log K} \cdot \frac{K^{-\lambda t} |X_{t \log K}^K + Y_{t \log K}^K - (x_{t \log K}^K + y_{t \log K}^K)|}{K^{-\lambda t} (x_{t \log K}^K + y_{t \log K}^K) \wedge K^{-\lambda t} (X_{t \log K}^K + Y_{t \log K}^K)} + \frac{C_*}{\log K} \\ &\leq \frac{1}{\log K} \sup_{t \leq T} \frac{K^{\eta + \lambda t}}{x_{t \log K}^K + y_{t \log K}^K - K^{\eta + \lambda t}} + \frac{C_*}{\log K} \\ &\leq \frac{C_*}{\log K} (K^{\eta - (\beta \vee \gamma)} + 1) \xrightarrow{K \rightarrow \infty} 0,\end{aligned} \quad (\text{A.10})$$

$$\leq \frac{C_*}{\log K} (K^{\eta - (\beta \vee \gamma)} + 1) \xrightarrow{K \rightarrow \infty} 0, \quad (\text{A.11})$$

where again C_* is a sufficiently large constant, which may change from line to line. Note that the denominator in (A.10) is well defined for K large enough since $\eta < \beta \vee \gamma$. Also, the first inequality

holds due to our choice of $\bar{\beta}_t$, which gives in combination with Lemma A.7 that

$$K^{-\bar{\beta}_t}(1 + x_{t \log K}^K + y_{t \log K}^K) \leq C_*(K^{-\bar{\beta}_t} + 1) \leq 2C_*$$

for K large enough since by assumption $\bar{\beta}_t > 0$ for all $t \in [0, T]$. Thus, we are in the same situation as in Step 1 of the proof of Theorem B.1 in [CMT21] with $\beta \vee \gamma$ instead of β and λ instead of r . For completeness we distinguish the same cases:

Case 1(a): $\lambda \geq 0$ and $a \leq 2\lambda$: In this case the condition (A.9) reduces to $\frac{\beta \vee \gamma}{2} < \eta < \beta \vee \gamma$, so choosing $\eta = \frac{3(\beta \vee \gamma)}{4}$ shows the claim from (A.9) and (A.11).

Case 1(b): $\lambda < 0$ and $a \leq \lambda$: Here the assumption (A.5) becomes $\inf_{t \in [0, T]} (\beta \vee \gamma) + \lambda t > 0$, which gives $T < \frac{\beta \vee \gamma}{|\lambda|}$. Now condition (A.9) becomes

$$\frac{((\beta \vee \gamma) - \lambda T) \vee (c + (a - 2\lambda)T)}{2} = \frac{(\beta \vee \gamma) - \lambda T}{2} < \eta < \beta \vee \gamma.$$

By the condition on T , such η exists and we can again conclude.

Case 1(c): $\lambda \geq 0$ and $a > 2\lambda$: In this case the restriction (A.9) can be satisfied as long as T is such that $c + (a - 2\lambda)T < 2(\beta \vee \gamma)$, that is $T < T^* := \frac{2(\beta \vee \gamma)}{a - 2\lambda}$. Hence we obtain the convergence on all intervals $[0, T]$ such that $T < T^*$. In Step 3 we will consider $T > T^*$ which still may satisfy (A.5).

Case 1(d): $\lambda < 0$, $a > \lambda$ and $c + \frac{a(\beta \vee \gamma)}{|\lambda|} \leq 0$: Here we easily see that $(\beta \vee \gamma) + \lambda t \geq c + at$ for all $t \leq \frac{\beta \vee \gamma}{|\lambda|}$. Thus assumption (A.5) is satisfied if and only if $T < \frac{\beta \vee \gamma}{|\lambda|}$. For these times T the condition (A.9) can be satisfied for suitable η since

$$((\beta \vee \gamma) - \lambda T) < (\beta \vee \gamma - \lambda \frac{\beta \vee \gamma}{|\lambda|}) = 2(\beta \vee \gamma)$$

and

$$(c + (a - 2\lambda)T) < c + (a - 2\lambda) \frac{\beta \vee \gamma}{|\lambda|} = 2(\beta \vee \gamma) + c + a \frac{\beta \vee \gamma}{|\lambda|} \leq 2(\beta \vee \gamma).$$

Case 1(e): $\lambda < 0$, $a > \lambda$ and $c + \frac{a(\beta \vee \gamma)}{|\lambda|} \geq 0$: For these parameters, the condition (A.9) can be satisfied as long as $T < T^* = \frac{\beta \vee \gamma}{|\lambda|} \wedge \frac{2(\beta \vee \gamma) - c}{a - 2\lambda}$. Note that

$$\begin{aligned} \frac{\beta \vee \gamma}{|\lambda|} > \frac{2(\beta \vee \gamma) - c}{a - 2\lambda} &\iff (a - 2\lambda)(\beta \vee \gamma) > c\lambda - 2(\beta \vee \gamma)\lambda \\ &\iff a(\beta \vee \gamma) > c\lambda, \end{aligned}$$

which is true since $a > \lambda$ and $(\beta \vee \gamma) > c$. Hence $T^* = \frac{2(\beta \vee \gamma) - c}{a - 2\lambda}$ as in Case 1(c). As before we get the convergence for all $T < T^*$ and we will show in Step 3 how to obtain convergence for $T \geq T^*$ which satisfy (A.5).

Step 2: Strong Immigration. Here, we will consider only the case $\beta \vee \gamma = c$ and $a > \lambda$. Similarly to Step 1, for $\eta > 0$ to be determined later, we consider the set

$$\Omega_2^K := \left\{ \sup_{t \in [0, T \log K]} |e^{-at}(X_t^K + Y_t^K - (x_t^K + y_t^K))| \leq K^\eta \right\}.$$

We experience the same difficulties as in Step 1: We are not able to use a supermartingale inequality, since the switching between X^K and Y^K complicates our process. However, proceeding in the same

manner as in Step 1, we get the inequality

$$\begin{aligned} & \mathbb{P} \left(\sup_{t \leq T \log K} |e^{-at}(X_t^K + Y_t^K - (x_t^K + y_t^K))| \geq K^\eta \right) \\ & \leq C_* K^{-2\eta} \left(\mathbb{E}[(\widehat{M}_{T \log K}^K)^2] + T \log K \int_0^{T \log K} e^{-2as} (\mathbb{V}(X_s^K) + \mathbb{V}(Y_s^K)) \, ds \right), \end{aligned}$$

where $\widehat{M}_t^K := \int_0^t e^{-as} \, d(M_s^K + N_s^K)$ is a martingale. Applying our estimates and the Itô Isometry from above gives with another calculation similar to the corresponding part in step 1 that

$$\begin{aligned} 1 - \mathbb{P}(\Omega_2^K) & \leq \mathbb{P} \left(\sup_{t \leq T \log K} |e^{-at}(X_t^K + Y_t^K - (x_t^K + y_t^K))| \geq K^\eta \right) \\ & \leq C_* K^{-2\eta} K^{\beta \vee \gamma \vee ((\beta \vee \gamma) + (2(\lambda - a) \vee (\lambda - 2a))T) \vee (c - aT)} (1 + T^3) \log^3(K) \\ & = C_* K^{-2\eta} K^{\beta \vee \gamma \vee ((\beta \vee \gamma) - aT)} (1 + T^3) \log^3(K), \end{aligned}$$

where we used $a > \lambda$ and $c = \beta \vee \gamma$ in the last equality. Indeed the exponent $2(\lambda - a)T$ is always negative and can therefore be omitted. On the other hand $\lambda - 2a < -a$ and thus $(\lambda - 2a)T$ may be replaced by $-aT$, which is accounted for in the last equality. Therefore, we now consider η such that

$$\frac{(\beta \vee \gamma) \vee ((\beta \vee \gamma) - aT)}{2} < \eta < \beta \vee \gamma, \quad (\text{A.12})$$

which ensures that $\mathbb{P}(\Omega_2^K) \rightarrow 1$ as $K \rightarrow \infty$. Again a calculation similar to step 1 shows that for this choice of η we have

$$\begin{aligned} & \sup_{t \leq T} \left| \frac{\log(1 + X_{t \log K}^K + Y_{t \log K}^K)}{\log K} - \bar{\beta}_t \right| \\ & \leq \sup_{t \leq T} \frac{1}{\log K} \cdot \frac{K^{-at} |X_{t \log K}^K + Y_{t \log K}^K - (x_{t \log K}^K + y_{t \log K}^K)|}{K^{-at} (x_{t \log K}^K + y_{t \log K}^K) \wedge K^{-\lambda t} (X_{t \log K}^K + Y_{t \log K}^K)} + \frac{C_*}{\log K} \\ & \leq \frac{1}{\log K} \sup_{t \leq T} \frac{K^{\eta + \lambda t}}{x_{t \log K}^K + y_{t \log K}^K - K^{\eta + at}} + \frac{C_*}{\log K} \\ & \leq \frac{C_*}{\log K} \left(K^{\eta - (\beta \vee \gamma)} + 1 \right) \xrightarrow{K \rightarrow \infty} 0. \end{aligned}$$

This computation allows us to show our convergence result for two more possible cases.

Case 2(a): $c = \beta \vee \gamma$, $a > \lambda$ and $a \geq 0$.: As in Case 1(a) we may choose $\eta = \frac{3(\beta \vee \gamma)}{4}$ and have shown convergence for this case.

Case 2(b): $c = \beta \vee \gamma$, $a > \lambda$ and $a < 0$.: Here, condition (A.5) on the final time T is satisfied if and only if $T < \frac{\beta \vee \gamma}{|a|}$. Hence

$$\frac{(\beta \vee \gamma) - aT}{2} < \frac{(\beta \vee \gamma) - a \frac{\beta \vee \gamma}{|a|}}{2} = \beta \vee \gamma$$

and thus we can find η such that (A.12) is satisfied.

Step 3: Completion of Step 1. It remains to extend the following two cases to $T > T^* = \frac{2(\beta \vee \gamma) - c}{a - 2\lambda}$:

- $\lambda \geq 0$, $a > 2\lambda$ and $c < \beta \vee \gamma$,
- $\lambda < 0$, $a > \lambda$, $c < \beta \vee \gamma$ and $c + \frac{a(\beta \vee \gamma)}{|\lambda|} \geq 0$.

This can be done exactly as in [CMT21] in Step 3 of the proof of Theorem B.1. In order to do so, we note that at time $t^* := \frac{(\beta \vee \gamma) - c}{a - \lambda}$ the lines $(\beta \vee \gamma) + \lambda t$ and $c + at$ intersect. Furthermore, we see that in both of the above cases $t^* < T^*$ since in the first case we may assume without loss of generality that $a > \lambda$ (otherwise t^* is negative) and therefore

$$\begin{aligned} \frac{2(\beta \vee \gamma) - c}{a - 2\lambda} > \frac{(\beta \vee \gamma) - c}{a - \lambda} &\iff 2(\beta \vee \gamma)(a - \lambda) - ac + \lambda c > (\beta \vee \gamma)(a - 2\lambda) - ac + 2\lambda c \\ &\iff (\beta \vee \gamma)a > \lambda c, \end{aligned}$$

which is true. In the second case we can perform a similar computation. Therefore we may apply Case 1(c) or Case 1(e) to our process in each case up to a time $T_1 \in (t^*, T^*)$. Note that at this time the limiting function satisfies $\bar{\beta}_{T_1} = c + aT_1$. Hence for all $\varepsilon > 0$ on a set Ω_3^K with $\mathbb{P}(\Omega_3^K) \rightarrow 1$ as $K \rightarrow \infty$ we have

$$K^{c+aT_1-\varepsilon} \leq X_{T_1 \log K}^K + Y_{T_1 \log K}^K \leq K^{c+aT_1+\varepsilon}.$$

We now couple the process $Z_{T_1 \log K+t}^K = (X_{T_1 \log K+t}^K, Y_{T_1 \log K+t}^K)$ in the following manner: Let $\hat{Z}_t^K = (\hat{X}_t^K, \hat{Y}_t^K)$ be a $BBPI_K(b_1, b_2, d_1, d_2, \sigma_1, \sigma_2, a, c+aT_1-\varepsilon, c+aT_1-\varepsilon, c+aT_1-\varepsilon)$ and let $\bar{Z}_t^K = (\bar{X}_t^K, \bar{Y}_t^K)$ be a $BBPI_K(b_1, b_2, d_1, d_2, \sigma_1, \sigma_2, a, c+aT_1+\varepsilon, c+aT_1+\varepsilon, c+aT_1+\varepsilon)$ such that

$$\hat{X}_t^K + \hat{Y}_t^K \leq X_{T_1 \log K+t}^K + Y_{T_1 \log K+t}^K \leq \bar{X}_t^K + \bar{Y}_t^K.$$

Indeed, the starting conditions of the bounding processes are justified by Remark A.11. Then, we can apply the convergence from Step 2 to \hat{Z}_t^K and \bar{Z}_t^K to show that

$$\frac{\log(1 + \hat{X}_{t \log K}^K + \hat{Y}_{t \log K}^K)}{\log K} \xrightarrow{K \rightarrow \infty} c + aT_1 - \varepsilon + ((\lambda \vee a)t) = c - \varepsilon + a(T_1 + t)$$

and

$$\frac{\log(1 + \bar{X}_{t \log K}^K + \bar{Y}_{t \log K}^K)}{\log K} \xrightarrow{K \rightarrow \infty} c + aT_1 + \varepsilon + ((\lambda \vee a)t) = c + \varepsilon + a(T_1 + t)$$

where $t \in [0, T - T_1]$. Note that for the second case, $a < 0$ and therefore the condition (A.5) is satisfied only for $T < \frac{c}{|a|}$. In particular, for ε small enough, we even have $T < \frac{c-\varepsilon}{|a|}$ and therefore $T - T_1 < \frac{c+aT_1-\varepsilon}{|a|}$, so we can indeed apply case 2(b). Using the Markov property at time T_1 and letting $\varepsilon \rightarrow 0$ finishes the proof. \square

As in [CMT21], we want to extend Theorem A.10 to further cases without needing the assumption $c \leq \beta \vee \gamma$ or the positivity condition (A.5). To obtain the convergence result for $c > \beta \vee \gamma$, we will use the next lemma. This, combined with using the Markov property and our previous Theorem A.10, already extends the convergence to all processes such that $c \in \mathbb{R}$, $\beta \vee \gamma > 0$ and the terminal time T satisfies condition (A.5)

$$\inf_{t \in [0, T]} ((\beta \vee \gamma) + \lambda t) \vee (c + at) > 0.$$

Lemma A.12. *Let $0 \leq \beta \vee \gamma < c$. Then for all $0 < \varepsilon < \frac{c}{4(|\lambda| \vee |a|)}$ and all $\bar{a} > |\lambda| \vee |a|$, the convergence*

$$\lim_{K \rightarrow \infty} \mathbb{P}(X_{\varepsilon \log K}^K + Y_{\varepsilon \log K}^K \in [K^{c-\bar{a}\varepsilon}, K^{c+\bar{a}\varepsilon}]) = 1$$

holds true.

Proof. Let $\varepsilon > 0$. Then we see from Lemma A.7 that for K sufficiently large

$$C_*^{-1} K^{c-(|\lambda| \vee |a|)\varepsilon} \leq \mathbb{E}(X_{\varepsilon \log K}^K + Y_{\varepsilon \log K}^K) \leq C_* \log(K) K^{c+(|\lambda| \vee |a|)\varepsilon}$$

and from Lemma A.9

$$\mathbb{V}(X_{\varepsilon \log K}^K + Y_{\varepsilon \log K}^K) \leq 2(\mathbb{V}(X_{\varepsilon \log K}^K) + \mathbb{V}(Y_{\varepsilon \log K}^K)) \leq C_* \log^2(K) K^{c+(2|\lambda| \vee |a|)\varepsilon}$$

for a suitable constant $C_* > 0$. Hence, using Markov's inequality for K large enough, we see that

$$\begin{aligned} \mathbb{P}(X_{\varepsilon \log K}^K + Y_{\varepsilon \log K}^K \leq K^{c+\bar{a}\varepsilon}) &= 1 - \mathbb{P}(X_{\varepsilon \log K}^K + Y_{\varepsilon \log K}^K > K^{c+\bar{a}\varepsilon}) \\ &\geq 1 - \frac{\mathbb{E}(X_{\varepsilon \log K}^K + Y_{\varepsilon \log K}^K)}{K^{c+\bar{a}\varepsilon}} \xrightarrow{K \rightarrow \infty} 1. \end{aligned}$$

For the lower bound we can use Chebyshev's inequality to obtain

$$\begin{aligned} &\mathbb{P}(X_{\varepsilon \log K}^K + Y_{\varepsilon \log K}^K \leq K^{c-\bar{a}\varepsilon}) \\ &\leq \mathbb{P}(X_{\varepsilon \log K}^K + Y_{\varepsilon \log K}^K - (x_{\varepsilon \log K}^K + y_{\varepsilon \log K}^K) \leq K^{c-\bar{a}\varepsilon} - C_*^{-1} K^{c-(|\lambda| \vee |a|)\varepsilon}) \\ &\leq \mathbb{P}(|X_{\varepsilon \log K}^K + Y_{\varepsilon \log K}^K - (x_{\varepsilon \log K}^K + y_{\varepsilon \log K}^K)| \geq C'_* K^{c-(|\lambda| \vee |a|)\varepsilon}) \\ &\leq C'_* \frac{\mathbb{V}(X_{\varepsilon \log K}^K + Y_{\varepsilon \log K}^K)}{K^{2c-2(|\lambda| \vee |a|)\varepsilon}} \\ &\leq C'_* \log^2(K) K^{-c+4(|\lambda| \vee |a|)\varepsilon} \xrightarrow{K \rightarrow \infty} 0 \end{aligned}$$

for some suitable changing constant $C'_* > 0$. \square

A.3. Proof of Theorem A.3. Henceforth, thanks to the previous Lemma, we are able to only consider the case $c \leq \beta \vee \gamma$ for the remainder of this section. In order to extend the convergence result Theorem A.10 to times T that do not satisfy the condition (A.5), we need another series of Lemmata.

Lemma A.13. *Let $\beta \vee \gamma = 0$ (that is, initially there are no individuals) and $c < 0$. Then, for all $T > 0$ with $c + aT < 0$ we have*

$$\lim_{K \rightarrow \infty} \mathbb{P}(X_t^K + Y_t^K = 0 \text{ for all } t \leq T \log K) = 1.$$

Proof. This proof is identical to the one of [CMT21, Lemma B.7]. \square

Lemma A.14. *Let $\beta \vee \gamma = 0$ and $c = -\varepsilon$ for some $\varepsilon > 0$ and let $a > 0$. Then for all $\eta > (1 \vee \frac{2\lambda}{a})\varepsilon$, the convergence*

$$\lim_{K \rightarrow \infty} \mathbb{P}(K^{\frac{\varepsilon}{2}} - 1 \leq X_{\frac{2\varepsilon}{a} \log K}^K + Y_{\frac{2\varepsilon}{a} \log K}^K \leq K^\eta - 1) = 1$$

holds.

Proof. We consider the one-dimensional branching process \tilde{X}_t^K with birth rate b_1 , death rate $d_1 + \sigma_1$, immigration at rate $K^c e^{at}$ and starting condition $X_0^K = K^\beta - 1 = 0$. Then, from the proof of Lemma B.8 from [CMT21], we have the convergence

$$\lim_{K \rightarrow \infty} \mathbb{P}(K^{\frac{\varepsilon}{2}} - 1 \leq \tilde{X}_{\frac{2\varepsilon}{a} \log K}^K) = 1.$$

Using a suitable coupling, we also have that $\tilde{X}_t^K \leq X_t^K + Y_t^K$ for all $t \geq 0$. Hence, it holds

$$\lim_{K \rightarrow \infty} \mathbb{P}(K^{\frac{\varepsilon}{2}} - 1 \leq X_{\frac{2\varepsilon}{a} \log K}^K + Y_{\frac{2\varepsilon}{a} \log K}^K) = 1.$$

For the upper bound, we know from Lemma A.7 with $x_0 = y_0 = 0$, that

$$\mathbb{E}(X_{\frac{2\varepsilon}{a} \log K}^K + Y_{\frac{2\varepsilon}{a} \log K}^K) \leq \begin{cases} C_* K^\varepsilon, & \text{if } a > \lambda \\ C_* K^\varepsilon \log K, & \text{if } a = \lambda \\ C_* K^{\frac{2\lambda}{a}\varepsilon}, & \text{if } a < \lambda. \end{cases}$$

With our choice of $\eta > (1 \vee \frac{2\lambda}{a})\varepsilon$ and Markov's inequality, we have

$$\lim_{K \rightarrow \infty} \mathbb{P}(X_{\frac{2\varepsilon}{a} \log K}^K + Y_{\frac{2\varepsilon}{a} \log K}^K \geq K^\eta) \leq \lim_{K \rightarrow \infty} K^{-\eta} \mathbb{E}(X_{\frac{2\varepsilon}{a} \log K}^K + Y_{\frac{2\varepsilon}{a} \log K}^K) = 0.$$

Thus the lemma is proven. \square

Lemma A.15. *There exists a constant $\bar{c} = \bar{c}(b_1, b_2, d_1, d_2, \sigma_1, \sigma_2, a)$, such that for all $\varepsilon > 0$ we have the convergence*

$$\lim_{K \rightarrow \infty} \mathbb{P}(K^{(\beta \vee \gamma) - \bar{c}\varepsilon} - 1 \leq X_t^K + Y_t^K \leq K^{(\beta \vee \gamma) + \bar{c}\varepsilon} \text{ for all } t \in [0, \varepsilon \log K]) = 1.$$

Proof. This result is very similar to [CMT21, Lemma B.9] and can be proven analogously. \square

Lemma A.16. *Suppose $\lambda < 0$, where λ is taken from (A.3).*

(i) *In addition, let $c < 0$ and $c + a \frac{\beta \vee \gamma}{|\lambda|} < 0$. Then, for all sufficiently small $\eta > 0$ it holds*

$$\lim_{K \rightarrow \infty} \mathbb{P} \left(\forall t \in \left[\left(\frac{\beta \vee \gamma}{|\lambda|} + \eta \right) \log K, \left(\frac{\beta \vee \gamma}{|\lambda|} + 2\eta \right) \log K \right] : X_t^K + Y_t^K = 0 \right) = 1.$$

(ii) *If in addition (independent of (i)) $a < 0$ and $c + a \frac{\beta \vee \gamma}{|\lambda|} > 0$, then for all $\eta > 0$ and all $T > \eta$, we have*

$$\lim_{K \rightarrow \infty} \mathbb{P} \left(\forall t \in \left[\left(\frac{c}{|a|} + \eta \right) \log K, \left(\frac{c}{|a|} + T \right) \log K \right] : X_t^K + Y_t^K = 0 \right) = 1.$$

Proof. This result is the bi-type analogue of [CMT21, Lemma B.10]. We start by proving part (i). Let η be small enough such that $c + a \left(\frac{\beta \vee \gamma}{|\lambda|} + 2\eta \right) < 0$. Define $T_1 := \frac{\beta \vee \gamma}{|\lambda|} + \eta$ and $T_2 := T_1 + \eta$. Then the probability of a migrant arriving during the interval $[0, T_2 \log K]$ converges to 0 as $K \rightarrow \infty$. This can be seen from the probability of immigration being bounded by $K^{c \vee (c + aT_2)} T_2 \log K$, which converges to 0 as $K \rightarrow \infty$ (cf. Lemma A.13). Hence, it suffices to show that, assuming no immigration occurs, the extinction time $T_{\text{ext}} := \inf\{t \geq 0 \mid X_t^K + Y_t^K = 0\}$ is asymptotically almost surely less than $T_1 \log K$, that is, our population itself is extinct before time $T_1 \log K$, and since no migrant arrives during the time interval $[T_{\text{ext}}, T_2 \log K]$, which would potentially resurrect the population, the claim follows.

Denote the event that a migrant arrives in the population during the time interval $[0, T_1 \log K]$ by Γ . Then, on the complement, the process $Z_t^K = (X_t^K, Y_t^K)$ behaves as a bi-type branching process $\tilde{Z}_t^K = (\tilde{X}_t^K, \tilde{Y}_t^K)$ with birth rates b_1, b_2 , death rates d_1, d_2 , switching rates σ_1, σ_2 , starting condition $\tilde{Z}_0^K = Z_0^K$ and no immigration. Hence, we obtain using Markov's inequality

$$\begin{aligned} \mathbb{P}(T_{\text{ext}} > T_1 \log K, \Gamma^c) &= \mathbb{P}(\tilde{X}_{T_1 \log K}^K + \tilde{Y}_{T_1 \log K}^K \geq 1) \\ &\leq \mathbb{E}(\tilde{X}_{T_1 \log K}^K + \tilde{Y}_{T_1 \log K}^K) \\ &\leq C_* e^{\lambda T_1 \log K} K^{\beta \vee \gamma} \\ &= C_* K^{(\beta \vee \gamma) + \lambda T_1}, \end{aligned} \tag{A.13}$$

where we can find the bound on the expected value for some constant $C_* > 0$ from the homogeneous solution to (A.1). Therefore, by our choice of T_1 , we obtain

$$\mathbb{P}(T_{\text{ext}} > T_1 \log K) \leq \mathbb{P}(\Gamma) + C_* K^{(\beta \vee \gamma) + \lambda T_1} \xrightarrow{K \rightarrow \infty} 0.$$

For part (ii) we may assume without loss of generality, that $c \leq \beta \vee \gamma < \frac{\eta|\lambda|}{4}$. To justify this, we apply Theorem A.10 until time $T \geq \frac{\beta \vee \gamma}{|\lambda|}$ which satisfies $r := ((\beta \vee \gamma) + \lambda T) \vee (c + aT) < \frac{\eta|\lambda|}{4}$. This allows us to consider the process $Z^K \sim BBPI_K(b_1, b_2, d_1, d_2, \sigma_1, \sigma_2, a, c + aT, r, r)$ instead by using the Markov property at time T , which satisfies our assumption. Indeed, by definition of T and r , we have $c + aT = r$.

Under the assumption $c \leq \beta \vee \gamma < \frac{\eta|\lambda|}{4}$, the condition $c + a \frac{\beta \vee \gamma}{|\lambda|} > 0$ implies $a > \lambda$. Now, let $\eta > 0$ and $T > \eta$ be arbitrary. As in case (i), we can prove that as $K \rightarrow \infty$ there is almost surely no immigrant arriving in the population on the time interval $[(\frac{c}{|a|} + \eta) \log K, (\frac{c}{|a|} + T) \log K]$. Denote the event of a migrant arriving during this time by Γ . From now on, we only consider the event Γ^c . Then we only need to show that on Γ^c , the process becomes extinct before time $(\frac{c}{|a|} + \eta) \log K$. Note that the number

of total families initially present in the population and those families started due to an immigration event up to time $(\frac{c}{|a|} + \frac{\eta}{2}) \log K$ is given by $K^{\beta \vee \gamma}$ (representing the families initially present) plus a Poisson random variable whose parameter is bounded from above by

$$\int_0^{\frac{c}{|a|} + \frac{\eta}{2} \log K} K^c e^{as} ds \leq \frac{K^c}{|a|},$$

which represents the number of families coming from immigration. In particular, the total number of families is less than $K^{\eta\lambda/3}$ with probability converging to 1. The size of such a family at time t is bounded from above by the size of a bi-type branching process $\tilde{Z} = (\tilde{X}, \tilde{Y})$ with birth rates b_1, b_2 , death rates d_1, d_2 , switching rates σ_1, σ_2 , no immigration and starting population $\tilde{Z}_0 = (1, 1)$. As in part (i), we see that the probability of such a process surviving for a time longer than $\frac{\eta}{2} \log K$ is dominated by

$$\mathbb{P}(\tilde{X}_{\frac{\eta}{2} \log K} + \tilde{Y}_{\frac{\eta}{2} \log K} \geq 1) \leq \mathbb{E}(\tilde{X}_{\frac{\eta}{2} \log K} + \tilde{Y}_{\frac{\eta}{2} \log K}) \leq C_* e^{\lambda \frac{\eta}{2} \log K} = C_* K^{\frac{\lambda \eta}{2}},$$

where we get the bound on the expectation from solving the homogeneous equation of (A.1) with initial condition $(\tilde{X}_0, \tilde{Y}_0) = (1, 1)$. Therefore, the probability of having one family alive at time $(\frac{c}{|a|} + \eta) \log K$ is given by the probability of at least one family alive at time $(\frac{c}{|a|} + \frac{\eta}{2}) \log K$ surviving for longer than $\frac{\eta}{2} \log K$, which is dominated by

$$(1 - C_* K^{\frac{\lambda \eta}{2}})^{K^{\frac{\eta \lambda}{3}}} \xrightarrow{K \rightarrow \infty} 0.$$

Therefore, the overall probability of having an individual alive at time $(\frac{c}{|a|} + \eta) \log K$ is dominated by

$$\mathbb{P}(\Gamma) + (1 - C_* K^{\frac{\lambda \eta}{2}})^{K^{\frac{\eta \lambda}{3}}} \xrightarrow{K \rightarrow \infty} 0.$$

Since the process is extinct at time $(\frac{c}{|a|} + \eta) \log K$ with probability converging to 1 and also with high probability there is no migrant arriving in the population after this time, the lemma is proven. \square

To end this section, we can now prove the general convergence from Theorem A.3 which we were looking for.

Proof of Theorem A.3. This proof is taken from [CMT21, Theorem B.5] and adapted to our case.

(iii) This is a direct consequence of Lemma A.13.

(ii) Let $\varepsilon > 0$. We can apply Lemma A.13 up to time $T_1 = \frac{|c|}{a} - \varepsilon$. This shows

$$\lim_{K \rightarrow \infty} \frac{\log(1 + X_{t \log K}^K + Y_{t \log K}^K)}{\log K} = 0 \quad \text{for all } t \in [0, T_1] \text{ almost surely.}$$

Applying the Markov property at time $T_1 \log K$, we can apply Lemma A.14 and Lemma A.15 to see that

$$\frac{\log(1 + X_{(T_1 + \delta \varepsilon) \log K}^K + Y_{(T_1 + \delta \varepsilon) \log K}^K)}{\log K} \in (\underline{c\varepsilon}, \bar{c\varepsilon})$$

and

$$\limsup_{K \rightarrow \infty} \sup_{t \in [T_1, T_1 + \delta \varepsilon]} \frac{\log(1 + X_{t \log K}^K + Y_{t \log K}^K)}{\log K} \leq c' \varepsilon,$$

where $\delta, \underline{c} > 0$ and $\bar{c} < c' < \infty$. Now, applying Theorem A.10 to a coupling from time $(T_1 + \delta \varepsilon) \log K$ with a lower bounding process $BBPI_K(b_1, b_2, d_1, d_2, \sigma_1, \sigma_2, a, \underline{c\varepsilon}, \underline{c\varepsilon}, \underline{c\varepsilon})$ and an upper bounding process $BBPI_K(b_1, b_2, d_1, d_2, \sigma_1, \sigma_2, a, \bar{c\varepsilon}, \bar{c\varepsilon}, \bar{c\varepsilon})$ shows that for times $t \in$

$[T_1 + \delta\varepsilon, T]$ it holds

$$\begin{aligned} c\varepsilon + (\lambda \vee a)(t - t_1 - \delta\varepsilon) &\leq \liminf_{K \rightarrow \infty} \frac{\log(1 + X_{t \log K}^K + Y_{t \log K}^K)}{\log K} \\ &\leq \limsup_{K \rightarrow \infty} \frac{\log(1 + X_{t \log K}^K + Y_{t \log K}^K)}{\log K} \\ &\leq \bar{c}\varepsilon + (\lambda \vee a)(t - t_1 - \delta\varepsilon). \end{aligned}$$

Now letting $\varepsilon \rightarrow 0$ proves the claim.

- (i) Note that this part follows immediately from Theorem A.10 in the case where $\lambda \geq 0$ or if $\lambda < 0$, $a \geq 0$ and $c + a \frac{\beta \vee \gamma}{|\lambda|} > 0$. The remaining cases are as follows.

Case(a): $\lambda < 0$ and $a < 0$:. Assume for now that $c + a \frac{\beta \vee \gamma}{|\lambda|} < 0$. Then, we can apply Theorem A.10 on the interval $[0, \frac{\beta \vee \gamma}{|\lambda|} - \varepsilon]$ for $\varepsilon > 0$. Using Lemma A.16 (i), we also see that the process converges on the interval $[\frac{\beta \vee \gamma}{|\lambda|} + \varepsilon, \frac{\beta \vee \gamma}{|\lambda|} + 2\varepsilon]$. Using Lemma A.15 together with the coupling argument from (ii) and letting $\varepsilon \rightarrow 0$ shows the convergence of the process on the interval $[0, T_1]$ for $T_1 > \frac{\beta \vee \gamma}{|\lambda|}$ sufficiently small against $\bar{\beta}$. Using the Markov property at time T_1 we can apply Lemma A.13 to obtain convergence on the entire interval $[0, T]$ towards $\bar{\beta}$.

If $c + a \frac{\beta \vee \gamma}{|\lambda|} > 0$, we obtain convergence on the interval $[0, \frac{c}{|a|} - \varepsilon]$ from Theorem A.10 and use Lemma A.16 (ii) instead. The remainder of the argument is still valid. In the case where $c + a \frac{\beta \vee \gamma}{|\lambda|} = 0$, we can use a coupling argument where $\tilde{Z}^K = (\tilde{X}^K, \tilde{Y}^K)$ is a $BBPI_K(b_1, b_2, d_1, d_2, \sigma_1, \sigma_2, a, c - \varepsilon, \beta, \gamma)$ and $\hat{Z}^K = (\hat{X}^K, \hat{Y}^K)$ is a $BBPI_K(b_1, b_2, d_1, d_2, \sigma_1, \sigma_2, a, c + \varepsilon, \beta, \gamma)$ such that

$$\tilde{X}_t^K + \tilde{Y}_t^K \leq X_t^K + Y_t^K \leq \hat{X}_t^K + \hat{Y}_t^K,$$

and let $\varepsilon \rightarrow 0$.

Case(b): $\lambda < 0$, $a = 0$ and $c < 0$:. This case can be argued as in the first paragraph of the proof of Case (a).

Case(c): $\lambda < 0$, $a > 0$, $c < 0$ and $\frac{\beta \vee \gamma}{|\lambda|} < \frac{|c|}{a}$:. In this case, the population becomes extinct at first and is then revived due to immigration. Hence, we apply Theorem A.10 up to time $\frac{\beta \vee \gamma}{|\lambda|} - \varepsilon$ and Lemma A.16 (i) on the interval $[\frac{\beta \vee \gamma}{|\lambda|} + \varepsilon, \frac{\beta \vee \gamma}{|\lambda|} + 2\varepsilon]$ for $\varepsilon > 0$. As in Case (a), using Lemma A.15 and letting $\varepsilon \rightarrow 0$ gives again convergence on the interval $[0, T_1]$ for $T_1 > \frac{\beta \vee \gamma}{|\lambda|}$ sufficiently small. Then, using the Markov property allows the application of Lemma A.13 until time $T_2 = \frac{|c|}{a}$. Again, we can use the Markov property and Lemma A.15 to apply Lemma A.14 on a sufficiently small interval $[T_2, T_3]$. Then, using the Markov property again and using another coupling argument as in Case (a), letting $T_3 \rightarrow T_2$, we can apply Theorem A.10 to obtain convergence towards $\bar{\beta}$ on the entire interval $[0, T]$.

Case(d): $\lambda < 0$, $a > 0$, $c < 0$ and $\frac{\beta \vee \gamma}{|\lambda|} = \frac{|c|}{a}$:. Here, we can couple as before with a lower bound process $BBPI_K(b_1, b_2, d_1, d_2, \sigma_1, \sigma_2, a, c - \varepsilon, \beta, \gamma)$, for which we can apply Case (c), and the upper bound process $BBPI_K(b_1, b_2, d_1, d_2, \sigma_1, \sigma_2, a, c + \varepsilon, \beta + \varepsilon, \gamma + \varepsilon)$, which satisfies $c + \varepsilon + a \frac{(\beta \vee \gamma) + \varepsilon}{|\lambda|} > 0$ and thus has already been treated as one of the trivial cases above. Letting $\varepsilon \rightarrow 0$ yields the claim.

Case(e): $\lambda < 0$, $c = a = 0$:. Here, we use a similar coupling argument as in Case (d). The lower bound process has distribution $BBPI_K(b_1, b_2, d_1, d_2, \sigma_1, \sigma_2, 0, -\varepsilon, \beta, \gamma)$, which satisfies the assumption of Case (b), and the upper bound process has distribution

$BBPI_K(b_1, b_2, d_1, d_2, \sigma_1, \sigma_2, 0, \varepsilon, \beta \vee \varepsilon, \gamma \vee \varepsilon)$, which satisfies $\varepsilon + 0 \cdot \frac{\beta \vee \gamma \vee \varepsilon}{|\lambda|} > 0$, so this case is treated as the second trivial case above. The claim follows by letting $\varepsilon \rightarrow 0$.

□

APPENDIX B. RESULTS ON LOGISTIC PROCESSES

In this section, we consider the bi-type logistic birth and death process $Z_t^K = (X_t^K, Y_t^K)$ where the transitions are given through

$$(n, m) \mapsto \begin{cases} (n+1, m) & \text{at rate } nb_1^K(\omega, t) + \gamma_1^K(\omega, t) \\ (n-1, m) & \text{at rate } n(d_1^K(\omega, t) + \frac{(1-p)C}{K}n) \\ (n, m-1) & \text{at rate } md_2^K(\omega, t) \\ (n-1, m+1) & \text{at rate } \frac{pC}{K}n^2 \\ (n+1, m-1) & \text{at rate } m\sigma_2 \end{cases}$$

with predictable, non-negative functions $b_1^K, d_1^K, d_2^K, \gamma_1^K: \Omega \times [0, \infty) \rightarrow \mathbb{R}$ and constants $C, \sigma_2 > 0$, $p \in (0, 1)$.

Lemma B.1. *Suppose that there are constants $b_1, d_1, d_2 \geq 0$ such that*

$$\sup_{0 \leq t \leq s \log K} \|b_1^K(t) - b_1\| + \|d_1^K(t) - d_1\| + \|d_2^K(t) - d_2\| + \left\| \frac{\gamma_1^K(t)}{K} \right\| \xrightarrow{K \rightarrow \infty} 0 \quad (\text{B.1})$$

in probability for some $s > 0$. If we have $\frac{Z_0^K}{K} \rightarrow (\varepsilon_1, \varepsilon_2)$ as $K \rightarrow \infty$ for fixed $\varepsilon_1, \varepsilon_2 > 0$, then the process $\frac{Z_t^K}{K}$ converges uniformly on compact intervals in probability towards the solution $(x(t), y(t))$ of the ordinary differential equation

$$\begin{aligned} \dot{x}(t) &= (b_1 - d_1)x(t) - Cx^2(t) + \sigma_2 y(t) \\ \dot{y}(t) &= -(d_2 + \sigma_2)y(t) + pCx^2(t) \end{aligned} \quad (\text{B.2})$$

with initial condition $(x(0), y(0)) = (\varepsilon_1, \varepsilon_2)$ as $K \rightarrow \infty$.

Proof. This is similar to [EK86, Theorem 11.2.1]. □

Notation B.2. *We denote processes Z^K as introduced above by $LBBI_K(b_1^K, d_1^K, d_2^K, \sigma_2, p, C, \gamma_1^K)$. In the case where the functions b_1^K, d_1^K and d_2^K are all constant and $\gamma_1^K \equiv 0$, we may refer to the process as a $LBBI_K(b_1, d_1, d_2, \sigma_2, p, C)$.*

We are interested in calculating a coordinatewise positive equilibrium of the system (B.2). Assume that the equilibrium (\bar{x}, \bar{y}) is positive, such that we can divide both sides of (B.2) by x to obtain

$$\frac{\bar{y}}{\bar{x}} = -\frac{b_1 - d_1 - C\bar{x}}{\sigma_2} = \frac{pC\bar{x}}{d_2 + \sigma_2}$$

Hence, we see that

$$\bar{x} = \frac{(b_1 - d_1)(d_2 + \sigma_2)}{C(d_2 + (1-p)\sigma_2)} \quad \text{and} \quad \bar{y} = \frac{p(b_1 - d_1)^2(d_2 + \sigma_2)}{C(d_2 + (1-p)\sigma_2)^2}. \quad (\text{B.3})$$

Thus, the assumption $b_1 > d_1$ is sufficient for obtaining a coordinatewise positive equilibrium. As is shown in [BT20, Section 2.2], this equilibrium is the only stable equilibrium and in fact the system converges towards this equilibrium for any initial condition $(\varepsilon_1, \varepsilon_2) \in (0, \infty)^2$. This can be seen from Lemma 4.6 in [BT20].

B.1. Problem of Exit and Entry of a Domain. We will now concern ourselves with estimating the length of time until a logistic bi-type branching process exits a neighbourhood of its equilibrium population size; and with the existence of a time such that the process enters for the first time a neighbourhood of its equilibrium. The important results in this section are Lemma B.5 for the entry into a domain around the equilibrium and Corollary B.8 for the exit of such a domain. Our first step is to generalize Lemma B.1.

Lemma B.3. *Let $T > 0$ and $b_1, b_2, d_1, d_2 \geq 0$, $\sigma_2 > 0$. Further let $C > 0$, $p \in (0, 1)$ and let \tilde{C} be a compact subset of $(0, \infty)^2$. Denote the solution of the differential equation*

$$\begin{aligned}\dot{\varphi}_1 &= (b_1 - d_1 - C\varphi_1)\varphi_1 + \sigma_2\varphi_2 \\ \dot{\varphi}_2 &= (b_2 - d_2 - \sigma_2)\varphi_2 + Cp(\varphi_1)^2\end{aligned}\tag{B.4}$$

with initial condition $(z_1, z_2) \in \tilde{C}$ by φ_{z_1, z_2} . Then for any $T > 0$,

$$r := \inf_{z \in \tilde{C}} \inf_{t \in [0, T]} \|\varphi_{z_1, z_2}(t)\| > 0 \quad \text{and} \quad R := \sup_{z \in \tilde{C}} \sup_{t \in [0, T]} \|\varphi_{z_1, z_2}(t)\| < \infty.$$

Denote the distribution of the Markov process with transition rates

$$\left(\frac{n}{K}, \frac{m}{K}\right) \mapsto \begin{cases} \left(\frac{n+1}{K}, \frac{m}{K}\right) & \text{at rate } nb_1 \\ \left(\frac{n}{K}, \frac{m+1}{K}\right) & \text{at rate } mb_2 \\ \left(\frac{n-1}{K}, \frac{m}{K}\right) & \text{at rate } n(d_1 + \frac{(1-p)C}{K}n) \\ \left(\frac{n}{K}, \frac{m-1}{K}\right) & \text{at rate } md_2 \\ \left(\frac{n-1}{K}, \frac{m+1}{K}\right) & \text{at rate } \frac{pC}{K}n^2 \\ \left(\frac{n+1}{K}, \frac{m-1}{K}\right) & \text{at rate } m\sigma_2 \end{cases}$$

by \mathbb{P}_z^K . Then for any $0 < \delta < r$, we have

$$\lim_{K \rightarrow \infty} \sup_{z \in \tilde{C}} \mathbb{P}_z^K \left(\sup_{t \in [0, T]} \|w_t - \varphi_{z_1, z_2}(t)\| \geq \delta \right) = 0,$$

where w_t is the canonical process on $D([0, \infty), \mathbb{R}^2)$, the space of càdlàg paths from $[0, \infty)$ to \mathbb{R}^2 .

Remark B.4. Note that this result holds also for processes where $b_2 > 0$. The rescaled process Z^K mentioned in the beginning of this section is a special case with $b_2 = 0$.

Proof. We use the techniques from [Cha06, Theorem 3]. Let $T > 0$. To show the boundedness properties, we first show that φ_2 is strictly larger than 0 for any positive initial condition. Indeed, we easily see that

$$\dot{\varphi}_2 \geq (b_2 - d_2 - \sigma_2)\varphi_2$$

which we can integrate directly to obtain

$$\varphi_2(t) \geq \varphi_2(0) \exp \left(\int_0^t (b_2 - d_2 - \sigma_2) \, ds \right) = \varphi_2(0) \exp((b_2 - d_2 - \sigma_2)t),$$

which is positive for all times t as soon as $\varphi_2(0) > 0$. In particular, the solution $(\varphi_{z_1, z_2})_1$ is bounded from below by the solution $\tilde{\varphi}$ of the differential equation

$$\dot{\tilde{\varphi}} = (b_1 - d_1 - C\tilde{\varphi})\tilde{\varphi}.$$

The function $\tilde{\varphi}$ is bounded from above by some constant \tilde{R} for any positive initial condition, since $\tilde{\varphi} < 0$ as soon as $\tilde{\varphi} > (b_1 - d_1)/C$. Hence, by integrating the differential equation, we obtain for any positive initial condition $\tilde{\varphi}(0) > 0$ the inequality

$$\tilde{\varphi}(t) = \tilde{\varphi}(0) \exp \left(\int_0^t (b_1 - d_1 - C\tilde{\varphi}(s)) \, ds \right) \geq \tilde{\varphi}(0) \exp((b_1 - d_1 - C\tilde{R})t) > 0.$$

Thus far we have shown $(\varphi_{z_1, z_2})_1(t), (\varphi_{z_1, z_2})_2(t) > 0$ for any time $t \geq 0$ and any initial condition $(z_1, z_2) \in \tilde{C}$. In particular, due to the continuity of $(z_1, z_2) \mapsto \varphi_{z_1, z_2}$ we have

$$\inf_{z \in \tilde{C}} \inf_{t \in [0, T]} \|\varphi_{z_1, z_2}(t)\| > 0.$$

For the upper bound, adding the two equations in (B.4) yields

$$\dot{\varphi}_1 + \dot{\varphi}_2 = (b_1 - d_1)\varphi_1 + (b_2 - d_2)\varphi_2 - C(1 - p)(\varphi_1)^2 \leq 2 \max(b_1 - d_1, b_2 - d_2)(\varphi_1 + \varphi_2),$$

where we used $\varphi_1, \varphi_2 \geq 0$ for any initial condition $(z_1, z_2) \in \tilde{C}$. Thus, Gronwall's inequality implies

$$\begin{aligned} \sup_{z \in \tilde{C}} \sup_{t \in [0, T]} \|\varphi_{z_1, z_2}(t)\|_1 &= \sup_{z \in \tilde{C}} \sup_{t \in [0, T]} (\varphi_{z_1, z_2})_1(t) + (\varphi_{z_1, z_2})_2(t) \\ &\leq \sup_{z \in \tilde{C}} (z_1 + z_2) \exp \left(\int_0^T 2 \max(b_1 - d_1, b_2 - d_2) dt \right) < \infty. \end{aligned}$$

By equivalence of norms on \mathbb{R}^2 , for any given norm, these bounds on the infimum and supremum can be chosen to hold.

Now, we can define a family of Markov processes with transitions

$$\left(\frac{i}{K}, \frac{j}{K}\right) \rightarrow \begin{cases} \left(\frac{i+1}{K}, \frac{j}{K}\right), & \text{at rate } Kp_1\left(\frac{i}{K}, \frac{j}{K}\right) \\ \left(\frac{i}{K}, \frac{j+1}{K}\right), & \text{at rate } Kp_2\left(\frac{i}{K}, \frac{j}{K}\right) \\ \left(\frac{i-1}{K}, \frac{j}{K}\right), & \text{at rate } Kq_1\left(\frac{i}{K}, \frac{j}{K}\right) \\ \left(\frac{i}{K}, \frac{j-1}{K}\right), & \text{at rate } Kq_2\left(\frac{i}{K}, \frac{j}{K}\right) \\ \left(\frac{i-1}{K}, \frac{j+1}{K}\right), & \text{at rate } Kr_1\left(\frac{i}{K}, \frac{j}{K}\right) \\ \left(\frac{i+1}{K}, \frac{j-1}{K}\right), & \text{at rate } Kr_2\left(\frac{i}{K}, \frac{j}{K}\right), \end{cases}$$

where $p_1, p_2, q_1, q_2, r_1, r_2: \mathbb{R}^2 \rightarrow \mathbb{R}$ are positive, bounded and Lipschitz functions. We denote the law of such a process by \mathbb{Q}_z^K , when the initial condition is given by $z \in \frac{1}{K}\mathbb{N}_0^2$. This choice of transition rates corresponds (using the notation from [DE97] in equation (10.1)) to the choice of $\varepsilon = \frac{1}{K}$ and the measure ν_x being given by

$$\begin{aligned} \nu_{\left(\frac{i}{K}, \frac{j}{K}\right)}(\{(1, 0)\}) &= p_1\left(\frac{i}{K}\right), & \nu_{\left(\frac{i}{K}, \frac{j}{K}\right)}(\{(0, 1)\}) &= p_2\left(\frac{j}{K}\right), & \nu_{\left(\frac{i}{K}, \frac{j}{K}\right)}(\{(-1, 0)\}) &= q_1\left(\frac{i}{K}\right), \\ \nu_{\left(\frac{i}{K}, \frac{j}{K}\right)}(\{(0, -1)\}) &= q_2\left(\frac{j}{K}\right), & \nu_{\left(\frac{i}{K}, \frac{j}{K}\right)}(\{(-1, 1)\}) &= r_1\left(\frac{i}{K}\right), & \nu_{\left(\frac{i}{K}, \frac{j}{K}\right)}(\{(1, -1)\}) &= r_2\left(\frac{j}{K}\right). \end{aligned}$$

The extension of $\nu_{x,y}$ for any vector $(x, y) \in \mathbb{R}^2$ is straightforward by replacing either $\frac{i}{K}$ by x or $\frac{j}{K}$ by y respectively. In addition, we choose the functions $b_1 = p_1 - q_1 - r_1 + r_2$, $b_2 = p_2 - q_2 - r_2 + r_1$ and $a \equiv 0$.

Since the functions involved are all bounded and continuous, Condition 10.2.2 from [DE97] is satisfied. In order to apply Theorem 10.2.6 of [DE97], we do not need any additional conditions. However, we only obtain the upper bound of the Laplace principle as can be seen from the remark preceding the Theorem. The good rate function I_T appearing in the Laplace principle writes for functions $\varphi: [0, T] \rightarrow \mathbb{R}^2$ as

$$I_T(\varphi) = \begin{cases} \int_0^T L(\varphi(t), \dot{\varphi}(t)) dt & \text{if } \varphi \text{ is absolutely continuous} \\ \infty & \text{otherwise,} \end{cases}$$

where for $y, z \in \mathbb{R}^2$ we define the function $L(y, z) = \sup_{\alpha \in \mathbb{R}^2} (\langle \alpha, z \rangle - H(y, \alpha))$ with

$$H(y, \alpha) = \int_{\mathbb{R}^2} (\exp(\langle \alpha, x \rangle) - 1) \nu_y(dx).$$

Hence, calculating the gradient of the function in the supremum with respect to α shows that $L(y, z) = 0$ if and only if $z_1 = p_1(y) - q_1(y) - r_1(y) + r_2(y)$ and $z_2 = p_2(y) - q_2(y) - r_2(y) + r_1(y)$. Therefore,

our rate function I_T satisfies

$$I_T(\varphi) = 0 \iff \begin{pmatrix} \dot{\varphi}_1 \\ \dot{\varphi}_2 \end{pmatrix} = \begin{pmatrix} p_1(\varphi) - q_1(\varphi) - r_1(\varphi) + r_2(\varphi) \\ p_2(\varphi) - q_2(\varphi) - r_2(\varphi) + r_1(\varphi) \end{pmatrix}. \quad (\text{B.5})$$

Since the upper bound of the Laplace principle is by [DE97, Corollary 1.2.5] equivalent to the upper bound in the large deviation principle, we obtain

$$\limsup_{K \rightarrow \infty} \frac{1}{K} \log \left(\sup_{z \in \tilde{C}} \mathbb{Q}_z^K(F) \right) \leq - \inf_{\psi \in F, \psi(0) \in \tilde{C}} I_T(\psi)$$

for any compact set $\tilde{C} \subseteq \mathbb{R}^2$ and any closed set $F \subseteq D([0, T], \mathbb{R}^2)$, that is the set of càdlàg functions on $[0, T]$ into \mathbb{R}^2 .

We define the cut-off function χ as the orthogonal projection from \mathbb{R}^2 onto $[r - \delta, R + \delta]^2$ with respect to the Euclidean norm. Then, we can define our functions

$$\begin{aligned} p_1(z) &= b_1 \chi_1(z), & p_2(z) &= b_2 \chi_2(z), & q_1(z) &= d_1 \chi_1(z) + C(1 - p) \chi_1^2(z), \\ q_2(z) &= d_2 \chi_2(z), & r_1(z) &= pC \chi_1^2(z), & r_2(z) &= \sigma_2 \chi_2(z). \end{aligned}$$

Now, the laws \mathbb{P}_z^K and \mathbb{Q}_z^K coincide as long as w_t^1 and w_t^2 are still inside the interval $[r - \delta, R + \delta]$. Thus, we have

$$\begin{aligned} & \limsup_{K \rightarrow \infty} \frac{1}{K} \log \sup_{z \in \tilde{C}} \mathbb{P}_z^K \left(\sup_{t \in [0, T]} \|w_t - \varphi_{z_1, z_2}(t)\| \geq \delta \right) \\ &= \limsup_{K \rightarrow \infty} \frac{1}{K} \log \sup_{z \in \tilde{C}} \mathbb{Q}_z^K \left(\sup_{t \in [0, T]} \|w_t - \varphi_{z_1, z_2}(t)\| \geq \delta \right) \leq - \inf_{\psi \in F} I_T(\psi), \end{aligned}$$

where

$$F := \left\{ \psi \in D([0, T], \mathbb{R}^2) \mid \psi(0) \in \tilde{C} \text{ and } \exists t \in [0, T] : \|\psi(t) - \varphi_{\psi(0)}\| \geq \delta \right\}.$$

Note that due to the continuity of φ_z the set F is closed with respect to the supremum norm, which allows us to use the large deviation principle. Furthermore, it is easy to see that absolutely continuous functions $\psi \in F$ cannot satisfy (B.5), as otherwise the distance between φ and ψ cannot become large. Since I_T is a good rate function, the infimum is attained for some function $\psi \in F$ and thus by (B.5) is non-zero. \square

This lemma has put us in a good position to show that the considered logistic bi-type branching process converges in finite time into a neighbourhood of its equilibrium.

Lemma B.5. *Let Z^K be a $LBB I_K(b_1^K, d_1^K, d_2^K, \sigma_2, p, C, \gamma_1^K)$ and assume that the convergence (B.1) holds with $b_1 > d_1$. Then, for all $\varepsilon_1, \varepsilon_2, \varepsilon'_1, \varepsilon'_2 > 0$ there exists a finite time $T(\varepsilon_1, \varepsilon_2, \varepsilon'_1, \varepsilon'_2)$ such that for all initial starting conditions $(2\bar{x}, 2\bar{y}) \geq \frac{Z_0^K}{K} = \frac{(X_0^K, Y_0^K)}{K} \geq (\varepsilon_1, \varepsilon_2)$, we have*

$$\lim_{K \rightarrow \infty} \mathbb{P} \left(\left\| \frac{Z_{T(\varepsilon_1, \varepsilon_2, \varepsilon'_1, \varepsilon'_2)}^K}{K} - (\bar{x}, \bar{y}) \right\| \leq \varepsilon'_1 + \varepsilon'_2 \right) = 1,$$

where (\bar{x}, \bar{y}) is given in (B.3).

Proof. For the proof of this claim, we use for K large enough the following coordinatewise coupling

$$\tilde{Z}^K \leq Z^K \leq \hat{Z}^K,$$

where \tilde{Z}^K is a $LBB I_K(b_1 - \varepsilon, d_1 + \varepsilon, d_2 + \varepsilon, \sigma_2, p, C)$ and the distribution of \hat{Z}^K is given by $LBB I_K(b_1 + \varepsilon, (d_1 - \varepsilon)_+, (d_2 - \varepsilon)_+, \sigma_2, p, C)$, and the initial conditions are $\tilde{Z}_0^K = Z_0^K = \hat{Z}_0^K$ with $\varepsilon > 0$. We need

to justify the coupling, specifically why we can increase the birth rate and at this cost neglect the immigration from outside in the upper bounding process. For this purpose, we need to show that Z^K is bounded component-wise from below in probability by $K\varepsilon$ for some $\varepsilon > 0$. Indeed, this can already be seen from the lower bounding process: Firstly there exists $\delta > 0$ and a time $T > 0$ such that the corresponding solutions $\tilde{\varphi}$ and $\hat{\varphi}$ to the differential equation for the processes \tilde{Z}^K and \hat{Z}^K satisfy

$$\|\tilde{\varphi}(T) - (x, y)_{\tilde{\varphi}}\| < \delta \quad \text{and} \quad \|\hat{\varphi}(T) - (x, y)_{\hat{\varphi}}\| < \delta,$$

where $(x, y)_{\varphi}$ denotes the unique coordinatewise positive stable equilibrium of the differential equation corresponding to φ . Note that we can use the same times, as we have convergence towards the equilibria as $T \rightarrow \infty$ from any starting condition and hence we can choose T sufficiently large so that both systems are close to their equilibrium. For this, we again refer the reader to [BT20, Lemma 4.6]. Now, Lemma B.3 implies that with $\tilde{C} := [\varepsilon_1, 2\bar{x}] \times [\varepsilon_2, 2\bar{y}]$ and $\tilde{Z}_0^K = z \in \tilde{C}$, we have for any $\delta > 0$ small enough

$$\sup_{z \in \tilde{C}} \mathbb{P} \left(\sup_{t \in [0, T]} \left\| \frac{\tilde{Z}_t^K}{K} - \tilde{\varphi}(t) \right\| \geq \delta \right) \xrightarrow{K \rightarrow \infty} 0.$$

Note, that

$$\inf_{z \in \tilde{C}} \inf_{t \in [0, T]} \|\tilde{\varphi}(t)\| > 0 \quad \text{and} \quad \inf_{z \in \tilde{C}} \inf_{t \in [0, T]} \|\tilde{\varphi}_1(t)\| > 0,$$

where the second inequality is due to the positive switching between components which can be seen from the proof of Lemma B.3. In particular, for $\varepsilon > 0$ small enough it holds $\inf_{t \in [0, T]} \left| \frac{(\tilde{Z}_t^K)_1}{K} \right| > \varepsilon$ with probability converging to 1 as $K \rightarrow \infty$. Thus for K large enough, we have

$$\inf_{t \in [0, T]} \frac{\varepsilon}{2} \cdot (\tilde{Z}_t^K)_1 > \sup_{t \in [0, T]} \gamma_1^K(\omega, t)$$

with probability converging to 1. Hence, we may dismiss the immigration component in the coupling and replace it by an increase in the birth rate as done above. Now, choosing ε small enough, we can achieve for any given $\delta > 0$ that

$$\|(x, y)_{\tilde{\varphi}} - (x, y)_{\hat{\varphi}}\| < \delta, \quad \|(x, y)_{\tilde{\varphi}} - (\bar{x}, \bar{y})\| < \delta \quad \text{and} \quad \|(x, y)_{\hat{\varphi}} - (\bar{x}, \bar{y})\| < \delta,$$

which can be seen from computing the equilibria similarly to (2.3). Hence, by Lemma B.3 we have with high probability, that

$$\begin{aligned} \left\| \frac{\hat{Z}_T^K - \tilde{Z}_T^K}{K} \right\| &\leq \left\| \frac{\hat{Z}_T^K}{K} - \hat{\varphi}(T) \right\| + \|\hat{\varphi}(T) - \tilde{\varphi}(T)\| + \left\| \frac{\tilde{Z}_T^K}{K} - \tilde{\varphi}(T) \right\| \\ &\leq \|\hat{\varphi}(T) - \tilde{\varphi}(T)\| + \delta \leq 4\delta. \end{aligned}$$

Because of Lemma B.3 for any starting condition the inequality

$$\left\| \frac{\hat{Z}_T^K}{K} - (\bar{x}, \bar{y}) \right\| \leq \left\| \frac{\hat{Z}_T^K}{K} - \hat{\varphi}(T) \right\| + \|\hat{\varphi}(T) - (x, y)_{\hat{\varphi}}\| + \|(x, y)_{\hat{\varphi}} - (\bar{x}, \bar{y})\| \leq 3\delta$$

is satisfied with probability converging to 1 as $K \rightarrow \infty$, and thus we obtain from the component-wise coupling that

$$\left\| \frac{Z_T^K}{K} - (\bar{x}, \bar{y}) \right\| \leq \left\| \frac{Z_T^K - \hat{Z}_T^K}{K} \right\| + \left\| \frac{\hat{Z}_T^K}{K} - (\bar{x}, \bar{y}) \right\| \leq \left\| \frac{\hat{Z}_T^K - \tilde{Z}_T^K}{K} \right\| + 3\delta \leq 7\delta$$

with high probability. Letting $\delta > 0$ small enough yields the claim. \square

Thus far, we have been considering the behaviour of our process when it is initially not close to its equilibrium size. We now turn to the question, how long it takes for a logistic process to exit a neighbourhood of its equilibrium. For this we first consider a process with constant rates.

Lemma B.6. *In the situation of Lemma B.3 let $b_1 > d_1$. Then, the unique asymptotically stable equilibrium of the system*

$$\begin{aligned}\dot{\varphi}_1 &= (b_1 - d_1 - C\varphi_1)\varphi_1 + \sigma_2\varphi_2 \\ \dot{\varphi}_2 &= (-d_2 - \sigma_2)\varphi_2 + Cp(\varphi_1)^2\end{aligned}$$

is given by $(\bar{\varphi}_1, \bar{\varphi}_2) = (\bar{x}, \bar{y})$ in (B.3). Let $\eta_1, \eta_2 > 0$ and set

$$T^K := \inf \{t \geq 0 \mid w_t^1 \notin [\bar{x} - \eta_1, \bar{x} + \eta_1] \text{ or } w_t^2 \notin [\bar{y} - \eta_2, \bar{y} + \eta_2]\},$$

where $w_t = (w_t^1, w_t^2)$ is the canonical process on $D([0, \infty), \mathbb{R}^2)$. Then there exists a constant $V > 0$ such that for all compact subsets \tilde{C} of $[\bar{x} - \frac{\eta_1}{2}, \bar{x} + \frac{\eta_1}{2}] \times [\bar{y} - \frac{\eta_2}{2}, \bar{y} + \frac{\eta_2}{2}]$ we have

$$\lim_{K \rightarrow \infty} \sup_{z \in \tilde{C}} \mathbb{P}_z^K(T^K < e^{KV}) = 0.$$

Proof. As in the proof of Lemma B.3 and similarly to [Cha06] we define the cut-off function χ as the orthogonal projection onto $C_1 := [\bar{x} - \eta_1, \bar{x} + \eta_1] \times [\bar{y} - \eta_2, \bar{y} + \eta_2]$. Also, we can similarly construct the functions p_1, q_1, q_2, r_1, r_2 , by which we obtain a family of laws \mathbb{Q}_z^K which coincide with \mathbb{P}_z^K on the time interval $[0, T^K]$. Furthermore, we obtain in a similar manner the good rate function $I_T(\varphi)$, where now

$$I_T(\varphi) = 0 \iff \begin{pmatrix} \dot{\varphi}_1 \\ \dot{\varphi}_2 \end{pmatrix} = \begin{pmatrix} p_1(\varphi) - q_1(\varphi) - r_1(\varphi) + r_2(\varphi) \\ -q_2(\varphi) - r_2(\varphi) + r_1(\varphi) \end{pmatrix}. \quad (\text{B.6})$$

Now, [FW98, p.157] and Theorem 4.4.2 therein imply, that there exists a constant $\bar{V} \geq 0$, such that for any $\delta > 0$ it holds

$$\lim_{K \rightarrow \infty} \inf_{z \in \tilde{C}} \mathbb{Q}_z^K(e^{K(\bar{V}-\delta)} < T^K < e^{K(\bar{V}+\delta)}) = 1.$$

Hence, proving that $\bar{V} > 0$ shows our claim. For this purpose, we note that \bar{V} is defined in [FW98, Theorem 4.4.1] by

$$\bar{V} := \inf_{(x,y) \in \partial C_1} V((\bar{x}, \bar{y}), (x, y)) = \min_{(x,y) \in \partial C_1} V((\bar{x}, \bar{y}), (x, y)),$$

where

$$V((x_1, x_2), (y_1, y_2)) := \inf_{t > 0, \varphi(0) = (x_1, x_2), \varphi(t) = (y_1, y_2)} I_t(\varphi),$$

that is $V((x_1, x_2), (y_1, y_2))$ is the infimum of our rate function over all possible paths connecting (x_1, x_2) and (y_1, y_2) . Since the function $(x, y) \mapsto V((\bar{x}, \bar{y}), (x, y))$ is continuous and the set ∂C_1 is compact, we can indeed replace the infimum by a minimum. In particular, there exists some $(x_0, y_0) \in \partial C_1$ where the infimum is attained. Theorem 5.4.3 from [FW98] shows that there exists an absolutely continuous function φ , which attains the infimum for $V((\bar{x}, \bar{y}), (x_0, y_0))$ over the rate function in the sense that for some $T > 0$ the function $\varphi: [0, T] \rightarrow \mathbb{R}^2$ satisfies the conditions $\varphi(0) = (\bar{x}, \bar{y})$, $\varphi(T) = (x_0, y_0)$ and $V((\bar{x}, \bar{y}), (x_0, y_0)) = I_T(\varphi)$ or that for some $T > -\infty$ the function $\varphi: (-\infty, T] \rightarrow \mathbb{R}^2$ satisfies $\lim_{t \rightarrow -\infty} \varphi(t) = (\bar{x}, \bar{y})$, $\varphi(T) = (x_0, y_0)$ and $V((\bar{x}, \bar{y}), (x_0, y_0)) = \int_{-\infty}^T L(\varphi(t), \dot{\varphi}(t)) dt$.

As long as φ satisfies the differential equation in (B.6) with initial condition $\varphi(0) = (\bar{x}, \bar{y})$, we have $\varphi(t) = (\bar{x}, \bar{y}) \neq (x_0, y_0)$. Thus, for the case where φ is defined on $[0, T]$, we have $I_T(\varphi) \neq 0$, as the differential equation in (B.6) must be violated at some time in order to leave the equilibrium state. For the second case, we already know that any solution to (B.6) started close to (\bar{x}, \bar{y}) stays close, since the equilibrium is asymptotically stable. Hence, choosing $0 < \varepsilon < \eta_1 \wedge \eta_2$ small enough and $T_1 := \sup\{t \leq T \mid \|\varphi(t) - (\bar{x}, \bar{y})\| < \frac{\varepsilon}{2}\}$, for all $t \geq T_1$ we have $\|\varphi(t) - (\bar{x}, \bar{y})\| < \varepsilon$. Therefore in this case connecting (\bar{x}, \bar{y}) to (x_0, y_0) requires violating (B.6). Hence we have $\bar{V} > 0$. \square

Using our usual coupling arguments to make the leap from fixed rates to some non-fixed rates, we can formulate and prove a very similar result for logistic processes with non-constant rates.

Lemma B.7. Let Z^K be a $LBB I_K(b_1^K, d_1^K, d_2^K, \sigma_2, p, C, \gamma_1^K)$ and assume that the convergence (B.1) holds with $b_1 > d_1$. Let $\eta_1, \eta_2 > 0$, $Z_0^K \in [\bar{x} - \frac{\eta_1}{2}, \bar{x} + \frac{\eta_1}{2}] \times [\bar{y} - \frac{\eta_2}{2}, \bar{y} + \frac{\eta_2}{2}]$ and set

$$T^K := \inf \left\{ t \geq 0 \mid \frac{(Z_t^K)_1}{K} \notin [\bar{x} - \eta_1, \bar{x} + \eta_1] \text{ or } \frac{(Z_t^K)_2}{K} \notin [\bar{y} - \eta_2, \bar{y} + \eta_2] \right\}.$$

Then there exists a constant $V > 0$ such that

$$\lim_{K \rightarrow \infty} \mathbb{P}(T^K < e^{KV}) = 0.$$

Proof. As in the proof of Lemma B.5, we couple our process with

$$\tilde{Z}^K \leq Z^K \leq \hat{Z}^K,$$

where \tilde{Z}^K is given as a $LBB I_K(b_1 - \varepsilon, d_1 + \varepsilon, d_2 + \varepsilon, \sigma_2, p, C)$ and the law of \hat{Z} is $LBB I_K(b_1 + \varepsilon, (d_1 - \varepsilon)_+, (d_2 - \varepsilon)_+, \sigma_2, p, C)$, where the initial conditions are $\tilde{Z}_0^K = Z_0^K = \hat{Z}_0^K$ and $\varepsilon > 0$. Choosing ε small enough and applying Lemma B.6 to \tilde{Z}^K and \hat{Z}^K yields the claim. \square

A very simple consequence is now the result that our process remains at least a time of order $\log K$ in a neighbourhood of its equilibrium.

Corollary B.8. Under the assumptions of Lemma B.7 it holds

$$\lim_{K \rightarrow \infty} \mathbb{P} \left(\forall t \in [0, T \log K], \frac{(Z_t^K)_1}{K} \in [\bar{x} - \eta_1, \bar{x} + \eta_1] \text{ and } \frac{(Z_t^K)_2}{K} \in [\bar{y} - \eta_2, \bar{y} + \eta_2] \right) = 1.$$

for any $T > 0$.

B.2. Competition Between two Bi-Type Processes with Transfer. Now, we consider a four-dimensional logistic branching process, which we will interpret as competition between two bi-type logistic branching processes each with active ('a') individuals and dormant ('d') individuals. Our main results in this section will be Propositions B.15 and B.17, where we show under suitable assumptions that the initially resident process declines below a small threshold, while the invading process reaches a neighbourhood of its equilibrium. The transfer rates of this process $(X_a^K, X_d^K, Y_a^K, Y_d^K)$ are

$$(i, j, k, \ell) \rightarrow \begin{cases} (i+1, j, k, \ell) & \text{at rate } ia_1^K(\omega, t) + \gamma_1^K(\omega, t) \\ (i, j, k+1, \ell) & \text{at rate } kb_1^K(\omega, t) + \gamma_2^K(\omega, t) \\ (i-1, j, k, \ell) & \text{at rate } i(d_1^K(\omega, t) + \frac{(1-p)C}{K}(i+k)) \\ (i, j, k-1, \ell) & \text{at rate } k(d_1^K(\omega, t) + \frac{(1-q)C}{K}(i+k)) \\ (i, j-1, k, \ell) & \text{at rate } jd_2^K(\omega, t) \\ (i, j, k, \ell-1) & \text{at rate } \ell d_2^K(\omega, t) \\ (i-1, j+1, k, \ell) & \text{at rate } i \frac{pC}{K}(i+k) \\ (i, j, k-1, \ell+1) & \text{at rate } k \frac{qC}{K}(i+k) \\ (i+1, j-1, k, \ell) & \text{at rate } j\sigma_2 \\ (i, j, k+1, \ell-1) & \text{at rate } \ell\sigma_2 \\ (i-1, j, k+1, \ell) & \text{at rate } \tau^K(\omega, t) \frac{ik}{i+k}, \end{cases}$$

with predictable, non-negative functions $a_1^K, b_1^K, d_1^K, d_2^K, \tau^K, \gamma_1^K, \gamma_2^K : \Omega \times [0, \infty) \rightarrow \mathbb{R}$ and constants $C, \sigma_2 > 0$, $p, q \in (0, 1)$. For now, we assume that the transition rates are constant with $a_1^K \equiv a_1$, $b_1^K \equiv b_1$, $d_1^K \equiv d_1$, $d_2^K \equiv d_2$, $\tau^K \equiv \tau$ and $\gamma_1^K = \gamma_2^K = 0$ and that $a_1, b_1 > d_1$. Then the process

$\frac{1}{K}(X_a^K, X_d^K, Y_a^K, Y_d^K)$ converges towards the unique solution of the differential equation

$$\begin{aligned}\dot{x}_a &= x_a(a_1 - d_1 - C(x_a + y_a)) + x_d\sigma_2 - \tau \frac{x_a y_a}{x_a + y_a} \\ \dot{x}_d &= pC x_a(x_a + y_a) - (d_2 + \sigma_2)x_d \\ \dot{y}_a &= y_a(b_1 - d_1 - C(x_a + y_a)) + y_d\sigma_2 + \tau \frac{x_a y_a}{x_a + y_a} \\ \dot{y}_d &= qC y_a(x_a + y_a) - (d_2 + \sigma_2)y_d.\end{aligned}\tag{B.7}$$

We are now interested in finding a suitable criterion for invasion of the process (Y_a^K, Y_d^K) into the initially resident population (X_a^K, X_d^K) . More specifically, we assume that initially the size of $(\frac{X_a^K}{K}, \frac{X_d^K}{K})$ is close to its equilibrium (\bar{x}_a, \bar{x}_d) with

$$\bar{x}_a = \frac{(a_1 - d_1)(d_2 + \sigma_2)}{C(d_2 + (1 - p)\sigma_2)} \quad \text{and} \quad \bar{x}_d = \frac{p(a_1 - d_1)^2(d_2 + \sigma_2)}{C(d_2 + (1 - p)\sigma_2)^2}\tag{B.8}$$

and the total size of the invasive species is $Y_a^K + Y_d^K = \lfloor \varepsilon K \rfloor$. Note that for such small population sizes we may approximate the transfer rate between X_a^K and Y_a^K by

$$\frac{X_a^K Y_a^K}{X_a^K + Y_a^K} \approx \frac{X_a^K Y_a^K}{X_a^K} = Y_a^K.$$

Further, we assume that the mean matrix

$$J = \begin{pmatrix} b_1 + \tau - d_1 - C\bar{x}_a & qC\bar{x}_a \\ \sigma_2 & -d_2 - \sigma_2 \end{pmatrix}\tag{B.9}$$

of the approximating process (\hat{Y}_a, \hat{Y}_d) of (Y_a^K, Y_d^K) given by the transitions

$$(n, m) \mapsto \begin{cases} (n+1, m) & \text{at rate } n(b_1 + \tau) \\ (n-1, m) & \text{at rate } n(d_1 + (1-q)C\bar{x}_a) \\ (n, m-1) & \text{at rate } md_2 \\ (n-1, m+1) & \text{at rate } nqC\bar{x}_a \\ (n+1, m-1) & \text{at rate } m\sigma_2 \end{cases}$$

has a positive eigenvalue λ , which means that the invasion fitness is positive and the process is supercritical. For one eigenvalue to be positive, the determinant must be negative, which in our case is equivalent to the inequality

$$-(b_1 + \tau - d_1 - C\bar{x}_a) < \frac{\sigma_2 q C \bar{x}_a}{d_2 + \sigma_2}.\tag{B.10}$$

In addition, we will assume that in a population, where (Y_a^K, Y_d^K) is resident and (X_a^K, X_d^K) is invasive, the approximating process (\hat{X}_a, \hat{X}_d) given by the transitions

$$(n, m) \mapsto \begin{cases} (n+1, m) & \text{at rate } na_1 \\ (n-1, m) & \text{at rate } n(d_1 + \tau + (1-p)C\bar{y}_a) \\ (n, m-1) & \text{at rate } md_2 \\ (n-1, m+1) & \text{at rate } npC\bar{y}_a \\ (n+1, m-1) & \text{at rate } m\sigma_2 \end{cases}$$

is sub-critical. This is the case if and only if both eigenvalues of the mean-matrix

$$\tilde{J} = \begin{pmatrix} a_1 - \tau - d_1 - C\bar{y}_a & pC\bar{y}_a \\ \sigma_2 & -d_2 - \sigma_2 \end{pmatrix}\tag{B.11}$$

are strictly negative. In particular, we must have a positive determinant, which is equivalent to

$$-(a_1 - \tau - d_1 - C\bar{y}_a) > \frac{\sigma_2 p C \bar{y}_a}{d_2 + \sigma_2}. \quad (\text{B.12})$$

Our first result is concerned with finding the equilibria of the dynamical system (B.7).

Lemma B.9. *Consider the system (B.7) and assume the matrix J has a positive eigenvalue and the matrix \tilde{J} only has negative eigenvalues. Then the systems only non-negative equilibria are $(0, 0, 0, 0)$, $(\bar{x}_a, \bar{x}_d, 0, 0)$ and $(0, 0, \bar{y}_a, \bar{y}_d)$, the latter of which is asymptotically stable, where*

$$\bar{y}_a = \frac{(b_1 - d_1)(d_2 + \sigma_2)}{C(d_2 + (1 - q)\sigma_2)} \quad \text{and} \quad \bar{y}_d = \frac{q(b_1 - d_1)^2(d_2 + \sigma_2)}{C(d_2 + (1 - q)\sigma_2)^2}$$

and (\bar{x}_a, \bar{x}_d) are as in (B.8).

Proof. It is easy to verify that the claimed vectors are indeed equilibria of the system. Also a quick calculation for all cases shows that any non-negative equilibrium with a zero component must coincide with one of the vectors above. Hence, it remains to show that there is no coordinatewise strictly positive equilibrium in this system. Towards a contradiction, let (x_a, x_d, y_a, y_d) be such an equilibrium. Then, rearranging the second line in (B.7) yields

$$x_d = \frac{pC x_a(x_a + y_a)}{d_2 + \sigma_2}.$$

Hence, the first line in (B.7) gives

$$0 = x_a(a_1 - d_1 - C(x_a + y_a)) + \frac{pC x_a(x_a + y_a)\sigma_2}{d_2 + \sigma_2} - \tau \frac{x_a y_a}{x_a + y_a}.$$

Since we assumed $x_a > 0$, we may divide by x_a to give

$$\begin{aligned} 0 &= a_1 - d_1 - C(x_a + y_a) + \frac{pC(x_a + y_a)\sigma_2}{d_2 + \sigma_2} - \tau \frac{y_a}{x_a + y_a} \\ \iff x_a + y_a &= \bar{x}_a - \tau \frac{y_a(d_2 + \sigma_2)}{C(x_a + y_a)(d_2 + (1 - p)\sigma_2)}. \end{aligned} \quad (\text{B.13})$$

Similarly, we obtain

$$y_d = \frac{qC y_a(x_a + y_a)}{d_2 + \sigma_2} \quad \text{and} \quad 0 = y_a(b_1 - d_1 - C(x_a + y_a)) + \sigma_2 y_d + \tau \frac{x_a y_a}{x_a + y_a}$$

as well as

$$x_a + y_a = \bar{y}_a + \tau \frac{x_a(d_2 + \sigma_2)}{C(x_a + y_a)(d_2 + (1 - q)\sigma_2)}. \quad (\text{B.14})$$

Furthermore, this shows

$$\frac{y_d}{y_a} = -\frac{b_1 - d_1 - C(x_a + y_a) + \tau \frac{x_a}{x_a + y_a}}{\sigma_2} = \frac{qC(x_a + y_a)}{d_2 + \sigma_2}$$

and substituting the expression (B.13) for $x_a + y_a$ shows that

$$-\frac{b_1 - d_1 - C\bar{x}_a}{\sigma_2} - \tau \left(\frac{y_a(d_2 + \sigma_2) + x_a(d_2 + (1 - p)\sigma_2)}{\sigma_2(x_a + y_a)(d_2 + (1 - p)\sigma_2)} \right) = \frac{qC\bar{x}_a}{d_2 + \sigma_2} - \tau \frac{qy_a}{(x_a + y_a)(d_2 + (1 - p)\sigma_2)},$$

which can be rearranged to read

$$\begin{aligned} &-(b_1 + \tau - d_1 - C\bar{x}_a) - \tau \left(\frac{y_a(d_2 + \sigma_2) + x_a(d_2 + (1 - p)\sigma_2)}{(x_a + y_a)(d_2 + (1 - p)\sigma_2)} - 1 \right) \\ &= \frac{\sigma_2 q C \bar{x}_a}{d_2 + \sigma_2} - \tau \frac{\sigma_2 q y_a}{(x_a + y_a)(d_2 + (1 - p)\sigma_2)}. \end{aligned}$$

If $\tau = 0$, this violates condition (B.10), and thus there is no such equilibrium. Assuming $\tau > 0$ and (B.10) the previous equality implies the inequality

$$\begin{aligned} \frac{y_a(d_2 + \sigma_2) + x_a(d_2 + (1-p)\sigma_2)}{(x_a + y_a)(d_2 + (1-p)\sigma_2)} - 1 &< \frac{\sigma_2 q y_a}{(x_a + y_a)(d_2 + (1-p)\sigma_2)} \\ \iff \sigma_2 p y_a &< \sigma_2 q y_a. \end{aligned}$$

Hence, in the case $p \geq q$ such a coordinatewise positive equilibrium cannot exist.

Now, consider the case $q > p$. We find in a similar manner

$$\frac{x_d}{x_a} = -\frac{a_1 - d_1 - C(x_a + y_a) - \tau \frac{y_a}{x_a + y_a}}{\sigma_2} = \frac{pC(x_a + y_a)}{d_2 + \sigma_2}$$

and, substituting $x_a + y_a$ with the right hand side of (B.14), we obtain

$$-(a_1 - d_1 - C\bar{y}_a) + \tau \left(\frac{x_a(d_2 + \sigma_2) + y_a(d_2 + (1-q)\sigma_2)}{(x_a + y_a)(d_2 + (1-q)\sigma_2)} \right) = \frac{\sigma_2 p C \bar{y}_a}{d_2 + \sigma_2} + \tau \frac{\sigma_2 p x_a}{(x_a + y_a)(d_2 + (1-q)\sigma_2)}.$$

In particular, condition (B.12) implies that

$$\begin{aligned} \frac{x_a(d_2 + \sigma_2) + y_a(d_2 + (1-q)\sigma_2)}{(x_a + y_a)(d_2 + (1-q)\sigma_2)} - 1 &< \frac{\sigma_2 p x_a}{(x_a + y_a)(d_2 + (1-q)\sigma_2)} \\ \iff \sigma_2 q x_a &< \sigma_2 p x_a. \end{aligned}$$

This contradicts the assumption $q > p$. Hence, there cannot be a coordinatewise positive equilibrium.

Now we turn towards the stability claim. For this we consider the Jacobian at $(0, 0, \bar{y}_a, \bar{y}_d)$, which is given by

$$A = \begin{pmatrix} a_1 - d_1 - C\bar{y}_a - \tau & \sigma_2 & 0 & 0 \\ pC\bar{y}_a & -d_2 - \sigma_2 & 0 & 0 \\ -C\bar{y}_a + \tau & 0 & b_1 - d_1 - 2C\bar{y}_a & \sigma_2 \\ qC\bar{y}_a & 0 & 2qC\bar{y}_a & -d_2 - \sigma_2 \end{pmatrix}.$$

It remains to show that all eigenvalues have negative real part. Notice that we recover the matrix \tilde{J}^T in the upper left corner. Since the matrix A is a block matrix, the eigenvalues of \tilde{J}^T are also eigenvalues of A . In particular, it suffices to show that

$$B = \begin{pmatrix} b_1 - d_1 - 2C\bar{y}_a & \sigma_2 \\ 2qC\bar{y}_a & -d_2 - \sigma_2 \end{pmatrix}$$

has only negative eigenvalues. By the same argument as for the matrix J in (B.10) it suffices to show that the determinant is positive. Using the definition of \bar{y}_a we compute the determinant to be $\det(B) = (d_2 + \sigma_2)(b_1 - d_1)$ (cf. [BT20, Section 2.2]), which is strictly positive by the assumption $b_1 > d_1$. Hence all eigenvalues of A are negative and thus $(0, 0, \bar{y}_a, \bar{y}_d)$ is an asymptotically stable equilibrium. \square

The next goal is to find a sufficient criterion for the convergence of our dynamical system towards the equilibrium of the process (Y_a^K, Y_d^K) . For this purpose, we need a suitable initial condition for the system (B.7). Indeed, the following lemmata give a first step towards this direction.

Lemma B.10. *Consider the system (B.7) and assume that the matrix J from (B.9) has a positive eigenvalue and the matrix \tilde{J} from (B.11) only has negative eigenvalues. If the initial condition $(x_a, x_d, y_a, y_d) = (x_a(0), x_d(0), y_a(0), y_d(0))$ satisfies*

$$\frac{qC(x_a + y_a)}{d_2 + \sigma_2} > \frac{y_d}{y_a} > \frac{d_1 - b_1 + C(x_a + y_a) - \tau \frac{x_a}{x_a + y_a}}{\sigma_2}, \quad (\text{B.15})$$

then

$$\lim_{t \rightarrow \infty} (x_a(t), x_d(t), y_a(t), y_d(t)) = (0, 0, \bar{y}_a, \bar{y}_d).$$

Proof. The proof of this claim can be easily adapted from [BT20, Lemma 4.7]. \square

Now, we are interested in finding a suitable condition such that the inequality (B.15) is satisfied. For this purpose, observe that the approximating process (\hat{Y}_a, \hat{Y}_d) is supercritical and hence by the Kesten-Stigum Theorem [GB03, Theorem 2.1] we have

$$\left(\frac{\hat{Y}_{a,t}}{\hat{Y}_{a,t} + \hat{Y}_{d,t}}, \frac{\hat{Y}_{d,t}}{\hat{Y}_{a,t} + \hat{Y}_{d,t}} \right) \xrightarrow{t \rightarrow \infty} (\pi_a, \pi_d),$$

where (π_a, π_d) is the unique left eigenvector of the matrix J from (B.9) for the principal eigenvalue λ with $\pi_a + \pi_d = 1$.

Lemma B.11. *Suppose that the initial condition (x_a, x_d, y_a, y_d) of the dynamical system (B.7) satisfies $x_a \in [\bar{x}_a - A\sqrt{\varepsilon}, \bar{x}_a + A\sqrt{\varepsilon}]$, $x_d \in [\bar{x}_d - A\sqrt{\varepsilon}, \bar{x}_d + A\sqrt{\varepsilon}]$ for some constant $A > 0$ and $y_a + y_d \in (0, \sqrt{\varepsilon})$ with $\frac{y_d}{y_a} = \frac{\pi_d}{\pi_a}$. Then for ε sufficiently small, (x_a, x_d, y_a, y_d) satisfies (B.15).*

Proof. The proof follows the proof of [BT20, Lemma 4.8]. Since (π_a, π_d) is an eigenvector for J , we easily see that

$$b_1 + \tau - d_1 - C\bar{x}_a + \sigma_2 \frac{\pi_d}{\pi_a} = \lambda = qC\bar{x}_a \frac{\pi_a}{\pi_d} - (d_2 + \sigma_2).$$

Since $\lambda > 0$, we obtain for $\varepsilon > 0$ small enough from the first of the two equalities that

$$\begin{aligned} \frac{\pi_d}{\pi_a} &= \frac{\lambda - b_1 - \tau + d_1 + C\bar{x}_a}{\sigma_2} \\ &> \frac{-b_1 + d_1 + C(\bar{x}_a + 2(A+1)\sqrt{\varepsilon}) - \tau \frac{\bar{x}_a - A\sqrt{\varepsilon}}{\bar{x}_a + (A+1)\sqrt{\varepsilon}}}{\sigma_2} \\ &> \frac{d_1 - b_1 + C(x_a + y_a) - \tau \frac{x_a}{x_a + y_a}}{\sigma_2} \end{aligned}$$

and similarly from the second equality we see

$$\frac{\pi_d}{\pi_a} = \frac{qC\bar{x}_a}{\lambda + d_2 + \sigma_2} < \frac{qC(\bar{x}_a - 2(A+1)\sqrt{\varepsilon})}{d_2 + \sigma_2} < \frac{qC(x_a + y_a)}{d_2 + \sigma_2}.$$

\square

The next series of lemmata shows that this initial condition is satisfied with high probability as $K \rightarrow \infty$ and $\varepsilon \rightarrow 0$. In order to show this, we need a couple of preparatory results on the hitting times of a certain population size.

Lemma B.12. *Assume that the matrix J from (B.9) has a positive eigenvalue. Let $K \mapsto (m_1^K, m_2^K)$ be a function from $(0, \infty)$ to $[0, \infty)$ such that $(m_1^K, m_2^K) \in \frac{1}{K}\mathbb{N}_0^2$ and $\lim_{K \rightarrow \infty} (m_1^K, m_2^K) = (\bar{x}_a, \bar{x}_d)$. Define the stopping times*

$$R_\varepsilon^K := \inf \left\{ t \geq 0 \mid \left| \frac{X_{a,t}^K}{K} - \bar{x}_a \right| > \varepsilon \text{ or } \left| \frac{X_{d,t}^K}{K} - \bar{x}_d \right| > \varepsilon \right\}$$

and

$$T_x^K := \inf \{ t > 0 \mid Y_{a,t}^K + Y_{d,t}^K = \lfloor xK \rfloor \}.$$

Then for any $\xi \in [\frac{1}{2}, 1]$ there exist some positive constants $A, \varepsilon_0 > 0$ such that for all $0 < \varepsilon \leq \varepsilon_0$ and any starting condition $\frac{1}{K}(X_{a,0}^K, X_{d,0}^K, Y_{a,0}^K, Y_{d,0}^K) = (m_1^K, m_2^K, \frac{n_1^K}{K}, \frac{n_2^K}{K})$ with n_1^K and n_2^K such that

$0 < n_1^K + n_2^K < K\varepsilon^\xi$, the convergence

$$\lim_{K \rightarrow \infty} \mathbb{P}(R_{2A\varepsilon^\xi}^K \leq T_{\varepsilon^\xi}^K \wedge T_0^K) = 0$$

holds.

Proof. The proof of [BT20, Lemma 4.2] can be modified to encompass our situation in a straightforward manner. \square

Our next goal is to show that for any given initial population size $Y_{a,0}^K + Y_{d,0}^K = \lfloor \varepsilon K \rfloor$ of (Y_a^K, Y_d^K) , we have the limit

$$\lim_{K \rightarrow \infty} \mathbb{P}(T_{\varepsilon^\xi}^K < T_0^K) = 1.$$

Lemma B.13. *Under the assumptions of Lemma B.12, for the initial population size given by $\mathbf{X}_0^K := (X_{a,0}^K, X_{d,0}^K, Y_{a,0}^K, Y_{d,0}^K)$, it holds that*

$$\lim_{K \rightarrow \infty} \mathbb{P} \left(T_{\varepsilon^\xi}^K < T_0^K \wedge R_{2A\varepsilon^\xi}^K \mid \mathbf{X}_0^K = K(m_1^K, m_2^K, \varepsilon_1^K, \varepsilon_2^K) \right) = 1,$$

where $\varepsilon_1^K \rightarrow \varepsilon_1$ and $\varepsilon_2^K \rightarrow \varepsilon_2$ such that $\varepsilon_1 + \varepsilon_2 = \varepsilon$ with $\varepsilon, \varepsilon_1, \varepsilon_2 > 0$.

Proof. The proof follows the ideas of [BT20, Proposition 4.1] and [CCL⁺21, Proposition 3.1]. We will consider the process on the event that the invading population is extinct or reaches a sufficient population size before the resident population exits a neighbourhood of its equilibrium, that is

$$E_\varepsilon := \{T_0^K \wedge T_{\varepsilon^\xi}^K < R_{2A\varepsilon^\xi}^K\},$$

whose probability converges to 1 for $\varepsilon > 0$ small enough as $K \rightarrow \infty$ by Lemma B.12. Thus, it suffices to show $T_{\varepsilon^\xi}^K < T_0^K$ with probability converging to 1 on the event E_ε . On this event, we can couple the process (Y_a^K, Y_d^K) up to the time $t_\varepsilon := T_0^K \wedge T_{\varepsilon^\xi}^K \wedge R_{2A\varepsilon^\xi}^K$ such that

$$\begin{aligned} Y_v^{\varepsilon,-} &\leq \widehat{Y}_v \\ Y_v^{\varepsilon,-} &\leq Y_v^K \end{aligned}$$

where $v \in \{a, d\}$ by defining the transition rates as follows: For the process $(Y_a^{\varepsilon,-}, Y_d^{\varepsilon,-})$ we choose the transition rates

$$(n, m) \mapsto \begin{cases} (n+1, m) & \text{at rate } n \left(b_1 + \tau \frac{\bar{x}_a - 2A\varepsilon^\xi}{\bar{x}_a + (2A+1)\varepsilon^\xi} \right) \\ (n-1, m) & \text{at rate } n(d_1 + (1-q)C(\bar{x}_a + 2A\varepsilon^\xi + \varepsilon^\xi) + qC(4A\varepsilon^\xi + \varepsilon^\xi)) \\ (n, m-1) & \text{at rate } md_2 \\ (n-1, m+1) & \text{at rate } nqC(\bar{x}_a - 2A\varepsilon^\xi) \\ (n+1, m-1) & \text{at rate } m\sigma_2. \end{cases}$$

Intuitively, the coupling is correct due to switching from active to dormant being more favourable over death, but not being better than not experiencing any competition at all. The precise reasoning for this coupling is identical to the one in the proof of [BT20, Proposition 4.1]. Also, we obtain the inequality $q_a \leq q_a^{\varepsilon,-} < 1$ for the extinction probabilities of each individual family started at an active individual in a similar manner. When starting the populations from one dormant individual, we obtain the extinction probabilities $q_d \leq q_d^{\varepsilon,-} < 1$. Note that $q_\diamond^{\varepsilon,-}$ is indeed strictly less than 1 for $\diamond \in \{a, d\}$, since the process $(Y_a^{\varepsilon,-}, Y_d^{\varepsilon,-})$ is supercritical for $\varepsilon > 0$ sufficiently small. Since the extinction probabilities are continuous in the transition rates – see [CCL⁺21, Lemma A.3] – we obtain

$$0 \leq \liminf_{\varepsilon \downarrow 0} |q^{\varepsilon,-} - q| \leq \limsup_{\varepsilon \downarrow 0} |q^{\varepsilon,-} - q| = 0.$$

Next, we define the corresponding hitting time for the coupled process. More specifically, for $x \geq 0$, let

$$T_x^{\varepsilon,-} := \inf\{t \geq 0 \mid Y_{a,t}^{\varepsilon,-} + Y_{d,t}^{\varepsilon,-} = \lfloor Kx \rfloor\}.$$

Then, due to the coupling on E_ε , we see that

$$\mathbb{P}(T_{\varepsilon\xi}^{\varepsilon,-} \leq T_0^{\varepsilon,-}, E_\varepsilon) \leq \mathbb{P}(T_{\varepsilon\xi}^K \leq T_0^K, E_\varepsilon).$$

Note that since $\mathbb{P}(E_\varepsilon) \rightarrow 1$ as $K \rightarrow \infty$, for any $\varepsilon > 0$ sufficiently small, we have for δ small enough

$$\begin{aligned} \liminf_{K \rightarrow \infty} \mathbb{P}(T_{\varepsilon\xi}^{\varepsilon,-} \leq T_0^{\varepsilon,-}, E_\varepsilon) &\geq \liminf_{K \rightarrow \infty} \mathbb{P}(T_{\varepsilon\xi}^{\varepsilon,-} \leq T_0^{\varepsilon,-}) - \delta \geq \liminf_{K \rightarrow \infty} \mathbb{P}(T_0^{\varepsilon,-} = \infty) - \delta \\ &\geq \liminf_{K \rightarrow \infty} \left(1 - (q^{\varepsilon,-})^{K(\varepsilon_1^K + \varepsilon_2^K)}\right) - \delta \\ &\geq \liminf_{K \rightarrow \infty} \left(1 - (q^{\varepsilon,-})^{K(\varepsilon - \delta)}\right) - \delta \\ &= 1 - \delta \xrightarrow{\delta \rightarrow 0} 1, \end{aligned}$$

where $q^{\varepsilon,-} := q_a^{\varepsilon,-} \vee q_d^{\varepsilon,-}$. Also due to $\mathbb{P}(E_\varepsilon) \rightarrow 1$ as $K \rightarrow \infty$, it therefore follows that

$$\lim_{K \rightarrow \infty} \mathbb{P}(T_{\varepsilon\xi}^K \leq T_0^K, E_\varepsilon) = \lim_{K \rightarrow \infty} \mathbb{P}(T_{\varepsilon\xi}^K \leq T_0^K) = 1.$$

□

Finally, we are able to show that with high probability as $\varepsilon \downarrow 0$, the assumptions of Lemma B.11 are satisfied.

Lemma B.14. *Assume that the matrix J from (B.9) has a positive eigenvalue and that the initial condition of the process satisfies $\frac{1}{K}(X_{a,0}^K, X_{d,0}^K, Y_{a,0}^K, Y_{d,0}^K) = (m_1^K, m_2^K, \varepsilon_1, \varepsilon_2)$ with $\varepsilon_1 + \varepsilon_2 \leq \varepsilon$ for some $\varepsilon, \varepsilon_1, \varepsilon_2 > 0$ and m_1^K, m_2^K as in Lemma B.12. Then for any $\delta > 0$ with $\pi_a \pm \delta \in (0, 1)$, it holds that*

$$\liminf_{K \rightarrow \infty} \mathbb{P}\left(\exists t \in [T_\varepsilon^K, T_{\sqrt{\varepsilon}}^K]: \pi_a - \delta < \frac{Y_{a,t}^K}{Y_{a,t}^K + Y_{d,t}^K} < \pi_a + \delta\right) \geq 1 - o_\varepsilon(1),$$

where $o_\varepsilon(1) \rightarrow 0$ as $\varepsilon \rightarrow 0$. In fact, the bounds on the frequency process $\frac{Y_a^K}{Y_a^K + Y_d^K}$ will be satisfied by the time $T_\varepsilon^K + \log \log(1/\varepsilon)$ with probability converging to 1 as $K \rightarrow \infty$.

Proof. This proof again is a simple adaptation of the proofs from [BT20, Proposition 4.4.] or [CCL⁺21, Proposition 3.2]. □

Now, we can show our general result on the competition with non-negative transfer. We do not assume the transition rates to be constant anymore, but instead assume that there are constants $a_1, b_1, d_1, d_2 > 0$, $\tau \geq 0$ with $a_1, b_1 > d_1$ such that for some $s > 0$

$$\begin{aligned} \sup_{0 \leq t \leq s \log K} &|a_1^K(t) - a_1| + |b_1^K(t) - b_1| + |d_1^K(t) - d_1| \\ &+ |d_2^K(t) - d_2| + |\tau^K(t) - \tau| + \left|\frac{\gamma_1^K(t)}{K}\right| + \left|\frac{\gamma_2^K(t)}{K}\right| \xrightarrow{K \rightarrow \infty} 0 \end{aligned} \quad (\text{B.16})$$

in probability. If $\tau = 0$, then we assume $\tau^K \equiv 0$ for all K .

Proposition B.15. *Assume that the conditions (B.10), (B.12) and (B.16) are true. Consider the process $(X_a^K, X_d^K, Y_a^K, Y_d^K)$ with initial condition $\frac{1}{K}(X_{a,0}^K, X_{d,0}^K) \in [\bar{x}_a - \varepsilon, \bar{x}_a + \varepsilon] \times [\bar{x}_d - \varepsilon, \bar{x}_d + \varepsilon]$ and $\frac{1}{K}(Y_{a,0}^K + Y_{d,0}^K) = m\varepsilon$ for some $\varepsilon > 0$ and $m > 0$ sufficiently small. Then, for any $\varepsilon' > 0$ there exists a finite time $T = T(m, \varepsilon, \varepsilon')$ such that*

$$\lim_{K \rightarrow \infty} \mathbb{P}\left(X_{a,T}^K + X_{d,T}^K \leq \varepsilon'K, \frac{Y_{a,T}^K}{K} \in [\bar{y}_a - \varepsilon', \bar{y}_a + \varepsilon'], \frac{Y_{d,T}^K}{K} \in [\bar{y}_d - \varepsilon', \bar{y}_d + \varepsilon']\right) \geq 1 - o_\varepsilon(1).$$

Remark B.16. The choice on the initial condition for (Y_a^K, Y_d^K) can be generalized to hold for an entire interval of initial conditions. With the proposition as stated, for $\varepsilon, \varepsilon', m > 0$ sufficiently small, it holds

$$Y_{a,T}^K + Y_{d,T}^K \geq K(\bar{y}_a + \bar{y}_d - 2\varepsilon') > 2m\varepsilon K$$

at time $T = T(m, \varepsilon, \varepsilon')$ with high probability as $\varepsilon \rightarrow 0$ and $K \rightarrow \infty$. Thus, with high probability we have $T_{2m\varepsilon}^K < T(m, \varepsilon, \varepsilon')$. Hence, with high probability, for all initial conditions from an interval $Y_{a,0}^K + Y_{d,0}^K \in [\frac{Km\varepsilon}{2}, Km\varepsilon]$, the time $T_{2m\varepsilon}^K$ is bounded, even as $K \rightarrow \infty$. In particular, we can adapt the proof of Proposition B.15 such that the claim holds for all initial conditions $Y_{a,0}^K + Y_{d,0}^K \in [\frac{Km\varepsilon}{2}, Km\varepsilon]$, where the only change is in the application of Lemma B.14, which now yields that with high probability the good initial condition is satisfied after a time shorter than $T_{2m\varepsilon}^K + \log \log(1/(2m\varepsilon))$. This time however is finite with high probability by the arguments made above.

Proof. The proof is based on a suitable coupling of the process in combination with the above lemmata. Let $\delta > 0$. We want to couple in such a way that coordinatewise both bi-type branching processes are bounded. That is, we want to find a coupling such that

$$(X_a^{\delta,1}, X_d^{\delta,1}, Y_a^{\delta,1}, Y_d^{\delta,1}) \leq (X_a^K, X_d^K, Y_a^K, Y_d^K) \leq (X_a^{\delta,2}, X_d^{\delta,2}, Y_a^{\delta,2}, Y_d^{\delta,2})$$

coordinatewise. As in Lemma B.5, this coupling can be achieved by subtracting and adding δ to the birth and death rates for K large enough, which we can do by the convergence criterion above. For now, we do not alter the horizontal transfer rate τ^K . Note that indeed we are allowed to neglect the immigration rate for to the same reason as in Lemma B.5. For $\delta > 0$ small enough, the corresponding equilibria $(\bar{x}_a^{\delta,*}, \bar{x}_d^{\delta,*})$ and $(\bar{y}_a^{\delta,*}, \bar{y}_d^{\delta,*})$ with $* \in \{1, 2\}$ are closer than $\frac{\varepsilon \wedge \varepsilon'}{2}$ to (\bar{x}_a, \bar{x}_d) and (\bar{y}_a, \bar{y}_d) respectively. If $\tau > 0$, then we further couple the processes $(X_a^{\delta,*}, X_d^{\delta,*}, Y_a^{\delta,*}, Y_d^{\delta,*})$ with $* \in \{1, 2\}$ with processes $(X_a^{\delta,*,\diamond}, X_d^{\delta,*,\diamond}, Y_a^{\delta,*,\diamond}, Y_d^{\delta,*,\diamond})$ where $\diamond \in \{+, -\}$. In the case $\diamond = -$ we set the horizontal transfer rate to be $\tau - \delta$ and for $\diamond = +$ it is set to be $\tau + \delta$. This definition yields the inequalities as displayed in Table 1. Now arguments analogous to Lemma B.3 show that for any initial condition

$$\begin{array}{cccc|cccc} X_a^{\delta,1,-} & \geq & X_a^{\delta,1} & \geq & X_a^{\delta,1,+} & X_a^{\delta,2,-} & \geq & X_a^{\delta,2} & \geq & X_a^{\delta,2,+} \\ X_d^{\delta,1,-} & \geq & X_d^{\delta,1} & \geq & X_d^{\delta,1,+} & X_d^{\delta,2,-} & \geq & X_d^{\delta,2} & \geq & X_d^{\delta,2,+} \\ Y_a^{\delta,1,-} & \leq & Y_a^{\delta,1} & \leq & Y_a^{\delta,1,+} & Y_a^{\delta,2,-} & \leq & Y_a^{\delta,2} & \leq & Y_a^{\delta,2,+} \\ Y_d^{\delta,1,-} & \leq & Y_d^{\delta,1} & \leq & Y_d^{\delta,1,+} & Y_d^{\delta,2,-} & \leq & Y_d^{\delta,2} & \leq & Y_d^{\delta,2,+} \end{array}$$

TABLE 1. An overview of the almost sure inequalities that we obtain by coupling.

from a compact set, the processes $(X_a^{\delta,*,\diamond}, X_d^{\delta,*,\diamond}, Y_a^{\delta,*,\diamond}, Y_d^{\delta,*,\diamond})$ converge in probability to the solutions of the respective differential equations with $K \rightarrow \infty$. Furthermore, Lemma B.14 implies that with high probability in $m\varepsilon$ the criterion for a good initial condition (B.15) is satisfied for each of the coupled processes after a time shorter than $\log \log(1/(m\varepsilon))$. Indeed, when applying the lemma, notice that we can substitute T_ε^K by 0 due to our choice of the starting condition. Hence, by Lemma B.10 the solutions of the differential equations converge towards the equilibria $(0, 0, \bar{y}_a^{\delta,*,\diamond}, \bar{y}_d^{\delta,*,\diamond})$. In particular, there exists a finite time such that for all initial conditions from a compact set as in the proposition, the process is in a neighbourhood of $(0, 0, \bar{y}_a^{\delta,*,\diamond}, \bar{y}_d^{\delta,*,\diamond})$ with high probability in $m\varepsilon$. Thus, the claim follows. \square

With a similar proof we also obtain the same result for negative transfer. In this situation we assume the process (Y_a^K, Y_d^K) to be initially resident and the process (X_a^K, X_d^K) to be invading.

Proposition B.17. *Assume that the reverse inequalities of (B.10) and (B.12) are true that is*

$$-(b_1 + \tau - d_1 - C\bar{x}_a) > \frac{\sigma_2 q C \bar{x}_a}{d_2 + \sigma_2} \quad \text{and} \quad -(a_1 - \tau - d_1 - C\bar{y}_a) > \frac{\sigma_2 p C \bar{y}_a}{d_2 + \sigma_2}, \quad (\text{B.17})$$

which indicates the approximating process (\hat{Y}_a, \hat{Y}_d) to be subcritical and the approximating process (\hat{X}_a, \hat{X}_d) to be supercritical. Further assume (B.16). Consider the process $(X_a^K, X_d^K, Y_a^K, Y_d^K)$ with initial condition $\frac{1}{K}(Y_{a,0}^K, Y_{d,0}^K) \in [\bar{y}_a - \varepsilon, \bar{y}_a + \varepsilon] \times [\bar{y}_d - \varepsilon, \bar{y}_d + \varepsilon]$ and $\frac{1}{K}(X_{a,0}^K + X_{d,0}^K) = m\varepsilon$ for some $\varepsilon > 0$ and $m > 0$ sufficiently small. Then for any $\varepsilon' > 0$, there exists a finite time $T = T(m, \varepsilon, \varepsilon')$ such that

$$\lim_{K \rightarrow \infty} \mathbb{P} \left(Y_{a,T}^K + Y_{d,T}^K \leq \varepsilon' K, \frac{X_{a,T}^K}{K} \in [\bar{x}_a - \varepsilon', \bar{x}_a + \varepsilon'], \frac{X_{d,T}^K}{K} \in [\bar{x}_d - \varepsilon', \bar{x}_d + \varepsilon'] \right) \geq 1 - o_\varepsilon(1).$$

The proof follows a very similar structure of the previous proposition. We give an outline of the necessary results.

Lemma B.18. *Consider the system (B.7) and assume (B.17). If the initial condition (x_a, x_d, y_a, y_d) satisfies*

$$\frac{pC(x_a + y_a)}{d_2 + \sigma_2} > \frac{x_d}{x_a} > \frac{d_1 - a_1 + C(x_a + y_a) + \tau \frac{y_a}{x_a + y_a}}{\sigma_2}, \quad (\text{B.18})$$

then

$$\lim_{t \rightarrow \infty} (x_a(t), x_d(t), y_a(t), y_d(t)) = (\bar{x}_a, \bar{x}_d, 0, 0).$$

Proof. The proof of this lemma is identical with the proof of Lemma B.10 where we reversed the roles of (x_a, x_d) and (y_a, y_d) . The validity holds due to the fact that we never explicitly use the term involving τ . \square

Proceeding in the same manner, we define (π_a, π_d) to be the unique normed left eigenvector of the matrix

$$J = \begin{pmatrix} a_1 - \tau - d_1 - C\bar{y}_a & pC\bar{y}_a \\ \sigma_2 & -d_2 - \sigma_2 \end{pmatrix} \quad (\text{B.19})$$

corresponding to the principal eigenvalue $\lambda > 0$, which exists due to our assumption (B.17).

Lemma B.19. *Suppose that the initial condition (x_a, x_d, y_a, y_d) of the dynamical system (B.7) satisfies $y_a \in [\bar{y}_a - A\sqrt{\varepsilon}, \bar{y}_a + A\sqrt{\varepsilon}]$, $y_d \in [\bar{y}_d - A\sqrt{\varepsilon}, \bar{y}_d + A\sqrt{\varepsilon}]$ for some constant $A > 0$ large enough and $x_a + x_d \in (0, \sqrt{\varepsilon})$ with $\frac{x_d}{x_a} = \frac{\pi_d}{\pi_a}$. Then, for ε sufficiently small, (x_a, x_d, y_a, y_d) satisfies (B.18).*

Proof. The proof in the case of positive transfer from Lemma B.11 is easily adapted to this case. \square

Next in our series of Lemmata, we had shown bounds on some exit times in Lemma B.12 in order to show that the assumptions from Lemma B.19 are satisfied with high probability. Here, there will be a major difference in the proof, as at one point we made use of the positive transfer.

Lemma B.20. *Let $K \mapsto (m_1^K, m_2^K)$ be a function from $(0, \infty)$ to $[0, \infty)^2$ such that $(m_1^K, m_2^K) \in \frac{1}{K}\mathbb{N}_0^2$ and $\lim_{K \rightarrow \infty} (m_1^K, m_2^K) = (\bar{y}_a, \bar{y}_d)$. Define the stopping times*

$$R_\varepsilon^K := \inf \left\{ t \geq 0 \mid \left| \frac{Y_{a,t}^K}{K} - \bar{y}_a \right| > \varepsilon \text{ or } \left| \frac{Y_{d,t}^K}{K} - \bar{y}_d \right| > \varepsilon \right\}$$

and

$$T_x^K := \inf \{ t > 0 \mid X_{a,t}^K + X_{d,t}^K = \lfloor xK \rfloor \}.$$

Then, for any $\xi \in [\frac{1}{2}, 1]$ and starting condition $\frac{1}{K}(X_{a,0}^K, X_{d,0}^K, Y_{a,0}^K, Y_{d,0}^K) = (\frac{n_1^K}{K}, \frac{n_2^K}{K}, m_1^K, m_2^K)$ with n_1^K and n_2^K such that $0 < n_1^K + n_2^K < K\varepsilon^\xi$, there exist some positive constants $A, \varepsilon_0 > 0$ such that for all $0 < \varepsilon \leq \varepsilon_0$

$$\lim_{K \rightarrow \infty} \mathbb{P}(R_{2A\varepsilon^\xi}^K \leq T_{\varepsilon^\xi}^K \wedge T_0^K) = 0.$$

Proof. This proof can be adapted from the proof of [BT21, Lemma C.2]. \square

From this lemma, we can show that with probability converging to 1, the invasive trait reaches a critical population size.

Lemma B.21. *Under the assumptions of Lemma B.20, denoting $\mathbf{X}_0^K := (X_{a,0}^K, X_{d,0}^K, Y_{a,0}^K, Y_{d,0}^K)$, it holds*

$$\lim_{K \rightarrow \infty} \mathbb{P}\left(T_{\varepsilon^\xi}^K < T_0^K \wedge R_{2A\varepsilon^\xi}^K \mid \mathbf{X}_0^K = K(\varepsilon_1^K, \varepsilon_2^K, m_1^K, m_2^K)\right) = 1,$$

where $\xi \in [\frac{1}{2}, 1]$ and $\varepsilon_1^K \rightarrow \varepsilon_1, \varepsilon_2^K \rightarrow \varepsilon_2$ are such that $\varepsilon_1 + \varepsilon_2 = \varepsilon$ with $\varepsilon, \varepsilon_1, \varepsilon_2 > 0$.

Proof. The proof now is almost identical with the one of Lemma B.13. The only difference arises in the coupling, where we now couple such that

$$X_v^{\varepsilon,-} \leq \widehat{X}_v \quad \text{and} \quad X_v^{\varepsilon,-} \leq X_v^K$$

where $v \in \{a, d\}$ by defining the transition rates as follows: For the process $(X_a^{\varepsilon,-}, X_d^{\varepsilon,-})$, we choose the transition rates

$$(n, m) \mapsto \begin{cases} (n+1, m) & \text{at rate } na_1 \\ (n-1, m) & \text{at rate } n \left(d_1 + \tau \frac{\bar{x}_a + 2A\varepsilon^\xi}{\bar{x}_a - (2A+1)\varepsilon^\xi} + (1-q)C(\bar{x}_a + 2A\varepsilon^\xi + \varepsilon^\xi) \right. \\ & \left. + qC(4A\varepsilon^\xi + \varepsilon^\xi) \right) \\ (n, m-1) & \text{at rate } md_2 \\ (n-1, m+1) & \text{at rate } nqC(\bar{x}_a - 2A\varepsilon^\xi) \\ (n+1, m-1) & \text{at rate } m\sigma_2, \end{cases}.$$

It is only important to note that for $\varepsilon > 0$ sufficiently small, the process $(X_a^{\varepsilon,-}, X_d^{\varepsilon,-})$ is again supercritical and thus the claim follows as in Lemma B.13 \square

Finally, we need to show that with high probability the assumptions of Lemma B.19 are satisfied, which is the analogous version of Lemma B.14.

Lemma B.22. *Assume (B.17) and assume that $\frac{1}{K}(X_{a,0}^K, X_{d,0}^K, Y_{a,0}^K, Y_{d,0}^K) = (\varepsilon_1, \varepsilon_2, m_1^K, m_2^K)$ is the initial condition with $\varepsilon_1 + \varepsilon_2 \leq \varepsilon$ for some $\varepsilon, \varepsilon_1, \varepsilon_2 > 0$ and m_1^K, m_2^K as in Lemma B.20. Then for any $\delta > 0$ with $\pi_a \pm \delta \in (0, 1)$, it holds*

$$\liminf_{K \rightarrow \infty} \mathbb{P}\left(\exists t \in [T_\varepsilon^K, T_{\sqrt{\varepsilon}}^K]: \pi_a - \delta < \frac{X_{a,t}^K}{X_{a,t}^K + X_{d,t}^K} < \pi_a + \delta\right) \geq 1 - o_\varepsilon(1).$$

In fact, the bound on the frequency process $\frac{X_a^K}{X_a^K + X_d^K}$ will be satisfied by the time $T_\varepsilon^K + \log \log(1/\varepsilon)$ with probability larger than $1 - o_\varepsilon(1)$ as $K \rightarrow \infty$.

Proof. The proof is largely analogous to the one of [BT20, Proposition 4.4], but there are subtle differences which we discuss here. We assume

$$\frac{X_{a,T_\varepsilon^K}^K}{X_{a,T_\varepsilon^K}^K + X_{d,T_\varepsilon^K}^K} \leq \pi_a - \delta.$$

Then, we introduce the event

$$\tilde{E}_\varepsilon := \left\{ T_{\sqrt{\varepsilon}}^K < T_0^K \wedge R_{2A\sqrt{\varepsilon}}^K \right\},$$

whose probability again converges to 1 as $K \rightarrow \infty$ by the previous lemma. Also, we define the stopping time

$$T_{\varepsilon, \varepsilon/M}^K := \inf \left\{ t \geq T_\varepsilon^K \mid X_{a,t}^K + X_{d,t}^K \leq \frac{\varepsilon K}{M} \right\}.$$

On the event \tilde{E}_ε , we can bound our process from below on the time interval $[T_\varepsilon^K, T_{\sqrt{\varepsilon}}^K]$ by a pure death process Z^K , with initial condition $Z_{T_\varepsilon^K}^K = \varepsilon K$ and death rate $(d_1 + \tau \frac{\sqrt{\varepsilon}(\bar{y}_a + (2A+1)\varepsilon)}{\bar{y}_a - 2A\varepsilon}) \vee d_2$. We are interested in $(\varepsilon - \frac{\varepsilon}{M})K$ individuals dying from the process Z^K . This takes longer than $\log \log(1/\varepsilon)$ with probability converging to 1 for $M > 1$. Therefore, we have

$$\lim_{K \rightarrow \infty} \mathbb{P}(T_{\varepsilon, \varepsilon/M}^K < T_\varepsilon^K + \log \log(1/\varepsilon) \mid T_\varepsilon^K + \log \log(1/\varepsilon) < T_{\sqrt{\varepsilon}}^K, \tilde{E}_\varepsilon) = 0.$$

Furthermore, the population size of $X_a^K + X_d^K$ can be bounded from above by a Yule process with birth rate a_1 . Hence, [CCL⁺21, Lemma A.2] implies

$$\lim_{K \rightarrow \infty} \mathbb{P}(T_{\sqrt{\varepsilon}}^K < T_\varepsilon^K + \log \log(\frac{1}{\varepsilon}) \mid \tilde{E}_\varepsilon) \leq \widetilde{M} \sqrt{\varepsilon} (\log(\frac{1}{\varepsilon}))^{a_1}.$$

Since the probability of \tilde{E}_ε converges to 1, we obtain

$$\lim_{K \rightarrow \infty} \mathbb{P}(T_{\varepsilon, \varepsilon/M}^K < T_\varepsilon^K + \log \log(1/\varepsilon)) \leq o_\varepsilon(1) \quad \text{and} \quad \lim_{K \rightarrow \infty} \mathbb{P}(T_{\sqrt{\varepsilon}}^K < T_\varepsilon^K + \log \log(\frac{1}{\varepsilon})) \leq o_\varepsilon(1). \quad (\text{B.20})$$

The next step in this proof is the semimartingale decomposition, for which we will satisfy ourselves with the abstract representation

$$\frac{X_{a,t}^K}{X_{a,t}^K + X_{d,t}^K} = \frac{X_{a,T_\varepsilon^K}^K}{X_{a,T_\varepsilon^K}^K + X_{d,T_\varepsilon^K}^K} + M^K(t) + V^K(t) \quad \text{for } t \geq T_\varepsilon^K,$$

where M^K is a martingale and V^K a process of finite variation. The remainder of the proof can be adapted from [BT20, Proposition 4.4], with the slight exception of now only obtaining

$$\begin{aligned} & \lim_{K \rightarrow \infty} \mathbb{P} \left(\sup_{T_\varepsilon^K \leq s \leq T_\varepsilon^K + \log \log(1/\varepsilon)} |M^K(t)| \geq \varepsilon \right) \\ & \leq \lim_{K \rightarrow \infty} \mathbb{P} \left(\sup_{T_\varepsilon^K \leq s \leq (T_\varepsilon^K + \log \log(1/\varepsilon)) \wedge T_{\varepsilon, \varepsilon/M}} |M^K(t)| \geq \varepsilon \right) + \mathbb{P}(T_{\varepsilon, \varepsilon/M} < T_\varepsilon^K + \log \log(\frac{1}{\varepsilon})) \\ & \leq \lim_{K \rightarrow \infty} \frac{1}{\varepsilon^2} \mathbb{E}[\langle M^K \rangle_{(T_\varepsilon^K + \log \log(1/\varepsilon)) \wedge T_{\varepsilon, \varepsilon/M}}] + o_\varepsilon(1) = o_\varepsilon(1), \end{aligned}$$

due to our weaker bound (B.20) and thus afterwards only having

$$\pi_a - \frac{\delta}{2} \geq \frac{X_{a,t}^K}{X_{a,t}^K + X_{d,t}^K} \geq \frac{\theta}{2} (\log \log(1/\varepsilon) \wedge (t_a^\varepsilon - T_\varepsilon^K)) - \varepsilon$$

with probability $1 - o_\varepsilon(1)$ as $K \rightarrow \infty$ for the corresponding time t_a^ε , which does not change the claim, since we are considering only high probabilities for small ε . \square

We can now discuss the proof of Proposition B.17, where the transition rates are not assumed to be constant.

Proof of Proposition B.17. We can copy the proof of Proposition B.15 until the point, where we have found the couplings $(X_a^{\delta,*,\diamond}, X_d^{\delta,*,\diamond}, Y_a^{\delta,*,\diamond}, Y_d^{\delta,*,\diamond})$. Now, the only difference is that we need to use our Lemmata applying to this case. The general argument is not changed. That is, from Lemma B.22 we see that with high probability in $m\varepsilon$ the convergence condition from Lemma B.19 is satisfied and therefore

Lemma B.18 implies convergence to the equilibrium $(\bar{x}_a^{\delta,*,\diamond}, \bar{x}_d^{\delta,*,\diamond}, 0, 0)$. In particular, there exists a finite time such that for all initial conditions from a compact set we have entered a neighbourhood of this equilibrium, which as before implies the claim. \square

B.3. Competition Between Bi-Type and Single-Type Processes with Transfer. We consider a three-dimensional branching process very similar to the one from the previous section. The transfer rates of this process (X_a^K, X_d^K, Y^K) are

$$(i, j, k) \rightarrow \begin{cases} (i+1, j, k) & \text{at rate } ia_1^K(\omega, t) + \gamma_1^K(\omega, t) \\ (i, j, k+1) & \text{at rate } kb_1^K(\omega, t) + \gamma_2^K(\omega, t) \\ (i-1, j, k) & \text{at rate } i(d_1^K(\omega, t) + \frac{(1-p)C}{K}(i+k)) \\ (i, j, k-1) & \text{at rate } k(d_1^K(\omega, t) + \frac{C}{K}(i+k)) \\ (i, j-1, k) & \text{at rate } jd_2^K(\omega, t) \\ (i-1, j+1, k) & \text{at rate } i\frac{pC}{K}(i+k) \\ (i+1, j-1, k) & \text{at rate } j\sigma_2 \\ (i-1, j, k+1) & \text{at rate } \tau^K(\omega, t)\frac{ik}{i+k} \end{cases}$$

with predictable, non-negative functions $a_1^K, b_1^K, d_1^K, d_2^K, \tau^K, \gamma_1^K, \gamma_2^K: \Omega \times [0, \infty) \rightarrow \mathbb{R}$ and constants $C, \sigma_2 > 0, p \in (0, 1)$. As before, we assume for some $s > 0$ the convergence

$$\sup_{0 \leq t \leq s \log K} |a_1^K(t) - a_1| + |b_1^K(t) - b_1| + |d_1^K(t) - d_1| + |d_2^K(t) - d_2| + |\tau^K(t) - \tau| + \left| \frac{\gamma_1^K(t)}{K} \right| + \left| \frac{\gamma_2^K(t)}{K} \right| \xrightarrow{K \rightarrow \infty} 0 \quad (\text{B.21})$$

in probability, where $a_1, b_1, d_1, d_2 > 0, \tau \geq 0$ and $a_1, b_1 > d_1$. If $\tau = 0$ we again assume $\tau^K \equiv 0$. Then, we can approximate the process Y^K in a population close to the equilibrium size of the process (X_a^K, X_d^K) which as in the previous section is given by $K(\bar{x}_a, \bar{x}_d)$ by the process \hat{Y} , which has the transitions

$$n \mapsto \begin{cases} n+1, & \text{at rate } nb_1 + \tau \\ n-1, & \text{at rate } n(d_1 + C\bar{x}_a). \end{cases}$$

We want the growth rate to be strictly positive, so that we have invasion of this trait. In the case of one dimensional branching processes, this is equivalent to

$$b_1 + \tau - d_1 - C\bar{x}_a > 0, \quad (\text{B.22})$$

which also coincides with (B.10) in the case $q = 0$. The approximation of the process (X_a^K, X_d^K) in a population, where Y^K is close to its equilibrium size $K\bar{y}$ with

$$\bar{y} = \frac{b_1 - d_1}{C},$$

can be done as before. That is, we approximate using the process (\hat{X}_a, \hat{X}_d) with transitions

$$(n, m) \mapsto \begin{cases} (n+1, m) & \text{at rate } na_1 \\ (n-1, m) & \text{at rate } n(d_1 + \tau + (1-p)C\bar{y}) \\ (n, m-1) & \text{at rate } md_2 \\ (n-1, m+1) & \text{at rate } npC\bar{y} \\ (n+1, m-1) & \text{at rate } m\sigma_2. \end{cases}$$

In order to guarantee a successful invasion, we want this process to be subcritical, which as before coincides with the criterion (B.12), that is

$$-(a_1 - \tau - d_1 - C\bar{y}) > \frac{\sigma_2 p C \bar{y}}{d_2 + \sigma_2}. \quad (\text{B.23})$$

We are now in a position to formulate our competition results for this case. Note, that a generalized form, in the sense that the invading process is initially of size contained in the interval $[\frac{m\varepsilon K}{2}, m\varepsilon K]$, can be proven as indicated in Remark B.16.

Proposition B.23. *Assume that conditions (B.21), (B.22) and (B.23) are true. Consider the process (X_a^K, X_d^K, Y^K) with initial condition $\frac{1}{K}(X_{a,0}^K, X_{d,0}^K) \in [\bar{x}_a - \varepsilon, \bar{x}_a + \varepsilon] \times [\bar{x}_d - \varepsilon, \bar{x}_d + \varepsilon]$ and $\frac{Y^K}{K} = m\varepsilon$ for some $\varepsilon > 0$ and $m > 0$ sufficiently small. Then for any $\varepsilon' > 0$, there exists a finite time $T = T(m, \varepsilon, \varepsilon')$ such that*

$$\lim_{K \rightarrow \infty} \mathbb{P} \left(X_{a,T}^K + X_{d,T}^K \leq \varepsilon' K, \frac{Y_T^K}{K} \in [\bar{y} - \varepsilon', \bar{y} + \varepsilon'] \right) = 1.$$

For the proof of this proposition, we use arguments from [BT21, Section 5]. The structure of the proof is as usual: First, we assume the population to satisfy a suitable initial condition and then approximate the dynamics of the system by a differential equation, whose solutions will converge to the corresponding fixed points.

Assume that the initial condition of the process $\frac{1}{K}(X_a^K, X_d^K, Y^K)$ is contained in the set

$$\mathcal{A}_\varepsilon^3 := [\bar{x}_a - 2A\varepsilon^\xi, \bar{x}_a + 2A\varepsilon^\xi] \times [\bar{x}_d - 2A\varepsilon^\xi, \bar{x}_d + 2A\varepsilon^\xi] \times [\varepsilon, \sqrt{\varepsilon}]$$

for fixed $\varepsilon > 0$. We want to show that the solution of the dynamical system

$$\begin{aligned} \dot{x}_a &= x_a(a_1 - d_1 - C(x_a + y)) + x_d\sigma_2 - \tau \frac{x_a y}{x_a + y} \\ \dot{x}_d &= pC x_a(x_a + y) - (d_2 + \sigma_2)x_d \\ \dot{y} &= y(b_1 - d_1 - C(x_a + y)) + \tau \frac{x_a y}{x_a + y} \end{aligned} \quad (\text{B.24})$$

converges towards the equilibrium $(0, 0, \bar{y})$ for any starting condition from $\mathcal{A}_\varepsilon^3$.

Lemma B.24. *Consider the dynamical system (B.24). If the initial condition $(x_a(0), x_d(0), y)$ is contained in the set $\mathcal{A}_\varepsilon^3$ and the inequalities (B.22) and (B.23) are satisfied, then*

$$\lim_{t \rightarrow \infty} (x_a(t), x_d(t), y(t)) = (0, 0, \bar{y}).$$

Proof. This proof is taken from [BT21, Proposition 5.4] and adapted to our case. Firstly, we notice that $y(t)$ is strictly increasing as long as

$$C(x_a + y) - \tau \frac{x_a}{x_a + y} < b_1 - d_1.$$

By our choice of the starting conditions, we see that

$$C(x_a + y) - \tau \frac{x_a}{x_a + y} \leq C(\bar{x}_a + (2A + 1)\sqrt{\varepsilon}) - \tau \frac{\bar{x}_a - 2A\sqrt{\varepsilon}}{\bar{x}_a + (2A + 1)\sqrt{\varepsilon}} \leq C\bar{x}_a - \tau + C_*\sqrt{\varepsilon} < b_1 - d_1,$$

where we used (B.22) and C_* is a sufficiently large constant. Hence, initially $y(t)$ is strictly increasing and will do so until

$$C(x_a + y) - \tau \frac{x_a}{x_a + y} = b_1 - d_1. \quad (\text{B.25})$$

We now consider the resident population dynamics (x_a, x_d) in the case where this equality is true and then we perform a perturbation argument. Assuming the equality, we easily compute

$$C(x_a + y) + \tau \frac{y}{x_a + y} = C(x_a + y) - \tau \frac{x_a}{x_a + y} + \tau = b_1 - d_1 + \tau$$

and similarly

$$pCx_a(x_a + y) = px_a \left(C(x_a + y) - \tau \frac{x_a}{x_a + y} + \tau \frac{x_a}{x_a + y} \right) \leq px_a(b_1 - d_1 + \tau).$$

Hence, we can dominate the dynamics of (x_a, x_d) under the assumption (B.25) by the solutions to the system

$$\begin{aligned} \dot{x}_a &= x_a(a_1 - b_1 - \tau) + \sigma_2 x_d \\ \dot{x}_d &= px_a(b_1 - d_1 + \tau) - (d_2 + \sigma_2)x_d. \end{aligned} \tag{B.26}$$

This system has the unique equilibrium $(0, 0)$ if the coefficient matrix of the corresponding linear system

$$A = \begin{pmatrix} a_1 - b_1 - \tau & \sigma_2 \\ p(b_1 - d_1 + \tau) & -(d_2 + \sigma_2) \end{pmatrix}$$

has non-zero determinant. Indeed, the determinant is 0 if and only if

$$a_1 = b_1 + \tau - \frac{p\sigma_2(C\bar{y} + \tau)}{d_2 + \sigma_2}.$$

However, this choice of a_1 would imply

$$\bar{x}_a = \frac{(a_1 - d_1)(d_2 + \sigma_2)}{C(d_2 + (1 - p)\sigma_2)} = \frac{(b_1 + \tau - d_1)(d_2 + \sigma_2) - p\sigma_2(b_1 - d_1 + \tau)}{C(d_2 + (1 - p)\sigma_2)} = \frac{b_1 + \tau - d_1}{C},$$

so

$$b_1 + \tau - d_1 - C\bar{x}_a = 0,$$

which contradicts (B.22) and hence the only equilibrium is $(0, 0)$.

Next, we will show that the system converges towards the equilibrium $(0, 0)$. For this, we need to show that both eigenvalues of the matrix A are negative. Note that for the existence of a positive eigenvalue we need a positive trace and hence $a_1 - b_1 - \tau > 0$. Then, we obtain

$$a_1 - \tau - d_1 - C\bar{y} = a_1 - \tau - b_1 > 0,$$

which contradicts the condition (B.23) since now

$$0 > -(a_1 - \tau - d_1 - C\bar{y}) > \frac{\sigma_2 p C \bar{y}}{d_2 + \sigma_2} \geq 0.$$

Hence, there cannot be a positive eigenvalue. Moreover, due to the determinant of A being non-zero, both eigenvalues must be strictly negative. Since the dynamics of the solutions are determined by the eigenvalues, the solution will converge for any positive starting condition to $(0, 0)$.

Now we turn again to the general dynamical system (B.24). We distinguish two cases:

Case 1: Monotonicity. In this case we assume that $y(t)$ eventually will be a monotone function. By boundedness of $y(t)$ this implies that $\lim_{t \rightarrow \infty} \dot{y}(t) = 0$ and since $y(t)$ is bounded away from 0, we must have

$$C(x_a(t) + y(t)) - \tau \frac{x_a(t)}{x_a(t) + y(t)} \xrightarrow{t \rightarrow \infty} b_1 - d_1.$$

In particular, for any $\delta > 0$ we can find a time $t_1(\delta) > 0$ such that for all starting conditions in $\mathcal{A}_\varepsilon^3$ we have

$$b_1 - d_1 - \delta < C(x_a(t) + y(t)) - \tau \frac{x_a}{x_y + a} < b_1 - d_1 + \delta$$

for all $t > t_1$. Then, we can consider the system (B.26) altered in the corresponding places by $\pm\delta$ to account for the bounds above. We still obtain convergence to $(0, 0)$ for both systems and therefore we must have

$$\lim_{t \rightarrow \infty} (x_a(t), x_d(t)) = (0, 0).$$

As $C(x_a(t) + y(t)) - \tau \frac{x_a(t)}{x_a(t) + y(t)} \rightarrow b_1 - d_1$ we have $y(t) \rightarrow \bar{y}$.

Case 2: Non-monotone convergence. It is possible that $t \mapsto y(t)$ is not monotone on any interval (s, ∞) for $s \geq 0$. In this case, due to the boundedness of s there must be a countable number of local extrema of y , for if there were only finitely many, then y would become monotone eventually. At any local minimum or maximum, since y is bounded away from 0, we must have

$$C(x_a + y) - \tau \frac{x_a}{x_a + y} = b_1 - d_1.$$

Using our observations from Case 1, we see that at any maximum and at any minimum the population of (x_a, x_d) is decreasing and converges to $(0, 0)$ even monotonically. Since this holds independently of the size of $y(t)$, we obtain that

$$\lim_{t \rightarrow \infty} (x_a(t), x_d(t)) = (0, 0).$$

As before, we derive $\lim_{t \rightarrow \infty} y(t) = \bar{y}$. □

Proof of Proposition B.23. As usual, we couple the process (X_a^K, X_d^K, Y^K) with two processes from above and below by

$$(X_a^{\delta,-}, X_d^{\delta,-}, Y^{\delta,-}) \leq (X_a^K, X_d^K, Y^K) \leq (X_a^{\delta,+}, X_d^{\delta,+}, Y^{\delta,+})$$

by accordingly increasing or decreasing the birth, death and switching rates by some term involving $\delta > 0$. We further couple with processes $(X_a^{\delta,*,\diamond}, X_d^{\delta,*,\diamond}, Y^{\delta,*,\diamond})$ as in Proposition B.15 by increasing or decreasing the transfer rate τ^K by δ . For $\varepsilon > 0$ sufficiently small, the initial condition from Lemma B.24 is satisfied and thus after a finite time, the solutions to the differential equations converge towards their equilibria as time tends to infinity. By arguments similar to Lemma B.3, the processes converge to the solutions of the corresponding differential equations. In particular, as in Proposition B.15 we see that for $\delta > 0$ sufficiently small, the processes are inside a neighbourhood of the equilibrium after a finite time with probability converging to 1. □

A similar result holds for the inverse invasion.

Proposition B.25. *Assume that (B.21) and the inverse inequalities of (B.22) and (B.23) are true, that is*

$$b_1 + \tau - d_1 - C\bar{x}_a < 0 \quad \text{and} \quad -(a_1 - \tau - d_1 - C\bar{y}) < \frac{\sigma_2 p C \bar{y}}{d_2 + \sigma_2}. \quad (\text{B.27})$$

Consider the process (X_a^K, X_d^K, Y^K) with initial condition $\frac{1}{K}(X_{a,0}^K + X_{d,0}^K) = m\varepsilon$ and $\frac{Y^K}{K} \in [\bar{y} - \varepsilon, \bar{y} + \varepsilon]$ for some $\varepsilon > 0$ and $m > 0$ sufficiently small. Then for any $\varepsilon' > 0$, there exists a finite time $T = T(m, \varepsilon, \varepsilon')$ such that

$$\lim_{K \rightarrow \infty} \mathbb{P} \left(Y_T^K \leq \varepsilon' K, \frac{X_{a,T}^K}{K} \in [\bar{x}_a - \varepsilon', \bar{x}_a + \varepsilon'], \frac{X_{d,T}^K}{K} \in [\bar{x}_d - \varepsilon', \bar{x}_d + \varepsilon'] \right) \geq 1 - o_\varepsilon(1).$$

Proof of Proposition B.25. Here, we are looking at the invasion of a population with dormancy which does not benefit from horizontal transfer. This is the same situation as in Proposition B.17, where the difference is only in the dormancy of the initially resident trait. However, since we have never used this aspect in the proof of Proposition B.17, the chain of arguments remains valid, where of course in the couplings we need to account for the lack of dormancy in the initially resident trait. \square

In addition, we also need the same results, but with inverted horizontal transfer. That is, we now consider the process (X_a^K, X_d^K, Y^K) which has the same transitions as in the beginning of the section except for the transition $(i, j, k) \rightarrow (i-1, j, k+1)$ to be replaced with the transition $(i, j, k) \rightarrow (i+1, j, k-1)$. Then we can still approximate the processes (X_a^K, X_d^K) and Y^K as before, but we need to switch the addition of τ in the rates from the birth to the death rate and vice versa. In particular, we want the inequalities

$$b_1 - \tau - d_1 - C\bar{x}_a > 0 \quad \text{and} \quad -(a_1 + \tau - d_1 - C\bar{y}) > \frac{\sigma_2 p C \bar{y}}{d_2 + \sigma_2} \quad (\text{B.28})$$

to hold. Then we obtain the same results as above.

Proposition B.26. *Assume that (B.21) and (B.28) are true. Consider the process (X_a^K, X_d^K, Y^K) with initial condition $\frac{1}{K}(X_{a,0}^K, X_{d,0}^K) \in [\bar{x}_a - \varepsilon, \bar{x}_a + \varepsilon] \times [\bar{x}_d - \varepsilon, \bar{x}_d + \varepsilon]$ and $\frac{Y^K}{K} = m\varepsilon$ for some $\varepsilon > 0$ and $m > 0$ sufficiently small. Then for any $\varepsilon' > 0$, there exists a finite time $T = T(m, \varepsilon, \varepsilon')$ such that*

$$\lim_{K \rightarrow \infty} \mathbb{P} \left(X_{a,T}^K + X_{d,T}^K \leq \varepsilon' K, \frac{Y_T^K}{K} \in [\bar{y} - \varepsilon', \bar{y} + \varepsilon'] \right) = 1.$$

The idea for proving Proposition B.26 is very similar to the one of Proposition B.23. The only difficulty arises in the preceding lemma, where we have used explicitly that the invading population is benefiting from horizontal transfer. As before, assume that the initial condition of the process $\frac{1}{K}(X_a^K, X_d^K, Y^K)$ is contained in the set

$$\mathcal{A}_\varepsilon^3 := [\bar{x}_a - 2A\varepsilon^\xi, \bar{x}_a + 2A\varepsilon^\xi] \times [\bar{x}_d - 2A\varepsilon^\xi, \bar{x}_d + 2A\varepsilon^\xi] \times [\varepsilon, \sqrt{\varepsilon}]$$

for fixed $\varepsilon > 0$. We want to show that the solution to the dynamical system

$$\begin{aligned} \dot{x}_a &= x_a(a_1 - d_1 - C(x_a + y)) + x_d\sigma_2 + \tau \frac{x_a y}{x_a + y} \\ \dot{x}_d &= pC x_a(x_a + y) - (d_2 + \sigma_2)x_d \\ \dot{y} &= y(b_1 - d_1 - C(x_a + y)) - \tau \frac{x_a y}{x_a + y} \end{aligned} \quad (\text{B.29})$$

converges towards the equilibrium $(0, 0, \bar{y})$ for any starting condition from $\mathcal{A}_\varepsilon^3$.

Lemma B.27. *Consider the dynamical system (B.29). If the initial condition $(x_a(0), x_d(0), y)$ is contained in the set $\mathcal{A}_\varepsilon^3$ and the inequalities (B.28) are satisfied, then*

$$\lim_{t \rightarrow \infty} (x_a(t), x_d(t), y(t)) = (0, 0, \bar{y}).$$

Proof. The concept of the proof stays the same. Initially the function $y(t)$ is increasing until eventually

$$C(x_a + y) + \tau \frac{x_a}{x_a + y} = b_1 - d_1.$$

We again consider the system for (x_a, x_d) under this condition and need to show that this implies $(x_a(t), x_d(t)) \rightarrow 0$ as $t \rightarrow \infty$. As before, we calculate

$$C(x_a + y) - \tau \frac{y}{x_a + y} = C(x_a + y) + \tau \frac{x_a}{x_a + y} - \tau = b_1 - d_1 - \tau$$

and

$$pC x_a(x_a + y) = p x_a \left(C(x_a + y) + \tau \frac{x_a}{x_a + y} - \tau \frac{x_a}{x_a + y} \right) \leq p x_a (b_1 - d_1).$$

Thus, we consider

$$\begin{aligned}\dot{x}_a &= x_a(a_1 - b_1 + \tau) + \sigma_2 x_d \\ \dot{x}_d &= p x_a(b_1 - d_1) - (d_2 + \sigma_2) x_d.\end{aligned}\tag{B.30}$$

Again, we compute the determinant of the coefficient matrix

$$A = \begin{pmatrix} a_1 - b_1 + \tau & \sigma_2 \\ p(b_1 - d_1) & -(d_2 + \sigma_2) \end{pmatrix}$$

to be 0 if and only if

$$a_1 = b_1 - \tau - \frac{p\sigma_2 C\bar{y}}{d_2 + \sigma_2}.$$

From (B.28) we know that

$$a_1 < -\frac{p\sigma_2 C\bar{y}}{d_2 + \sigma_2} - \tau + d_1 + C\bar{y} = -\frac{p\sigma_2 C\bar{y}}{d_2 + \sigma_2} - \tau + b_1,$$

so $(0, 0)$ is again the only equilibrium and the eigenvalues of A are negative, so for any positive initial condition the solution of the system (B.29) converges to $(0, 0)$. The remainder of the proof can be taken from the proof of Lemma B.24. \square

Now, we can prove Proposition B.26.

Proof of Proposition B.26. The proof is identical with the one of Proposition B.23, except for the need of using Lemma A.7 in this section instead of Lemma B.24. \square

Proposition B.28. *Assume that (B.21) and the inverse inequalities of (B.28) are true, that is*

$$b_1 - \tau - d_1 - C\bar{x}_a < 0 \quad \text{and} \quad -(a_1 + \tau - d_1 - C\bar{y}) < \frac{\sigma_2 p C\bar{y}}{d_2 + \sigma_2}.\tag{B.31}$$

Consider the process (X_a^K, X_d^K, Y^K) described above before Proposition B.26 with initial condition $\frac{1}{K}(X_{a,0}^K + X_{d,0}^K) = m\varepsilon$ and $\frac{Y^K}{K} \in [\bar{y} - \varepsilon, \bar{y} + \varepsilon]$ for some $\varepsilon > 0$ and $m > 0$ sufficiently small. Then for any $\varepsilon' > 0$, there exists a finite time $T = T(m, \varepsilon, \varepsilon')$ such that

$$\lim_{K \rightarrow \infty} \mathbb{P} \left(Y_T^K \leq \varepsilon' K, \frac{X_{a,T}^K}{K} \in [\bar{x}_a - \varepsilon', \bar{x}_a + \varepsilon'], \frac{X_{d,T}^K}{K} \in [\bar{x}_d - \varepsilon', \bar{x}_d + \varepsilon'] \right) \geq 1 - o_\varepsilon(1).$$

Proof of Proposition B.28. In this case, we are looking at the invasion of a population with dormancy which benefits from horizontal transfer. This is the same situation as in Proposition B.15, where the difference is only in the dormancy of the initially resident trait. As above, we have never used this aspect in the proof and hence the chain of arguments is still valid, where again the couplings with the initially resident population need to be slightly adapted. \square

Acknowledgements. JB was supported by DFG Priority Programme 1590 ‘‘Probabilistic Structures in Evolution’’, project 1105/5-1, and by Berlin Mathematics Research Center MATH+, project EF 4-7. TP was supported by Berlin Mathematics Research Center MATH+, project EF 4-7. AT was supported by DFG Priority Programme 1590 ‘‘Probabilistic Structures in Evolution’’.

REFERENCES

- [Bal13] D. A. Baltrus. Exploring the costs of horizontal gene transfer. *Trends Ecol. Evol.*, 28(8):489 – 495, 2013.
- [BB18] M. Baar and A. Bovier. The polymorphic evolution sequence for populations with phenotypic plasticity. *Elect. J. Prob.*, 23(72):1–27, 2018.
- [BCF⁺16] S. Billiard, P. Collet, R. Ferrière, S. Méléard, and V. C. Tran. The effect of competition and horizontal trait inheritance on invasion, fixation, and polymorphism. *J. Theoret. Biol.*, 411:48 – 58, 2016.

- [BCF⁺18] S. Billiard, P. Collet, R. Ferrière, S. Méléard, and V. C. Tran. Stochastic dynamics for adaptation and evolution of microorganisms. *J. Euro. Math. Soc., special issue for the Proceedings ECM2016*, pages 527–552, 2018.
- [BCS19] A. Bovier, L. Coquille, and C. Smadi. Crossing a fitness valley as a metastable transition in a stochastic population model. *Ann. Appl. Prob.*, 29(6):3541–3589, 2019.
- [BEGC⁺15] J. Blath, B. Eldon, A. González-Casanova, N. Kurt, and M. Wilke Berenguer. Genetic variability under the seed bank coalescent. *Genetics*, 200:921–934, 2015.
- [Ben08] P. M. Bennett. Plasmid encoded antibiotic resistance: acquisition and transfer of antibiotic resistance genes in bacteria. *Br. J. Pharmacol.*, 153:Suppl 1: S347–57, 2008.
- [BGCKWB16] J. Blath, A. González-Casanova, N. Kurt, and M. Wilke Berenguer. A new coalescent for seed-bank models. *Ann. Appl. Prob.*, 26:857–891, 04 2016.
- [BGCKWB20] J. Blath, A. González-Casanova, N. Kurt, and M. Wilke Berenguer. The seed bank coalescent with simultaneous switching. *Elect. J. Prob.*, 25(27):21 pp., 2020.
- [BP14] F. Baumdicker and P. Pfaffelhuber. The infinitely many genes model with horizontal gene transfer. *Elect. J. Prob.*, 19(115):1–27, 2014.
- [BT20] J. Blath and A. Tóbiás. Invasion and fixation of microbial dormancy traits under competitive pressure. *Stoch. Process. Their Appl.*, 130(12):7363–7395, 2020.
- [BT21] J. Blath and A. Tóbiás. The interplay of dormancy and transfer in bacterial populations: Invasion, fixation and coexistence regimes. *Theoret. Pop. Biol.*, 139:18–49, 2021.
- [CCL⁺21] C. Coron, M. Costa, F. Laroche, H. Leman, and C. Smadi. Emergence of homogamy in a two-loci stochastic population model. *ALEA, Lat. Am. J. Probab. Math. Stat.*, 18(469-508), 2021.
- [Cha06] N. Champagnat. A microscopic interpretation for adaptive dynamics trait substitution sequence models. *Stoch. Process. Their Appl.*, 116:1127–1160, 2006.
- [CKS21] L. Coquille, A. Kraut, and C. Smadi. Stochastic individual-based models with power law mutation rate on a general finite trait space. *Electron. J. Probab.*, 26:1–37, 2021.
- [CMT21] N. Champagnat, S. Méléard, and V. C. Tran. Stochastic analysis of emergence of evolutionary cyclic behavior in population dynamics with transfer. *Ann. Appl. Prob.*, 31:1820–1867, 2021.
- [Coh66] D. Cohen. Optimizing reproduction in a randomly varying environment. *J. Theoret. Biol.*, 12:119–129, 1966.
- [DE97] P. Dupuis and R. S. Ellis. *A weak convergence approach to the theory of large deviations*. John Wiley & Sons, 1997.
- [DM11] R. Durrett and J. Mayberry. Traveling waves of selective sweeps. *Ann. Appl. Prob.*, 21(2):699–744, 2011.
- [EK86] S. N. Ethier and T. G. Kurtz. *Markov Processes: Characterization and convergence*. Wiley, New York, 1986.
- [FM04] N. Fournier and S. Méléard. A microscopic probabilistic description of a locally regulated population and macroscopic approximations. *Ann. Appl. Probab.*, 14(4):1880–1919, 2004.
- [FW98] M. I. Freidlin and A. D. Wentzell. *Random Perturbations of Dynamical Systems*. Springer, second edition, 1998.
- [GB03] H.-O. Georgii and E. Baake. Supercritical multitype branching processes: The ancestral types of typical individuals. *Adv. Appl. Prob.*, 35(4):1090–1110, 2003.
- [KKL01] I. Kaj, S. M. Krone, and M. Lascoux. Coalescent theory for seed bank models. *J. Appl. Prob.*, 38(2):285–300, 2001.
- [KL05] E. Kussell and S. Leibler. Phenotypic diversity, population growth, and information in fluctuating environments. *Science*, 309:2075–2078, 2005.
- [KW12] E. V. Koonin and Y. I. Wolf. Evolution of microbes and viruses: a paradigm shift in evolutionary biology? *Front. Cell. Inf. Microbiol.*, 2(119), 2012.
- [LdHWBB21] J. T. Lennon, F. T. W den Hollander, M. Wilke Berenguer, and J. Blath. Principles of seed banks: complexity emerging from dormancy. *Nat. Commun.*, 12:1–16, 2021.
- [Lew10] K. Lewis. Persister cells. *Ann. Rev. Microbiol.*, 64:357–372, 2010.
- [LFL17] K. J. Locey, M. C. Fisk, and J. T. Lennon. Microscale insight into microbial seed banks. *Front. Microbiol.*, 7:2040, 2017.
- [LJ11] J. T. Lennon and S. E. Jones. Microbial seed banks: the ecological and evolutionary implications of dormancy. *Nat. Rev. Microbiol.*, 9:119–130, 2011.
- [LT46] J. Lederberg and E. L. Tatum. Gene recombination in escherichia coli. *Nature*, 158:558, 1946.
- [SL77] F.M. Stewart and B.R. Levin. The population biology of bacterial plasmids: A priori conditions for the existence of conjugationally transmitted factors. *Genetics*, 2(87), 1977.
- [Sma17] C. Smadi. The effect of recurrent mutations on genetic diversity in a large population of varying size. *Acta Applic. Math.*, 149(1):11–15, 2017.