

# A stochastic adaptive dynamics model for bacterial populations with mutation, dormancy and transfer

Jochen Blath, Tobias Paul and András Tóbiás

Goethe-Universität Frankfurt, Robert-Mayer-Straße 10, 60325 Frankfurt am Main, Germany

*E-mail address:* [blath@math.uni-frankfurt.de](mailto:blath@math.uni-frankfurt.de)

*URL:* [https://www.uni-frankfurt.de/117025373/Prof\\_\\_Dr\\_\\_Jochen\\_Blath](https://www.uni-frankfurt.de/117025373/Prof__Dr__Jochen_Blath)

HU Berlin, Rudower Chaussee 25, 12489 Berlin, Germany

*E-mail address:* [t.paul@math.hu-berlin.de](mailto:t.paul@math.hu-berlin.de)

Budapest University of Technology and Economics, Műegyetem rkp. 3., 1111 Budapest and Alfréd Rényi Institute of Mathematics, Reáltanoda utca 13-15., 1053 Budapest, Hungary

*E-mail address:* [tobias@cs.bme.hu](mailto:tobias@cs.bme.hu)

*URL:* <https://cs.bme.hu/tobias/>

**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.

## 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. [Champagnat \(2006\)](#); [Fournier and Méléard \(2004\)](#); [Billiard et al. \(2016, 2018\)](#); [Locey et al.](#)

---

*Received by the editors December 18th, 2021; accepted January 23rd, 2023.*

2010 *Mathematics Subject Classification.* 60J85, 92D25.

*Key words 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.

This work was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany’s Excellence Strategy MATH+: The Berlin Mathematics Research Center, EXC-2046/1 project ID EF 4-7. JB was supported by DFG Priority Programme 1590 “Probabilistic Structures in Evolution”, project BL1105/5-1 ID 285659567. AT was supported by DFG Priority Programme 1590 “Probabilistic Structures in Evolution”.

(2017); Baar and Bovier (2018)). 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 Lederberg and Tatum (1946). 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 conjugation. However, it is known that carrying a large quantity of plasmids slows down cell division and as such reduces the reproduction rate (cf. Baltrus (2013)). 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. Baumdicker and Pfaffelhuber (2014); Koonin and Wolf (2012); Stewart and Levin (1977)).

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 Lennon and Jones (2011); Lennon et al. (2021) 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. Bennett (2008)), dormancy in the form of *persistor cells* can be the cause of chronic infections, since these dormant cells with their vanishing metabolism seem to be protected from antibiotic treatment (Lewis (2010)).

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. (2016, 2018) 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 Champagnat et al. (2021), the approach is transferred and extended into an adaptive dynamics setting with moderately large mutation rates (as previously considered in Durrett and Mayberry (2011), see also Coquille et al. (2021)), 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 [Kaj et al. \(2001\)](#), and become a topic of investigation in coalescence theory (cf. [Blath et al. \(2016, 2015, 2020\)](#)). In ecology, dormancy and seed banks have been investigated for several decades, starting with [Cohen \(1966\)](#), and this lead to a rich (mostly deterministic) theory, see e.g. [Lennon et al. \(2021\)](#) for many further references. Traditional seed bank theory is complemented by quantitative research on phenotypic switches in microbial communities, cf. e.g. [Kussell and Leibler \(2005\)](#). 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 [Blath and Tóbiás \(2020\)](#), 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 [Blath and Tóbiás \(2021\)](#), 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 [Champagnat et al. \(2021\)](#). 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 carrying capacity will dictate the size of the total population in our model. 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 [Champagnat et al. \(2021, Theorem 2.1\)](#) or [Coquille et al. \(2021, Theorem 2.2\)](#). 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 [Blath and Tóbiás \(2020\)](#), 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 Champagnat (2006), 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 Durrett and Mayberry (2011) in a model with clonal interference. In the adaptive dynamics literature, this scaling of mutations occurred before in Smadi (2017); Bovier et al. (2019). From a mathematical point of view, the main novelty of the paper Champagnat et al. (2021) is the systematic study of logistic birth-and-death processes with non-constant immigration, as it was also noted in Coquille et al. (2021).

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 Champagnat et al. (2021). 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 Coquille et al. (2021) in the case of asymmetric competition without HGT. In the model studied in Durrett and Mayberry (2011), 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 Champagnat et al. (2021, 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.9, 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 Blath and Tóbiás (2020), 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 Blath and Tóbiás (2021) 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 Champagnat et al. (2021). 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  $\kappa$  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.

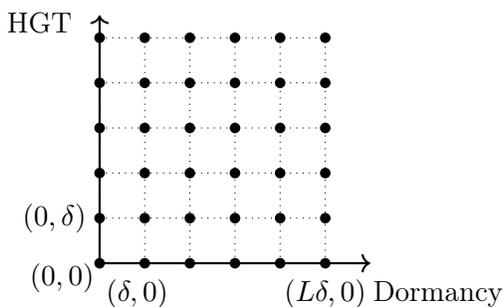


FIGURE 2.1. A visualization of the trait space  $\mathcal{X}$ . The strength of dormancy in a trait increases as the first component increases and the strength of HGT increases with the second component.

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

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

The reason for considering this quantity lies in our auxiliary result for bi-type branching processes in Theorem A.3. 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 rates 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 to obtain our result. We will however also consider cases where we only have a dominant trait. For this, see Section 3.2.

Now, assume that the trait  $(x, y) = (m\delta, n\delta)$  is resident. Then, for large  $K$ , we can approximate the dynamics of  $(N_{m,n}^{K,a}(t), N_{m,n}^{K,d}(t))$  on finite time intervals by  $K(z^a(t), z^d(t))$ , 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 Ethier and Kurtz (1986, Theorem 11.2.1). We want to calculate a stable equilibrium of this system, which has already been done in Blath and Tóbiás (2020, 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

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

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

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

if  $n\alpha < 1$  and  $(m + n)\alpha < 1$  respectively and 0 otherwise. Indeed, this choice is consistent with [Champagnat et al. \(2021, Lemma B.4\)](#), 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

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

**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. Here, we set

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

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.* Our model in principle allows for the possibility of long-term coexistence in the sense that  $\text{sign}(S((\tilde{x}, \tilde{y}), (x, y))) = -\text{sign}(S((x, y), (\tilde{x}, \tilde{y})))$ . 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. We do this to avoid difficulties in the stability of equilibria of approximating dynamical systems as in Appendix B.

Using the above definitions of the invasion fitness, we can state our convergence result, which is similar to [Champagnat et al. \(2021, Theorem 2.1\)](#).

**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 entirely 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 Champagnat et al. (2021, 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 on the resident trait  $(m_{k+1}^* + n_{k+1}^*)\delta < 6$  is needed for the invading trait to have a positive equilibrium size.
- (c) If there is a trait which, at the time the resident trait changes, consists of very few individuals e.g. 1 or 2 individuals, we are not able to determine against which of the two traits of size  $K$  there is competition. Hence, we cannot determine precisely whether this trait will be able to survive without incoming mutations. Therefore, we want to ensure each trait with small population size of order 1 to be 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 Champagnat et al. (2021, Section 3) is not feasible. While we can obtain (eventually) periodic functions as in Example 3.1, we also see possibly non-periodic functions in Example 3.3. Especially this example poses the problem of determining whether it is truly non-periodic or becomes periodic eventually.

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 graphs as in Figure 3.2.

In this case we see that a similar behaviour as in Champagnat et al. (2021) 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.

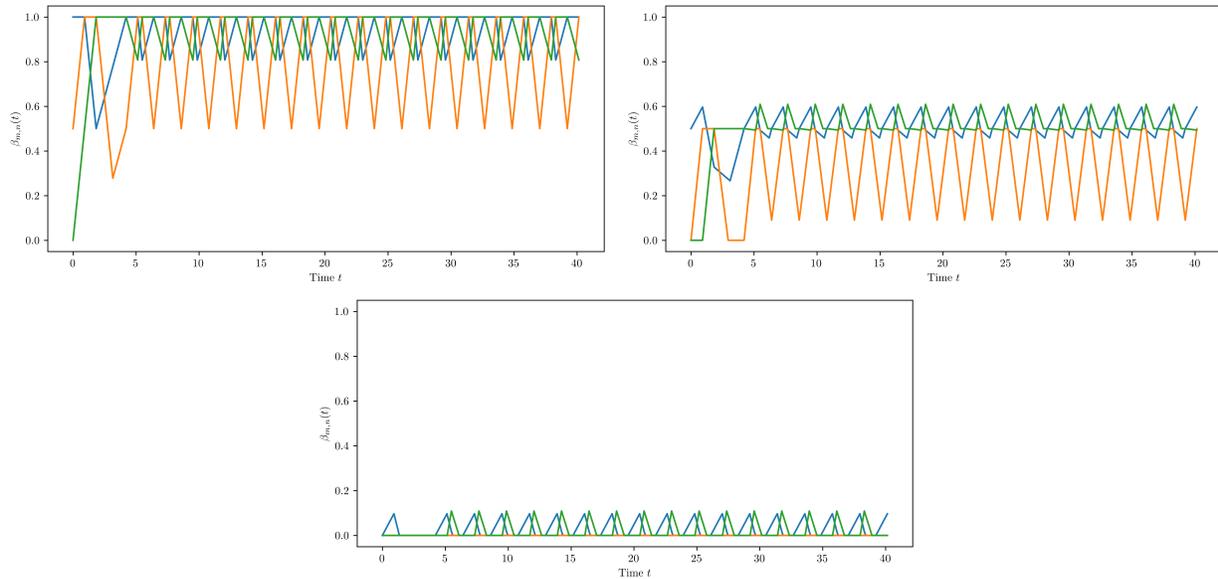


FIGURE 3.2. ( $p = 0.21$ ) Top left: The dynamics of  $\beta_{0,0}$  (blue),  $\beta_{0,1}$  (orange) and  $\beta_{0,2}$  (green). Top right: The dynamics of  $\beta_{1,0}$  (blue),  $\beta_{1,1}$  (orange) and  $\beta_{1,2}$  (green). Bottom: The dynamics of  $\beta_{2,0}$  (blue),  $\beta_{2,1}$  (orange) and  $\beta_{2,2}$  (green).

*Example 3.2* ( $p = 0.22$ ). In this case, the resulting dynamics are given in Figure 3.3.

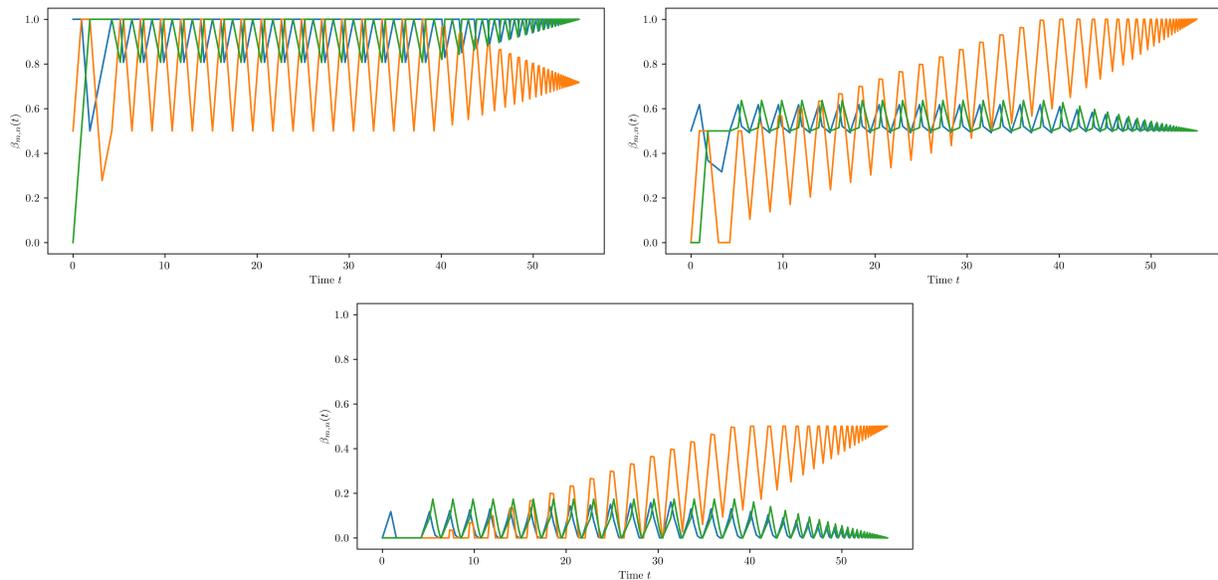


FIGURE 3.3. ( $p = 0.22$ ) Top left: The dynamics of  $\beta_{0,0}$  (blue),  $\beta_{0,1}$  (orange) and  $\beta_{0,2}$  (green). Top right: The dynamics of  $\beta_{1,0}$  (blue),  $\beta_{1,1}$  (orange) and  $\beta_{1,2}$  (green). Bottom: The dynamics of  $\beta_{2,0}$  (blue),  $\beta_{2,1}$  (orange) and  $\beta_{2,2}$  (green).

Here, two phases are to be distinguished: At first, we observe a 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  $(\delta, \delta)$  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 Champagnat et al. (2021) 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  $(0, 0)$ ,  $(\delta, \delta)$  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  $(\delta, \delta)$  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 3.4. We calculate the individual fitnesses as determined by the fitness function. Recall that  $\kappa = 0$  which simplifies the calculations significantly. We obtain

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

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

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

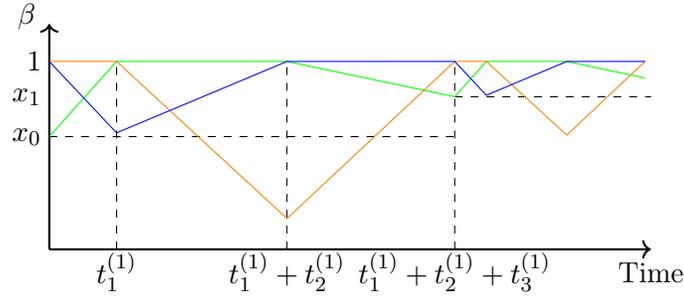


FIGURE 3.4. Illustration for the successive construction of the times  $t_i^{(1)}$ . Blue represents  $\beta_{0,0}$ , orange is  $\beta_{1,1}$  and green is  $\beta_{0,2}$ .

We 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  $(\delta, \delta)$  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

$$\beta_{0,0}(t_1^{(1)} + t_2^{(1)} + t_3^{(1)}) = 1, \quad \beta_{1,1}(t_1^{(1)} + t_2^{(1)} + t_3^{(1)}) = 1$$

and

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

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.

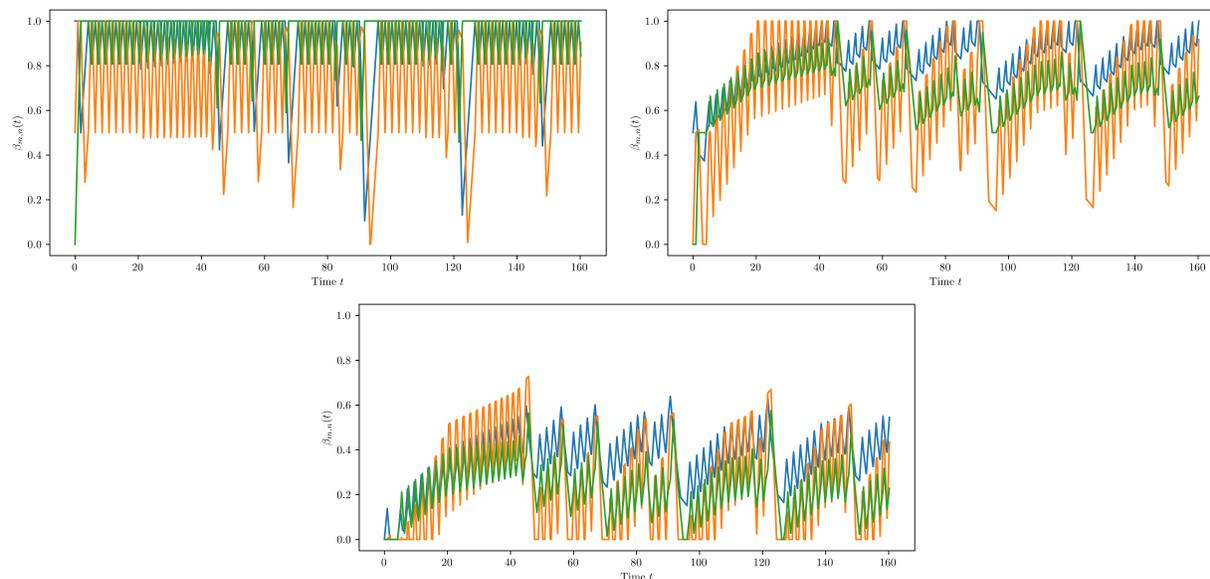


FIGURE 3.5. ( $p = 0.23$ ) Top left: The dynamics of  $\beta_{0,0}$  (blue),  $\beta_{0,1}$  (orange) and  $\beta_{0,2}$  (green). Top right: The dynamics of  $\beta_{1,0}$  (blue),  $\beta_{1,1}$  (orange) and  $\beta_{1,2}$  (green). Bottom: The dynamics of  $\beta_{2,0}$  (blue),  $\beta_{2,1}$  (orange) and  $\beta_{2,2}$  (green).

*Example 3.3* ( $p = 0.23$ ). We obtain the functions in Figure 3.5

Here, there are even more phases to distinguish: At first we have the growth phase of the trait  $(\delta, \delta)$  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  $(\delta, \delta)$ , which in turn implies that while  $(\delta, \delta)$  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 as displayed in Figure 3.6. 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  $(\delta, \delta)$ , 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.

*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  $(\delta, \ell\delta)$ . We show the functions in Figure 3.7.

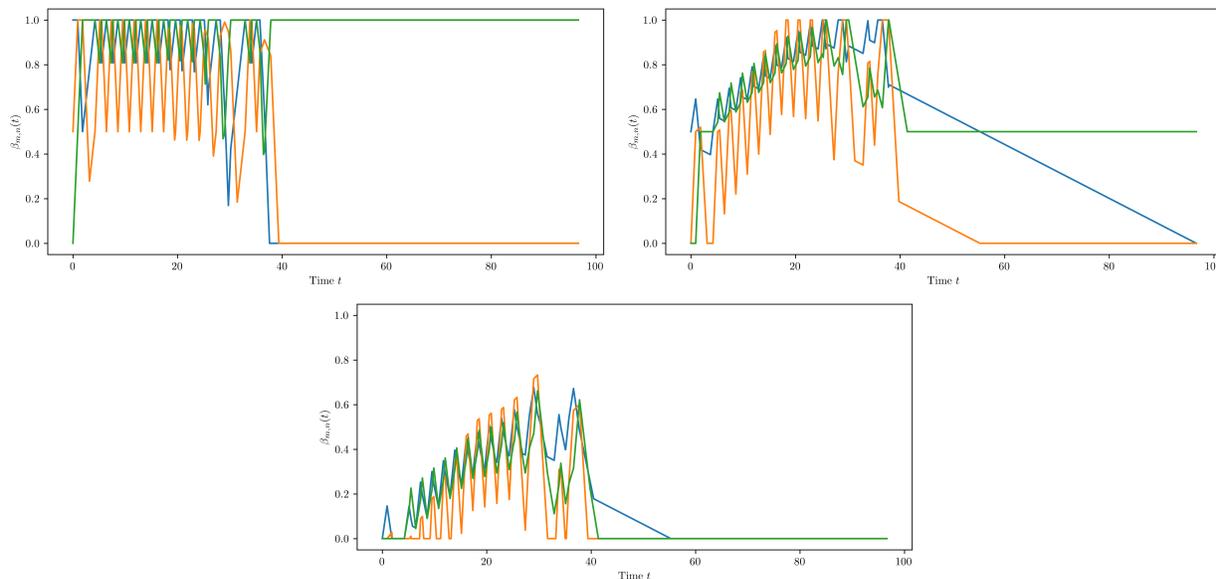


FIGURE 3.6. ( $p = 0.234$ ) Top left: The dynamics of  $\beta_{0,0}$  (blue),  $\beta_{0,1}$  (orange) and  $\beta_{0,2}$  (green). Top right: The dynamics of  $\beta_{1,0}$  (blue),  $\beta_{1,1}$  (orange) and  $\beta_{1,2}$  (green). Bottom: The dynamics of  $\beta_{2,0}$  (blue),  $\beta_{2,1}$  (orange) and  $\beta_{2,2}$  (green).

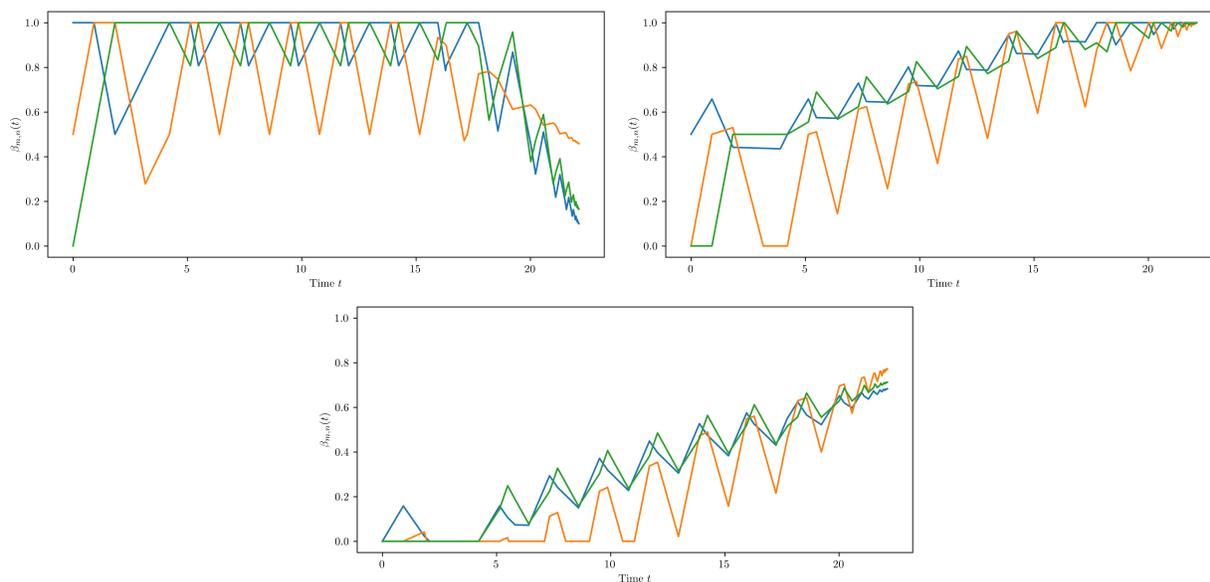


FIGURE 3.7. ( $p = 0.24$ ) Top left: The dynamics of  $\beta_{0,0}$  (blue),  $\beta_{0,1}$  (orange) and  $\beta_{0,2}$  (green). Top right: The dynamics of  $\beta_{1,0}$  (blue),  $\beta_{1,1}$  (orange) and  $\beta_{1,2}$  (green). Bottom: The dynamics of  $\beta_{2,0}$  (blue),  $\beta_{2,1}$  (orange) and  $\beta_{2,2}$  (green).

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. 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  $\alpha = 0.5$ .

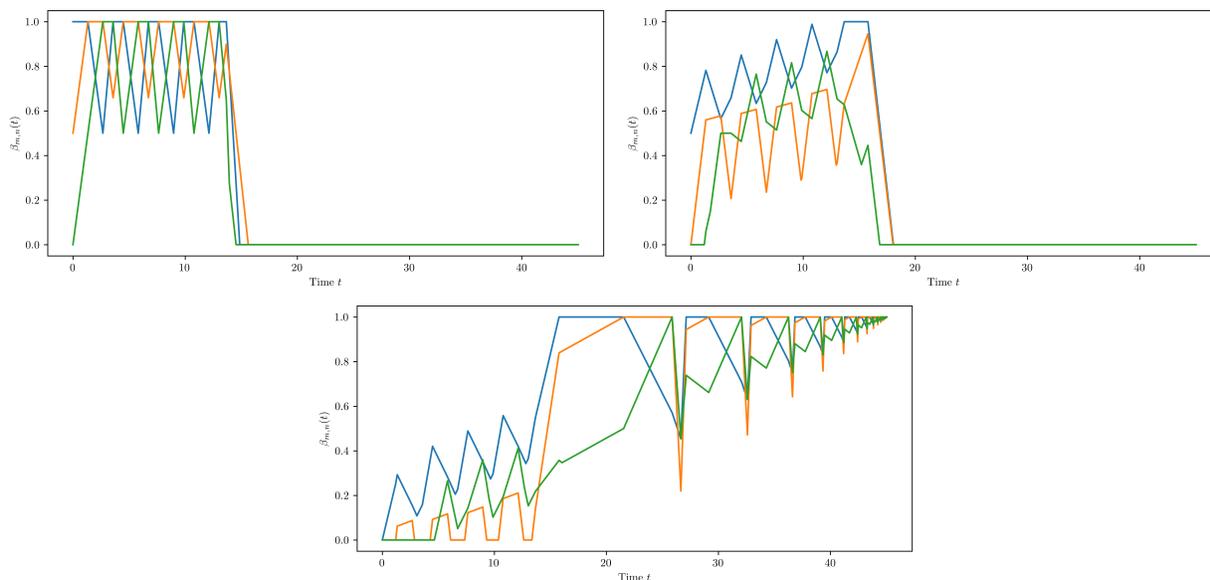


FIGURE 3.8. Top left: The dynamics of  $\beta_{0,0}$  (blue),  $\beta_{0,1}$  (orange) and  $\beta_{0,2}$  (green). Top right: The dynamics of  $\beta_{1,0}$  (blue),  $\beta_{1,1}$  (orange) and  $\beta_{1,2}$  (green). Bottom: The dynamics of  $\beta_{2,0}$  (blue),  $\beta_{2,1}$  (orange) and  $\beta_{2,2}$  (green).

Interestingly, as can be seen in Figure 3.8 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  $\delta$ , then a similar but simultaneously new behaviour emerges (Figure 3.9). It is similar in the sense that we again have convergence in a finite time horizon, but the limiting exponent of the population size for all traits is less than 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.

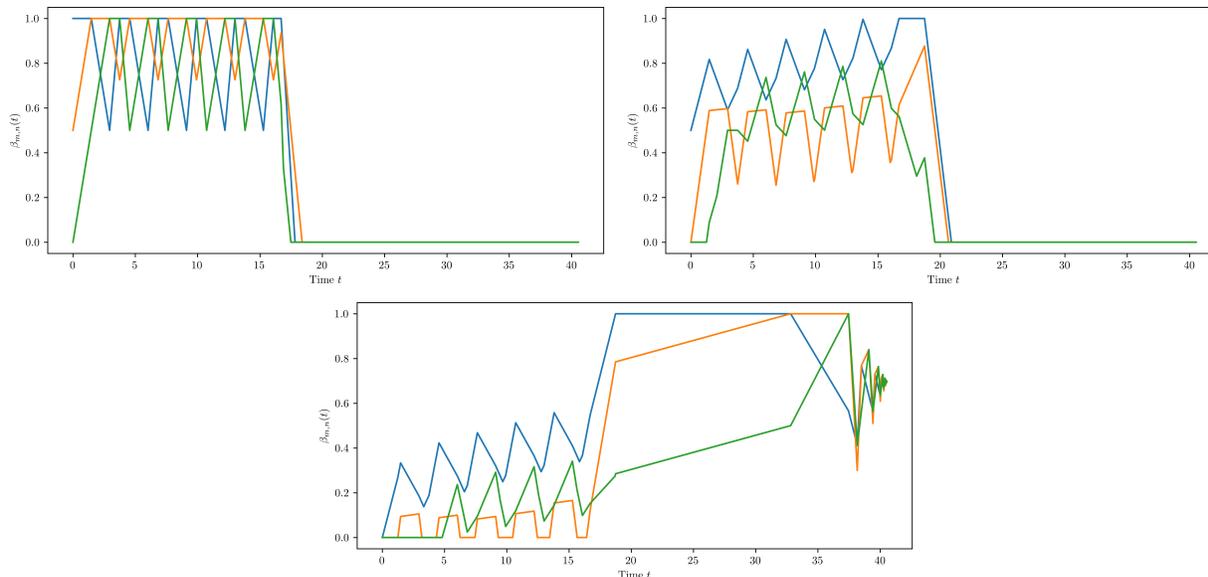


FIGURE 3.9. Top left: The dynamics of  $\beta_{0,0}$  (blue),  $\beta_{0,1}$  (orange) and  $\beta_{0,2}$  (green). Top right: The dynamics of  $\beta_{1,0}$  (blue),  $\beta_{1,1}$  (orange) and  $\beta_{1,2}$  (green). Bottom: The dynamics of  $\beta_{2,0}$  (blue),  $\beta_{2,1}$  (orange) and  $\beta_{2,2}$  (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  $\sigma$  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 value 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.

From our simulations with  $K = 10^5$  in Figure 3.10 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 3.10 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 mutation

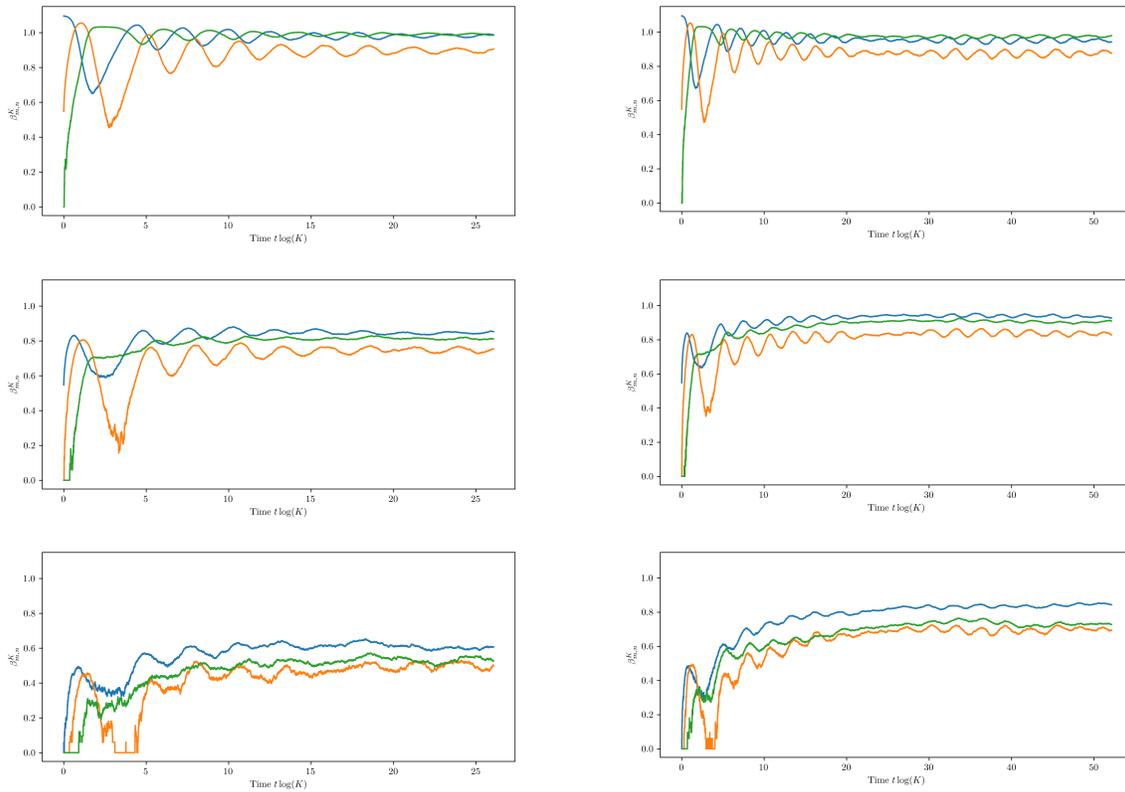


FIGURE 3.10. Left: Simulations with the parameters as in Example 3.1 and  $C = 1$ ,  $K = 10^5$ . Right: Simulations with parameters as in Example 3.2 and  $C = 1$ ,  $K = 10^5$ . From top to bottom we are increasing the index  $m$  of  $\beta_{m,n}^K$  by 1 and in each plot  $n = 0$  is blue,  $n = 1$  is orange and  $n = 2$  is green.

can be approximated by the solution of the differential equation

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

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 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  $\Delta t$  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  $\Delta t$  is too small compared with  $1/K$ , then it may happen that the immigration during a time step of length  $\Delta t$  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.

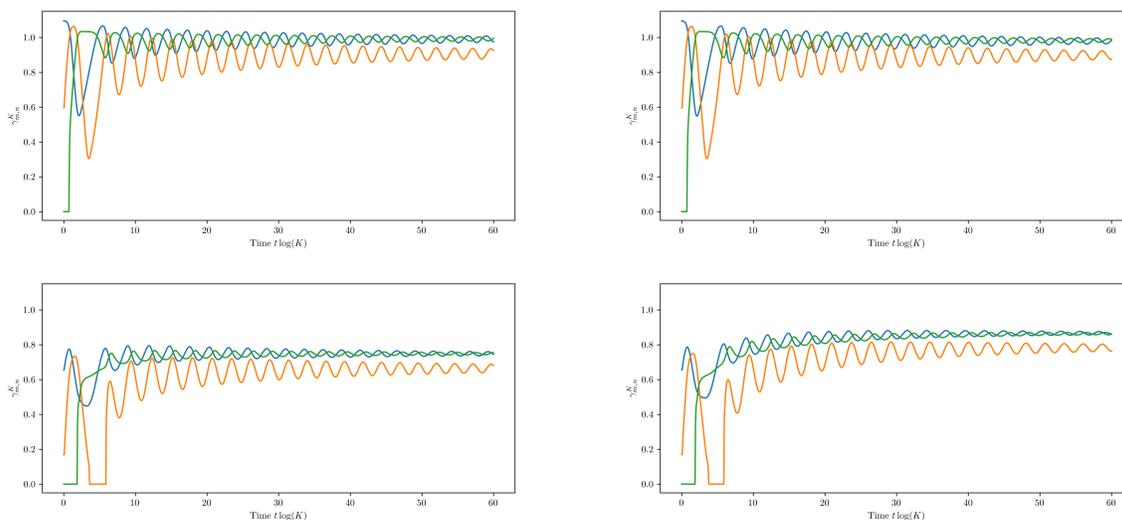


FIGURE 3.11. Left: Solving the ODE with Euler scheme,  $K = 10^5$  and  $\Delta t = T \log(K)K^{-1}$  with parameters as in Example 3.1 and  $C = 1$ . Right: Solving the ODE with Euler scheme,  $K = 10^5$  and  $\Delta t = T \log(K)K^{-1}$  with parameters as in Example 3.2 and  $C = 1$ . From top to 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.

Comparing 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 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 3.12 with  $K = 10^{15}$ , the limiting functions are a much better approximation of the exponents, although  $\Delta t$  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 usual cycle of residency between  $(0, 0)$ ,  $(\delta, \delta)$  and  $(0, 2\delta)$  in which these three traits change being resident in a periodic manner. 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  $(\delta, \delta)$  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

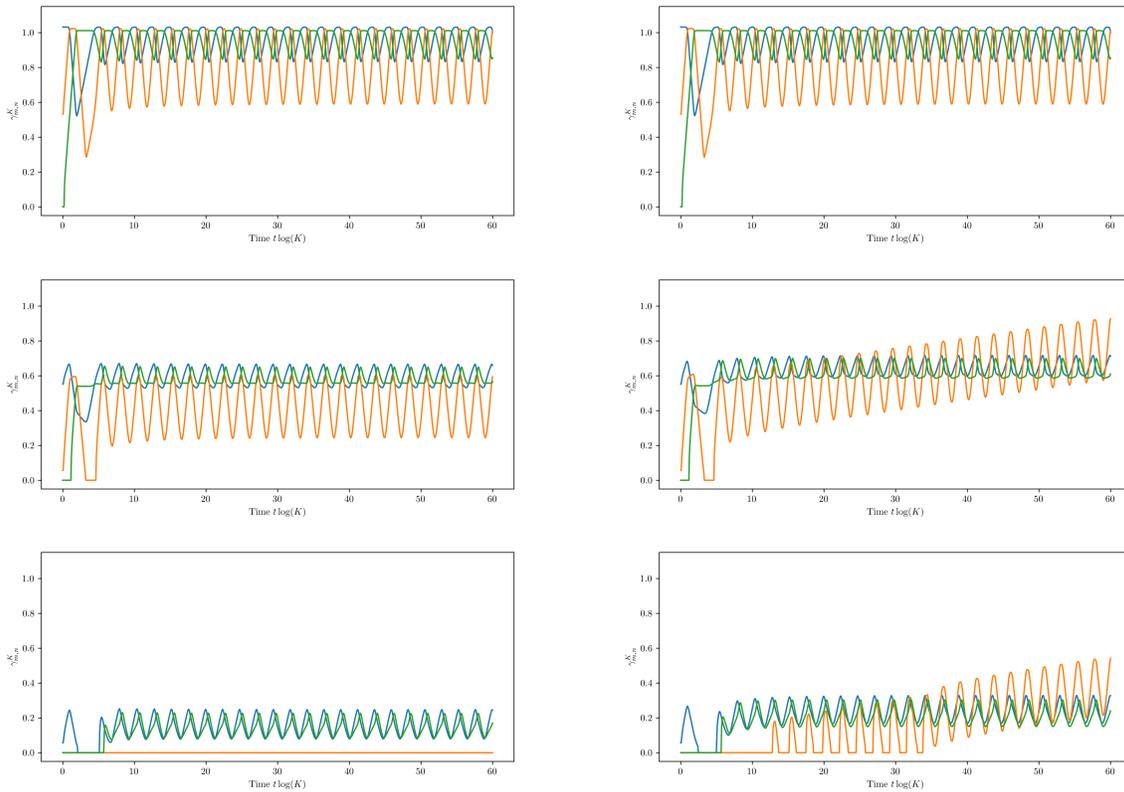


FIGURE 3.12. Left: Solving the ODE with Euler scheme,  $K = 10^{15}$  and  $\Delta t = T \log(K)10^{-5}$  with parameters as in Example 3.1 and  $C = 1$ . Right: Solving the ODE with Euler scheme,  $K = 10^{15}$  and  $\Delta t = T \log(K)10^{-5}$  with parameters as in Example 3.2 and  $C = 1$ . From top to bottom we have the usual arrangement of the Exponents  $\gamma_{m,n}^K$ .

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.

#### 4. Proof of Theorem 2.2

We will give a short sketch of the proof, which is similar to Champagnat et al. (2021, 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 Champagnat et al. (2021). 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  $\delta$ -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 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  $(m\delta, 0)$ .

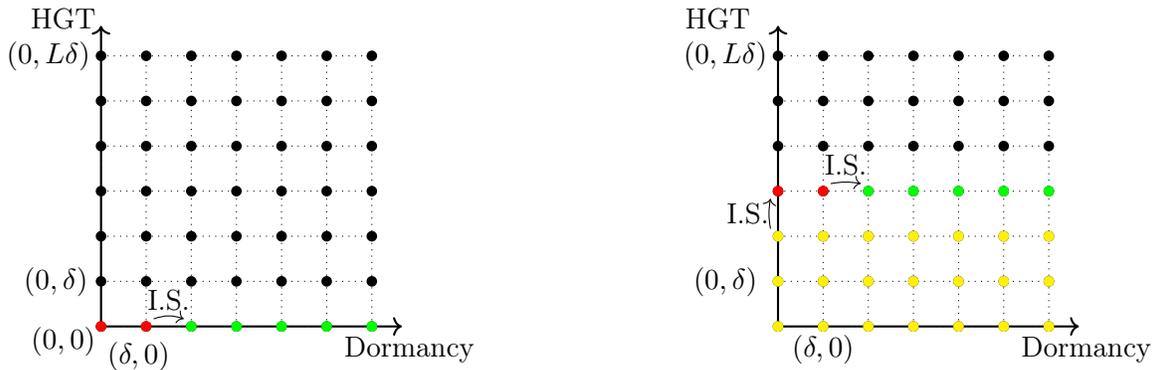


FIGURE 4.13. The schematic induction: In the left picture, red are base cases, and the convergence for the remaining green traits are proven via the induction step. Then in the right picture, we assume the result to be shown for all yellow traits. We have the new base cases of the nested induction in red and subsequently another induction step again in green.

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 Champagnat et al. (2021, Appendix A,B) for results concerning these processes. With the notation  $BBPI_K(b_1, b_2, d_1, d_2, \sigma_1, \sigma_2, a, c, \beta, \gamma)$  we denote a two-dimensional branching process with birth rates  $b_1, b_2$ , death rates  $d_1, d_2$ , switching rates  $\sigma_1, \sigma_2$ , 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 Champagnat et al. (2021, Appendix C). Also  $LBBI_K(b_1, d_1, d_2, \sigma_2, p, C, \gamma_1)$  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.

*Proof of Theorem 2.2:* We will only discuss the idea for the induction step over  $k$ . The complete proof can be found in [Blath et al. \(2021\)](#).

4.1. *Step k.* We will 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  $\sigma_k^K$  with the convergence property mentioned above such that for the resident population of trait  $(m_k^* \delta, n_k^* \delta)$  the bounds

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

and

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

hold for some  $\varepsilon_k > 0$  sufficiently small. 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 that we have shown  $N_{m,n}^K(\sigma_k^K \log K) = 0$  if  $\beta_{m,n}(s_{k-1}) = 0$  and otherwise we assume to have shown

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

in our induction to this step.

We introduce the time  $\theta_k^K$ , 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

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

*Step 0: Deriving bounds on the rates.* We can derive bounds on the birth, death and migration rates on the time interval  $[\sigma_k^K, \theta_k^K]$ . The bounds for the birth and immigration 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 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.* 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 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 suitable constant  $a_{m,n} \in \mathbb{R}$ . We will not fully carry out the induction over the traits, but give a broad idea for the case  $(m_k^*, n_k^*) \neq (0, 0)$ . Assume  $\beta_{0,0}(s_{k-1}) > 0$ . Since there is no incoming immigration into the trait  $(0, 0)$ , 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 [Champagnat et al. \(2021, Lemma A.1\)](#) 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 observing 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

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

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  $BBPI_K(4 - \frac{(m+n)\delta}{2} + \tau \mathbb{1}_{n > n_k^*} \mp C_*\varepsilon_k, 0, 1 + (1 - pm\delta) \bar{z}_{m_k^*, n_k^*}^a + \tau \mathbb{1}_{n < n_k^*} \pm 2C_*\varepsilon_k, \kappa, pm\delta \mp C_*\varepsilon_k, \sigma, a_{m,n-1}, \beta_{m,n-1}(s_{k-1}) - \alpha \mp C_*\varepsilon_k, (\beta_{m,n}(s_{k-1}) \mp C_*\varepsilon_k)_\times)$ . Here, we use the notation

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

Applying Theorem [A.1](#) or [A.3](#) accordingly shows the bounds (4.1) by definition of our fitness function. Continuing this process for all time intervals on which all  $\beta_{m,n}$  are affine shows the bounds (4.1) 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  $\theta_k^K$ .* We want to show that  $(s_k - \eta) \wedge T < \theta_k^K$  with probability converging to 1. 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 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  $\theta_k^K$ . Again we need to distinguish between the possibility of becoming dormant or not. If  $m_k^* = 0$ , we can choose  $Z_{m_k^*, n_k^*, 1}$  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  $Z_{m_k^*, n_k^*, 2}$  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. Without loss of generality, we assume the immigration to be dominated by  $N_{m_k^*, n_k^* - 1}$ . Then we can choose the process  $Z_{m_k^*, n_k^*, 1}$  to be determined by a  $LBDI_K((4 - \frac{(m_k^* + n_k^*)\delta}{2})(1 - \varepsilon), 1 + C\varepsilon, \kappa, \sigma, p, C, K^{-\alpha}N_{m_k^*, n_k^* - 1})$  and  $Z_{m_k^*, n_k^*, 2}$  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 [Champagnat et al. \(2021, Lemma C.1\)](#) to the first case and [Corollary B.5](#) 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. Intermediate Phase $k$ .

*Step 1: Convergence of  $\theta_k^K \rightarrow s_k$ .* This part of the proof can be taken from [Champagnat et al. \(2021\)](#) and is not repeated here.

*Step 2: Emergence of a new population.* This part is also similar. However, we may need to couple with logistic bi-type branching processes instead of single type. We obtain that

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

for some suitable  $\rho > 0$  sufficiently small 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 define corresponding birth, death, migration, switching and horizontal transfer rates for suitable couplings which then allow us to apply one of the Propositions from [B.10](#), [B.12](#), [B.13](#), [B.15](#), [B.16](#) and [B.17](#) in conjunction with [Remark B.11](#) or [Champagnat et al. \(2021, Lemma C.3\)](#), which in each case gives 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

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

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

and

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

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

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

if  $\beta_{m,n}(s_k) > 0$  and  $N_{m,n}^K(\sigma_{k+1}^K \log K) = 0$  otherwise. 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 [Champagnat et al. \(2021\)](#).

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  $\frac{\log(1+Z_t^K)}{\log K}$  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 [Champagnat et al. \(2021\)](#).  $\square$

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 also define  $r_1 := b_1 - d_1 - \sigma_1$  and  $r_2 := b_2 - d_2 - \sigma_2$ .

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 [Champagnat et al. \(2021\)](#). 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*

$$\lambda := \frac{r_1 + r_2 + \sqrt{(r_1 - r_2)^2 + 4p\sigma_1\sigma_2}}{2}.$$

Then for all  $T \geq 0$  the process

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

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  $Y_t^K$ .

**Lemma A.4.** 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. □

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.5.** The expected value  $(x_t^K, y_t^K)$  of  $(X_t^K, Y_t^K)$  solves the ordinary differential equation

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

*Proof:* This is a direct consequence of Lemma A.4, 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). □

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

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

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

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

where  $\Delta = \sqrt{(r_1 - r_2)^2 + 4\sigma_1\sigma_2} \neq 0$  and hence  $\lambda > \tilde{\lambda}$ . Note that this is exactly the definition we already introduced in Theorem A.3. In particular we are now able to give a characterization of the expected values for  $X_t^K$  and  $Y_t^K$ .

**Lemma A.6.** 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

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

An explicit computation shows the claim. □

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.7.** *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*

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

*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

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

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

$$\begin{aligned} \overline{M}_t^K &= (X_t^K)^2 - (X_0^K)^2 - \int_0^t (2X_s^K + 1)(b_1 X_s^K + \sigma_2 Y_s^K + K^c e^{as}) + (\sigma_1 + d_1)(-2X_s^K + 1) X_s^K ds, \\ \underline{M}_t^K &= (X_t^K)^2 - (X_0^K)^2 - \int_0^t 2X_s^K ((b_1 - d_1 - \sigma_1) X_s^K + \sigma_2 Y_s^K + K^c e^{as}) ds - [M^K]_t, \end{aligned}$$

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

$$[M^K]_t = \int_0^t (b_1 + d_1 + \sigma_1) X_s^K + \sigma_2 Y_s^K + K^c e^{as} ds.$$

□

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

**Lemma A.8.** *There exists a constant  $C_* \geq 0$  independent of  $K$  such that the variances  $\mathbb{V}(X_t^K)$ ,  $\mathbb{V}(Y_t^K)$  satisfy*

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

*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.7 to  $(X_t^K - x_t^K)^2$  and  $(Y_t^K - y_t^K)^2$  as well as using Integration by Parts for the product  $(X_t^K - x_t^K)(Y_t^K - y_t^K)$  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.6. 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.6 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. □

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.9.** *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  $\lambda$  from (A.3). Then the following convergence in probability holds in  $L^\infty([0, T])$ :

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

*Remark A.10.* Note that due to the strictly positive switching rates, the same convergence also holds for the processes  $\left\{s \mapsto \frac{\log(1 + X_{s \log K}^K)}{\log K}\right\}$  and  $\left\{s \mapsto \frac{\log(1 + Y_{s \log K}^K)}{\log K}\right\}$  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.9 is possible.

*Proof:* For the proof we make extensive use of ideas from Champagnat et al. (2021, 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  $\eta$  such that  $\mathbb{P}(\Omega_1^K) \rightarrow 1$  as  $K \rightarrow \infty$ . For this, in Champagnat et al. (2021, Lemma B.3) it is shown, that the process of which the absolute value is taken in  $\Omega_1^K$  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.4 and A.5 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

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

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.7 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.6, 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

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

We now turn to the integral. Using Lemma A.8 and A.6 again, as well as  $c \leq \beta \vee \gamma$ , we see that

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

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  $\eta$  such that

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

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

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

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

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.6 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 Champagnat et al. (2021) with  $\beta \vee \gamma$  instead of  $\beta$  and  $\lambda$  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  $\eta$  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  $\eta$  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  $\eta$  such that

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

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  $\eta$  we have

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

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  $\eta$  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 Champagnat et al. (2021) 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  $\Omega_3^K$  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  $\widehat{Z}_t^K = (\widehat{X}_t^K, \widehat{Y}_t^K)$  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

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

Indeed, the starting conditions of the bounding processes are justified by Remark A.10. Then, we can apply the convergence from Step 2 to  $\widehat{Z}_t^K$  and  $\bar{Z}_t^K$  to show that

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

and

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

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  $\varepsilon$  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.  $\square$

### A.3. Proof of Theorem A.3.

*Proof of Theorem A.3:* The proof can now be taken from the proof of Champagnat et al. (2021, Theorem B.5), where the remaining auxiliary results are easily extended for bitype processes. The only significant challenge is found in the extension of the weaker version which is our Theorem A.9.  $\square$

## 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*

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

*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*

$$\begin{aligned} \dot{x}(t) &= (b_1 - d_1)x(t) - Cx^2(t) + \sigma_2y(t) \\ \dot{y}(t) &= -(d_2 + \sigma_2)y(t) + pCx^2(t) \end{aligned} \tag{B.2}$$

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

*Proof:* This is similar to Ethier and Kurtz (1986, Theorem 11.2.1).  $\square$

**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

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

Thus, the assumption  $b_1 > d_1$  is sufficient for obtaining a coordinatewise positive equilibrium. As is shown in Blath and Tóbiás (2020, 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 Blath and Tóbiás (2020).

**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.3 for the entry into a domain around the equilibrium and Corollary B.5 for the exit of such a domain. Both results are straightforward extensions of the results of Champagnat (2006). For detailed proofs, we refer to Blath et al. (2021).

**Lemma B.3.** *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).

**Lemma B.4.** *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

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

A 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.5.** *Under the assumptions of Lemma B.4 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 4-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.10 and B.12, 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

$$(i, j, k, \ell) \rightarrow \begin{cases} (i+1, j, k, \ell) & \text{at rate } ia_1^K(\omega, t) + \gamma_1^K(\omega, t) \\ (i, j, k+1, \ell) & \text{at rate } kb_1^K(\omega, t) + \gamma_2^K(\omega, t) \\ (i-1, j, k, \ell) & \text{at rate } i(d_1^K(\omega, t) + \frac{(1-p)C}{K}(i+k)) \\ (i, j, k-1, \ell) & \text{at rate } k(d_1^K(\omega, t) + \frac{(1-q)C}{K}(i+k)) \\ (i, j-1, k, \ell) & \text{at rate } jd_2^K(\omega, t) \\ (i, j, k, \ell-1) & \text{at rate } ld_2^K(\omega, t) \\ (i-1, j+1, k, \ell) & \text{at rate } i\frac{pC}{K}(i+k) \\ (i, j, k-1, \ell+1) & \text{at rate } k\frac{qC}{K}(i+k) \\ (i+1, j-1, k, \ell) & \text{at rate } j\sigma_2 \\ (i, j, k+1, \ell-1) & \text{at rate } \ell\sigma_2 \\ (i-1, j, k+1, \ell) & \text{at rate } \tau^K(\omega, t)\frac{ik}{i+k}, \end{cases}$$

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

$$\begin{aligned} \dot{x}_a &= x_a(a_1 - d_1 - C(x_a + y_a)) + x_d\sigma_2 - \tau\frac{x_a y_a}{x_a + y_a} \\ \dot{x}_d &= pCx_a(x_a + y_a) - (d_2 + \sigma_2)x_d \\ \dot{y}_a &= y_a(b_1 - d_1 - C(x_a + y_a)) + y_d\sigma_2 + \tau\frac{x_a y_a}{x_a + y_a} \\ \dot{y}_d &= qCy_a(x_a + y_a) - (d_2 + \sigma_2)y_d. \end{aligned} \tag{B.4}$$

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

$$\bar{x}_a = \frac{(a_1 - d_1)(d_2 + \sigma_2)}{C(d_2 + (1-p)\sigma_2)} \quad \text{and} \quad \bar{x}_d = \frac{p(a_1 - d_1)^2(d_2 + \sigma_2)}{C(d_2 + (1-p)\sigma_2)^2} \tag{B.5}$$

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

$$J = \begin{pmatrix} b_1 + \tau - d_1 - C\bar{x}_a & qC\bar{x}_a \\ \sigma_2 & -d_2 - \sigma_2 \end{pmatrix} \tag{B.6}$$

of the approximating process  $(\widehat{Y}_a, \widehat{Y}_d)$  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  $\lambda$ , 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

$$-(b_1 + \tau - d_1 - C\bar{x}_a) < \frac{\sigma_2 q C \bar{x}_a}{d_2 + \sigma_2}. \quad (\text{B.7})$$

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  $(\widehat{X}_a, \widehat{X}_d)$  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

$$\tilde{J} = \begin{pmatrix} a_1 - \tau - d_1 - C\bar{y}_a & pC\bar{y}_a \\ \sigma_2 & -d_2 - \sigma_2 \end{pmatrix} \quad (\text{B.8})$$

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

$$-(a_1 - \tau - d_1 - C\bar{y}_a) > \frac{\sigma_2 p C \bar{y}_a}{d_2 + \sigma_2}. \quad (\text{B.9})$$

Our first result is concerned with finding the equilibria of the dynamical system (B.4).

**Lemma B.6.** *Consider the system (B.4) 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.5).

*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.4) yields

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

Hence, the first line in (B.4) gives

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

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

$$\begin{aligned} 0 &= a_1 - d_1 - C(x_a + y_a) + \frac{pC(x_a + y_a)\sigma_2}{d_2 + \sigma_2} - \tau \frac{y_a}{x_a + y_a} \\ \iff x_a + y_a &= \bar{x}_a - \tau \frac{y_a(d_2 + \sigma_2)}{C(x_a + y_a)(d_2 + (1-p)\sigma_2)}. \end{aligned} \quad (\text{B.10})$$

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

$$x_a + y_a = \bar{y}_a + \tau \frac{x_a(d_2 + \sigma_2)}{C(x_a + y_a)(d_2 + (1-q)\sigma_2)}. \quad (\text{B.11})$$

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.10) 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.7), and thus there is no such equilibrium. Assuming  $\tau > 0$  and (B.7) 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.11), 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.9) 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.7) 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. Blath and Tóbiás (2020, 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.4). Indeed, the following lemmata give a first step towards this direction.

**Lemma B.7.** *Consider the system (B.4) and assume that the matrix  $J$  from (B.6) has a positive eigenvalue and the matrix  $\tilde{J}$  from (B.8) 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*

$$\frac{qC(x_a + y_a)}{d_2 + \sigma_2} > \frac{y_d}{y_a} > \frac{d_1 - b_1 + C(x_a + y_a) - \tau \frac{x_a}{x_a + y_a}}{\sigma_2}, \tag{B.12}$$

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 Blath and Tóbiás (2020, Lemma 4.7).  $\square$

Now, we are interested in finding a suitable condition such that the inequality (B.12) 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 Georgii and Baake (2003, 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  $(\pi_a, \pi_d)$  is the unique left eigenvector of the matrix  $J$  from (B.6) for the principal eigenvalue  $\lambda$  with  $\pi_a + \pi_d = 1$ .

**Lemma B.8.** *Suppose that the initial condition  $(x_a, x_d, y_a, y_d)$  of the dynamical system (B.4) 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  $\varepsilon$  sufficiently small,  $(x_a, x_d, y_a, y_d)$  satisfies (B.12).*

*Proof:* The result follows from the proof of Blath and Tóbiás (2020, Lemma 4.8).  $\square$

Finally, we are able to show that with high probability as  $\varepsilon \downarrow 0$ , the assumptions of Lemma B.8 are satisfied.

**Lemma B.9.** *Assume that the matrix  $J$  from (B.6) 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  $(m_1^K, m_2^K) \rightarrow (\bar{x}_a, \bar{x}_d)$  as  $K \rightarrow \infty$ . 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 Blath and Tóbiás (2020, Proposition 4.4.) or Coron et al. (2021, Proposition 3.2).  $\square$

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  $s > 0$

$$\begin{aligned} \sup_{0 \leq t \leq s \log K} & |a_1^K(t) - a_1| + |b_1^K(t) - b_1| + |d_1^K(t) - d_1| \\ & + |d_2^K(t) - d_2| + |\tau^K(t) - \tau| + \left| \frac{\gamma_1^K(t)}{K} \right| + \left| \frac{\gamma_2^K(t)}{K} \right| \xrightarrow{K \rightarrow \infty} 0 \end{aligned} \tag{B.13}$$

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

**Proposition B.10.** *Assume that the conditions (B.7), (B.9) and (B.13) 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.11.* 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.10 such that the claim holds for all  $Y_{a,0}^K + Y_{d,0}^K \in [\frac{Km\varepsilon}{2}, Km\varepsilon]$ , where the only change is in the application of Lemma B.9, 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.3, this coupling can be achieved by subtracting and adding  $\delta$  to the birth and death rates for  $K$  large enough, which we can do by the convergence criterion (B.13) above. For now, we do not alter the horizontal transfer rate  $\tau^K$ . Note that indeed we are allowed to neglect the immigration rate for the same reason as in Lemma B.3. For  $\delta > 0$

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

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

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 B.1. Hence it suffices to show the existence of such a finite time as in the proposition for these coupled processes. Now, arguments analogous to Champagnat (2006, Theorem 3) show that for any initial condition 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.9 implies that with high probability in  $m\varepsilon$  the criterion for a good initial condition (B.12) 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  $T_\varepsilon^K$  by 0 due to our choice of the starting condition. Hence, by Lemma B.7 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. While there are differences in the details, we leave the proof to the reader and refer to Blath et al. (2021). 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.12.** *Assume that the reverse inequalities of (B.7) and (B.9) are true that is*

$$-(b_1 + \tau - d_1 - C\bar{x}_a) > \frac{\sigma_2 q C \bar{x}_a}{d_2 + \sigma_2} \quad \text{and} \quad -(a_1 - \tau - d_1 - C\bar{y}_a) > \frac{\sigma_2 p C \bar{y}_a}{d_2 + \sigma_2}, \tag{B.14}$$

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.13). 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

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

**B.3. Competition Between Bi-Type and Single-Type Processes with Transfer.** We consider a three-dimensional branching process 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

$$\sup_{0 \leq t \leq s \log K} |a_1^K(t) - a_1| + |b_1^K(t) - b_1| + |d_1^K(t) - d_1| + |d_2^K(t) - d_2| + |\tau^K(t) - \tau| + \left| \frac{\gamma_1^K(t)}{K} \right| + \left| \frac{\gamma_2^K(t)}{K} \right| \xrightarrow{K \rightarrow \infty} 0 \tag{B.15}$$

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

$$b_1 + \tau - d_1 - C\bar{x}_a > 0, \tag{B.16}$$

which also coincides with (B.7) 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.9), that is

$$-(a_1 - \tau - d_1 - C\bar{y}) > \frac{\sigma_2 p C \bar{y}}{d_2 + \sigma_2}. \tag{B.17}$$

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  $[\frac{m\epsilon K}{2}, m\epsilon K]$ , can be proven as indicated in Remark B.11.

**Proposition B.13.** *Assume that conditions (B.15), (B.16) and (B.17) 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 Blath and Tóbiás (2021, 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

$$\begin{aligned} \dot{x}_a &= x_a(a_1 - d_1 - C(x_a + y)) + x_d\sigma_2 - \tau \frac{x_a y}{x_a + y} \\ \dot{x}_d &= pCx_a(x_a + y) - (d_2 + \sigma_2)x_d \\ \dot{y} &= y(b_1 - d_1 - C(x_a + y)) + \tau \frac{x_a y}{x_a + y} \end{aligned} \tag{B.18}$$

converges towards the equilibrium  $(0, 0, \bar{y})$  for any starting condition from  $\mathcal{A}_\varepsilon^3$ .

**Lemma B.14.** *Consider the dynamical system (B.18). If the initial condition  $(x_a(0), x_d(0), y)$  is contained in the set  $\mathcal{A}_\varepsilon^3$  and the inequalities (B.16) and (B.17) are satisfied, then*

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

*Proof:* This proof can be adapted from the proof of Blath and Tóbiás (2021, Proposition 5.4) and is left to the reader. We refer to Blath et al. (2021). □

*Proof of Proposition B.13:* 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.10 by increasing or decreasing the transfer rate  $\tau^K$  by  $\delta$ . For  $\varepsilon > 0$  sufficiently small, the initial condition from Lemma B.14 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 Champagnat (2006, Theorem 3), the processes converge to the solutions of the corresponding differential equations. In particular, as in Proposition B.10 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.15.** *Assume that (B.15) and the inverse inequalities of (B.16) and (B.17) are true, that is*

$$b_1 + \tau - d_1 - C\bar{x}_a < 0 \quad \text{and} \quad -(a_1 - \tau - d_1 - C\bar{y}) < \frac{\sigma_2 p C \bar{y}}{d_2 + \sigma_2}. \tag{B.19}$$

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.15:* 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.12, 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.12, 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  $\tau$  in the rates from the birth to the death rate and vice versa. In particular, we want the inequalities

$$b_1 - \tau - d_1 - C\bar{x}_a > 0 \quad \text{and} \quad -(a_1 + \tau - d_1 - C\bar{y}) > \frac{\sigma_2 p C \bar{y}}{d_2 + \sigma_2} \tag{B.20}$$

to hold. Then we obtain the same results as above.

**Proposition B.16.** Assume that (B.15) and (B.20) 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

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

*Proof:* The idea for proving Proposition B.16 is similar to the one of Proposition B.13 and left to the reader. We refer to Blath et al. (2021).  $\square$

**Proposition B.17.** Assume that (B.15) and the inverse inequalities of (B.20) are true, that is

$$b_1 - \tau - d_1 - C\bar{x}_a < 0 \quad \text{and} \quad -(a_1 + \tau - d_1 - C\bar{y}) < \frac{\sigma_2 p C \bar{y}}{d_2 + \sigma_2}. \tag{B.21}$$

Consider the process  $(X_a^K, X_d^K, Y^K)$  described above before Proposition B.16 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.17:* 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.10, 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

The authors would like to thank an anonymous referee for insightful comments.

## References

- Baar, M. and Bovier, A. The polymorphic evolution sequence for populations with phenotypic plasticity. *Electron. J. Probab.*, **23**, Paper No. 72, 27 (2018). [MR3835478](#).
- Baltrus, D. A. Exploring the costs of horizontal gene transfer. *Trends Ecol. Evol.*, **28** (8), 489–495 (2013). [DOI: 10.1016/j.tree.2013.04.002](#).
- Baumdicker, F. and Pfaffelhuber, P. The infinitely many genes model with horizontal gene transfer. *Electron. J. Probab.*, **19**, no. 115, 27 (2014). [MR3296531](#).
- Bennett, P. M. Plasmid encoded antibiotic resistance: acquisition and transfer of antibiotic resistance genes in bacteria. *Br. J. Pharmacol.*, **153** (Suppl 1), S347–S357 (2008). [DOI: 10.1038/sj.bjp.0707607](#).
- Billiard, S. et al. The effect of competition and horizontal trait inheritance on invasion, fixation, and polymorphism. *J. Theoret. Biol.*, **411**, 48–58 (2016). [MR3571027](#).
- Billiard, S. et al. Stochastic dynamics for adaptation and evolution of microorganisms. In *European Congress of Mathematics*, pp. 525–550. Eur. Math. Soc., Zürich (2018). [MR3887784](#).
- Blath, J., Eldon, B., González Casanova, A., Kurt, N., and Wilke-Berenguer, M. Genetic Variability Under the Seedbank Coalescent. *Genetics*, **200** (3), 921–934 (2015). [DOI: 10.1534/genetics.115.176818](#).
- Blath, J., González Casanova, A., Kurt, N., and Wilke-Berenguer, M. A new coalescent for seed-bank models. *Ann. Appl. Probab.*, **26** (2), 857–891 (2016). [MR3476627](#).
- Blath, J., González Casanova, A., Kurt, N., and Wilke-Berenguer, M. The seed bank coalescent with simultaneous switching. *Electron. J. Probab.*, **25**, Paper No. 27, 21 (2020). [MR4073688](#).
- Blath, J., Paul, T., and Tóbiás, A. A stochastic adaptive dynamics model for bacterial populations with mutation, dormancy and transfer. *ArXiv Mathematics e-prints* (2021). [arXiv: 2105.09228](#).
- Blath, J. and Tóbiás, A. Invasion and fixation of microbial dormancy traits under competitive pressure. *Stochastic Process. Appl.*, **130** (12), 7363–7395 (2020). [MR4167209](#).
- Blath, J. and Tóbiás, A. The interplay of dormancy and transfer in bacterial populations: Invasion, fixation and coexistence regimes. *Theoret. Pop. Biol.*, **139**, 18–49 (2021). [DOI: 10.1016/j.tpb.2021.05.001](#).
- Bovier, A., Coquille, L., and Smadi, C. Crossing a fitness valley as a metastable transition in a stochastic population model. *Ann. Appl. Probab.*, **29** (6), 3541–3589 (2019). [MR4047987](#).
- Champagnat, N. A microscopic interpretation for adaptive dynamics trait substitution sequence models. *Stochastic Process. Appl.*, **116** (8), 1127–1160 (2006). [MR2250806](#).
- Champagnat, N., Méléard, S., and Tran, V. C. Stochastic analysis of emergence of evolutionary cyclic behavior in population dynamics with transfer. *Ann. Appl. Probab.*, **31** (4), 1820–1867 (2021). [MR4312848](#).
- Cohen, D. Optimizing reproduction in a randomly varying environment. *J. Theoret. Biol.*, **12**, 119–129 (1966). [DOI: 10.1016/0022-5193\(66\)90188-3](#).
- Coquille, L., Kraut, A., and Smadi, C. Stochastic individual-based models with power law mutation rate on a general finite trait space. *Electron. J. Probab.*, **26**, Paper No. 123, 37 (2021). [MR4317717](#).
- Coron, C. et al. Emergence of homogamy in a two-loci stochastic population model. *ALEA Lat. Am. J. Probab. Math. Stat.*, **18** (1), 469–508 (2021). [MR4219672](#).
- Durrett, R. and Mayberry, J. Traveling waves of selective sweeps. *Ann. Appl. Probab.*, **21** (2), 699–744 (2011). [MR2807971](#).
- Ethier, S. N. and Kurtz, T. G. *Markov processes: Characterization and convergence*. Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons, Inc., New York (1986). ISBN 0-471-08186-8. [MR838085](#).
- Fournier, N. and Méléard, S. A microscopic probabilistic description of a locally regulated population and macroscopic approximations. *Ann. Appl. Probab.*, **14** (4), 1880–1919 (2004). [MR2099656](#).

- Georgii, H.-O. and Baake, E. Supercritical multitype branching processes: the ancestral types of typical individuals. *Adv. in Appl. Probab.*, **35** (4), 1090–1110 (2003). [MR2014271](#).
- Kaj, I., Krone, S. M., and Lascoux, M. Coalescent theory for seed bank models. *J. Appl. Probab.*, **38** (2), 285–300 (2001). [MR1834743](#).
- Koonin, E. and Wolf, Y. Evolution of microbes and viruses: a paradigm shift in evolutionary biology? *Front. Cell. Inf. Microbiol.*, **2** (119) (2012). DOI: [10.3389/fcimb.2012.00119](#).
- Kussell, E. and Leibler, S. Phenotypic Diversity, Population Growth, and Information in Fluctuating Environments. *Science*, **309** (5743), 2075–2078 (2005). DOI: [10.1126/science.1114383](#).
- Lederberg, J. and Tatum, E. L. Gene Recombination in Escherichia Coli. *Nature*, **158**, 558 (1946). DOI: [10.1038/158558a0](#).
- Lennon, J. T., den Hollander, F. T. W., Wilke-Berenguer, M., and Blath, J. Principles of seed banks: complexity emerging from dormancy. *Nat. Commun.*, **12**, 1–16 (2021). DOI: [10.1038/s41467-021-24733-1](#).
- Lennon, J. T. and Jones, S. E. Microbial seed banks: the ecological and evolutionary implications of dormancy. *Nat. Rev. Microbiol.*, **9**, 119–130 (2011). DOI: [10.1038/nrmicro2504](#).
- Lewis, K. Persister Cells. *Ann. Rev. Microbiol.*, **64** (1), 357–372 (2010). DOI: [10.1146/annurev.micro.112408.134306](#).
- Locey, K. J., Fisk, M. C., and Lennon, J. T. Microscale Insight into Microbial Seed Banks. *Front. Microbiol.*, **7**, 2040 (2017). DOI: [10.3389/fmicb.2016.02040](#).
- Smadi, C. The effect of recurrent mutations on genetic diversity in a large population of varying size. *Acta Appl. Math.*, **149**, 11–51 (2017). [MR3647031](#).
- Stewart, F. M. and Levin, B. R. The population biology of bacterial plasmids: a priori conditions for the existence of conjugationally transmitted factors. *Genetics*, **87** (2), 209–228 (1977). [MR456605](#).