Stochasticity and Randomness in Eco-Evolutionary Modelling

Xiaoyuan Liu

PhD

University of York
Mathematics

May 2024

Abstract

Evolution is an inherently stochastic process comprised of many randomly occurring events. The evolutionary fate of a population depends largely on its underlying ecological interactions, while ecological interactions can also be influenced by evolutionary change in turn. This phenomenon is known as an eco-evolutionary feedback loop. Ecological models tend to have a strong focus on how complex ecological features such as interaction structures influence the behaviour of ecosystems, rather than their consequences on long-term evolutionary fates. Evolutionary models on the other hand tend to overlook the complexities associated with their underlying ecological features. This is of particular importance since many evolutionary problems in nature, particularly those associated to the evolution of sexual reproduction, are underpinned by myriad ecological factors. The aim of this thesis is to develop mathematical models for ecological and evolutionary problems in biology, with particular focus on problems surrounding the evolution of sexual reproduction. We begin in chapter 2 by developing an analytical prediction for the stability of generalised Lotka-Volterra systems with biologically motivated interaction structures. In chapter 3, we develop an eco-evolutionary model for the evolution of gamete size and motility to study the evolution of male and female sexes. Chapter 4 repurposes the model of chapter 3 to look at how binary cell fusion can evolve in response to environmental stress. Chapter 5 investigates how genetic recombination evolves in response to environmental stress using an integrative mathematical model that incorporates aspects of population dynamics, population genetics and eco-evolutionary feedback. This allows us to explain analytically how recombination and hibernation evolved to occur together, as well as why they both occur shortly before the onset of environmental stress.

Contents

List of Tables

List of Figures

Acknowledgments

I would like to thank my supervisors George Constable and Jon Pitchford for their continuous support throughout my PhD journey. Thank you to George for the sharing of ideas, long discussions and guidance at each stage of my research. I express my gratitude towards Jon for the patience, long discussions and pastoral support. It is through both your combined efforts that enabled me to develop into the researcher I am today.

I would like to acknowledge EPSRC (Engineering and Physical Sciences Research Council) DTP for the financial support that enabled me to undertake my PhD. Thanks also to the department research committee for providing the financial support that made it possible for me to travel to my conferences.

A huge thank you to my parents, for their relentless unconditional love and support. Despite being 6000 miles away from my location of study, they made sure I could reunite with them every so often.

Cheers to Farzad and Brennen for their helpful discussions and friendship.

Author's declaration

I declare that the work presented in this thesis, except where otherwise stated, is based on my own research carried out at the University of York and has not been submitted previously for any degree at the University of York. All sources are acknowledged by explicit references.

Chapter 2 corresponds to the published paper: X. Liu, G. W. A. Constable, J. W. Pitchford, **Feasibility and stability in large lotka volterra systems with interaction structure**, Physical Review E 107 (5) (2023) 054301. https://doi.org/10.1103/PhysRevE.107.054301

Chapter 3 corresponds to the paper: X. Liu, J. W. Pitchford, G. W. A. Constable, **Adaptive dynamics, switching environments and the origin of the sexes** bioRxiv. 2024, which is currently in submission with Theoretical Population Biology. Preprint can be found on bioRxiv https://doi.org/10.1101/2023.06.09.544312

Chapters 4 and 5 have not been submitted for publication yet. Chapter 4 is written with the intention of submission to PLOS Computational biology as an article.

Accompanying Material

All MATLAB and Mathematica code used for numerical simulations and analytical derivations of equations are found on Github repositories on `https://github.com/phil-liu1020?tab=repositories`. Codes for Chapter 2 are in the repository `feasibility_and_stability`, codes for Chapter 3 are in the repository `ADSW_origin_sexes`, codes for Chapter 4 are in the repository `CellSize_BinaryCellFusion` and codes for Chapter 5 are in the repository `Evol_Env_triggered_sex`.

— 1 —

Introduction

The main focus of this thesis is on the development of eco-evolutionary models for problems in biology. The motivation of the thesis title is as follows. Stochasticity refers to processes comprised of many random events that happen over time. This results in uncertainty in the time series of the process. Evolution can be thought of as a stochastic process as it results from many random processes that happen over time. Randomness refers to the uncertainty in the value of a quantity i.e a random variable. A quantity that is a function of many random variables is also a random variable. Ecological features such as ecosystem stability may be considered a random variable, as it is governed by many random processes including interspecific interations.

This thesis adopts a journal style presentation and contains 6 chapters. Chapter 1 is the introduction chapter, and each chapter from Chapters 2 to 5 corresponds to an individual research project. Chapter 6 is our discussion chapter which provides a synthesis of our results along with some future perspectives. In Chapter 1, we detail some of the main tenets of eco-evolutionary modelling and its applications. This chapter also includes an overview of the current literature in the field of mathematical ecology and the evolution of the sexes and sexual reproduction, which are areas where we wish to develop eco-evolutionary models in. In Section 1.4, we summarise what we do in each of the following chapters.

1.1 INTRODUCTION TO ECO-EVOLUTIONARY MODELLING

Ecology is the study of interactions between organisms and their environment as well as all the biotic and abiotic factors that make their environment habitable [102]. In ecology, it is often assumed that groups of organisms interact as a single unit, known as an ecosystem. This unit may either be comprised of many species, in which case it is called a community, or many strains of a single species, in which case it is called a population. It is also possible for an ecosystem to be comprised of many communities. Ecological modelling has useful applications in many fields such as conservation, where it is important to study the factors that lead to habitat loss and the maintenance of species diversity [175, 210]. Due to climate change and the unsustainable demands of natural resources such as food, there is an ever greater need to understand the ecological factors that constitute species extinctions and the maintenance of biodiversity.

Evolution on the other hand studies how traits of organisms change over time when populations are subject to selective pressures [27]. Rather than considering a static configuration of interactions and traits in the ecosystem, evolution is about how populations modify their physiological traits over time to better adapt to their environment. Evolutionary change results from a combination of selection pressures, mutations, invasions and extinctions of species. As all these factors occur randomly in nature, the evolutionary outcome of the population is a result of a series of random events. Evolution can thus be thought of as a stochastic phenomenon [116, 157], where its outcome cannot be predicted precisely. Nevertheless, mathematical modelling enables us to predict the expected evolutionary outcome amid the stochasticity.

Although ecological factors have important implications for the dynamical behaviour of ecosystems, it is necessary to be aware that ecosystems are subject to evolution over time due to factors such as mutations and invasions of new species [170]. Such factors may influence how organisms interact with one another in the ecosystem, for instance, the emergence of a mutant strain that is better adapted to the environment will impact how strains in the population interact with each other. Conversely, the fate of an evolving ecosystem can also be influenced by its underlying ecological factors [91]. This complex feedback is known as an eco-evolutionary feedback loop [53, 57]. Indeed, there has been a growing awareness of the importance of such a feedback loop in the past few decades, with [120, 181] to be the first few authors to come across it. Despite this increasing awareness, many models in

evolutionary biology tend to overlook the complexity associated with eco-evolutionary feedbacks, while purely ecological models [5, 67, 146] tend to pay little attention to how their complex features may govern their evolutionary fates.

The main aim of this thesis is to apply eco-evolutionary modelling to study various evolutionary problems in biology. For this thesis, we focus on evolutionary problems in sexual reproduction, an aspect underpinned by numerous ecological factors, such as cheating [230], sexual conflict [29], gamete competition [124] and the twofold [123] ecological cost of sex incurred by the requirement of two parents to produce an offspring. First, we include a summary of some of the current gaps in the literature of mathematical ecology and evolutionary modelling to provide context for our research.

1.2 CURRENT LITERATURE IN MATHEMATICAL ECOLOGY

Mathematical ecologists have had a longstanding interest in how the dynamical properties and behaviour of complex ecosystems are influenced by their ecological features, prompting the development of macroecological models [24]. Such models enable us to elucidate how ecological stability may in general be influenced by various system parameters, such as species diversity [146, 188, 210], interspecific interaction strengths and interaction structures [4, 71]. Classical models for macroecological systems assume a large number of species and randomly distributed interaction strengths, which allowed the use of random matrix theory (RMT) techniques to study analytically the effect that system parameters had on ecological stability in general [70, 146, 229].

One of the most famous models that studies the stability of macroecological systems is May's RMT model [146], which posits that the stability is governed by a single compound parameter that is a function of three system parameters; diversity, interaction strength and connectance of the ecological network. This compound parameter is coined as "complexity". Although May's model has been celebrated for its elegance, it lacks biological realism in a few ways. First, it is based on a linearisation of some non-linear ecological model [107], and can only capture the linear stability of the system at best. As May's model is not associated to any specific ecological model, it cannot account for species abundances, and the system parameters that give rise to biologically meaningful species abundances.

A biologically meaningful species abundance is one where the abundance of every species takes positive values, known as feasible. In addition, the model does not account for biases in interaction structure (e.g ecosystems with mere predation, mutualism or competition).

Subsequent attempts have been made to either relate May's linear ecological model to specific ecological models [69, 188, 210], or to generalise May's model to account for biologically motivated interaction structures [4, 75]. Stone [210] was the first to directly link May's model [146] to the generalised Lotka-Volterra (GLV) model. Here it investigated the parameter regions that most likely gave rise to biologically meaningful equilibrium abundances. Analytical results by [4] have shown that stability is reduced by mutualistic and hierarchical interaction structures, including the niche [231] and cascade [35] models. Separate analytical results have also shown how stability can be increased by competitive interactions [39, 49] and modularity (block structure) in the ecological interaction network [75].

Analyses of models that capture the effect of interaction structures on ecosystem stability tend to be based on linear stability, without accounting for specific ecological models. However, analyses that relate random interaction matrices to specific ecological models tend to neglect specific interaction structures [49, 210]. There is thus scope for making analytical progress in the stability analysis of models that both account for biologically motivated interaction structures and relate to specific ecological models. Our project in Chapter 2 develops an analytical approximation of the stability of a random GLV model with ecologically motivated interaction structures. Details of May's RMT model along with other RMT based models will be presented in Chapter 2.

1.3 CURRENT LITERATURE IN EVOLUTIONARY MODELLING OF SEXUAL REPRODUCTION AND SEXES

Evolutionary biologists have had a long interest in the origin of the sexes and the selection pressures that gave rise to sexual reproduction itself. Interests in the origins of the sexes can be traced back to Darwin [43] in his famous book "The Descent of Man, and Selection in Relation to Sex" that focused on sexual selection, where selection acts on the sexes themselves. Males with the most desirable sex specific traits are likely to produce more offspring and pass on their traits, for instance,

vigorous stags with larger horns stand a greater chance of producing more offspring than their hornless counterparts. There has also been curiosities into the evolutionary advantages of sexual reproduction itself, dating back to Weismann [228], who first proposed that sex may be favorable due to its potential to create genetic diversity. These lead naturally to the question of what gave rise to the male and female sexes in the first place.

Within the past few decades, the topic on the evolution of the sexes have received great attention [109, 173, 196]. The male sex is characterised by individuals that produce small motile gametes (sperms) while the female sex is characterised by those that produce large immotile gametes (eggs). An early model known as the Parker-Baker-Smith (PBS) model [173] investigated the evolution of the sexes based on the assumption that the survival of gametes depends on their mass, with gametes of greater mass having a greater chance of survival [222]. Their model showed that if individuals have a fixed energy budget for gamete production, selection favours the production of either a large number of small, poorly adapted gametes (males) or a small number of large well adapted gametes (females). This is known as a "quality-quantity" tradeoff.

The quality-quantity tradeoff theory for the evolution of the sexes paved the way for subsequent work that examined the various conditions under which the two sexes can evolve. By applying the eco-evolutionary modelling framework of adaptive dynamics, Maire [143] showed that anisogamy (the size dimorphism of gametes that characterises the sexes themselves) can evolve from isogamy even in the presence of one self-compatible mating type. However, in addition to size dimorphism, the gametes of the two sexes also exhibit motility dimorphism, with highly motile sperms and immotile eggs, known as oogamy. Recent eco-evolutionary models have investigated the conditions in which anisogamy can evolve under the assumption that gametes can develop without fertilisation (parthenogenetic development) should they fail to fertilise [37, 130]. Although, none of the aforementioned models explain how oogamy or sexual reproduction evolved at first place.

There is thus the scope to investigate the selective pressures that gave rise to sexual reproduction at first place. As fertilisation is an intrinsic mechanism of sexual reproduction, we could investigate the selective pressures for sexual reproduction by considering the mechanistic advantage conferred by fertilisation. In particular, what conditions allow fertilisation to evolve from a state of no fertilisation, and if fertilisation evolves, what size and motility should the gametes evolve to. This can

be done by utilising the modelling framework of [37, 130], and will be the focus of Chapters 3 and 4. Allowing for parthenogenic development also enables us to gain better insights into the mechanisms leading to the evolution of facultative sex, which is frequently observed in organisms such as the green algae *C. reinhardtii* and the fission yeast *S. pombe* [77, 197].

Much of the existing literature on the evolutionary origins of sexual reproduction are based on genetic arguments, such as the rate at which genetic diversity is conferred during genetic recombination, which facilitates adaptation to novel environments [65, 154, 221]. Now we outline some literature for the evolution of sexual reproduction that focuses on the selective advantages of genetic recombination, which sets the scene for our project in Chapter 5. An early theory for the genetic benefits of sex is the Fisher-Muller hypothesis [54], which states that genetic recombination allows multiple advantageous genes to be combined into the genome of one individual on a faster timescale than mutations alone. Another theory known as Muller's ratchet asserts that the advantage of genetic recombination lies in its ability to prevent the buildup of deleterious mutations [154], thereby enhancing the survival of the organism. Subsequent theories have argued that genetic recombination increases the rate at which genetic diversity, and thus well adapted genotypes, are created which hastens adaptation in unpredictable environments [82, 106, 221].

All of the theories above imply a strong genetic advantage to recombination, although it is worth noting that they are based solely on population genetics, with no account for the numerous ecological factors at play in sexually reproducing populations. As mentioned in Section 1.1, these ecological factors have significant consequences for the evolutionary fate of the population. Another overlooked ecological factor that can even influence the cost of sex itself is population density [47, 193]; as the population density increases, the population growth rate naturally reduces due to resource limitation, which reduces the cost of sex.

There is the opportunity to develop models for the evolution of sex that combine population genetic approaches with eco-evolutionary modelling. In Chapter 5, we develop an eco-evolutionary model that incorporates population genetics to investigate how genetic recombination can evolve as a response to environmental stress.

1.4 OVERVIEW OF THESIS

In Chapter 2, we develop an analytical approximation that allows us to study the feasibility, stability and resilience of a GLV model with biologically motivated interaction structures. Here, we specifically consider the effects of predator-prey, mutualistic and competitive interactions on the stability of ecosystems. Stone [210] investigated the feasibility probability of a GLV system where its underlying interaction matrix is May's interaction matrix [146]. Here, we generalise Stone's result to the case where the underlying interaction matrix is Allesina and Tang's interaction matrix [5], which accounts for predator-prey, competitive and mutualistic interactions.

In Chapter 3, we investigate the early evolution of sexual reproduction by considering the mechanistic benefits conferred by fertilisation of gametes due to the increased mass. To do this, we coevolve gamete mass and fertilisation rate by suitably modifying an existing eco-evolutionary model [37]. We assume self-compatible mating types for simplicity. By accounting for gamete mass, we simultaneously explore the evolution of different modes of sexual reproduction. For instance, anisogamy is a reproductive mode that characterises the sexes themselves, where males are microgametes and females are macrogametes. First, we find the parameter regions where sexual reproduction evolves, followed by the masses that gametes evolve to. As we are coevolving mass and fertilisation rate, we can also examine whether the population evolves any heterogeneity in fertilisation rates, such as in oogamy. In line with empirical observations and recent models [37, 130], we assume that gametes can undergo parthenogenic development should they fail to fertilise.

In Chapter 4, we use the same model as in Chapter 3 to investigate how cell fusion may evolve as a response to environmental stress. Here, we assume that organisms exhibit phenotypic plasticity, where they can evolve different strategies in different environments. By neglecting any genetic factors associated with sex or cell fusion, our work may provide an alternative mechanistic hypothesis for how sex evolves as a response to environmental stress, in contrast to some existing population genetics hypotheses [62, 221]. In addition to sex, our work in this chapter will have important implications for how facultative multicellularity evolves as a response to environmental stress [76, 215]. Note well that the phrase "fusion" which we use in this chapter and "fertilisation" which we use in Chapter 3 both refer to the same thing in our model, but are phrased differently, to match the context of the papers

corresponding to the chapters.

In Chapter 5 we develop an eco-evolutionary model that incorporates population genetics aspects of genetic recombination to study how sex evolves as a response to environmental stress. This partly complements the work in Chapter 4, where we investigate stress induced sex in terms of the mechanistic benefits of cell fusion. By accounting for population dynamics as well, we are able to factor in the ecological costs of sex. Furthermore, by suitably modifying our model, we investigate how hibernation can be triggered by environmental stress as well as sex.

Chapter 6 contains a general discussion which seeks to synthesise the results presented in this thesis within the broader scope of eco-evolutionary modelling. Some future research questions are described and prioritised, with suggestions as to the methods which may be useful in seeking their answers.

— 2 —

Feasibility and Stability of large Lotka Volterra Systems with Interaction Structure

Abstract

Complex system stability can be studied via linear stability analysis using Random Matrix Theory (RMT) or via feasibility (requiring positive equilibrium abundances). Both approaches highlight the importance of interaction structure. Here we show, analytically and numerically, how RMT and feasibility approaches can be complementary. In generalised Lotka-Volterra (GLV) models with random interaction matrices, feasibility increases when predator-prey interactions increase; increasing competition/mutualism has the opposite effect. These changes have crucial impact on the stability of the GLV model.

2.1 INTRODUCTION

In the 1950s, ecologists such as Odum and MacArthur argued [141, 166] that ecosystems with a larger number of species tend to be more stable than less biodiverse systems. This idea was famously mathematised by May in 1972, who applied random matrix theory (RMT) to the problem [146]. May considered perturbations in the abundances of n species communities, $\boldsymbol{\zeta}$, where $\boldsymbol{\zeta}$ was linearised about a hypothetical fixed point of some unspecified nonlinear ecological model, with near-equilibrium dynamics described by

$$\frac{\mathrm{d}\boldsymbol{\zeta}}{\mathrm{d}t} = A\boldsymbol{\zeta} \tag{2.1}$$

where he suggested parameterising A according to

$$A_{ii} = -1\,, \quad A_{ij} = \sigma c a_{ij} \tag{2.2}$$

with A_{ii} representing the species self-regulation at equilibrium and $a_{ij} \sim \mathcal{N}(0,1)$ represents the interaction coefficients which are distributed according to standard normal distribution, and $c \sim \mathrm{B}(1,C)$ is a Bernoulli random variable representing the probability that a pair of species interact. Here A_{ij} represents random species interactions that are non-zero with probability C (referred to as connectance) and when present have standard deviation σ (referred to as interaction strength). Since the asymptotic stability of Eq. (2.1) is governed solely by its eigenvalues, system-level stability is determined by characterising the eigenvalues of random matrix A.

The eigenvalue distribution of A is uniform across a circle in the complex plane, centered on $(-1,0)$ and with radius $\sigma\sqrt{nC}$ as $n \to \infty$ [146, 214, 229].

Thus the stability criterion for Eq. (2.1) is $\sigma\sqrt{nC} < 1$ (see Fig. 2.1(a)). This suggests that more diverse ecosystems with more interspecific interactions are less likely to be stable for a given variance in interaction strength.

Allesina and Tang [4] added ecologically-motivated structure to May's approach, choosing elements of A pairwise by imposing a correlation, ρ, between A_{ij} and A_{ji} for $j \neq i$,

$$\begin{aligned} &(A_{ij}, A_{ji}) = \sigma c(a_{ij}, a_{ji}) \ \text{ where} \\ &(a_{ij}, a_{ji}) \sim \mathcal{N}(\mathbf{0}, \Sigma) \ \text{ with } \ \Sigma = [(1,\rho),(\rho,1)] \end{aligned} \tag{2.3}$$

where again $c \sim B(1,C)$. Ecologically, $\rho < 0$ implies more predator-prey interactions in the ecosystem (A_{ij} and A_{ji} are more likely to have opposite signs), while $\rho > 0$ implies more mutualistic and competitive interactions (A_{ij} and A_{ji} are more likely to have the same sign). Utilising another RMT result [71, 201] they generalised May's stability criterion to

$$\sigma\sqrt{nC}(1+\rho) < 1\,. \tag{2.4}$$

Thus, increasing the proportion of predator-prey interactions increases stability, whilst increasing the proportion of competitive and mutualistic interactions reduces stability in Eq. (2.1) (see Fig. 2.1(a)). Eq. (2.4) implies that in the extreme limit $\rho \to -1$, ecosystems are stable as long as there is sufficient self-regulation, since the real part of all the eigenvalues of A will be equal to the mean of its diagonal elements A_{ii} when $\rho \to -1$, which is negative in the presence of sufficient self-regulation.

These analytic results are independent of the underlying non-linear model from which they are hypothetically derived. However, this apparent generality conceals an implicit assumption that the fixed point about which the non-linear system is linearised (to arrive at Eq. (2.1)) exists and is biologically meaningful. Such biologically meaningful fixed points, where every species is present at a positive abundance, are termed feasible equilibria [188].

As in the case of [210, 67, 34], we use the generalised Lotka-Volterra model (GLV)

$$\frac{\mathrm{d}\boldsymbol{x}}{\mathrm{d}t} = \boldsymbol{x} \odot (\boldsymbol{r} + A\boldsymbol{x}) \,, \tag{2.5}$$

to explore the links between the parameterisations of the interaction matrix A in Eqs. (2.2-2.3) and feasibility. Here x_i is the abundance of species i, r_i is its intrinsic growth rate, A the interaction matrix, and $\odot$ the Hadamard product, where x_i and r_i are components of the vectors $\boldsymbol{x}$ and $\boldsymbol{r}$ in Eq. (2.5) respectively. Eq. (2.5) has a single non-zero fixed point, $\boldsymbol{x}^*$, with a Jacobian, J, such that

$$\boldsymbol{x}^* = -A^{-1}\boldsymbol{r} \,, \qquad J = \mathrm{diag}(\boldsymbol{x}^*)A. \tag{2.6}$$

Note that if the elements of A are drawn from a random distribution, then $\boldsymbol{x}^*$ is also a random variable (see, for instance Fig. 2.2). We denote the multivariate distribution of $\boldsymbol{x}^*$ as $P(\boldsymbol{x}^*)$. In particular, there is nothing intrinsic about the structure of $\boldsymbol{x}^*$ in Eq. (2.6) that guarantees that it is feasible (i.e. that $x_i^* > 0 \,\forall\, i$). Instead, for any given randomly sampled A, there is a probability that the fixed point is feasible, which we denote P_{feas}. The relationships between feasibility, stability and different system constraints such as interaction structure is a central theme in theoretical ecology [26].

Early analytic insight into the feasibility of $\boldsymbol{x}^*$ in Eq. (2.6) assumed that A had interaction coefficients with fixed strengths, or with randomly generated signs [72, 69, 188]. Stone [210] linked this to May's approach by considering the probability that $\boldsymbol{x}^*$ is feasible given an ensemble of random interaction matrices parameterised according to Eq. (2.2). Under the condition that $r_i = 1 \;\forall\; i \in [i, n]$, the distribution of x_i^* can be determined using some mathematical arguments as outlined in Section 2.2 (see also Appendix VIII and [210]). Those arguments show that under the condition that $\sigma\sqrt{n} < 1$, x_i^* tends to a normal distribution in the limit of large n. Stone thus assumed that such a parameterisation of interaction matrices gives rise to a normally distributed x_i^* (see Figure 2.2 and Appendix VIII).

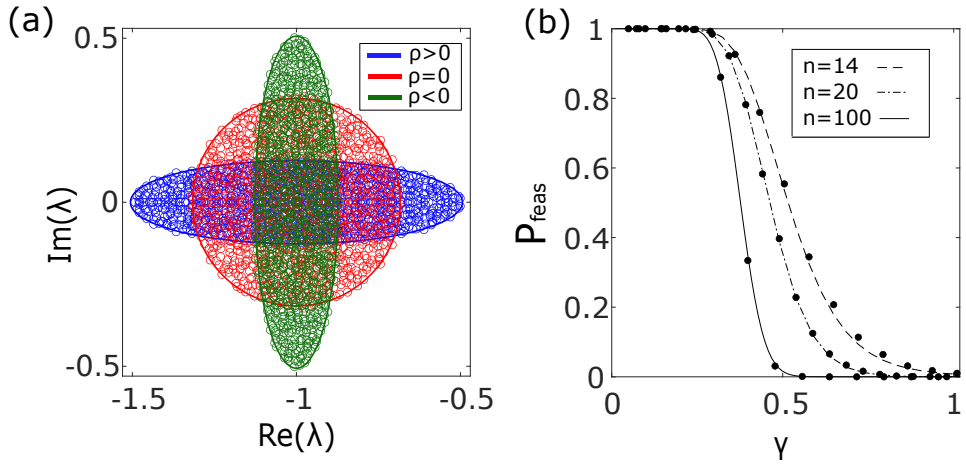

Figure 2.1: Panel (a): Eigenvalue distributions of interaction matrix A parameterised according to Eq. (2.2) (red, $\rho = 0$, see [146]) and Eq. (2.3) (blue and green, $\rho \neq 0$, see [4]), used to infer the stability of the linear model proposed in Eq. (2.1). Parameter values are $\sigma = 0.01$, $n = 1000$, $C = 1$ and $|\rho| = 0.6$. Panel (b): Feasibility probability, P_{feas}, for an ensemble of random fixed points from the non-linear GLV model, Eq. (2.5), with interaction matrices parameterised according to Eq. (2.2) ($\rho = 0$, see [210]). P_{feas} is plotted as a function of May's complexity parameter $\gamma = \sigma\sqrt{nC}$, for community sizes ranging from $n = 14$ to $n = 100$. In this panel $C = 1$. Curves are analytical predictions and markers are numerical simulations, obtained by sampling 10^4 random interaction matrices A parameterised according to Eq. (2.2) and calculating the proportion of those that give rise to a feasible equilibrium solution of the GLV model (see Appendix IV).

Stone showed that under the assumption that x_i^* is normal, the probability of feasibility for a fully connected system $C = 1$ is

$$P_{\text{feas}} = 2^{-n}\left(1 + \text{erf}\left(\frac{1}{\gamma_S\sqrt{1+\gamma_S^2+\gamma_S^4}}\right)\right)^n, \tag{2.7}$$

where $\gamma_S = \sigma\sqrt{n}$ is known as the disturbance in Stone's analysis, which is equivalent to May's definition of complexity for the case $C = 1$. We see that P_{feas} drops sharply at a critical value of γ_S, and also has an additional dependence on system size n (see Fig. 2.1(b)). By working in the limit $n \to \infty$, [19, 34] determined a threshold interaction strength above which feasibility is lost in GLV models with interaction matrices parameterised according to Eq. (2.2). An analytical prediction for the relationship between P_{feas} and the complexity $\gamma = \sigma\sqrt{nC}$ which accounts for C was obtained by Dougoud *et al.* [49]. Akjouj *et al.* [3] investigated the feasibility of

sparse ecosystems with interaction matrices that are block structured and d-regular (where each species interacts with d other species). Together these results suggest that feasibility is the more critical measure of complex system stability; compared to linear stability, feasibility is lost at smaller values of complexity.

Here we seek to strengthen the links between RMT [5, 146] and feasibility analyses by calculating how the feasibility of an ecosystem changes with complexity [3, 49, 208, 210] when additional species interaction structure is accounted for [4, 5]. It was shown by Bunin [26] that feasible systems lose stability above a certain interaction strength by transition to a phase with multiple attractors. The interaction strength of this phase transition increases as predator-prey interactions increase. Numerical results by Clenet *et al.* [34] also show that systems biased towards predator-prey interactions lose feasibility at larger interaction strengths than systems without interaction structure, and those biased towards competition and mutualism lose feasibility at smaller interaction strengths than systems without interaction structure. They also obtained an analytical result for the interaction strength above which feasibility is lost, in the limit of large n. In this limit the effect of the correlation parameter ρ, the parameter that governs the proportion of predator-prey or competition/mutualistic interactions, disappears [34]. In this chapter, we instead work in the large but finite n limit in order to explore the effect of ρ on the probability of feasibility, P_{feas}. In order to calculate P_{feas}, we must also obtain an approximation for the distribution of fixed points. This approximation opens up the possibility of leveraging recent results [9, 67] to determine the probability of stability of the GLV model with interaction structure.

2.2 ANALYSIS

Following Stone [209] we obtain an analytical approximation of $P_{\text{feas}}(\gamma)$ via the distribution of equilibrium species abundances $P(\boldsymbol{x}^*)$. In particular Stone [210] applied the Central limit theorem to $\boldsymbol{x}^*$ in Eq. (2.9) to argue that $P(\boldsymbol{x}^*)$ is normal as $n \to \infty$, and this normality remains a good approximation when n is large but finite. Indeed we have shown numerically in Appendix VIII that for $n = 25$, this normal approximation holds. The task of calculating the feasibility probability is then equivalent to calculating

$$P_{\text{feas}} = \int_{\boldsymbol{x}^*=\boldsymbol{0}}^{\infty} P(\boldsymbol{x}^*)\mathrm{d}\boldsymbol{x}^* \approx \int_{\boldsymbol{x}^*=\boldsymbol{0}}^{\infty} \mathcal{N}(\boldsymbol{\mu}_{\boldsymbol{x}^*}\Sigma_{\boldsymbol{x}^*})\mathrm{d}\boldsymbol{x}^* \tag{2.8}$$

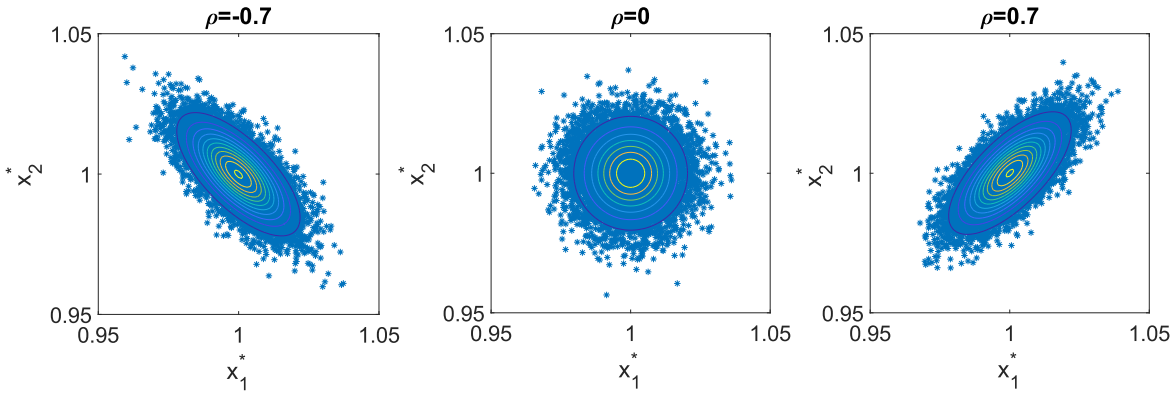

Figure 2.2: Plots showing the joint distribution of x_1^* and x_2^* for the GLV model Eq. (2.5) with $n = 2$, $\sigma = 0.01$ and $C = 1$. Blue markers represent 10^4 numerical solutions of the GLV model, obtained as described in Appendix IV. Contours are analytical predictions for the joint distribution of x_1^* and x_2^* calculated using Eqs. (2.13-2.15).

where $\boldsymbol{\mu}_{\boldsymbol{x}^*}$ and $\Sigma_{\boldsymbol{x}^*}$ are respectively the mean and covariance matrix of the species abundances at equilibrium. Note that by symmetry, we can see that for interaction matrices randomly generated according to Eq. (2.3), $\boldsymbol{\mu}_{\boldsymbol{x}^*}$ and $\Sigma_{\boldsymbol{x}^*}$ are themselves highly symmetric, with $[\boldsymbol{\mu}_{\boldsymbol{x}^*}]_i = [\boldsymbol{\mu}_{\boldsymbol{x}^*}]_j$, $[\Sigma_{\boldsymbol{x}^*}]_{ii} = [\Sigma_{\boldsymbol{x}^*}]_{jj}$ and $[\Sigma_{\boldsymbol{x}^*}]_{ij} = [\Sigma_{\boldsymbol{x}^*}]_{ji}$ for all $i, j \in [1, n]$ (i.e. $\boldsymbol{\mu}_{\boldsymbol{x}^*}$ is a constant vector and the variance-covariance matrix $\Sigma_{\boldsymbol{x}^*}$ is a double constant matrix [165]).

We now calculate approximations for $\boldsymbol{\mu}_{\boldsymbol{x}^*}$ and $\Sigma_{\boldsymbol{x}^*}$. For simplicity we focus on the case $r_i = 1 \; \forall \; i$ in Eq. (2.5). Recall that following [5], the elements of the interaction matrix A_{ij} and A_{ji} have correlation ρ. Writing $A = \sigma\mathcal{E} - \mathbf{I}$, our fixed point in Eq. (2.6) can be expressed as a Neumann series [119] for $||\sigma\mathcal{E}|| < 1$:

$$\boldsymbol{x}^* = (\mathbf{I} - \sigma\mathcal{E})^{-1}\mathbf{r} \equiv \left(\sum_{j=0}^{\infty}(\sigma\mathcal{E})^j\right)\mathbf{r}. \tag{2.9}$$

where its elements x_i^* may alternatively be expressed as

$$x_i^* = 1 + \sigma c \sum_{i=1}^{n} a_{ij} + \dots \tag{2.10}$$

As Eq. (2.10) involves sums over n random variables, the Central limit theorem can be applied to show that x_i^* is normally distributed as $n \to \infty$. However, the normality in the distribution of $x_i^* \; \forall i \in [1, n]$ does not necessarily imply that the

multivariate distribution of $\boldsymbol{x}^*$ is normal. Given that x_i^* is normal $\forall i \in [1, n]$, a necessary condition for $\boldsymbol{x}^*$ to be a multivariate normal distribution is for any linear combination of all the x_i^* $\forall i \in [1, n]$ to also be normal [6]. As x_i^* has the same distribution $\forall i \in [1, n]$, the Central limit theorem can be applied to see that any linear combination of the abundances of all species is normally distributed [6], which justifies that the distribution of $\boldsymbol{x}^*$ is a multivariate normal distribution.

This enables us, in principle, to calculate x_i^* up to an arbitrary order in σ. In our work, we approximate $E(x_i^*)$, $Var(x_i^*)$ and $Cov(x_i^*, x_j^*)$ taking into account ρ and C. Using Eq. (2.9), we approximate $E(x_i^*)$ and $Var(x_i^*)$ up to and including order σ^6. Using the fact that the product of an odd number of normal random variables with zero mean has zero expectation, we know that all terms of $E(x_i^*)$ at odd orders of σ vanish. From Eq. (2.9), we find that the expression for the expectation of x_i^* up to this given order is

$$E(x_i^*) = E\left(1 + \sigma^2 \sum_{\substack{j=1 \\ j\neq i}}^{n} \sum_{\substack{k=1 \\ k\neq j}}^{n} \kappa a_{ij} a_{jk}\right) + e_4\sigma^4 + e_6\sigma^6 \tag{2.11}$$

where e_4 and e_6 are coefficients of σ^4 and σ^6 respectively in the expectation of x_i^*, and

$$\kappa = \begin{cases} C & \text{if} \quad i = k\,, \\ C^2 & \text{if} \quad i \neq k\,, \end{cases} \tag{2.12}$$

since $i = k$ corresponds to the case where $a_{jk} = a_{ji}$, which corresponds to the case where A_{ij} and A_{ji} are both nonzero with probability C (see Eq. (2.3) and Allesina and Tang [5]). We use Eq. (2.11) to illustrate how we obtain our approximation of $E(x_i^*)$. Since $E(a_{ij}a_{ji}) = \rho$, $E(a_{ij}) = 0$ and $E(a_{ij}a_{jk}) = 0$ if $k \neq i$, Eq. (2.11) is equal to

$$E(x_i^*) = 1 + (n-1)\rho C\sigma^2 + e_4\sigma^4 + e_6\sigma^6 \tag{2.13}$$

where through direct calculation, it can be shown that $e_4 = (n-1)(C + \rho^2(2C + 2C^2(n-2)))$, given by Eq. (2.28). Similarly we can calculate e_6, which is given by Eq. (2.66) of the Appendix.

An analogous approach can be used to obtain an approximation for $Var(x_i^*)$ and $Cov(x_i^*, x_j^*)$ (see Appendix I), with $Var(x_i^*)$ given by

$$Var(x_i^*) = (n-1)C\sigma^2 + v_4\sigma^4 + v_6\sigma^6 + O(\sigma^8) \tag{2.14}$$

where v_4 and v_6 are the coefficients of σ^4 and σ^6 respectively, which depend on n, ρ and C. Specifically, v_4 is the coefficient of σ^4 in Eq. (2.35) and v_6 is given by Eq. (2.73) in the Appendix. The formulas for v_4 and v_6 are too lengthy to produce here, however of particular note is the fact that they, along with coefficients e_4 and e_6, are nontrivial polynomials that do not preserve the simple dependence on the complexity parameter γ observed in [4] or [146]. $Cov(x_i^*, x_j^*)$ is given by

$$Cov(x_i^*, x_j^*) = \rho C\sigma^2 + c_4\sigma^4 + O(\sigma^6) \tag{2.15}$$

where $c_4 = (3 + (6 + C(5n - 11))\rho^2)$. While we could extend this approximation to order σ^6, we note that this makes little quantitative difference to the approximation. In the expression for $Cov(x_i^*, x_j^*)$, the coefficient of each order of σ is a factor of n smaller than the corresponding coefficients in the expression for $E(x_i^*)$ and $Var(x_i^*)$ (see Appendix VII). This implies that for a fixed value of large but finite n, $Cov(x_i^*, x_j^*)$ increases more slowly with σ than $E(x_i^*)$ and $Var(x_i^*)$, and thus $Cov(x_i^*, x_j^*)$ plays a smaller role in governing how $P(\boldsymbol{x}^*)$, and similarly P_{feas}, varies with σ. It is therefore possible to approximate $Cov(x_i^*, x_j^*)$ to order σ^4 without sacrificing the accuracy of the analytical prediction of P_{feas}. The slower increase in $Cov(x_i^*, x_j^*)$ with σ is verified numerically in Figure S7. Since an analytical approximation of $Cov(x_i^*, x_j^*)$ to order σ^6 requires considerably more algebra (see Appendix 2.5.4.5) without conferring significant improvements to the accuracy of P_{feas}, we restrict our analysis to the order σ^4 approximation given in Eq. (2.15).

Eqs. (2.13-2.15) are then used to construct $\boldsymbol{\mu}_{\boldsymbol{x}^*}$ and $\Sigma_{\boldsymbol{x}^*}$ in Eq. (2.8). Note that we expect our approximation to hold when n is large (such that $P(\boldsymbol{x}^*)$ is approximately normal, see Eq. (2.8)) and when σ is small (such that the expansions in Eqs. (2.13-2.15) remain sufficient). When these conditions are not met, the approximations given in Eqs. (2.13-2.15) break down at lower values of $|\rho|$. For instance in a 25 species ($n = 25$) system, the analytical approximation of $Var(x_i^*)$ in Eq. (2.14) loses accuracy when $|\rho| > 0.25$), while for a 100 species system $Var(x_i^*)$ remains accurate up to $|\rho| = 0.5$ (see Appendix II).

The fact that our normal distributions feature such a high degree of symmetry, with $\boldsymbol{\mu}_{x^*}$ a constant vector and $\Sigma_{\boldsymbol{x}^*}$ a double constant matrix, allows us to further simplify the calculation of P_{feas}. This provides ease of computation for large systems. Using the results of [41] which expresses integrals over the cubic region of the variable

space, Eq. (2.8) can be reduced to an expression involving a single integral, given by

$$P_{\text{feas}} = \int_{-\infty}^{\infty} \left\{ \prod_{i=1}^{n} \Phi\left(\frac{y_i - b_i u}{(1-b_i^2)^{1/2}}\right) \right\} \phi(u) du \tag{2.16}$$

where $\phi(u)$ is the density function of a standard normal random variable u and $\Phi(v)$ denotes the cumulative distribution function of a standard normal random variable v. In our analytical prediction of P_{feas}, we have that $y_i = \frac{E(x_i^*)}{\sqrt{Var(x_i^*)}}$ and $b_i = \frac{\sqrt{Cov(x_i^*, x_j^*)}}{Var(x_i^*)}$ (see Appendix III). In other words, P_{feas} is the expression obtained by substituting these expressions for y_i and b_i into Eq. (2.16). (see Appendix III). Interestingly, note that in the results of [5, 146], C appears as a compound parameter with σ^2, but in Eqs. (2.13-2.15), C appears in a complicated polynomial form. The analytical prediction of $P_{\text{feas}}(\gamma)$ is shown in Figure 2.3 (a)-(b). Moreover, the fact that $Cov(x_i^*, x_j^*)$ is a factor of n smaller than $Var(x_i^*)$ partly explains the observation of Clenet [34] that as $n \to \infty$, the effect of ρ on P_{feas} completely disappears.

2.3 RESULTS

2.3.1 PREDATOR-PREY INTERACTIONS INCREASE THE FEASIBILITY OF RANDOM ECOSYSTEMS

The qualitative difference in how P_{feas} changes with the complexity γ as the correlation ρ is varied is shown analytically in Figure 2.3. For a given value of n, when ρ is positive (blue), feasibility is lost at a smaller complexity compared to the case where $\rho = 0$ (red). However when ρ is negative (green), we observe the opposite effect whereby feasibility is lost at a larger complexity than the case $\rho = 0$. In other words, increasing ρ reduces the feasibility probability for a given complexity in a GLV system of finite n.

It can be seen in Figure 2.3 that the magnitude of the difference between $P_{\text{feas}}(\gamma, \rho)$ and $P_{\text{feas}}(\gamma, 0)$ also varies with γ. For instance when γ is sufficiently small, there is no difference between $P_{\text{feas}}(\gamma, \rho)$ and $P_{\text{feas}}(\gamma, 0)$, since P_{feas} is 1 regardless of ρ. The bottom panels of Figure 2.3 below plot this difference, demonstrating how it varies with γ. The difference between $P_{\text{feas}}(\gamma, \rho)$ and $P_{\text{feas}}(\gamma, 0)$ is the greatest for intermediate values of complexity γ, where the system is transitioning rapidly away from feasibility. For a given system size n, the magnitude of this difference ($|P_{\text{feas}}(\gamma, \rho)$-$P_{\text{feas}}(\gamma, 0)|$) also increases with the magnitude of ρ.

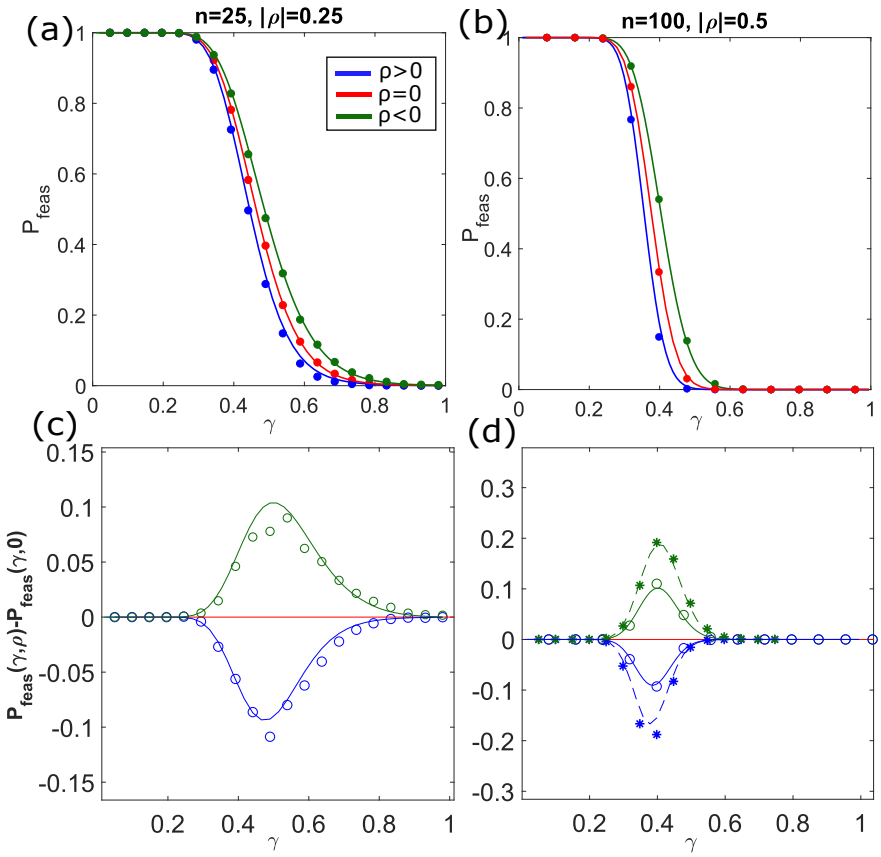

Figure 2.3: Panels (a) and (b) plot the feasibility probability P_{feas} as a function of complexity γ for systems with ecologically motivated interaction structure: blue ($\rho > 0$) biased toward competitive/mutualistic interactions; red ($\rho = 0$) unbiased interactions; green ($\rho < 0$) biased towards predator-prey interactions. Panels (c) and (d) plot the difference between P_{feas} in systems with $\rho \neq 0$ and P_{feas} in systems where $\rho = 0$ ($P_{\text{feas}}(\gamma, \rho) - P_{\text{feas}}(\gamma, 0)$) as a function of γ, with lines the prediction derived from Eq. (2.16) and markers the results of numerical simulation. In panel (c), $n = 25$ and hollow circles show the results of numerical simulations for the case $|\rho| = 0.25$. In panel (d), where $n = 100$ (and our approximations are valid for larger values of ρ) hollow circles again represent the case the case $|\rho| = 0.25$, while asterisks are numerical simulations for the case $|\rho| = 0.5$. Numerical simulations are obtained by sampling 10^4 random interaction matrices A parameterised according to Eq. (2.2) and calculating the proportion of those that give rise to a feasible equilibrium solution of the GLV model Eq. (2.5) (see Appendix IV)

In Appendix I.E, we see that for all values of ρ, the loss of feasibility in the GLV model with Allesina and Tang type interaction matrices occurs at a smaller complexity than the loss of stability in the corresponding linear model. As an extreme example,

in linear systems comprising all predator-prey interactions ($\rho = -1$) stability is guaranteed regardless of ecosystem complexity (see Eq. (2.4)); conversely feasibility is still lost above a critical value of the complexity parameter γ (see Figure S2 of Appendix). Figure 2.3 demonstrates that the analytical results in Eq. (2.13-2.15) can be used to accurately predict P_{feas} as a function of γ in the case where $C = 1$. Furthermore, Appendix V shows that the same analytical results remain highly accurate for predicting P_{feas} as a function of γ in the case where $C = 0.3$. By comparing the feasibility probabilities of such a system with that of a fully connected system, we see that a sparsely connected system of $n = 100$ shows an almost identical feasibility-complexity relation as a fully connected system.

Most importantly, in Eqs. (2.13-2.15) we have analytically approximated the distributions of x_i^* for non-linear GLV models Eq. (2.5) where the underlying interaction matrix A is constructed according to Eq. (2.3). This opens up the possibility to extend these results to predict the stability of GLV models with ecologically motivated interaction structures. Such a stability analysis is beyond the scope of this work, but would be attainable through detailed analysis of the GLV Jacobian. In the next section we investigate how this might be achieved within the scope of existing methods.

2.3.2 Comparing RMT predictions with GLV Jacobian matrices

Gibbs *et al.* [67] studied the eigenvalue distribution of a matrix that is assumed to be of the same structure as the GLV Jacobian (Eq. (2.6) right), where J is decomposed into a product of an interaction matrix A and fixed points $\boldsymbol{x}^*$. However, for simplicity, they assume that the distribution from which $\boldsymbol{x}^*$ is drawn is independent of A, whereas this is clearly not the case (see Eq. (2.6) left).

Gibbs' assumption of independence between the random elements of A and $\boldsymbol{x}^*$ means that cross correlations between them need-not be considered, thereby simplifying the analysis. We test whether this assumption holds, in order to determine whether Gibbs' method may be applicable to calculating the eigenvalue distribution of the GLV Jacobian (Eq. (2.6)). To do so, we first calculate the eigenvalue distribution of $J = \boldsymbol{x}^* A$ where the elements of $\boldsymbol{x}^*$ are sampled independently to those of A. The distribution from which we sample the elements of $\boldsymbol{x}^*$ is a normal distribution with $E(x_i^*)$, $Var(x_i^*)$ and $Cov(x_i^*, x_j^*)$ given by Eq. (2.13-2.15), which we approximated. A is constructed according to Eq. (2.3). We then compare this eigenvalue distribution

(shown in Figure 2.4 bottom panels) to that of the GLV Jacobian where the exact $\boldsymbol{x}^*$ corresponding to each given A is used (shown in black markers of Figure 2.4 top panels).

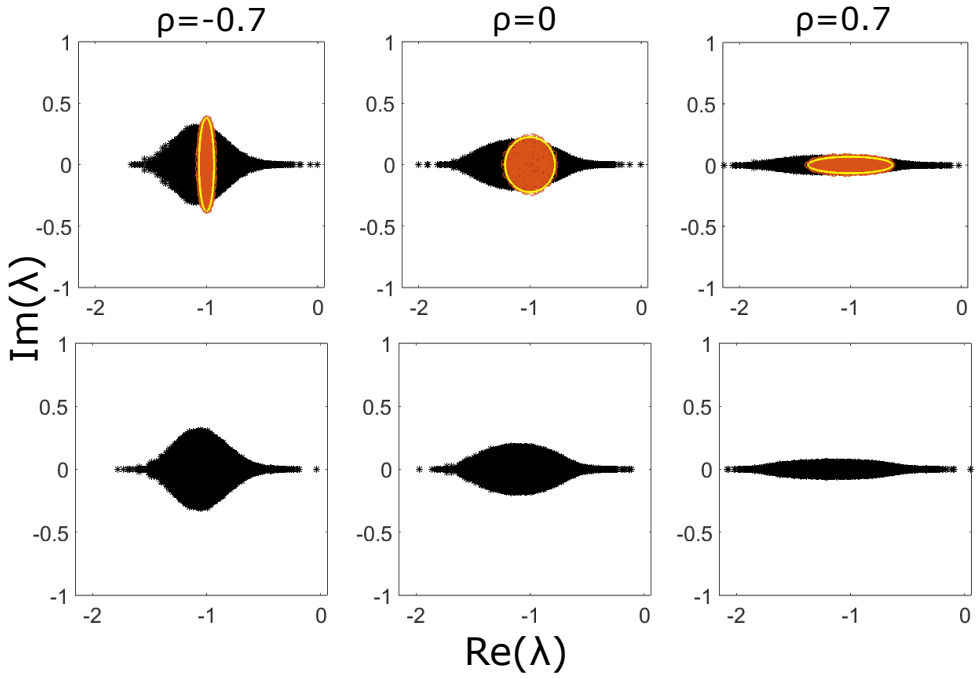

Figure 2.4: Top row: Orange ellipses are eigenvalue distributions of A where A is parameterised according to Eqs. (2.2-2.3). Yellow boundaries are predicted by Allesina and Tang. Black markers represent 50 realisations of the eigenvalue distribution of the GLV Jacobian $J = \boldsymbol{x}^* A$ where the exact $\boldsymbol{x}^*$ corresponding to each given A is used. Bottom row: 50 realisations of the eigenvalue distribution of $J = \boldsymbol{x}^* A$ where elements of $\boldsymbol{x}^*$ are sampled independently of A, from the multivariate normal distribution characterised by Eqs. (2.13-2.15). Parameter values are $\sigma = 0.01$, $n = 500$ and $C = 1$. Given these parameters, Eqs. (2.13-2.15) predict that in the left panel $P_{\text{feas}} = 0.993$, middle panel $P_{\text{feas}} = 0.997$ and right panel $P_{\text{feas}} = 1.000$.

By comparing the black markers on the top panels with those of the bottom panels of Figure 2.4, we see that our method of sampling $\boldsymbol{x}^*$ independently of A from our distribution of $\boldsymbol{x}^*$ works well in predicting the eigenvalue distribution of the GLV Jacobian. This comparison is conducted in a region where feasibility is almost surely guaranteed. From the top panels, we see that when the correlation parameter is negative i.e $\rho < 0$, the bulk eigenvalue distribution of J gets stretched in the $Im(\lambda)$ plane, and when $\rho > 0$ in the $Re(\lambda)$ plane. This qualitative effect is consistent with the result of Allesina and Tang [5]. It is shown numerically in Appendix VI that

increasing ρ decreases the average resilience of the GLV model, where resilience is a measure of the ability for communities to recover following perturbations in species abundance from their equilibrium abundance [210]. Increasing the real part of the eigenvalue with the largest real part decreases the resilience of the GLV model.

The average maximum outlier eigenvalue (averaged over multiple realisations of the interaction matrix A) is also correctly predicted by our theory, which relies on the assumption of statistical independence between A and our calculated distribution of $\boldsymbol{x}^*$ (see Eqs. (2.13-2.15)), as illustrated in Figure S6 (a). However, our theory does not correctly predict the maximum outlier eigenvalue of individual realisations of the GLV Jacobian. This suggests that cross-correlations between the entries of A and $\boldsymbol{x}^*$ may be quantitatively important in calculating the stability of individual realisations of the GLV model. As the stability of a system is governed solely by the eigenvalue with the largest real part, a stability analysis of the GLV model must be preceded via calculating such an eigenvalue. Below, we provide an insight into some possible techniques for calculating the stability of the GLV model with Allesina and Tang type interaction matrices.

Stone [208] showed that provided that $||\sigma\mathcal{E}||$ is sufficiently small, the eigenvalue with the largest real part (outlier eigenvalue of J) is approximately equal to minus the abundance of the least abundant species i.e $\lambda_{max} \approx -\min_{i\in\{1,n\}} x_i^*$; in which case we have the weak condition whereby feasibility corresponds to the local asymptotic stability of the GLV model. In the case where $\rho = 0$ or $|\rho|$ is small, $-\min_{i\in\{1,n\}} x_i^*$ is an accurate estimate of the outlier eigenvalue of J, however this accuracy breaks down as we increase $|\rho|$ (see Appendix VI).

Relying on Gibbs' assumption allows us to accurately capture the bulk eigenvalue distribution of J and the effect that the correlation parameter ρ has on the average resilience over a large number of realisations (see Figure S6 (a)), although it fails to accurately calculate the outlier eigenvalue of J corresponding to a specific realisation of A.

2.4 DISCUSSION

We have obtained an analytical prediction of the feasibility probability as a function of complexity $\gamma = \sigma\sqrt{nC}$ for random GLV models with interaction matrices of Allesina and Tang type [5]. By extending the analytical result of [34] to the case of large, but

finite n, we have shown that a positive value of ρ reduces the feasibility probability for a given complexity, while a negative value of ρ increases the corresponding feasibility probability, an effect not quantifiable in the infinite n limit. We have also accounted for the connectance C. Since natural ecological systems are sparsely connected [60], both these generalisations mentioned above add biological realism to the result of Stone 2016 [209]. Relationships between complexity and feasibility have also been studied by [74], where they characterised feasibility by how freely one could choose the intrinsic growth rate vectors to allow the system to remain feasible. As a whole, these results strengthen connections between feasibility and RMT systems, whilst also adding biological realism.

Along the way, we managed to analytically approximate the distribution of $\boldsymbol{x}^*$ as a function of the system parameters n, C, σ and ρ. In doing so, we emphasise how the small covariance between the abundances of species can partly explain the observation of [34] that the effect of interaction structure on feasibility completely disappears as $n \to \infty$. Most importantly, our approximation of the distribution of $\boldsymbol{x}^*$ has allowed us to check the utility of Gibbs' assumption of independence between $\boldsymbol{x}^*$ and A in predicting the eigenvalue distribution of the GLV Jacobian for systems with Allesina and Tang type interaction matrices [67, 208]. Figure 2.4 shows that Gibbs' assumption can be used to accurately predict the effect of interaction structure [5] on the eigenvalue distribution of feasible random GLV models. It is worth noting that since A is random, the outlier eigenvalue of the GLV Jacobian would vary with each realisation of A, which makes the stability and resilience of the GLV model a random variable as well. Nevertheless, we have shown in Appendix VI that Gibbs' assumption allows us to accurately predict the effect that interaction structure has on the average resilience over multiple realisations of A.

It is of note that our method for calculating the feasibility probability relies on several assumptions on the parameter values to ensure accuracy (see Appendix I.E and II). We also assumed that x_i^* is normally distributed. Since the Neumann series approximation for x_i^* is normal in the limit $n \to \infty$, and is convergent if and only if $\sigma\sqrt{nC} < 1$, our method is accurate for large n and small σ (see Appendix VIII). Since the Neumann series expansion is precise, it is straightforward to extend our analysis to arbitrary orders of precision by working to higher orders in σ (see Eq. (2.9)).

The concept of feasibility has been associated with the extinction probability. It was summarised by Stone (1988) [210] that a higher feasibility probability is linked

to the reduction in the probability of extinction following structural disturbances, which are perturbations to the interaction coefficients that cause changes to the dynamical mode of the GLV model. Our results imply that increasing predator-prey interactions reduces the chance of extinction following structural disturbances.

We have used the assumption of May (1972) that all species are self-regulating. This is representative of natural ecosystems since they require 50 percent of species to self-regulate to allow for stability [8]. However, the assumption that $r_i = 1 \; \forall i \in [1, n]$ may not be biologically realistic, as natural ecosystems contain consumer species which do not grow in isolation. This is an interesting area for future investigation, however it was suggested by Song *et al.* [202] that this assumption gives the parameter region where feasible systems are likely to be present.

Having generalised the distribution of $\boldsymbol{x}^*$ to account for arbitrary ρ, we have opened up the possibility for extending the results of Gibbs *et al.* [67] to analytically predict the boundary of the eigenvalue distribution of the GLV Jacobian of such systems. This would enable us to calculate the stability of such GLV models. One potential method to perform this calculation is by applying the cavity method as detailed in [67]. It may also be possible to calculate the expected value of $-\min_{i\in\{1,n\}} x_i^*$ by applying order statistics as detailed in [175], and thus the expected resilience of a GLV model with a given value of ρ, although this is only applicable to systems where $|\rho|$ is small. We note, also, that the analytical approaches central to this study lead to predictions of normal distributions of steady-state species abundances. Empirical evidence is typically scale-dependent and points to a range of more complex possible species-abundance distributions [7] and the development of scale-dependent theory to bridge this gap with models may be a fruitful line of further enquiry.

Overall, our analyses, combined with [5, 34, 175] show that increasing the proportion of predator-prey interactions not only increases feasibility, but also the resilience of feasible GLV models. This provides greater support to Allesina and Tang's [5] conclusion that predator-prey interactions are stabilising whilst competitive/mutualistic interactions are destabilising.

2.5 APPENDIX I: ANALYTICALLY APPROXIMATING $Var(x_i^*)$ TO ORDER σ^6

2.5.1 COEFFICIENT OF σ^4 IN $Var(x_i^*)$

We first approximate the coefficient of σ^4 in $Var(x_i^*)$. To do this we specify the Taylor expansion of x_i^* in σ in index notation. In matrix form, the Taylor expansion of $\boldsymbol{x}^*$ to order σ^4 is

$$\boldsymbol{x}^* = (\mathbf{I} + \sigma\mathcal{E} + \sigma^2\mathcal{E}^2 + \sigma^3\mathcal{E}^3 + \sigma^4\mathcal{E}^4 + O(\sigma^5))\mathbf{r} \tag{2.17}$$

where

$$\mathcal{E} = \begin{pmatrix} 0 & a_{12} & \dots & a_{1n} \\ & \ddots & & \\ & & 0 & \\ a_{n1} & & & 0 \end{pmatrix} \quad , \quad \mathbf{r} = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}. \tag{2.18}$$

In index notation, Eq. (2.17) can be expressed as

$$x_i^* = 1 + \sigma\mathcal{E}_{ij} + \sigma^2(\mathcal{E}^2)_{ij} + \sigma^3(\mathcal{E}^3)_{ij} + \sigma^4(\mathcal{E}^4)_{ij} + O(\sigma^5) \tag{2.19}$$

since $r_i = 1$ for all $i \in [1, n]$. All terms in Eq. (2.19) represent terms to be summed over. The subscript i in Eq. (2.19) is the free index while all other indicies are dummy indicies. Please note that terms such as $\mathcal{E}_{ij}$ denote vectors and not matrices, since i is the free index. $Var(x_i^*)$ is defined by the equation

$$Var(x_i^*) = E(x_i^{*2}) - E(x_i^*)^2 \tag{2.20}$$

so we need to we seek the second moment of x_i^*, which can be found using Eq. (2.19). The expression for x_i^{*2} is deduced by squaring Eq. (2.19). Since the expectation of all terms of odd powers of σ is 0, we can safely ignore them, which gives

$$x_i^{*2} = 1+\sigma^2\Big(2(\mathcal{E}^2)_{ij}+\mathcal{E}_{ij}\mathcal{E}_{ik}\Big)+\sigma^4\Big(2(\mathcal{E}^4)_{ij}+2(\mathcal{E}^3)_{ij}(\mathcal{E})_{ik}+(\mathcal{E}^2)_{ij}(\mathcal{E}^2)_{ik}\Big)+O(\sigma^6) \tag{2.21}$$

which we can apply to calculate the second moment of x_i^*. We see from Eq. (2.21) that we need to determine $E((\mathcal{E}^4)_{ij})$, $E((\mathcal{E}^3)_{ij}(\mathcal{E})_{ik})$ and $E((\mathcal{E}^2)_{ij}(\mathcal{E}^2)_{ik})$. The expression for $(\mathcal{E}^4)_{im}$ can be expressed as a sum of terms involving products of interaction coefficients

$$(\mathcal{E}^4)_{im} = \sum_{k=1}^{n} \sum_{j\neq i, j\neq k}^{n} \sum_{l\neq k, l\neq m}^{n} a_{ij}a_{jk}a_{kl}a_{lm} \tag{2.22}$$

and the expression for $(\mathcal{E}^4)_{ij}$ is

$$(\mathcal{E}^4)_{ij} = (\mathcal{E}^4)_{ii} + (\mathcal{E}^4)_{ij}\mathbf{1}_{\{j\neq i\}} \tag{2.23}$$

we first determine the expectation of $(\mathcal{E}^4)_{ii}r_i$. We see from Eq. (2.22) that the expression for $(\mathcal{E}^4)_{ii}$ is

$$(\mathcal{E}^4)_{ii} = \sum_{k=1}^{n}\sum_{j\neq i,j\neq k}^{n}\sum_{l\neq k}^{n} a_{ij}a_{jk}a_{kl}a_{li}. \tag{2.24}$$

It is also possible for Eq. (2.24) to have terms where $i = k$, $j = l$ or both, which give rise to terms of Eq. (2.24) with nonzero expectation. To represent the case where $i = k$ but $j \neq l$, Eq. (2.24) is multiplied by δ_{ik} and to represent the case where $j = l$ but $i \neq k$, Eq. (2.24) is multiplied by δ_{jl}. To represent the case where both $i = k$ and $j = l$, we multiply Eq. (2.24) by $\delta_{ik}\delta_{jl}$. The expectation of Eq. (2.24) is

$$\begin{aligned} E((\mathcal{E}^4)_{ii}) &= E\left(\sum_{k=1}^{n}\sum_{j\neq i,j\neq k}^{n}\sum_{l\neq k}^{n} a_{ij}a_{jk}a_{kl}a_{li}(\delta_{ik}+\delta_{jl}+\delta_{ik}\delta_{jl})\right) &(2.25)\\ &= \sum_{j\neq i,j\neq l}^{n}\sum_{l\neq i}^{n} E(a_{ij}a_{ji}a_{il}a_{li}) + \sum_{j\neq i}^{n}\sum_{k\neq j,k\neq i}^{n} E(a_{ij}a_{jk}a_{kj}a_{ji}) + \sum_{j\neq i}^{n} E(a_{ij}^2 a_{ji}^2)\\ &= (n-1)(1+2(n-1)\rho^2). &(2.26)\end{aligned}$$

It is also proven in Section 2.5.2 below that the second summation term of Eq. (2.23) $(\mathcal{E}^4)_{ij}\mathbf{1}_{\{j\neq i\}}$ has an expectation of 0, which implies that

$$\begin{aligned} E((\mathcal{E}^4)_{ij}) &= E((\mathcal{E}^4)_{ii}) &(2.27)\\ &= 2(n-1)(n-2)\rho^2 + (n-1)(1+2\rho^2)\\ &= (n-1)(1+2(n-1)\rho^2).\end{aligned}$$

When generalised to account for C, this expression becomes

$$E((\mathcal{E}^4)_{ij}) = (n-1)(C+\rho^2(2C+2C^2(n-2))). \tag{2.28}$$

This generalisation is done by counting the number of pairwise uncorrelated terms in the expectation e.g in $E(a_{ij}a_{ji}a_{il}a_{li})$, the variables (a_{ij}, a_{ji}) and (a_{il}, a_{li}) are uncorrelated, so a factor of C^2 would be present in $E(a_{ij}a_{ji}a_{il}a_{li})$. It is crucial to note that Eq. (2.28) is also the coefficient of σ^4 in the expression for $E(x_i^*)$. The expression for $(\mathcal{E}^3)_{ij}\mathcal{E}_{ik}$ can be expressed as a sum of terms involving products of interaction coefficients

$$(\mathcal{E}^3)_{ij}\mathcal{E}_{ik} = \sum_{k=1}^{n}\sum_{j\neq i,j\neq k}^{n}\sum_{l\neq k,l\neq m}^{n} a_{ij}a_{ik}a_{kl}a_{lm}. \tag{2.29}$$

The terms that give rise to nonzero expectation of Eq. (2.29) include those where $m = k$ and $i = l$, $l = i$ and $m = j$ and those where $m = k$ only.

$$\begin{aligned} E((\mathcal{E}^3)_{ij}\mathcal{E}_{ik}) &= E\Bigg(\sum_{k=1}^{n}\sum_{j\neq i,j\neq k}^{n}\sum_{l\neq k}^{n} a_{ij}a_{ik}a_{kl}a_{lm}(\delta_{il}\delta_{mk} + \delta_{il}\delta_{mj} + \delta_{mk})\Bigg) \qquad (2.30)\\ &= \sum_{j\neq i,j\neq l}^{n}\sum_{l\neq i}^{n} E(a_{ij}^2 a_{ik}a_{ki}) + \sum_{j\neq i}^{n}\sum_{k\neq j,k\neq i}^{n} E(a_{ik}^3 a_{ki}) + \sum_{j\neq i}^{n} E(a_{ik}^2 a_{kl}a_{lk})\\ &= 3(n-1)\rho + 2(n-1)(n-2)\rho. \qquad (2.31)\end{aligned}$$

The expression for $(\mathcal{E}^2)_{ij}(\mathcal{E}^2)_{ik}$ can be expressed as a sum of terms involving products of interaction coefficients

$$(\mathcal{E}^2)_{ij}(\mathcal{E}^2)_{ik} = \sum_{k=1}^{n}\sum_{j\neq i,j\neq k}^{n}\sum_{l\neq k,l\neq m}^{n} a_{ij}a_{jk}a_{il}a_{lm}. \qquad (2.32)$$

The terms that give rise to nonzero expectation of Eq. (2.32) include those where $k = i$ and $l = j$ and $m = i$, $k = m$ and $j = l$ and those where $k = i$ and $m = i$.

$$\begin{aligned} E((\mathcal{E}^2)_{ij}(\mathcal{E}^2)_{ik}) &= E\Bigg(\sum_{k=1}^{n}\sum_{j\neq i,j\neq k}^{n}\sum_{l\neq k}^{n} a_{ij}a_{jk}a_{il}a_{lm}(\delta_{ki}\delta_{lj}\delta_{mi} + \delta_{km}\delta_{jl} + \delta_{ki}\delta_{mi})\Bigg)\\ &= \sum_{j\neq i,j\neq l}^{n}\sum_{l\neq i}^{n} E(a_{ij}^2 a_{ji}^2) + \sum_{j\neq i}^{n}\sum_{k\neq j,k\neq i}^{n} E(a_{ij}^2 a_{jk}^2) + \sum_{j\neq i}^{n} E(a_{ij}a_{ji}a_{il}a_{li})\\ &= (n-1)\Big((n-1) + n\rho^2\Big). \qquad (2.33)\end{aligned}$$

We have also ensured that no terms in all the summations above are double counted. Substituting all the results derived here into Eq. (2.21), we see that the coefficient of $Var(x_i^*)$ at order σ^4 is $(n-1)^2 + \rho(n-1)(4n-2) + \rho^2(n-1)$, and combined with the coefficient of $Var(x_i^*)$ at order σ^2 we get

$$Var(x_i^*) = (n-1)\sigma^2 + \sigma^4\Bigg((n-1)^2 + \rho(n-1)(4n-2) + \rho^2(n-1)\Bigg) + O(\sigma^6). \qquad (2.34)$$

We can generalise the expression for $Var(x_i^*)$ to account for the connectance C. Since a_{ij} and a_{ji} are zero with probability $(1-C)$ and are sampled from a bivariate normal distribution with probability C and $Corr(a_{ij}, a_{ji}) = \rho$, we have that

$$\begin{aligned} Var(x_i^*) = (n-1)C\sigma^2 + \sigma^4\Bigg(&[C(n-1) + C^2(n-1)(n-2)] +\\ &\rho[4C^2(n-1)(n-2) + 6C(n-1)] +\\ &\rho^2[2C(n-1) + C^2(n-1)(n-2) - (n-1)^2C^2]\Bigg) + O(\sigma^6). \qquad (2.35)\end{aligned}$$

The analytics in Eq. (2.35) show that for the case where $C < 1$, sampling a_{ij} and a_{ji} from a bivariate distribution with probability C gives a different $Var(x_i^*)$ to when A is sampled randomly with connectance C, as was done by May [146].

2.5.2 PROOF THAT $\sum_{j\neq i}^{n}(\mathcal{E}^4)_{ij}r_j$ HAS ZERO EXPECTATION

In the particular scenario where $i \neq m$ in Eq. (2.22), the possible scenarios are $m = j$, $m = k$ and $m = l$. In the case where $m = j$, Eq. (2.22) becomes

$$(\mathcal{E}^4)_{ij} = \sum_{k=1}^{n}\sum_{j\neq i,j\neq k}^{n}\sum_{l\neq k,l\neq j}^{n} a_{ij}a_{jk}a_{kl}a_{lj} \tag{2.36}$$

and it is possible that $i = k$, $j = l$ or $i = l$. If $j = l$, Eq. (2.36) is 0 since $\mathcal{E}_{ii} = 0$ for all $i \in [1, n]$. We therefore consider the cases $i = l$ and $i = k$ in the case where $m = j$. The expression for the expectation of $(\mathcal{E}^4)_{ij}$ is

$$\begin{aligned} E((\mathcal{E}^4)_{ij}) &= E\left(\sum_{k=1}^{n}\sum_{j\neq i,j\neq k}^{n}\sum_{l\neq k}^{n} a_{ij}a_{jk}a_{kl}a_{lj}(\delta_{ik}+\delta_{il}+\delta_{ik}\delta_{il})\right) \\ &= \sum_{j\neq i,j\neq l}\sum_{l\neq i} E(a_{ij}a_{ji}a_{il}a_{lj}) + \sum_{j\neq i}\sum_{k\neq j,k\neq i} E(a_{ij}a_{jk}a_{ki}a_{ij}) + 0 \\ &= 0. \end{aligned} \tag{2.37}$$

Now we consider the case where $m = k$. In this case, it is possible that $j = l$ and $i = l$. It is not possible for $i = k$ since we are considering the scenario in Eq. (2.22) where $i \neq m$. The expression for the expectation of $(\mathcal{E}^4)_{ik}$ is

$$\begin{aligned} E((\mathcal{E}^4)_{ik}) &= E\left(\sum_{k=1}^{n}\sum_{j\neq i,j\neq k}^{n}\sum_{l\neq k}^{n} a_{ij}a_{jk}a_{kl}a_{lk}(\delta_{il}+\delta_{jl}+\delta_{jl}\delta_{il})\right) \\ &= \sum_{j\neq i,j\neq l}\sum_{l\neq i} E(a_{ij}a_{jk}a_{ki}a_{ik}) + \sum_{j\neq i}\sum_{k\neq j,k\neq i} E(a_{ij}a_{jk}a_{kj}a_{jk}) + 0 \\ &= 0. \end{aligned} \tag{2.38}$$

Finally, in the case where $m = l$, we have that $E((\mathcal{E}^4)_{il}) = 0$ since $(\mathcal{E}^4)_{il}$ involves a product of $\mathcal{E}_{lm}$ which equals 0 if $l = m$. We have therefore proven that $\sum_{j\neq i}^{n}(\mathcal{E}^4)_{ij}r_j$ has an expectation of 0.

2.5.3 COEFFICIENT OF σ^4 IN $Cov(x_i^*, x_j^*)$

Using the same technique as entailed in Section 2.5.1, here we provide the calculation of the σ^4 coefficient of $Cov(x_i^*, x_j^*)$. Firstly, $Cov(x_i^*, x_j^*)$ is defined by the equation

$$Cov(x_i^*, x_j^*) = E(x_i^* x_j^*) - E(x_i^*)E(x_j^*). \tag{2.39}$$

Due to symmetry in the expression for $\boldsymbol{x}^*$, $E(x_i^*)$ is identical to $E(x_j^*)$ and so Eq. (2.39) simplifies to

$$Cov(x_i^*, x_j^*) = E(x_i^* x_j^*) - E(x_i^*)^2. \tag{2.40}$$

Here, the only extra number we need to calculate is $E(x_i^* x_j^*)$. Again in index notation, the expression for x_i^* is given by

$$x_i^* = 1 + \sigma\mathcal{E}_{ik} + \sigma^2(\mathcal{E}^2)_{ik} + \sigma^3(\mathcal{E}^3)_{ik} + \sigma^4(\mathcal{E}^4)_{ik} + O(\sigma^5) \tag{2.41}$$

and x_j^* is given by

$$x_j^* = 1 + \sigma\mathcal{E}_{jl} + \sigma^2(\mathcal{E}^2)_{jl} + \sigma^3(\mathcal{E}^3)_{jl} + \sigma^4(\mathcal{E}^4)_{jl} + O(\sigma^5) \tag{2.42}$$

$$x_i^* x_j^* = 1 + \sigma^2\Big(2(\mathcal{E}^2)_{ik} + \mathcal{E}_{ik}\mathcal{E}_{jl}\Big) + \sigma^4\Big(2(\mathcal{E}^4)_{ik} + 2(\mathcal{E}^3)_{ik}\mathcal{E}_{jl} + (\mathcal{E}^2)_{ik}(\mathcal{E}^2)_{jl}\Big). \tag{2.43}$$

We can calculate $E(x_i^* x_j^*)$ directly from Eq. (2.43). From Eq. (2.43), we see that the extra terms we need to determine are $E[(\mathcal{E}^3)_{ik}\mathcal{E}_{jl}]$ and $E[(\mathcal{E}^2)_{ik}(\mathcal{E}^2)_{jl}]$. $E[(\mathcal{E}^4)_{ik}]$ is already determined from our calculation of the σ^4 coefficient of $Var(x_i^*)$ in Section 2.5.1. Firstly, the expression for $(\mathcal{E}^3)_{ik}\mathcal{E}_{jl}$ can be expressed as

$$(\mathcal{E}^3)_{ik}\mathcal{E}_{jl} = \sum_{k=1,k\neq o}^{n}\sum_{m\neq i,m\neq o}^{n}\sum_{l\neq k,l\neq m}^{n}\sum_{o\neq m,o\neq k}^{n} a_{jl}a_{im}a_{mo}a_{ok}. \tag{2.44}$$

The terms that give rise to nonzero expectation of Eq. (2.44) include those where $j = o$ and $k = l$, $o = l$ and $k = j$, $o = i$ and $m = k$, and simply $m = k$. The expectation of Eq. (2.44) is

$$\begin{aligned} E[(\mathcal{E}^3)_{ik}\mathcal{E}_{jl}] &= E\Bigg(\sum_{k=1,k\neq o}^{n}\sum_{m\neq i,m\neq o}^{n}\sum_{l\neq k,l\neq m}^{n}\sum_{o\neq m,o\neq k}^{n} a_{jl}a_{im}a_{mo}a_{ok}\Big(\delta_{jo}\delta_{kl} + \delta_{lo}\delta_{kj} + \delta_{oi}\delta_{mk} + \delta_{mk}\Big)\Bigg) \\ &= \sum_{l\neq j}^{n}\sum_{m\neq i,m\neq j}^{n} E(a_{jl}a_{im}a_{mj}a_{jl}) + \sum_{l\neq j,l\neq m}^{n}\sum_{m\neq i,m\neq l}^{n} E(a_{jl}a_{im}a_{ml}a_{lj}) \\ &+ \sum_{k\neq i}^{n}\sum_{l\neq j}^{n} E(a_{jl}a_{ki}a_{ik}a_{ki}) + \sum_{l\neq j}^{n}\sum_{k\neq i,k\neq o}^{n}\sum_{o\neq k}^{n} E(a_{jl}a_{ik}a_{ko}a_{ok}). \end{aligned} \tag{2.45}$$

We now calculate the expectations of each of the individual summation terms in the second and third rows of Eq. (2.45).

$$\sum_{l\neq j}^{n}\sum_{m\neq i,m\neq j}^{n} E(a_{jl}a_{im}a_{mj}a_{jl}) = 0 \tag{2.46}$$

$$\begin{aligned}\sum_{l\neq j}^{n}\sum_{m\neq i,m\neq j}^{n} E(a_{jl}a_{im}a_{ml}a_{lj}) &= \sum_{m\neq i}^{n} E(a_{ji}a_{im}a_{mi}a_{ij}) + E(a_{ij}^2a_{ji}^2) \qquad (2.47)\\ &= (1+2\rho^2)+(n-2)\rho^2\end{aligned}$$

$$\sum_{l\neq j}^{n}\sum_{m\neq i,m\neq j}^{n} E(a_{jl}a_{ik}a_{ki}a_{ik}) = E(a_{ij}^2a_{ji}^2) = (1+2\rho^2) \qquad (2.48)$$

$$\sum_{l\neq j}^{n}\sum_{k\neq i,l\neq n}^{n}\sum_{o\neq i}^{n} E(a_{jl}a_{ik}a_{ko}a_{ok}) = \sum_{o\neq i}^{n} E(a_{ij}a_{ji}a_{io}a_{oi}) = (n-2)\rho^2. \qquad (2.49)$$

Noting that Eq. (2.48) is already counted into Eq. (2.47), we neglect Eq. (2.48) to prevent double counting. Summing Eq. (2.46), Eq. (2.47) and Eq. (2.49) we see that

$$E[(\mathcal{E}^3)_{ik}\mathcal{E}_{jl}] = 1+2(n-1)\rho^2. \qquad (2.50)$$

The expression for $(\mathcal{E}^2)_{ik}(\mathcal{E}^2)_{jl}$ is given by

$$(\mathcal{E}^2)_{ik}(\mathcal{E}^2)_{jl} = \sum_{k=1,k\neq m}^{n}\sum_{m\neq i,m\neq k}^{n}\sum_{l\neq o}^{n}\sum_{o\neq j,o\neq l}^{n} a_{im}a_{mk}a_{jo}a_{ol}. \qquad (2.51)$$

The terms that give rise to nonzero expectation of Eq. (2.51) include those where $m=j$ and $i=k$, $l=j$, and $m=j$ and $i=o$. The expectation of Eq. (2.51) is

$$\begin{aligned}E[(\mathcal{E}^2)_{ik}(\mathcal{E}^2)_{jl}] &= E\left(\sum_{k=1,k\neq m}^{n}\sum_{m\neq i,m\neq k}^{n}\sum_{l\neq o}^{n}\sum_{o\neq j,o\neq l}^{n} a_{im}a_{mk}a_{jo}a_{ol}(\delta_{oi}\delta_{lm}+\delta_{mo}\delta_{li}+\delta_{lj})\right)\\ &= \sum_{m\neq i,m\neq k}^{n}\sum_{k\neq m}^{n} E(a_{im}a_{mk}a_{ji}a_{im}) + \sum_{m\neq i,m\neq j}^{n}\sum_{k\neq m}^{n} E(a_{im}a_{mk}a_{jm}a_{mi})\\ &+ \sum_{m\neq i,m\neq k}^{n}\sum_{k\neq m}^{n}\sum_{o\neq j}^{n} E(a_{im}a_{mk}a_{jo}a_{oj}). \qquad (2.52)\end{aligned}$$

We now calculate the expectations of each individual summation term in the second line of Eq. (2.52).

$$\sum_{m\neq i,m\neq k}^{n}\sum_{k\neq m}^{n} E(a_{im}a_{mk}a_{ji}a_{im}) = E(a_{ij}^2a_{ji}^2) = (1+2\rho^2) \qquad (2.53)$$

$$\begin{aligned}\sum_{m\neq i,m\neq k}^{n}\sum_{k\neq m}^{n} E(a_{im}a_{mk}a_{jm}a_{mi}) &= \sum_{m\neq i,m\neq j}^{n} E(a_{im}a_{mj}a_{jm}a_{mi}) \qquad (2.54)\\ &= (n-2)\rho^2\end{aligned}$$

$$\begin{aligned}\sum_{m\neq i,m\neq k}^{n}\sum_{k\neq m}^{n}\sum_{o\neq j}^{n} E(a_{im}a_{mk}a_{jo}a_{oj}) &= \sum_{o\neq j}^{n}\sum_{m\neq i}^{n} E(a_{im}a_{mi}a_{jo}a_{oj}) \qquad (2.55)\\ &= ((n-1)^2-1)\rho^2\end{aligned}$$

Summing Eq. (2.53), Eq. (2.54) and Eq. (2.55) we see that

$$E[(\mathcal{E}^2)_{ik}(\mathcal{E}^2)_{jl}] = 1 + n(n-1)\rho^2. \tag{2.56}$$

Substituting Eq. (2.56) and Eq. (2.50) into the expression or the expectation of Eq. (2.43), we can determine the σ^4 coefficient of $Cov(x_i^*, x_j^*)$ which is given by $3 + 5(n-1)\rho^2$. The expression for $Cov(x_i^*, x_j^*)$ up to and including order σ^4 is

$$Cov(x_i^*, x_j^*) = \rho\sigma^2 + (3 + 5(n-1)\rho^2)\sigma^4 \tag{2.57}$$

and for the case where $C < 1$, it can be generalised to

$$Cov(x_i^*, x_j^*) = \rho C\sigma^2 + (3 + (6 + C(5n-11))\rho^2)\sigma^4. \tag{2.58}$$

2.5.4 COEFFICIENT OF σ^6 IN $Var(x_i^*)$

To obtain the coefficient of σ^6 in $Var(x_i^*)$, we apply the Taylor expansion of x_i^* up to order σ^6, given by

$$\boldsymbol{x}^* = (\mathbf{I} + \sigma\mathcal{E} + \sigma^2\mathcal{E}^2 + \sigma^3\mathcal{E}^3 + \sigma^4\mathcal{E}^4 + \sigma^5\mathcal{E}^5 + \sigma^6\mathcal{E}^6 + O(\sigma^7))\mathbf{r} \tag{2.59}$$

which in index notation is

$$x_i^* = 1 + \sigma\mathcal{E}_{ij} + \sigma^2(\mathcal{E}^2)_{ij} + \sigma^3(\mathcal{E}^3)_{ij} + \sigma^4(\mathcal{E}^4)_{ij} + \sigma^5(\mathcal{E}^5)_{ij} + \sigma^6(\mathcal{E}^6)_{ij} + O(\sigma^7). \tag{2.60}$$

Again, we need to apply Eq. (2.20), which requires us to calculate the second moment of x_i^* as well as $E(x_i^*)^2$. To calculate $E(x_i^*)$, we need to calculate $E((\mathcal{E}^6)_{ij})$. To calculate $E(x_i^{*2})$, we need to calculate the expectation of the square of Eq. (2.60) with terms of order higher than σ^6 truncated. At order σ^6, it is convenient to adopt a different notation to that detailed in Section 2.5. Here, we re-express the matrix $\mathcal{E}$ in the form

$$\mathcal{E} = \sum_{i=1}^{n-1} \sum_{j=2, j>i}^{n} \Phi_{ij} \tag{2.61}$$

where Φ_{ij} represents matrices such that their (i,j)-th component is a_{ij} (i.e. $(\Phi_{ij})_{ij} = a_{ij}$ and $(\Phi_{ij})_{ji} = a_{ji}$ for $j > i$) and all other entries are zero. The coefficients of x_i^{*2} at order σ^6 are given in Eq. (2.62) below.

$$2(\mathcal{E}^6)_{ij} + 2(\mathcal{E}^5)_{ij}\mathcal{E}_{ik} + 2(\mathcal{E}^4)_{ij}(\mathcal{E}^2)_{ik} + (\mathcal{E}^3)_{ij}(\mathcal{E}^3)_{ik} \tag{2.62}$$

and the coefficients of $E(x_i^*)^2$ at order σ^6 are

$$2E[(\mathcal{E}^6)_{ij}] + 2E[(\mathcal{E}^4)_{ii}]E[(\mathcal{E}^2)_{ii}] \tag{2.63}$$

and the coefficient of $Var(x_i^*)$ at order σ^6 referred to as v_6 in the main text is given by

$$2E[(\mathcal{E}^5)_{ij}\mathcal{E}_{ik}] + 2E[(\mathcal{E}^4)_{ij}(\mathcal{E}^2)_{ik}] + E[(\mathcal{E}^3)_{ij}(\mathcal{E}^3)_{ik}] - 2E[(\mathcal{E}^4)_{ii}]E[(\mathcal{E}^2)_{ii}] \tag{2.64}$$

obtained by subtracting Eq. (2.63) from Eq. (2.62). To calculate the coefficient of σ^6 in $Var(x_i^*)$, it is necessary to calculate the expectation of $(\mathcal{E}^5)_{ij}(\mathcal{E})_{ik}$, $(\mathcal{E}^4)_{ij}(\mathcal{E}^2)_{ik}$ and $(\mathcal{E}^3)_{ij}(\mathcal{E}^3)_{ik}$ in Eq. (2.64). To calculate the expectations of these terms, we consider the expression for the sixth power of Eq. (2.61) in terms of the Φ_{ij} terms. In Sections 2.5.4.1 to 2.5.4.4, we describe the method of calculating the expectations of each of the terms in Eq. (2.64). We note that to obtain an approximation of $E(x_i^*)$ to order σ^6, it is necessary to calculate $E[(\mathcal{E}^6)_{ij}]$. The principles used to calculate this are identical to those used to calculate all the terms in Eq. (2.64), and we begin this subsection by explaining the procedures for calculating $E[(\mathcal{E}^6)_{ij}]$.

2.5.4.1 Calculating $E((\mathcal{E}^6)_{ij})$

Only terms in $(\mathcal{E}^6)_{ij}$ involving products of certain Φ_{ij} terms have nonzero expectation, such as $\Phi_{12}^2\Phi_{13}^2\Phi_{23}^2$ which is the product of the square of three Φ_{ij} terms that involve mutually uncorrelated variables. As another example, terms of the form $\Phi_{ij}^4\Phi_{ik}^2$ for $j \neq k$ also has nonzero expectation. Another term in $(\mathcal{E}^6)_{ij}$ with nonzero expectation is Φ_{ij}^6. To count the number of terms of each form, we consider in Table 2.1 below the set of all forms in which the terms having nonzero expectation can take. Let A, B and C denote the Φ_{ij} matrices with variables that are all mutually uncorrelated when $\rho \neq 0$ e.g. $A = \Phi_{12}$, $B = \Phi_{13}$ and $C = \Phi_{23}$, we have that

$A^2B^2C^2$	A^4B^2	A^6
$\Phi_{ij}^2\Phi_{ik}^2\Phi_{jk}^2$	$\Phi_{ij}^4\Phi_{ik}^2$	Φ_{ij}^6
$\Phi_{ij}^2\Phi_{ik}^2\Phi_{il}^2$	$\Phi_{ij}^4\Phi_{jk}^2$	
$\Phi_{ij}^2\Phi_{ik}^2\Phi_{kl}^2$	$\Phi_{ij}^4\Phi_{kl}^2$	
$\Phi_{ij}^2\Phi_{jk}^2\Phi_{kl}^2$		
$\Phi_{ij}^2\Phi_{ik}^2\Phi_{jl}^2$		
$\Phi_{ij}^2\Phi_{jk}^2\Phi_{jl}^2$		
$\Phi_{ij}^2\Phi_{jk}^2\Phi_{il}^2$		

Table 2.1: Set of all forms of different terms of $\mathcal{E}^6$ which have nonzero expectations. We specifically do not allow any pair of indicies to be equal i.e. $i \neq j$, $j \neq k$. The first column ($A^2B^2C^2$) involves products of 3 different matrices that contain mutually uncorrelated variables, while the second column (A^4B^2) involves products of 2 different matrices that contain mutually uncorrelated variables. i is assumed to be the free index throughout Section 2.5.4, with all other indicies dummy indicies.

Since matrix multiplication is not commutative, we now consider the set of all permutations of each entry of Table 2.1 above. First, we seek the set of all permutations of terms of each form that gives rise to a nonzero sum of first row (first row sum) (we take the first row $i = 1$ without loss of generality). $\Phi_{ij}^2\Phi_{ik}^2\Phi_{jk}^2$ contains 6 permutations that give rise to a nonzero first row sum. The expectation of the first row sum of each matrix of this form (i.e expectation of the first row sum of $\Phi_{12}\Phi_{13}\Phi_{23}$) is ρ^3. Since two indicies (j and k) are summed over, there exists $(n-1)(n-2)$ terms in the sum. There exists 6 permutations with nonzero first row sums, but three of these 6 permutations are double countings of the other three. The expectation of $\Phi_{ij}^2\Phi_{ik}^2\Phi_{jk}^2$ is thus $3(n-1)(n-2)\rho^3$.

$\Phi_{ij}^2\Phi_{ik}^2\Phi_{il}^2$ contains 2 distinct permutations that give rise to nonzero first row sum, and with 3 indicies summed over, there are $((n-1)^3 - (n-1)^2 - 2(n-1)(n-2))$ terms in the sum. Subtraction of $(n-1)^2 + 2(n-1)(n-2)$ from $(n-1)^3$ is to ensure no double countings in the sum over 3 indices. Since the expectation of the first row sum of each matrix of this form is also ρ^3, the expectation of $\Phi_{ij}^2\Phi_{ik}^2\Phi_{il}^2$ is thus $2((n-1)^3 - (n-1)^2 - 2(n-1)(n-2))\rho^3$. The expectation of each entry of Table 2.1 is deduced by following these steps.

1. Checking the number of permutations of each entry of Table 2.1 that has a nonzero sum of (first) row. We do this for the first row if we set our dummy index i to 1 without loss of generality Check if any pair of rows are double countings of each other.

2. Calculating the expectation of the first row sum of each permutation that has a nonzero first row sum.

3. Count the number of terms in the sum over its set of dummy indicies. We specifically do not allow any index to equal each other, not even the free index.

4. Check if any terms in the sums over its dummy indicies are double counted by relabelling dummy indicies. If there are double countings, subtract the number of double counted terms from the quantity deduced in 3. For example, if summing over 2 indicies and we specifically do not allow any index to equal each other, then there are $(n-1)^2-(n-1)$ terms, since $(n-1)^2$ terms are summed over and $(n-1)$ of them are double counted.

Finally we take the product of all quantities deduced in 1. to 3. above to give the expectation of each entry of Table 2.1. Repeating this procedure for all entries of Table 2.1 and summing, we find that

$$\begin{aligned} E((\mathcal{E}^6)_{ij}) &= (n-1)(n-2)+(n-1)(9\rho+6\rho^3) \\ &\quad +6(n-1)(n-2)\rho(1+2\rho^2) \\ &\quad +3(n-1)(n-2)\rho^3+5((n-1)^3-(n-1)^2-2(n-1)(n-2))\rho^3. \end{aligned} \tag{2.65}$$

If A has a connectance C and a_{ij} and a_{ji} are sampled from a bivariate distribution, then they are both nonzero with probability C, and so we have

$$\begin{aligned} E((\mathcal{E}^6)_{ij}) &= (n-1)(n-2)C^3+(n-1)(9\rho+6\rho^3)C \\ &\quad +6(n-1)(n-2)\rho(1+2\rho^2)C^2 \\ &\quad +3(n-1)(n-2)\rho^3C^3+5((n-1)^3-(n-1)^2-2(n-1)(n-2))\rho^3C^3. \end{aligned} \tag{2.66}$$

2.5.4.2 Calculating $E((\mathcal{E}^5)_{ij}\mathcal{E}_{ik})$

Now we seek to find $E((\mathcal{E}^5)_{ij}\mathcal{E}_{ik})$. Table 2.2 shows the set of all forms in which the terms of $(\mathcal{E}^5)_{ij}\mathcal{E}_{ik}$ having nonzero expectation can take.

$A^2B^2C^2$	A^4B^2	A^6
$(\Phi_{ij}^2\Phi_{jk}^2\Phi_{ik})\Phi_{ik}$	$(\Phi_{ij}^4\Phi_{ik})\Phi_{ik}$	$(\Phi_{ij}^5)\Phi_{ij}$
$(\Phi_{ij}^2\Phi_{ik}^2\Phi_{il})\Phi_{il}$	$(\Phi_{ij}^2\Phi_{ik}^3)\Phi_{ik}$	
$(\Phi_{ij}\Phi_{jk}^2\Phi_{jl}^2)\Phi_{ij}$	$(\Phi_{ij}^3\Phi_{jk}^2)\Phi_{ij}$	
$(\Phi_{ij}^2\Phi_{jk}^2\Phi_{il})\Phi_{il}$	$(\Phi_{ij}\Phi_{jk}^4)\Phi_{ij}$	
$(\Phi_{ij}\Phi_{ik}^2\Phi_{jl}^2)\Phi_{ij}$		
$(\Phi_{ij}\Phi_{ik}^2\Phi_{kl}^2)\Phi_{ij}$		

Table 2.2: Set of all forms of different terms of $(\mathcal{E}^5)_{ij}\mathcal{E}_{ik}$ which have nonzero expectations. Here, all the Φ_{ij} matrices inside the brackets represent matrices that comprise the term $(\mathcal{E}^5)_{ij}$ and the matrix outside the brackets represent the Φ_{ij} matrices that comprise $\mathcal{E}_{ik}$.

Next, we repeat step 1. detailed in the page above, but only for the terms inside the brackets. This gives $(\mathcal{E}^5)_{ij}$. After that, multiply the first row sums of all matrices with nonzero first row sum by the first row sum of the Φ outside the bracket. In other words, multiplying $(\mathcal{E}^5)_{ij}$ with $\mathcal{E}_{ik}$. We next calculate the expectation of this quantity. Then we repeat step 3 and 4. We find that

$$\begin{aligned} E((\mathcal{E}^5)_{ij}\mathcal{E}_{ik}) &= 2(n-1)(n-2)(1+2\rho^2) + 5(3\rho^2(n-1)(n-2)) + (n-1)(n-2)\rho \\ &\quad +5((n-1)^3 - (n-1)^2 - 2(n-1)(n-2))\rho^2 \\ &\quad +(n-1)(3+12\rho^2) \end{aligned} \tag{2.67}$$

and if A has connectance C with a_{ij} and a_{ji} sampled from a bivariate distribution, then we have

$$\begin{aligned} E((\mathcal{E}^5)_{ij}\mathcal{E}_{ik}) &= 2(n-1)(n-2)(1+2\rho^2)C^2 + 4(3\rho^2(n-1)(n-2))C^2 + \\ &\quad (3\rho^2(n-1)(n-2))C^3 + (n-1)(n-2)\rho C^3 + \\ &\quad 5((n-1)^3 - (n-1)^2 - 2(n-1)(n-2))\rho^2 C^3 + (n-1)(3+12\rho^2)C. \end{aligned} \tag{2.68}$$

2.5.4.3 Calculating $E((\mathcal{E}^4)_{ij}(\mathcal{E}^2)_{ik})$

Now we seek to find $E((\mathcal{E}^4)_{ij}(\mathcal{E}^2)_{ik})$. Table 2.3 shows the set of all forms in which the terms of $(\mathcal{E}^4)_{ij}(\mathcal{E}^2)_{ik}$ having nonzero expectation can take.

$A^2B^2C^2$	A^4B^2	A^6
$(\Phi_{ij}^2\Phi_{jk}^2)(\Phi_{ik}^2)$	$(\Phi_{ij}^2\Phi_{ik}^2)(\Phi_{ij}^2)$	$(\Phi_{ij}^4)(\Phi_{ij}^2)$
$(\Phi_{ij}\Phi_{ik}\Phi_{il}^2)(\Phi_{ij}\Phi_{ik})$	$(\Phi_{ij}^2\Phi_{jk}^2)(\Phi_{ij}^2)$	
$(\Phi_{ij}\Phi_{ik}\Phi_{jk}^2)(\Phi_{ij}\Phi_{ik})$	$(\Phi_{ij}^3\Phi_{ik})(\Phi_{ij}\Phi_{ik})$	
$(\Phi_{ij}\Phi_{ik}\Phi_{jl}^2)(\Phi_{ij}\Phi_{ik})$	$(\Phi_{ij}^3\Phi_{jk})(\Phi_{ij}\Phi_{jk})$	
$(\Phi_{ij}\Phi_{ik}\Phi_{kl}^2)(\Phi_{ij}\Phi_{ik})$	$(\Phi_{ij}\Phi_{jk}^3)(\Phi_{ij}\Phi_{jk})$	
$(\Phi_{ij}\Phi_{jk}\Phi_{ik}^2)(\Phi_{ij}\Phi_{jk})$	$(\Phi_{ij}^4)(\Phi_{ik}^2)$	
$(\Phi_{ij}\Phi_{jk}\Phi_{il}^2)(\Phi_{ij}\Phi_{jk})$		
$(\Phi_{ij}\Phi_{jk}\Phi_{jl}^2)(\Phi_{ij}\Phi_{jk})$		
$(\Phi_{ij}\Phi_{jk}\Phi_{kl}^2)(\Phi_{ij}\Phi_{jk})$		
$(\Phi_{ij}^2\Phi_{jk}^2)(\Phi_{il}^2)$		
$(\Phi_{ij}^2\Phi_{ik}^2)(\Phi_{il}^2)$		

Table 2.3: Set of all forms of different terms of $(\mathcal{E}^4)_{ij}(\mathcal{E}^2)_{ik}$ with nonzero expectations. Here, all the Φ_{ij} matrices inside the left bracket represent those that comprise the $(\mathcal{E}^4)_{ij}$ term and all the Φ_{ij}s in the right bracket represents all those in the $(\mathcal{E}^2)_{ik}$ term.

Now, it is necessary to seek the set of all $(\mathcal{E}^4)_{ij}(\mathcal{E}^2)_{ik}$ such that both $(\mathcal{E}^4)_{ij}$s and $(\mathcal{E}^2)_{ik}$s have nonzero first row sums. In other words, all the entries in Table 2.3 in which both their left and right brackets have nonzero first row sums. To do this, we repeat step 1 for both the left and right brackets of each entry of table 2.3, discarding any permutations of matrices in each bracket that have zero first row sums. Next, calculate the expectation of the product of the first row sum of the left bracket and the first row sum of the right bracket. Finally, repeat steps 3. and 4. We find that

$$\begin{aligned} E((\mathcal{E}^4)_{ij}(\mathcal{E}^2)_{ik}) &= 4(n-1)(n-2)\rho(1+2\rho^2) + 8\rho(n-1)(n-2) \\ &\quad +3\rho((n-1)^3-(n-1)^2-2(n-1)(n-2)) + (n-1)(n-2)\rho^3 \\ &\quad +(n-1)(n-2)\rho^2 + 2((n-1)^3-(n-1)^2-2(n-1)(n-2))\rho^3 \\ &\quad +(n-1)(9\rho+6\rho^3) \end{aligned} \tag{2.69}$$

and if A has connectance C with a_{ij} and a_{ji} sampled from a bivariate distribution, then we have

$$\begin{aligned} E((\mathcal{E}^4)_{ij}(\mathcal{E}^2)_{ik}) &= 4(n-1)(n-2)\rho(1+2\rho^2)C^2 + 6\rho(n-1)(n-2)C^2 \\ &\quad +2\rho(n-1)(n-2)C^3 + (n-1)(9\rho+6\rho^3)C \\ &\quad +3\rho((n-1)^3-(n-1)^2-2(n-1)(n-2))C^3 \\ &\quad +(n-1)(n-2)\rho^3C^3 + (n-1)(n-2)\rho^2C^3 \\ &\quad +2((n-1)^3-(n-1)^2-2(n-1)(n-2))\rho^3C^3. \end{aligned} \tag{2.70}$$

Now, we find $E((\mathcal{E}^3)_{ij}(\mathcal{E}^3)_{ik})$. Since $(\mathcal{E}^3)_{ij}(\mathcal{E}^3)_{ik}$ involves the product of two terms that contain a third power of $\mathcal{E}$, each one of which has a range of permutations that give rise to nonzero first row, $(\mathcal{E}^3)_{ij}(\mathcal{E}^3)_{ik}$ will have more permutations that give nonzero first row. For instance, the first $(\mathcal{E}^3)_{ij}$ can contain a product of either two or three different Φ_{ij}s. We therefore follow an alternative set of steps.

1. Find the set of all permutations of the tuples $(1,2,3,1,2,3)$ and $(1,1,1,1,2,2)$. Each number in the tuple represents a distinct Φ_{ij} matrix e.g. 1 representing Φ_{ij}, 2 for Φ_{ik}, 3 for Φ_{il}. Tuple $(1,2,3,1,2,3)$ represents the order of products of three distinct matrices $(A^2B^2C^2)$ while $(1,1,1,1,2,2)$ represents the order of products of two distinct matrices (A^4B^2). The permutations of each of these tuples should be expressed as a table.
2. Split the tables of permutations found in step 1. into two tables by separating columns 1 to 3 and 4 to 6. In each of these two tables, discard any repeated rows if any. The table containing columns 1 to 3 represents the set of all permutations of $(\mathcal{E}^3)_{ij}$ and the table containing columns 4 to 6 represents the set of all permutations of $(\mathcal{E}^3)_{ik}$.
3. Check which rows of both tables represent permutations that give nonzero first row sum in BOTH $(\mathcal{E}^3)_{ij}$ and $(\mathcal{E}^3)_{ik}$.
4. Repeat Steps 3 and 4 in the list above Eq. (2.65).

2.5.4.4 Calculating $E((\mathcal{E}^3)_{ij}(\mathcal{E}^3)_{ik})$

The entries of Table 2.4 represents the set of all Φ_{ij} matrices that comprise the $(\mathcal{E}^4)_{ij}(\mathcal{E}^2)_{ik}$s which have nonzero expectations.

$A^2B^2C^2$	A^4B^2	A^6
$\Phi_{ij}\Phi_{ik}\Phi_{il}$	$\Phi_{ij}\Phi_{ik}$	Φ_{ij}
$\Phi_{ij}\Phi_{ik}\Phi_{jk}$	$\Phi_{ij}\Phi_{jk}$	
$\Phi_{ij}\Phi_{ik}\Phi_{jl}$		
$\Phi_{ij}\Phi_{jk}\Phi_{jl}$		
$\Phi_{ij}\Phi_{jk}\Phi_{kl}$		
$\Phi_{ij}\Phi_{ik}\Phi_{kl}$		

Table 2.4: Each entry of the table contains the set of Φ_{ij} matrices that make up the $(\mathcal{E}^4)_{ij}(\mathcal{E}^2)_{ik}$ with nonzero expectations. The column $A^2B^2C^2$ are the set of products of three different Φ_{ij} matrices that can give nonzero first row sums depending on the permutation. A^4B^2 are the set of products of two different Φ_{ij} matrices that can give nonzero first row sums, and A^6 is the product of a one such matrix e.g. Φ_{ij}^6.

Using the method outlined immediately above, we find that

$$\begin{aligned} E((\mathcal{E}^3)_{ij}(\mathcal{E}^3)_{ik}) &= 4\rho^2((n-1)^3-(n-1)^2-2(n-1)(n-2)) \\ &\quad +2(n-1)(n-2)\rho^2+(n-1)(n-2)\rho^3+(n-1)(n-2) \\ &\quad +((n-1)^3-(n-1)^2-2(n-1)(n-2)) \\ &\quad +4(n-1)(n-2)3\rho^2+2(n-1)(n-2)(1+2\rho^2) \\ &\quad +(n-1)(3+12\rho^2) \end{aligned} \tag{2.71}$$

and if A has connectance C with a_{ij} and a_{ji} sampled from a bivariate distribution, then we have

$$\begin{aligned} E((\mathcal{E}^3)_{ij}(\mathcal{E}^3)_{ik}) &= 4\rho^2((n-1)^3-(n-1)^2-2(n-1)(n-2))C^3 \\ &\quad +2(n-1)(n-2)\rho^2C^3+(n-1)(n-2)\rho^3C^3+(n-1)(n-2)C^3 \\ &\quad +((n-1)^3-(n-1)^2-2(n-1)(n-2))C^3 \\ &\quad +4(n-1)(n-2)3\rho^2C^2+2(n-1)(n-2)(1+2\rho^2)C^2 \\ &\quad +(n-1)(3+12\rho^2)C. \end{aligned} \tag{2.72}$$

We can use Eq. (2.65) to Eq. (2.72) to calculate Eq. (2.64). Putting this together, we see that the order σ^6 coefficient of $Var(x_i^*)$ is equal to

$$\begin{aligned} &C(n-1)(9-12C-2C^2(n-2)+6Cn+C^2(n-2)n) \\ &+C(n-1)\Big(18-38C-12C^2(n-2)+18Cn+6C^2(n-2)n\Big)\rho \\ &+C(n-1)(36-96C-32C^2(n-2)+48Cn+14C^2(n-2)n)\rho^2 \\ &+C(n-1)(12-28C-5C^2(n-2)+12Cn)\rho^3. \end{aligned} \tag{2.73}$$

2.5.4.5 Number of Extra terms required to calculate $Cov(x_i^*, x_j^*)$ to order σ^6

We note that the fact that the expression for $Cov(x_i^*, x_j^*)$ involves an extra (free) index j makes the algebra required for its calculation lengthier. We have the additional term Φ_{ij} where both i and j are the free indices. For clarity, in this section we will denote the second free index as i' instead of j. The number of indices inside each bracket of the tables in Section 2.5.4 is at most the number of unique Φ plus 1 (i.e $(\Phi_{ij}\Phi_{ik}\Phi_{il}^2)$ in Section 2.5.4.3 has 3 unique Φs Φ_{ij}, Φ_{ik} and Φ_{il}, and 4 indices i,j,k and l). We formalise this as

$$N_{ind}^{max} = N_{uniq\Phi} + 1. \tag{2.74}$$

It is possible for the index i' to occur in the bracket a various number of times. The maximum number of times i' can occur in the left bracket is

$$max_{(\#i')} = N_{ind}^{max} - 1 \tag{2.75}$$

since the left bracket has to include the free index i. Just for terms of the form $(A^2B^2)C^2$ (see Table in Section 2.5.4.3), the number of terms needed is equal to 1 plus the number of terms of the form $(A^2B^2)C^2$ already present in $Var(x_i^*)$ ($N_{(A^2B^2)C^2}^{Var(x_i^*)}$) all times $max_{(\#i')}$, in other words

$$(N_{(A^2B^2)C^2}^{Var(x_i^*)} + 1)max_{(\#j)}. \tag{2.76}$$

When we have two free indices i and i', $N_{ind}^{max} = 3$ and $max_{(\#i')} = 2$. In the expression for $Var(x_i^*)$, there are 3 terms of the form $(A^2B^2)C^2$ so $N_{(A^2B^2)C^2}^{Var(x_i^*)} = 3$. Since $max_{(\#i')} = 2$, we have $(N_{(A^2B^2)C^2}^{Var(x_i^*)} + 1)max_{(\#i')} = 8$, and so we require 8 terms of the form $(A^2B^2)C^2$. Since it is also possible for the index i' to not occur inside the bracket (A^2B^2), we have an extra $(N_{(A^2B^2)C^2}^{Var(x_i^*)} + 1)$ terms which brings the total number of $(A^2B^2)C^2$ type terms to 12. The terms of the form $(A^2B^2)C^2$ are $(\Phi_{ij}^2\Phi_{jk}^2)(\Phi_{i'j}^2)$, $(\Phi_{ij}^2\Phi_{ik}^2)(\Phi_{i'j}^2)$, $(\Phi_{ij}^2\Phi_{ii'}^2)(\Phi_{i'j}^2)$, $(\Phi_{ii'}^2\Phi_{ij}^2)(\Phi_{i'k}^2)$, $(\Phi_{ii'}^2\Phi_{i'j}^2)(\Phi_{i'k}^2)$, $(\Phi_{ij}^2\Phi_{i'j}^2)(\Phi_{i'k}^2)$, $(\Phi_{ij}^2\Phi_{ik}^2)(\Phi_{i'l}^2)$, $(\Phi_{ij}^2\Phi_{jk}^2)(\Phi_{i'l}^2)$, $(\Phi_{ij}^2\Phi_{jk}^2)(\Phi_{i'i}^2)$, $(\Phi_{ij}^2\Phi_{ik}^2)(\Phi_{i'i}^2)$, $(\Phi_{ij}^2\Phi_{i'j}^2)(\Phi_{i'i}^2)$.

Terms of the form $(AB^2C^2)(A)$ contain 3 variables in the left bracket, and so it can have 4 indices in that bracket. This means that $N_{ind}^{max} = 4$ and $max_{(\#j)} = 3$. In the expression for $Var(x_i^*)$, there are 6 terms of the form $(AB^2C^2)A$, so $N_{ind}^{max} = 6$, and we require $(6 + 1) \times 3 = 21$ terms.

2.5.5 Accuracy of Analytical Prediction

The analytical prediction of $P_{feas}(\gamma, \rho)$ remains accurate up to $|\rho| = 0.5$ for $n = 100$ and $|\rho| = 0.25$ for $n = 25$.

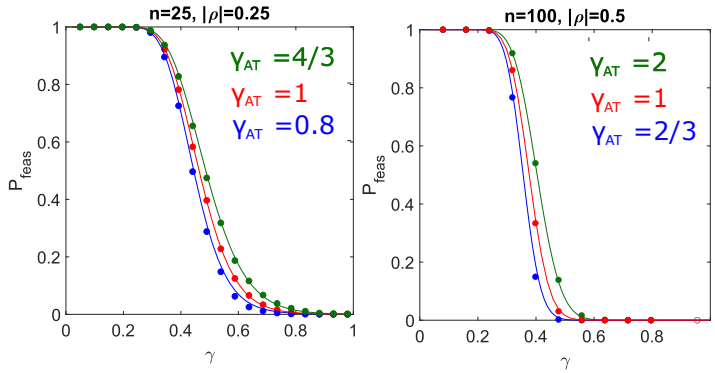

Figure S1: Dots are numerical simulations of $P_{feas}(\gamma, \rho)$. Solid curves are analytical predictions of $P_{feas}(\gamma, \rho)$ using $E(x_i^*)$ approximated up to and including order σ^3 and $Var(x_i^*)$ up to and including order σ^6. The system has $C = 1$. Blue $\rho < 0$, red $\rho = 0$ and green $\rho > 0$. γ_{AT} is the complexity above which linear stability is lost in the Allesina and Tang 2015 model. For any value of ρ, feasibility is lost at smaller complexities than linear stability in large systems.

We see from Figure S1 that feasibility is lost at a smaller complexity compared to linear stability. For systems with $\rho < 0$, feasibility is lost at a much smaller complexity compared to linear stability than systems with $\rho > 0$, implying that increasing the proportion of predator-prey interactions will have a more modest effect on stability than predicted by Allesina and Tang. Figure S2 plots numerical simulations of $P_{feas}(\gamma, \rho)$ for the case where $\rho = \pm 1$. For the case where $\rho = -1$, P_{feas} still decreases to 0 above a sufficiently large complexity even though the system is linearly stable for all complexities.

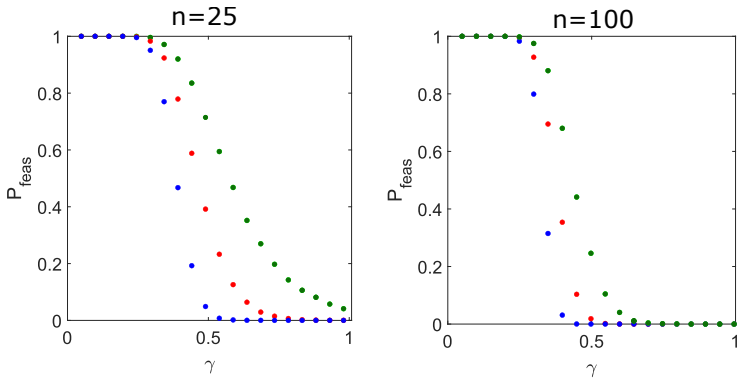

Figure S2: Numerical simulations of $P_{feas}(\gamma, \rho)$ for the case where $|\rho| = 1$. Analytical predictions break down for this magnitude of ρ, so only numerical simulations are included. The system has $C = 1$.

2.6 APPENDIX II: PARAMETER REGIONS WHERE APPROXIMATION OF $Var(x_i^*)$ BEYOND ORDER σ^6 IS REQUIRED

In Figure S4, we show examples of plots in parameter regions where the approximation of $P_{feas}(\gamma, \rho)$ breaks down. It is apparent from Figure S3 that for large magnitudes of ρ, the analytical approximation of $Var(x_i^*)$ to order σ^6 breaks down at a smaller value of σ. Since systems with a large negative ρ also has a higher P_{feas} at a given value of σ, the analytical prediction of $P_{feas}(\gamma, \rho)$ becomes inaccurate before $P_{feas}(\gamma, \rho)$ transitions to 0. These points indicate how the analytical approximation would break down given a sufficiently large negative ρ. We also show in Figure S4 that for systems such as $n = 25$, the analytical prediction of $P_{feas}(\gamma, \rho)$ also breaks down for small C.

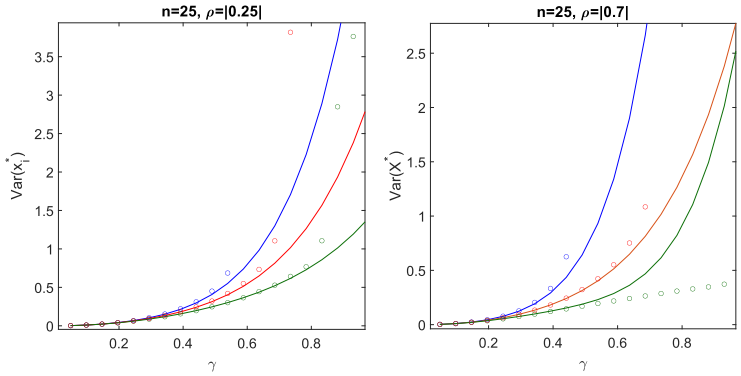

Figure S3: $Var(x_i^*)$ as a function of γ for systems with $n = 25$. Panel (a) is for system with $|\rho| = 0.25$ and panel (b) for $|\rho| = 0.7$. It is evident from panel (b) that for large negative ρ such as $\rho = -0.7$, the analytical prediction of $Var(x_i^*)$ breaks down at around $\gamma = 0.5$, which is a smaller value than for the case $\rho = -0.25$ where the analytical prediction breaks down at a around $\gamma = 0.8$. For $\rho = 0$ and $\rho > 0$, outliers in x_i^* begin to emerge above a sufficiently large γ, causing numerical data of $Var(x_i^*)$ to become noisy.

In the right panel ($n = 25$, $\rho = -0.7$), the feasibility probability at value of γ at which the order σ^6 approximation of $Var(x_i^*)$ breaks down is 0.632, whereas in the left panel ($n = 25$, $\rho = -0.25$), this corresponding feasibility probability is 0.021. In Figure S4, we show that the analytical prediction of $P_{feas}(\gamma, \rho)$ can break down if either ρ is sufficiently large or if C is sufficiently small for a fixed community size n.

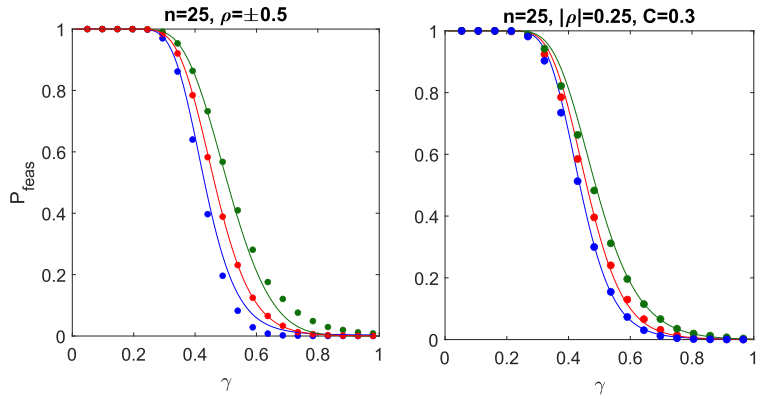

Figure S4: Left: The analytical prediction of $P_{feas}(\gamma, \rho)$ breaks down when ρ is sufficiently large and negative given a sufficiently small n. Here $C = 1$. Right: The analytical prediction of $P_{feas}(\gamma, \rho)$ begins to break down when C is sufficiently small.

It is worth noting that when ρ becomes sufficiently large and negative, the variance-covariance matrix ceases to be positive semi-definite if $Var(x_i^*)$ is approximated up to order σ^6.

2.7 APPENDIX III: ANALYTICAL PREDICTION OF FEASIBILITY PROBABILITY AS A FUNCTION OF COMPLEXITY

The analytical prediction of P_{feas} as a function of complexity γ is determined by integrating the joint density function of $\boldsymbol{x}^*$ from $x_i^* = 0$ to $x_i^* = \infty$ for all $i \in [1, n]$. For systems of $n \leq 25$, this is done using the *mvncdf* command in MATLAB. For systems where $n > 25$, *mvncdf* is no longer applicable, therefore we obtain the analytical prediction of P_{feas} as a function of γ by reducing the multivariate normal integral to a single integral using the method detailed in [41]. The multivariate normal distribution function is given by

$$F_{\mathbf{X}}(\mathbf{X}) = \int_{-\infty}^{x_i} \dots \int_{-\infty}^{x_n} f_{\mathbf{X}}(\mathbf{X}, \Sigma^{\mathbf{X}}) d\mathbf{X} \tag{2.77}$$

where $\Sigma^{\mathbf{X}}$ is the variance-covariance matrix of $\mathbf{X}$. Here $\mathbf{X}$ is a random variable such that $X_i = -x_i^*$. Define y_i as a standardised normal random variable $y_i = \frac{X_i - \mu_{X_i}}{\sigma_{X_i}}$. If $Corr(x_i^*, x_j^*)$ can be expressed in the form $Corr(x_i^*, x_j^*) = b_i b_j$ where $b_i, b_j \in \mathbb{R}$, then $F_{\mathbf{X}}(\mathbf{X})$ can be expressed as

$$F_{\mathbf{X}}(\mathbf{X}) = \int_{-\infty}^{\infty} \left\{ \prod_{i=1}^{n} \Phi\left(\frac{y_i - b_i u}{(1 - b_i^2)^{1/2}}\right) \right\} \phi(u) du \tag{2.78}$$

where $\phi(u)$ is the density function of a standard normal random variable u and $\Phi(v)$ denotes the cumulative distribution function of a standard normal random variable v [41]. The condition of feasibility $\boldsymbol{x}^* > \mathbf{0}$ is equivalent to the condition that $\mathbf{X} < \mathbf{0}$. Since $y_i = \frac{X_i - E(X_i)}{\sigma_{X_i}}$, the condition that $X_i < 0$ is equivalent to the condition that $y_i < -E(X_i)/\sigma_{X_i}$ and thus $y_i < E(x_i^*)/\sigma_{x_i^*}$. In our analytical prediction of P_{feas}, we have that $y_i = \frac{E(x_i^*)}{\sqrt{Var(x_i^*)}}$ and $b_i = \frac{\sqrt{Cov(x_i^*, x_j^*)}}{Var(x_i^*)}$. In other words, P_{feas} is the expression you get by plugging these expressions for y_i and b_i into Eq. (2.78). The integral Eq. (2.78) is also applicable for $Cov(x_i^*, x_j^*) < 0$, however the integrand is complex [41]. As a result, we approximated $P_{feas}(\gamma, \rho)$ for the case $\rho < 0$ by numerically integrating Eq. (2.78) using the *Nintegrate* command in mathematica. The lower and upper limits of this integral are set to -20 and 20 respectively. This numerical integral works well since the imaginary part is of magnitude 10^{-18} and the real part matches the numerical simulations of $P_{feas}(\gamma, \rho)$ closely, as can be seen in Figure S1.

2.8 APPENDIX IV: OBTAINING NUMERICAL SIMULATIONS OF FEASIBILITY PROBABILITY

In fully connected systems $C = 1$, we obtained 10000 numerical solutions of $\boldsymbol{x}^*$ by doing 10000 successive runs of the GLV equation (Eq. (2.5) in main text). In particular, the numerical solutions were obtained by generating 10000 random interaction matrices A parameterised according to Eq. (2.3) of the main text and calculating each of their corresponding $\boldsymbol{x}^*$ using Eq. (2.5). The feasibility probability is found by calculating the fraction of numerical solutions of $\boldsymbol{x}^*$ out of the 10000 realisations that contain all non-negative entries (i.e. the fraction of numerical solutions of $\boldsymbol{x}^*$ such that $x_i^* > 0 \ \forall \ i \in [1, n]$). For each value of n, ρ and C we considered, the step described above is repeated for a range of values of σ, where this range falls within the interval $\gamma \in [0, 1]$ (where $\gamma = \sigma\sqrt{nC}$). For the $n = 25$ case (Figure 2.3 (a)), the feasibility probability is calculated for values of σ ranging from 0.01 and 0.2, with increment of 0.01. In other words, the width between each marker in Figure 2.3 (a) is $0.01\sqrt{25}$. For the $n = 100$ case (Figure 2.3 (b)), the values of σ range from 0.008 to 0.096, with increment of 0.008, where the width between each marker is $0.008\sqrt{100}$. For predator-prey, mutualistic and competitive interactions ($\rho \neq 0$), the interaction matrix A is constructed according to Eq. (2.3) of the main text. For fully connected ($C = 1$) systems with random interaction structure (i.e

Figure (3) red in main text), A is constructed according to Eq. (2.2) of the main text.

When considering sparsely connected matrices ($C < 1$), we exclude such matrices A that contain disconnected components. In particular, we obtain 10000 realisations of $\boldsymbol{x}^*$ which correspond to sparse interaction matrices constructed according to Eq. (2.3) and contain no disconnected components. Even in the case where $\rho = 0$, we construct A according to Eq. (2.3) rather than Eq. (2.2). This is done because for the case where $C < 1$, constructing A according to Eq. (2.3) would mean that if A_{ij} is nonzero, then A_{ji} is also nonzero, which is not true if A is constructed according to Eq. (2.2). As a result, these two different ways of constructing A would give different P_{feas}. However in the case where $C = 1$, constructing A according to Eq. (2.3) with $\rho = 0$ would give the same feasibility probability as when A is constructed according to Eq. (2.2).

2.9 APPENDIX V: ANALYTICAL PREDICTION FOR SPARSELY CONNECTED SYSTEMS

Empirical ecological networks may be sparsely connected [60], so it would be useful to generalise our feasibility calculations to account for connectance C. Recent work by Akjouj and Najim [3] have shown that in sparse random GLV models with block structure, P_{feas} also exhibits a rapid transition to 0 above a critical interaction strength, hinting that it could be viable to generalise Stone's analytical prediction to account for C. In this section, we demonstrate the success of this generalisation by providing an example in Figure S5. In Figure S5, we show that for a system with $n = 100$, analytical predictions for $P_{feas}(\gamma, \rho)$ remain highly accurate even in systems with connectance as low as $C = 0.3$.

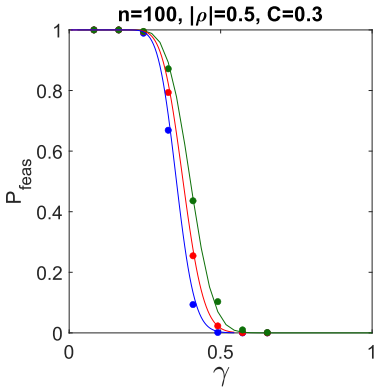

Figure S5: We show that the analytical prediction of feasibility probability as a function of complexity [210] can be generalised for C as well as ρ (i.e. P_{feas} as a funtion of γ where $\gamma = \sigma\sqrt{nC}$). All labels and parameters for this figure are the same as that of Figure 3 right, except $C = 0.3$. Our analysis of n species systems concerns systems comprising n species that interact as a single unit, so interaction matrices of $C < 1$ that contain disconnected components are excluded from our analysis (see Supplemental Information 2.8).

By comparing the analytical and numerical data in Figure S5 with those of Figure S1, we see that the system with $C = 0.3$ shows an almost identical feasibility profile with a system of $C = 1$. In other words, we get the same value of P_{feas} for a given value of γ_M in both systems. This implies that increasing σ and decreasing C to give the same complexity has negligible effect on the feasibility profile.

2.10 APPENDIX VI: EFFECT OF ρ ON OUTLIER EIGENVALUE

This section shows the effect of ρ on the stability of GLV models by looking at how the outlier eigenvalue of J changes with ρ. It also shows that the abundance of the least abundant species is a good predictor of the outlier eigenvalue of J, and thus its stability.

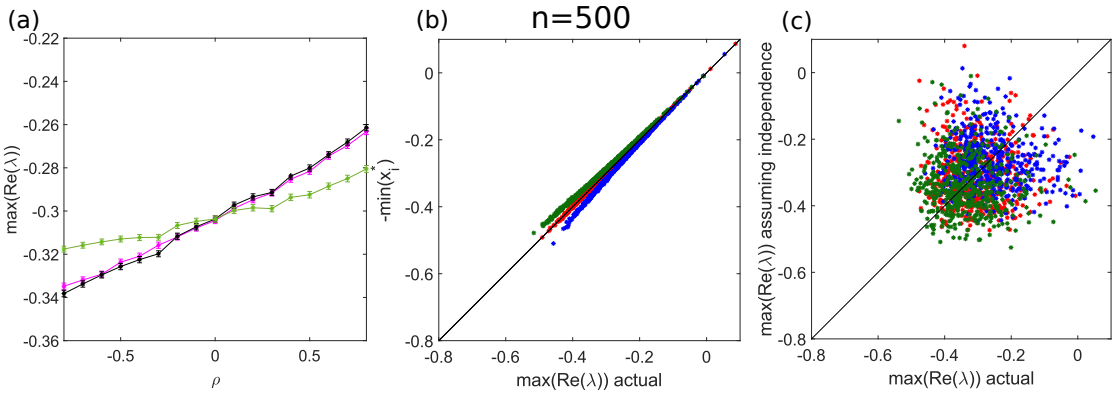

Figure S6: Panel (a) shows the effect of ρ on the outlier eigenvalue of $J = \boldsymbol{x}^* A$, averaged over 3500 realisations. Black represents the outlier eigenvalue of the actual GLV Jacobian ($\max(Re(\lambda))$ actual), where each realisation possesses a feasible $\boldsymbol{x}^*$. Light green represents the outlier eigenvalue approximated by the relation $\max(Re(\lambda)) = -\min_{i\in\{1,n\}} x_i^*$ and pink is the outlier eigenvalue of J constructed by sampling $\boldsymbol{x}^*$ and A independently. Error bars represent the standard error about the mean. Panel (b) plots $-\min_{i\in\{1,n\}} x_i^*$ against $\max(Re(\lambda))$ actual for 500 realisations of the GLV model, with $\gamma = 0.01\sqrt{500}$. Blue $\rho = 0.7$, red $\rho = 0$ and green $\rho = -0.7$. Black line is the line on which $max(Re(\lambda)) = -\min_{i\in\{1,n\}} x_i^*$. Panel (c) plots $\max(Re(\lambda))$ actual against that of J constructed by sampling $\boldsymbol{x}^*$ independently of A.

Panel (a) shows that the outlier of the Jacobian constructed by sampling x^* independently of A (Grilli's assumption) correctly captures the qualitative effect of ρ on stability. Although it is shown in panel (c) that constructing the Jacobian by adopting Grilli's assumption fails to accurately calculate the outlier eigenvalue of each Jacobian. Panel (b) shows that $-\min_{i\in\{1,n\}} x_i^*$ is a highly accurate predictor of stability of the GLV model corresponding to a given realisation of A, since most green markers sit close to the diagonal line. Notice that for systems where $\rho > 0$, the markers lie below the diagonal line. This implies that the stability is marginally overestimated by the relation $max(Re(\lambda)) = -\min_{i\in\{1,n\}} x_i^*$ and for $\rho < 0$, this relation underestimates the stability slightly. Since $Corr(\max(Re(\lambda)), -\min_{i\in\{1,n\}} x_i^*) = 0.9999$, when $\rho = 0$ $Corr(\max(Re(\lambda)), -\min_{i\in\{1,n\}} x_i^*) = 0.9996$ when $\rho = -0.7$ and $Corr(\max(Re(\lambda)), -\min_{i\in\{1,n\}} x_i^*) = 0.9995$ when $\rho = 0.7$, $-\min_{i\in\{1,n\}} x_i^*$ is an accurate predictor of stability for all regimes of ρ. It is of note that for large magnitudes of ρ, $-\min_{i\in\{1,n\}} x_i^*$ becomes a poor predictor of the outlier eigenvalue of J statistically, and thus a poor estimator of stability.

For smaller n systems, the accuracy of $-\min_{i\in\{1,n\}} x_i^*$ at predicting the outlier eigenvalue of J is reduced to such an extent that it ceases to accurately predict the effect of ρ on resilience.

2.11 APPENDIX VII: EFFECT OF $E(x_i^*)$, $Var(x_i^*)$ AND $Cov(x_i^*, x_j^*)$ ON PROBABILITY OF FEASIBILITY

For a multivariate normal distribution with a given $E(x_i^*)$ and $Cov(x_i^*, x_j^*)$, Increasing $Var(x_i^*)$ decreases feasibility probability. Increasing $E(x_i^*)$ acts to increase P_{feas}, given fixed values of $Cov(x_i^*, x_j^*)$ and $Var(x_i^*)$ and increasing $Cov(x_i^*, x_j^*)$ also acts to increase P_{feas} given fixed values of $Var(x_i^*)$ and $E(x_i^*)$. In Figure S7 below, we show that out of these three quantities, $Cov(x_i^*, x_j^*)$ increases most slowly with σ (and thus γ). As a result, $Cov(x_i^*, x_j^*)$ plays the smallest part in governing how P_{feas} changes with γ. In Figure S8, we justify how this argument holds true by quantifying the effects of $E(x_i^*)$, $Var(x_i^*)$ and $Cov(x_i^*, x_j^*)$ on P_{feas}.

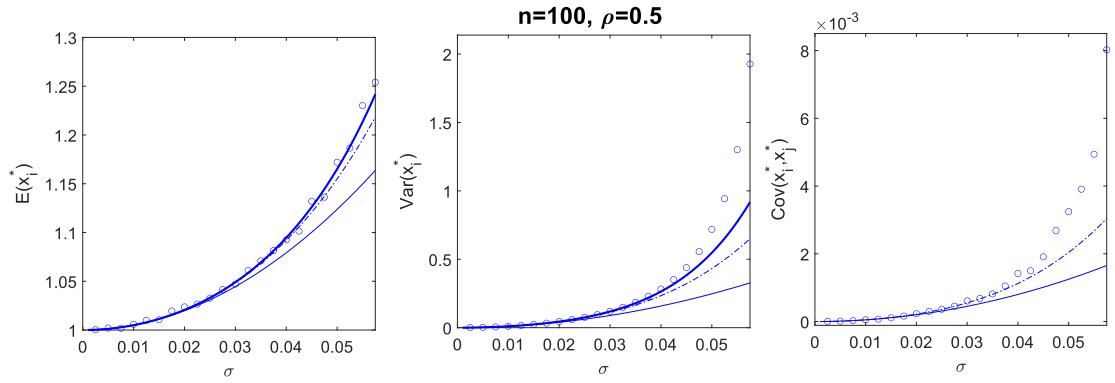

Figure S7: Analytical approximations of $E(x_i^*)$, $Var(x_i^*)$ and $Cov(x_i^*, x_j^*)$ as a function of σ at various orders of σ. Fine solid curve order σ^2, dash dotted curve order σ^4 and bold solid curve order σ^6. Circles are numerical simulations of these quantities, obtained from 10000 numerical solutions of main text Eq. (2.5), which are acquired as described in Section IV. Here, $C = 1$. Since we do not have an analytical approximation of $Cov(x_i^*, x_j^*)$ to order σ^6, there is no bold solid curve in the right panel. For values of σ such that $\sigma\sqrt{nC} > 1/(1+\rho)$, the Neumann approximation of $\boldsymbol{x}^*$ Eq. (2.9) breaks down given fixed n and C. This condition is equivalent to $\sigma > 0.0670$ here. Since n is finite, the normality in distribution of $\boldsymbol{x}^*$ breaks down at some point where $\sigma\sqrt{nC} < 1/(1+\rho)$ (see Section 2.12 for explanation). Due to this, numerical results for $Var(x_i^*)$ and $Cov(x_i^*, x_j^*)$ no longer converges upon increase in sample size. σ is plotted up to the largest value in which these numerical results still converge, which is 0.0575 here.

In Figure S8 below, we show how varying each of the quantities $E(x_i^*)$, $Var(x_i^*)$ and $Cov(x_i^*, x_j^*)$ individually impacts P_{feas}. In each panel, we vary one of the three quantities while keeping the other two fixed. In the middle panel of Figure S8, we see that if we vary $Var(x_i^*)$ by an amount equal to the difference between its approximation at order σ^2 and σ^6, P_{feas} changes by 0.5493, which is significantly large. Even if we vary $Var(x_i^*)$ by the difference between its approximation at order σ^4 and σ^6, P_{feas} still changes by 0.1097. If we vary $Cov(x_i^*, x_j^*)$ by an amount equal to the difference between its approximation at order σ^2 and σ^4, an even smaller change in P_{feas} manifests (change of 0.0019). Since we do not have the approximation of $Cov(x_i^*, x_j^*)$ at order σ^6, we showed what happens if we vary $Cov(x_i^*, x_j^*)$ by an amount equal to the difference between its approximation at order σ^2 and its actual value. In this case, P_{feas} exhibits a change of 0.0037. All of the above demonstrate that going to higher orders in the approximation $Cov(x_i^*, x_j^*)$ confers negligible effect

on P_{feas}.

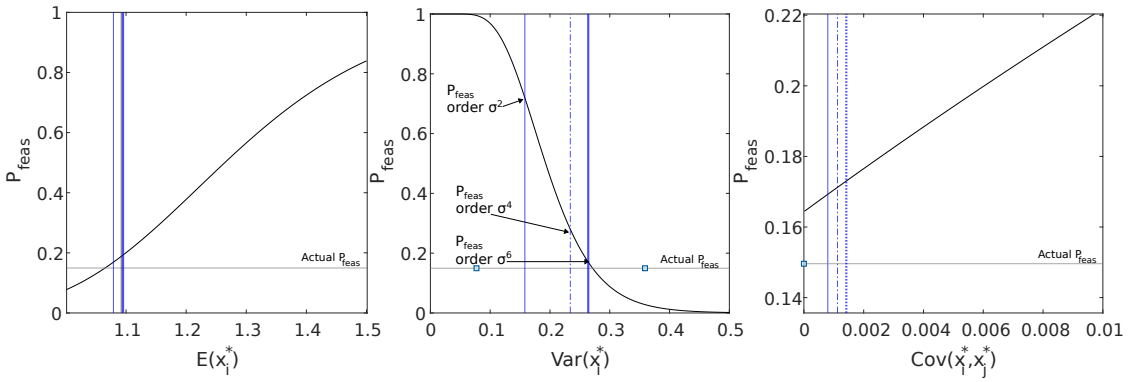

Figure S8: Panels showing how P_{feas} changes as each quantity ($E(x_i^*)$, $Var(x_i^*)$ or $Cov(x_i^*, x_j^*)$) is varied, provided that the other two quantities are fixed, e.g. if $E(x_i^*)$ is varied, we fix $Var(x_i^*)$ and $Cov(x_i^*, x_j^*)$. The two fixed quantities are set to the analytically approximated value (Eqs. (2.13-2.15) in main text) they would take if $\sigma = 0.04$, $n = 100$ and $C = 1$. Black curve shows how P_{feas} varies with each of these quantities and vertical lines show the varying quantity analytically approximated at different orders of σ. Notations for the line textures are consistent with all other Figures in this section, except here, we have a dotted line which represents the actual numerically simulated $Cov(x_i^*, x_j^*)$, since we do not have the analytically approximation of $Cov(x_i^*, x_j^*)$ at order σ^6.

Below, we show analytically that $Cov(x_i^*, x_j^*)$ increases more slowly with σ compared to $E(x_i^*)$ and $Var(x_i^*)$.

ANALYTICAL ILLUSTRATION OF THE MAGNITUDES OF $E(x_i^*)$, $Var(x_i^*)$ AND $Cov(x_i^*, x_j^*)$

Here, we illustrate analytically how $Cov(x_i^*, x_j^*)$ increases more slowly with σ than $E(x_i^*)$ and $Var(x_i^*)$, and therefore plays the smalest part in governing how P_{feas} changes with σ. In the expression for $Var(x_i^*)$, the coefficient of σ^2 contains a term of order n, and the corresponding coefficient in the expression for $E(x_i^*)$ contains a term of order $n\rho$. The corresponding coefficient in $Cov(x_i^*, x_j^*)$ contains a term of order ρ. At order σ^4, the coefficient for $Var(x_i^*)$ includes terms of order n^2, ρn^2 and $\rho^2 n$ while the coefficient of $E(x_i^*)$ includes terms of order n and $\rho^2 n^2$. Finally, the coefficient of $Cov(x_i^*, x_j^*)$ includes terms of order 1 and $\rho^2 n$. We see that at both

order σ^2 and σ^4, the coefficients of $Cov(x_i^*, x_j^*)$ are a factor n smaller than those of $E(x_i^*)$ and $Var(x_i^*)$, which implies that $Cov(x_i^*, x_j^*)$ increases slowly with σ given fixed values of n and C. The small increase in $Cov(x_i^*, x_j^*)$ with σ is also shown numerically in Figure S8.

2.12 APPENDIX VIII: DISTRIBUTION OF x_i^*

It is crucial to note that the Neumann series in Eq. (2.9) of main text is convergent if $||\sigma\mathcal{E}|| < 1$. This corresponds to the spectral radius of A being less than 1. For systems of large n, this spectral radius is analytically approximated by $\sigma\sqrt{nC}(1+\rho)$ (see Eq. (2.4) of main text), and thus the Neumann propagation of x_i^* is applicable only under the condition $\sigma\sqrt{nC} < 1/(1+\rho)$ (or equivalently $\gamma < 1/(1+\rho)$). Stone [210] argued using the Central limit theorem (CLT) that Eq. (2.9) is normally distributed as $n \to \infty$. This implies that the statement that x_i^* is normal for all cases satisfying $\sigma\sqrt{nC} < 1/(1+\rho)$ is only exact in the limit as $n \to \infty$. Here we show using numerical solutions to Eq. (2.5) that the distribution of x_i^* is normal for small σ, and this normality breaks down for larger values of σ. The numerical solutions are obtained as described in Section 2.8.

This agrees with the CLT argument above which infers that the larger the value of n, the smaller the value of σ at which the boundary $\sigma\sqrt{nC} = 1/(1+\rho)$ is reached (provided a given C), and thus the distribution of x_i^* is more likely to be normal at or near the boundary $\gamma = 1/(1+\rho)$ as we increase σ. As we increase n, the normality in the distribution of x_i^* breaks down at a larger value of γ such that $\gamma < 1/(1+\rho)$ (see Figure S9 and S10). In theory, as $n \to \infty$, the distribution of x_i^* would remain normal up to $\gamma = 1/(1+\rho)$.

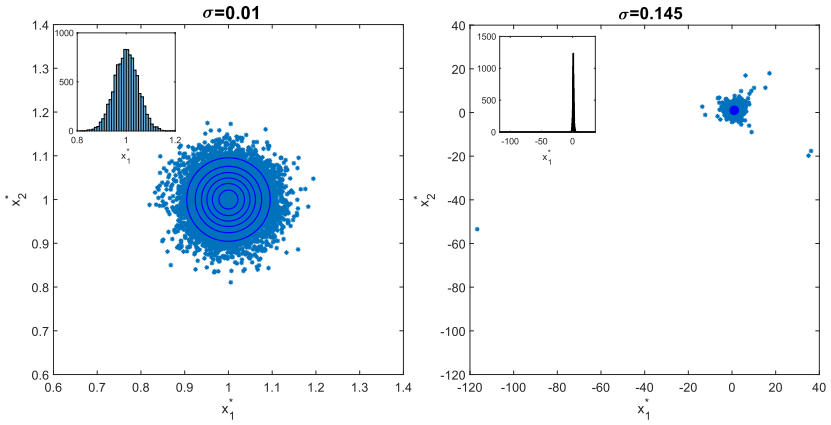

Figure S9: For an $n = 25$ system, x_i^* is normally distributed for small values of σ, such as $\sigma = 0.01$. For this given system size $n = 25$, the normality breaks down when $\sigma = 0.145$, which corresponds to $\gamma = 0.725$. Light blue markers represent x_1^* and x_2^* values of 10000 numerical solutions of $\boldsymbol{x}^*$, obtained as described in Section 2.8. Panel insets show histograms for the distribution of x_1^*. Other parameters are $\rho = 0$ and $C = 1$.

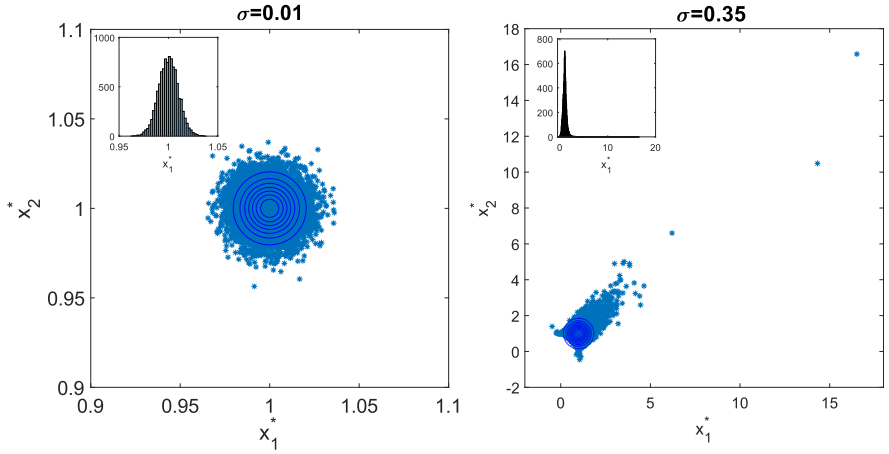

Figure S10: As in Figure S9 but for an $n = 2$ system. For an $n = 2$ system, the normality breaks down when $\sigma = 0.35$, which corresponds to $\gamma = 0.495$.

We see from Figure S9 and S10 that as n increases, the value of γ at which normality in x_i^* is lost increases. Similarly for the equivalent $n = 100$ system as in the two Figures above, numerical results show that normality is lost at $\sigma = 0.085$ which corresponds to $\gamma = 0.85$. We see that the larger the value of n, the larger the value of γ up to which the distribution of x_i^* remains normal. In theory, as $n \to \infty$, x_i^* would remain normally distributed up to $\gamma = 1/(1 + \rho)$, which corresponds to

$\gamma = 1$ for the scenario in the two Figures above, since we have $\rho = 0$. Our analytical results for P_{feas} in Figure (3) of the main text are highly accurate since for systems where $n \geq 25$, P_{feas} already becomes approximately 0 before the point at which normality in $\boldsymbol{x}^*$ is lost (see Figure S1). At $\gamma = 0.725$ where normality in $\boldsymbol{x}^*$ breaks down for an $n = 25$, $\rho = 0$ system, $P_{feas} \approx 0.02$.

— 3 —

Adaptive dynamics, switching environments and the origin of the sexes

Abstract

In the face of varying environments, organisms exhibit a variety of reproductive modes, from asexuality to obligate sexuality. Should reproduction be sexual, the morphology of the sex cells (gametes) produced by these organisms has important evolutionary implications; these cells can be the same size (isogamy), one larger and one smaller (anisogamy), and finally the larger cell can lose its capacity for motility (oogamy, the familiar sperm-egg system). Understanding the origin of the sexes, which lies in the types of gametes they produce, thus amounts to explaining these evolutionary transitions. Here we extend classic results in this area by exploring these transitions in a model in which organisms can reproduce both sexually and asexually (a reproductive mode present in many unicellular species). In particular, we investigate the co-evolution of gamete cell size and fertilization rate, which is a proxy for motility but usually held constant in such models. Using adaptive dynamics generalized to the case of switching environments, we find that isogamy can evolve to anisogamy through evolutionary branching, and that anisogamy can evolve to oogamy through a further branching driven by sexual conflict. We also derive analytic conditions on the model parameters required to arrest evolution on this isogamy-oogamy trajectory, with low fertilization rates and stochastically switching environments stabilizing isogamy under a bet-hedging strategy, and low fertilization costs stabilizing anisogamy.

3.1 INTRODUCTION

The origin of the sexes lies in the types of sex cells they produce; males produce small microgametes while females produce larger macrogametes [95]. This gamete size dimorphism is referred to as a state of anisogamy [219]. Should the microgamete be motile (sperm) and the macrogamete sessile (egg), the population is said to be oogamous [112]. While anisogamy is the most commonly observed mode of sexual reproduction in eukaryotic organisms [145], having evolved several times in evolutionary history [84], it is now widely accepted to be derived from isogamy (equal gamete sizes) [12]. While rarer, isogamous species include study organisms such as the yeast *Saccharomyces cerevisiae* [160] and the green algae *Chlamydomonas reinhardtii* [103], where self-incompatible mating types play the role of ancestral versions of the true sexes [121]. Indeed, the volvocine algae (of which *C. reinhardtii* is a member), provide neat empirical examples of these transitions, with phylogenetic analysis indicating that numerous independent lineages have undergone the transition from isogamy, to anisogamy, and finally to oogamy [81, 112]. Explaining the evolutionary mechanisms behind these transitions has been the focus of much work in evolutionary theory.

While theoretical investigations of the evolution of anisogamy date back to the 1930's [109] and were developed into the 1960's [110, 196], it was arguably the Parker-Baker-Smith (PBS) model [173] that synthesised these ideas into a complete evolutionary model that is now widely accepted as providing an explanation for the evolution of anisogamy [128, 172, 218]. They assumed that a fixed mass or energy budget is allocated by individuals to gamete production, such that microgametes can be produced in larger quantities than macrogametes. However, while microgametes may be more numerous, they contribute a lower fitness than macrogametes to a fertilized zygote due to their low cytoplasmic volume [222]. In this way the PBS model was able to show how anisogamy was the result of a quality-quantity trade-off, with disruptive selection acting on gamete size [218]. In simple terms, the PBS model posits that selection simultaneous favours the production of large gametes to enhance survival, and a large number of gametes to compete for fertilisation [128].

The key elements of the PBS model have since been set in a game theoretic [25] and population genetic [32, 105] context, as well as extended to account for more general reproductive modes such as hermaphrodism [92]. Models using adaptive dynamics [113] in particular have been useful. It has been shown analytically that

even in the absence of mating types, anisogamy can evolve from isogamy through evolutionary branching in mass [143]. Meanwhile, accounting for self-incompatibility of mating types and investigating the effect of varying fertilization rate, [124] showed that anisogamy can evolve from isogamy through both gamete competition and gamete limitation. Altogether, the results described above suggest that fertilization rate is a crucial factor that may impact gamete survival, and selection is likely to act on the fertilization rate between microgametes and macrogametes.

As addressed, oogamy, the loss of motility in eggs and specialization for motility in sperm, is often seen as the "last step in the evolution of the egg–sperm dichotomy" [133], a view supported by empirical analyses that suggest oogamy is derived from anisogamy [198]. However the theory of this transition is comparatively less studied than the earlier transition from isogamy to anisogamy. Ghiselin [66] provided an argument based on the physiological division of labour between macro and microgametes, with females specialising in provisioning and males in motility. Most other work has considered the evolution of oogamy as a strategy to maximise gamete encounter rate. This can be achieved by having a population of pheromone emitters and receivers [79, 96] and by having a large stationary target egg and small motile sperm [38, 50]. Although the assumption of an inverse speed-size relationship in [38] has justification in some gametic systems, it is also worth noting that positive speed-size relationship have been observed in *C. reinhardtii*, due to larger cells having greater propulsive forces; at scales such as these the precise speed-size relationship is complicated by details of cell morphology [204]. Lastly, the prior evolution of internal fertilization has been proposed as a mechanism that could generate selection for oogamy [132], consistent with empirical evidence from *volvox* (external fertilization and anisogamous) and its sister lineage *platydorina* (internal fertilization and oogamous) [164].

The work described above all assumes obligate sexual reproduction (unfertilized gametes die at the end of each generation) and a fixed fertilization rate (the rate at which gametes encounter one another and fertilize to form zygotes) and a static environment [25, 124, 133]. Recently, however, inspired by the life-histories of green and brown algae such as *Blidingia minima* (isogamous [216]), *Urospora neglecta* (anisogamous [140]), and *Saccharina japonica* (oogamous [111]), whose gametes can develop asexually through parthenogenesis should they fail to find a mate (see [136, 140, 216], respectively), the first of these assumptions was relaxed in two theoretical papers [37, 130]. In [130], extra survival costs were incurred by gametes developing

parthenogenetically, while in [37] extra survival costs on either the parthenogenetic or the sexual reproductive route were considered. They found that in the presence of two self-incompatible mating types, isogamy can be stabilized under low costs to parthenogenesis, while anisogamy is the evolutionarily stable state when fertilization is favoured.

The studies described above on parthenogenesis and the evolution of anisogamy lead to a natural question; if the sexual route (via fertilization) and asexual route (via parthenogenesis) carry different survival costs, how should the fertilization rate evolve to account for this? Should this rate increase (to minimise the number of unfertilized gametes taking a potentially perilous route to survival) or should it decrease (to avoid potential costs incurred during cell fusion)? The modelling of fertilization kinetics is an interesting topic in its own right [126, 129], and plays a role in models for the evolution of anisogamy from isogamy [124]. However fertilization kinetics are also clearly a key element of the evolution of oogamy, with selection on sperm to increase their encounter rate with eggs and selection on eggs to remain sessile.

In this chapter, we modify the adaptive dynamics models of [37, 130, 143] to study the co-evolution of gamete size and fertilization rate in species capable of parthenogenesis under external fertilization. For simplicity, we assume an absence of self-incompatible mating types [12, 143]. We show that under such assumptions, anisogamy can evolve from an initial state of isogamy, followed by the subsequent evolution to oogamy under sexual conflict between microgametes and macrogametes. In Section 5.2, we introduce the model, describing its key behaviour in Section 5.3. These behaviours include the evolution of oogamy (Section 3.3.2), conditions that stabilize isogamy (Section 3.3.3), conditions that stabilize anisogamy (Section 3.3.4), and the emergence of isogamy as a bet-hedging strategy in switching environments (Section 3.3.5).

3.2 MODEL

In this section we describe the specifics of the models we use, paying careful attention to the various time scales involved. We begin by considering the evolutionary model in a fixed environment (Section 3.2.1), before generalizing to the case of a under which bet-hedging strategies can evolve (Section 3.2.2).

3.2.1 MODEL DYNAMICS IN A FIXED ENVIRONMENT

The evolutionary dynamics of the model are built from a hierarchy of timescales, which are particularly important to keep in mind once environmental switching is introduced in later sections. The shortest timescale is the generational timescale. The intermediate timescale is that over which the invasion of a rare mutant (taking place over many generations) can occur. The longest timescale is the evolutionary timescale, representing the cumulative effect of multiple mutations and invasions.

Dynamics within each generation

At the start of each discrete generation, a number of adults with haploid mass M (energy budget allocated to the production of haploid cells) produce gametes (mass m), such that mass/energy budget is conserved (i.e. each adult produces M/m gametes). Note that this implicitly assumes, for simplicity, that a continuous (rather than discrete) number of gametes is possible. Gametes then enter a pool and the fertilization process takes place. After a finite time window, the resultant cells face a round of survival dependent on their mass. The surviving cells form the basis of the next generation, completing the generational cycle, as illustrated in Figure 3.1. The assumption of a continuous number of gametes is particularly suitable for large populations [83], such as the population of gametes in a pool of gametes.

Fertilization Kinetics For simplicity, we assume that all gametes may fertilize each other (i.e. there are no self-incompatible mating types). Given a total of A adults, the population is initially comprised of $N = (AM)/m$ single gametes produced by adults through meiosis. We assume fertilization is external, with cells fertilizing according to mass action dynamics at rate α, such that the number of single (unfertilized) cells, N, is given by the solution to

$$\frac{\mathrm{d}N}{\mathrm{d}t} = -\alpha N^2\,, \qquad N(0) = \frac{AM}{m} \tag{3.1}$$

(see also [127]). At the end of the fertilization window, which is assumed to have a duration T, there are therefore $N(T)$ single cells remaining, and $(N(0) - N(T))/2$ fertilized cells.

The parameter α is variously referred to as the fertilization rate, the collision rate [126], the "'aptitude' for union" [196], or the "bimolecular reaction constant" [224]. We will refer to α as the fertilization rate, and treat it as a trait subject to evolution. Overall, α captures the compound effect of the propensity for

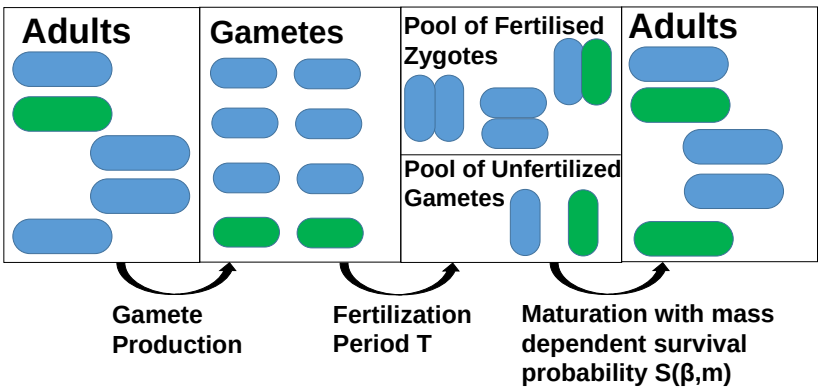

Figure 3.1: Schematic of dynamics within each generation. Mature cells (adults) produce gametes at the start of a generation. All the gametes are given a fixed time period T in which to complete the fertilization process. At the end of the fertilization period, there will be a pool of fertilized zygotes and unfertilized gametes, both of which are capable of maturation. Each cell survives according to its independent survival functions ($(1-C_z)S(\beta, m_1+m_2)$ and $(1-C_p)S(\beta, m_i)$ respectively) to produce a number of mature cells in the subsequent generation. The pool of gametes consist of resident (blue) and mutant gametes (green), where the mutation occurs in either the mass m or fertilization rate α.

fertilization between cells encountering each other, as well as additional mechanisms to enhance cell encounter rate such as increased emission or response to cheomoattractants [79] or cell motility [220]. In practice there is a likely upper-bound on α, brought about by energy trade-offs, diffusion in aquatic environments, or dispersal in terrestrial environments. To account for this we introduce a ceiling on α, and restrict its evolutionary dynamics to the range $\alpha_{\max} \geq \alpha \geq 0$.

Survival Probability At the end of a finite fertilization window at time T, the population will consist of both fertilized and unfertilized gametes. Fertilized gametes produce diploid zygotes, while unfertilized gametes can develop as haploids parthenogenetically (e.g. parthenosporophytes [151]). We assume that the probability that either of these cell types survives is given by the Vance survival function [222], which is a common assumption in the literature [25, 124, 135]. Given a cell size m_c (for either fertilized or unfertilized cells), the survival probability is given by

$$S(\beta, m_c) = \exp\left(-\frac{\beta}{m_c}\right). \tag{3.2}$$

Note that this is an increasing function of cell size, and we do not account for gamete mortality during the fertilization period. Thus, although both fertilized and unfertil-

ized gametes are exposed to the same survival function, fertilized cells (with a mass around twice the size of unfertilized cells in a monomorphic isogamous population) will have a greater survival probability than unfertilized cells. Meanwhile for a given mass m_c, increasing β will decrease the survival probability. We therefore refer to β as the resistance to survival, with high β corresponding to harsh environments in which survival is difficult, and low β corresponding to more benign environments in which even cells of modest mass have a high probability of surviving.

In addition to the benefits explained in the introduction, fertilization may also carry risks: generally these include cell-fusion failure [80], selfish extra-genomic elements in the cytoplasm [86] and cytoplasmic conflict [100, 101]. However costs may also arise in multicellular organisms if the reproductive output of diploids (e.g. sporophytes) is less than twice that of haploids (e.g. parthenosporophytes), as described in Appendix 3.5. We therefore introduce an additional fixed cost, $1 \geq C_z \geq 0$, applied to fertilized zygotes independent of their mass. Similarly, there can be costs associated to parthenogenetic development such as reduced fitness due to lack of genetic diversity [1] and the possibility of failure of parthenogenetic development [130]. The final probability of survival for a zygote formed from the fertilization of two gametes of sizes m_1 and m_2 is then given by $(1 - C_z)\exp\left[-\beta/(m_1 + m_2)\right]$, while the probability of survival for an unfertilized cell of size m is $(1 - C_p)\exp\left[-\beta/m\right]$.

Invasion Dynamics

We assume that haploid gametes are characterised by two genetically determined non-recombining traits; their mass, m, and their fertilization rate, α. We next consider a monomorphic resident population to which a mutant individual is introduced at rate μ. This mutant may produce gametes of a different mass to its ancestor, $m \pm \delta m$, where δm represents the size of a mutational step. Under this scenario the mutant may produce more or fewer gametes than its ancestor (see Appendix 3.6.1), but the survival probability of its unfertilized cells (mass $m \pm \delta m$), mutant-resident fertilized cells (mass $2m \pm \delta m$), and mutant-mutant fertilized cells (mass $2(m \pm \delta m)$), will also be simultaneously decreased or increased (see Eq. (3.2) and Appendix 3.6.3). The cumulative effect of this quality-quantity trade-off will either lead to selection for or against the mutant over subsequent generations.

Alternatively, the mutant may engage in an increased or decreased fertilization rate relative to its ancestor, at a rate $\alpha \pm \delta\alpha$ (see Appendix 3.6.2). Under this scenario the mutant fertilizes with residents at their average fertilization rate $(2\alpha \pm \delta\alpha)/2$

and other mutants at a rate $\alpha \pm \delta\alpha$. Mutants will either contribute to more or fewer fertilized cells and depending on the resistance to survival, β, the costs to fertilization, C_z, and the costs to parthenogenesis, C_p, may experience a selective advantage over the resident by devoting more of its gametes to one of the reproductive routes (see Appendix 3.6.4).

In order to mathematically characterise the invasion dynamics (which occur over discrete generations), we assume that δm and $\delta\alpha$ are small, so that we can approximate the dynamics continuously. Denoting by $\hat{f}_m$ the frequency of mutants of size $m \pm \delta m$ in the population, and t_g the number of generations, we find (see Appendix 3.7.1)

$$\frac{\mathrm{d}\hat{f}_m}{\mathrm{d}t_g} = h_m(m, \alpha, \beta, C_z, C_p)\hat{f}_m(1 - \hat{f}_m)\,, \tag{3.3}$$

where $h_m(m, \alpha, \beta, C_z, C_p)$ is a constant selective pressure that depends on the parameters $m, \alpha, \beta, C_z, C_p$ (see Eq. (3.32)). Similarly, denoting by $\hat{f}_\alpha$ the frequency of mutants with fertilization rate $\alpha \pm \delta\alpha$ in the population (see Appendix 3.7.2), we find

$$\frac{\mathrm{d}\hat{f}_\alpha}{\mathrm{d}t_g} = h_\alpha(m, \alpha, \beta, C_z, C_p)\hat{f}_\alpha(1 - \hat{f}_\alpha) \tag{3.4}$$

where $h_\alpha(m, \alpha, \beta, C_z, C_p)$ is a constant that depends on the parameters $m, \alpha, \beta, C_z, C_p$ (see Eq. (3.33)). Our analytical results Eqs. (3.3-3.4) show that in the case of a single mutant, we have frequency-independent selection for both traits, and whether selection acts to increase the frequency of the mutant depends on the resistance to survival β, the cost to fertilization and parthenogenesis C_z and C_p, and the traits of the resident population, m and α. We assume for the remainder of this section that mutants encounter a strictly monomorphic resident population, allowing us infer a simple trait substitution process that considerably simplifies the evolutionary analysis. However we will return to the possibility of mutants coexisting with residents in Section 3.3.2 of the Results. Here, we have assumed for simplicity that t_g is a continuous parameter, which is a reasonable assumption since invasion of a mutant takes place over a large number of generations (typically of the order 10^3). In Figures 3.10 and 3.11 in Appendix 3.8, we indeed see that this assumption allows us to capture the invasion dynamics with great accuracy.

Evolutionary Dynamics

Following standard approaches in adaptive dynamics [150] (see also Appendix 3.8), we construct the evolutionary equations for the gamete mass, m, and the fertilization rate, α. Denoting by τ the evolutionary timescale over which mutations appear and trait substitutions occur, we find

$$\begin{aligned}\frac{\mathrm{d}m}{\mathrm{d}\tau} &= H_m(m,\alpha;\beta,C_z,C_p)\\ &= -\frac{4(1-C_p)m(m-\beta)+AM\alpha T(1-C_z)e^{\frac{\beta}{2m}}(4m-\beta)}{4m^2((1-C_p)m+AM\alpha T(1-C_z)e^{\frac{\beta}{2m}})}\\ \frac{\mathrm{d}\alpha}{\mathrm{d}\tau} &= H_\alpha(m,\alpha;\beta,C_z,C_p)\\ &= -m\frac{((1-C_p)-(1-C_z)e^{\frac{\beta}{2m}})\ln(1+\frac{AM\alpha T}{m})}{2\alpha((1-C_p)m+AM\alpha T(1-C_z)e^{\frac{\beta}{2m}})}.\end{aligned} \tag{3.5}$$

Note that these equations involve highly nonlinear functions, with some evident non-physical behaviour near the boundaries (e.g. when cell masses are zero). This feature is common in simplified models of anisogamy evolution [37].

3.2.2 EVOLUTIONARY DYNAMICS IN SWITCHING ENVIRONMENTS

We now wish to consider the case of a population characterised by gamete mass (m) and fertilization rate (α) traits, evolving subject to changing environmental conditions. We may again employ techniques from adaptive dynamics, but now generalized to dynamic environments [155]. Explicitly, we allow the resistance to survival to alternate between two values β_1 and β_2; recall from Eq. (3.2) that if $\beta_1 > \beta_2$ then β_1 represents a comparatively "harsh environment", where cells have a lower survival probability than in environment β_2. This form of environmental fluctuations is frequently used in the literature concerning evolution in switching environments [155, 15], and enables us to capture how the evolutionary dynamics may differ with the rate of environmental switching.

Switching between these two environments is modelled as a discrete stochastic telegraph process [15, 99]; the time spent in each environment is distributed geometrically (a discrete analogue of the exponential distribution), spending an average period $\tau_1 \approx 1/\lambda_{1\to 2}$ in environment 1 and $\tau_2 \approx 1/\lambda_{2\to 1}$ in environment 2, where $\lambda_{i\to j}$ is the transition rate from environment i to j. We must carefully consider the magnitude of these timescales in comparison with the other timescales at work in the model (see Section 3.2.1).

First consider the case where the environmental switching timescales, τ_1 and τ_2, are larger than the generational timescale (t_g), but much smaller than the invasion timescale (characterised by the inverse of the strength of selection, proportional to $1/\delta m$ and $1/\delta\alpha$) and the mutational timescale ($1/\mu$). We call this the 'fast relative to invasion' switching regime (FRTI). In this scenario, the population does not switch environments during a single round of fertilization kinetics, but typically switches between the two environmental states many times before an invasion has time to complete i.e. once every few generations. When switching occurs this frequently, we can approximate the dynamics mathematically by observing that the population experiences the weighted average of the selection pressures in the two environments [99]. Denoting by P_1 and P_2 the probability of finding the population in either of the respective environments, we have

$$P_1 = \frac{\tau_1}{\tau_1 + \tau_2}\,, \quad P_2 = \frac{\tau_2}{\tau_1 + \tau_2} = (1 - P_1)\,. \tag{3.6}$$

The effective selection pressure on mutants with mass $m+\delta m$ in a resident population of mass m during an invasion is then given by

$$P_1 h_m(m, \alpha, \beta_1, C_z, C_p) + P_2 h_m(m, \alpha, \beta_2, C_z, C_p) \tag{3.7}$$

which can be contrasted against the selection pressure $h_m(m, \alpha, \beta, C_z, C_p)$ in Eq. (3.3). An analogous approach allows us to approximate the invasion dynamics for mutants with a different fertilization rate to their ancestors in this FRTI regime (see Appendix 3.11.1).

With equations for $\mathrm{d}\hat{f}_m/\mathrm{d}t_g$ and $\mathrm{d}\hat{f}_\alpha/\mathrm{d}t_g$ in hand, we can proceed to apply the same standard techniques from adaptive dynamics as were used to derive Eq. (3.5) from Eqs. (3.3-3.4) (see Appendix 3.11). We obtain the effective evolutionary dynamics

$$\begin{aligned} \frac{\mathrm{d}m}{\mathrm{d}\tau} &= P_1 H_m(m, \alpha; \beta_1, C_z, C_p) + P_2 H_m(m, \alpha; \beta_2, C_z, C_p) \\ \frac{\mathrm{d}\alpha}{\mathrm{d}\tau} &= P_1 H_\alpha(m, \alpha; \beta_1, C_z, C_p) + P_2 H_\alpha(m, \alpha; \beta_2, C_z, C_p) \end{aligned} \tag{3.8}$$

where $H_m(m, \alpha; \beta_1, C_z, C_p)$ and $H_\alpha(m, \alpha; \beta_1, C_z, C_p)$ retain the functional forms given in Eq. (3.5), and P_1 and P_2 are taken from Eq. (3.6).

In contrast to the FRTI regime discussed above, we can also investigate the regime in which environmental switching occurs on a comparable or slower timescale than invasion, but still occurs fast relative to the evolutionary timescale. In this

scenario, which we call the "fast relative to evolution" switching regime (FRTE), environmental switching occurs on a similar rate to that at which new mutations are introduced, but much faster than the combined effect of mutation and selection (e.g. $\lambda_{i\to j} \ll \mu \times \delta m$), such that only a small number of mutations can fixate in either environment before the population switches to the alternate environment. In this regime, the environment would typically switch once every few mutation events rather than every few generations as in the FRTI regime. Although we shall show via simulations in Section 3.3.5 that this FRTE regime leads to quantitatively different evolutionary trajectories compared to the FRTI regime, we show mathematically in Appendix 3.11.2 that the evolutionary dynamics can be approximated by the same equations (see Eq. (3.8)).

3.3 RESULTS

In this section we proceed to analyse the evolutionary dynamics derived in Sections 3.2.1-3.2.2 and compare our results to numerical simulations of the full stochastic simulations. In these simulations, the fertilisation kinetics are run for a fixed time period and solved numerically using the ODE solver *ode45*. The time at which mutations and environmental switching events occur as well as the trait that mutates at each mutation event and whether they mutate up or down are random, giving rise to stochasticity in the evolutionary trajectories. In other words, our evolutionary dynamics become a piecewise deterministic Markov process (PDMP) [44] with deterministic dynamics between the randomly occurring events. The mean stochastic evolutionary trajectory averaged over a large number of realisations is thus expected to converge to the deterministic trajectories as predicted by Eq. (3.5). Repeating the simulations for the FRTI regime under periodic switching, we find that the results are in general qualitatively similar, however some interesting differences emerge. An initial investigation into periodic environmental switching in this context is presented in Appendix 3.13, and may form the basis for valuable future research [15, 212].

3.3.1 INITIAL EVOLUTION OF FERTILIZATION RATE

In Figure 3.2, we see two potential evolutionary outcomes for the co-evolutionary dynamics of m and α in a single fixed environment that are dependent on the initial

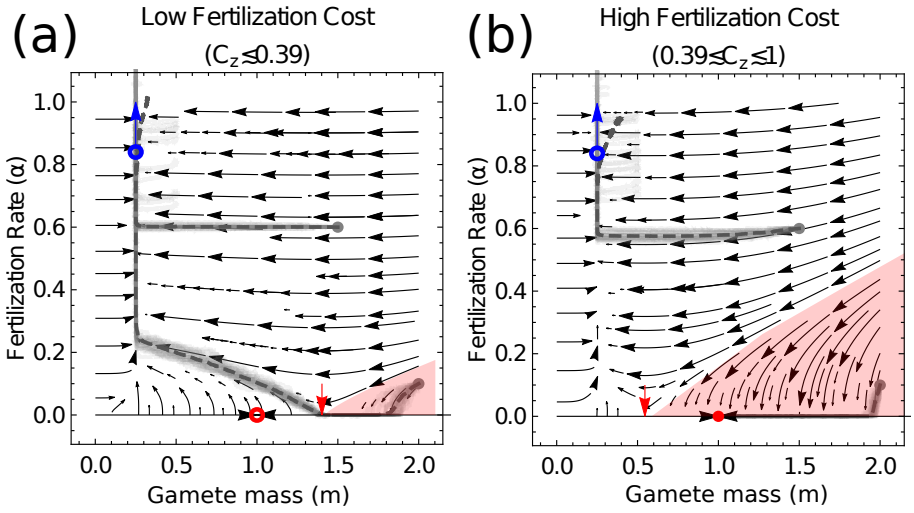

Figure 3.2: Phase portraits for the co-evolutionary dynamics in a fixed environment (see Eq. (3.5)). High fertilization rates are the only evolutionary outcome in panel (a), while high and zero fertilization rate are both evolutionary outcomes in panel (b) (see Eq. (3.9)). The red region shows trajectories leading to points on the $\alpha = 0$ boundary for which evolution selects for decreasing fertilization rate ($\mathrm{d}\alpha/\mathrm{d}\tau < 0$) (the point at which $\mathrm{d}\alpha/\mathrm{d}\tau = 0$ is marked by a red arrow). Red filled (stable) and unfilled (unstable) circles mark a fixed point in the evolutionary dynamics of m ($m^* = \beta$, see Eq. (3.35)). Blue circles and arrows illustrate the high fertilization rate attractor ($(m^*, \alpha^*) \to (\beta/4, \infty)$, see Eq. (3.37)). Average population trait trajectories, $(\langle m \rangle(t), \langle \alpha \rangle(t))$, from simulation of the stochastic model (see Appendix 3.8.3) are plotted in light gray, and their mean over multiple realisations are given dark gray. The time at which mutations occur, whether the trait values of m or α increase/decrease at each mutation event and the trait that undergoes mutation at each mutation event are random, giving rise to stochasticity in the evolutionary trajectories. The cost to fertilization is $C_z = 0.3$ (panel (a)) and $C_z = 0.6$ (panel (b)). In both panels $\beta = 1$ and $C_p = 0$. Remaining model parameters are given in 3.12.

conditions and parameters; the population can either evolve to large fertilization rates (limited only by $\alpha_{\max}$) or to zero fertilization rates.

When the cost to fertilization, C_z, is low (Figure 3.2, panel (a)), there exists a smaller region of initial conditions that drive α towards zero (pink shaded region). When $\alpha = 0$ within this region, selection on gamete mass, m, drives the population towards the point $m = \beta$ (red dot, see also Appendix 3.8.2). As this point exists outside the region in which $\mathrm{d}\alpha/\mathrm{d}\tau < 0$, selection for increased α can again manifest along the evolutionary trajectory. Thus when costs are sufficiently low, high fertilization rates are the only evolutionary outcome.

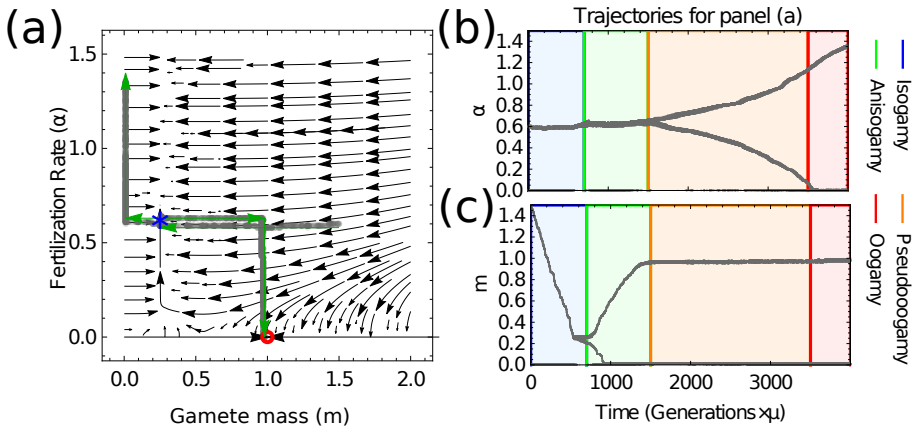

Figure 3.3: Numerical illustration of evolutionary branching in Figure 3.2(b). Panel (a): Analytic predictions for the early evolutionary dynamics (as Figure 3.2(b)) overlaid with trajectories $(m_i(t), \alpha_i(t)))$ for each i^{th} trait. Evolutionary branching is observed along the $m \approx \beta/4$ manifold, indicated by the blue star. Green arrows show the temporal progression of the branching. Panels (b) and (c): The temporal trajectories of the traits $\alpha_i(t)$ and $m_i(t)$ respectively, showing that the evolutionary trajectory passes from isogamy to oogamy. Parameters used are $A = 100$, $M = 1$, $T = 1$, $C_z = 0.6$, $C_p = 0$, $\beta = 1$, $\delta = 0.01$, $\mu = 10^{-3}$, $f_0 = 2 \times 10^{-3}$ and simulation run for 3.5×10^6 generations.

Conversely when costs to fertilization, C_z, are intermediate (Figure 3.2, panel (b)), there exists a larger region of initial conditions that drive α towards zero (pink shaded region). The point $m = \beta$ (red dot), towards which the population evolves when $\alpha = 0$, is now contained within this region in which $\mathrm{d}\alpha/\mathrm{d}\tau < 0$, and so is a stable fixed point. Thus when costs are sufficiently high, there are two evolutionary outcomes, depending on the initial conditions; either high fertilization rates or zero fertilization rates.

In Appendix 3.8.2 we conduct a mathematical and numerical analysis to formalise the arguments above. In summary, the possible early evolutionary attractors, (m^*, α^*), are given by:

$$(m^*, \alpha^*) \to \begin{cases} (\beta/4, \alpha_{\max}) & & \text{if } 1 - \frac{1-C_p}{\sqrt{\mathrm{e}}} > C_z \geq 0 \\ (\beta/4, \alpha_{\max}) \quad \text{or} \quad (\beta, 0) & & \text{if } 1 \gtrapprox C_z > 1 - \frac{1-C_p}{\sqrt{\mathrm{e}}} \\ (\beta, 0) & & \text{if } C_z \approx 1 \end{cases} \tag{3.9}$$

where we note that in the absence of parthenogenesis costs i.e $C_p = 0$, the condition in which a high fertilization rate is the only evolutionary outcome is $0 \leq C_z <$

$1 - 1/\sqrt{e} \approx 0.39$ (see Figure 3.2). Since situations in which the fertilization rate initially tends to zero are not of interest in the current analysis, we work in the remainder of this paper in the regime in which increasing fertilization rate is selected for; that is, $1 \gtrapprox C_z$ and with initial conditions that do not lead to a stable $\alpha = 0$ fixed point (see Figure 3.2, unshaded regions).

3.3.2 Evolutionary branching can lead to anisogamy, followed by oogamy

In Figure 3.2 we see that the approximation obtained for the co-evolutionary dynamics of m and α in a single fixed environment, Eq. (3.5), accurately captures the dynamics of the full model realised via numerical simulation at early times. One point of departure is that at long times as α increases along the $m \approx \beta/4$ manifold, we see the mean mass trait value from simulations increasing to higher values than those predicted analytically. In Figure 3.3, we show that this is a result of evolutionary branching in the simulations.

Should the evolutionary trajectories reach a sufficiently large fertilization rate, α, along the $m \approx \beta/4$ manifold, anisogamy can evolve through evolutionary branching in mass (see Figure 3.3), which gives rise to a dimorphic gamete population that contains one large gamete (macrogamete) with $m_{macro} \approx \beta - \delta m$ and one small gamete (microgamete) with $m_{micro} = \delta m$. Analytically, this is shown to be due to disruptive selection in mass in the vicinity of the $m \approx \beta/4$ manifold at high values of α. In Section 3.9, we show that this region is an approximate evolutionary singularity. This initial branching has been noted in other models that also do not account for self-incompatible mating types but that consider the case of fixed α and obligate sexual reproduction [143].

As the mass of the macrogamete becomes larger (and that of the microgamete smaller), a branching in fertilization rates can also occur (see Figure 3.3 (b)). Near this mass $m_{macro} \approx \beta - \delta m$, the macrogamete has a high survival rate under parthenogenesis, and the benefits of fertilization (in particular with microgametes of very small mass) can be outweighed by the costs of fertilization. Selection can thus act to lower the fertilization rate of macrogametes, α_{macro}, towards zero. This in turn leads to an increased selection pressure for the microgametes to increase their fertilization rate, with $\alpha_{\text{micro}} \to \alpha_{\max}$, to increase the probability of microgametes fertilizing macrogametes (averting the low survival probabilities of microgametes

under parthenogenesis). This is a situation in which macrogametes are still fertilized by microgametes (at a rate $\alpha_{\text{micro}}/2$) but do not fertilize themselves. Biologically, this situation can be interpreted as the evolution of oogamy, with $\alpha_{\text{micro}} \to \alpha_{\text{max}}$ and obligate sexual reproduction in the limit $\alpha_{\text{max}} \to \infty$.

Above we have shown that when costs to fertilization are accounted for, a continuous evolutionary trajectory can exist that takes the population from a state of isogamy to oogamy. In the following sections we demonstrate how each of these transitions can be arrested under various parameter regimes.

3.3.3 A low ceiling on the fertilization rate can stabilize isogamy

If the maximum possible fertilization rate α_{max} is limited to a low value, for instance due to energetic constraints on motility or environmental constraints such as turbulence, then it is possible to prevent the transition from isogamy to anisogamy. In Section 3.9, we derive the minimum value of α_{max} at which evolutionary branching in mass can occur, and find that if

$$\frac{48(1-C_p)\beta}{35AMT(1-C_z)e^{\frac{7}{6}}} > \alpha_{\text{max}} \tag{3.10}$$

evolutionary branching in mass cannot occur. These results are illustrated in Figure 3.4. Should α_{max} lie below this threshold (equivalent in approximately 70% of cells gametes fertilized in Figure 3.4, see Figure 3.9) the population is held in a state of isogamy. Alternatively, we can see in Eq. (3.10) that increasing the cost to parthenogenesis C_p and reducing the resistance to survival β can decrease the minimum value of α_{max} at which anisogamy can evolve, meaning that anisogamy evolves more readily under costly parthenogenesis.

3.3.4 High costs of parthenogenesis relative to zygote formation can stabilize anisogamy

Analytically, we show in Appendix 3.8.4 that when costs for zygote formation are relatively low and costs for parthenogenesis relatively high, true oogamy (in which the macrogamete's fertilization rate evolves towards zero, $\alpha_{\text{macro}} \to 0$) does not evolve following the transition to anisogamy. The costs to fertilization must instead be sufficiently high, such that

$$C_z > 1 - (1-C_p)\exp\left(-\frac{\delta m}{\beta - \delta m}\right) \tag{3.11}$$

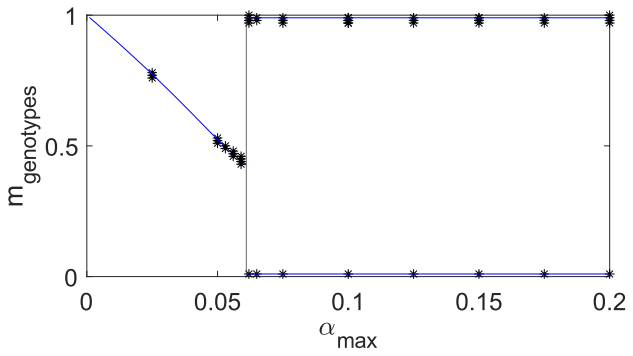

Figure 3.4: Analytical prediction (blue) and numerical illustration of the range of α_{max} in which branching to anisogamy is possible. Black markers represent the mass of each gamete genotype within the population after 2.5×10^6 generations. Once branching to anisogamy has occurred, a dimorphic gamete population, characterised by the presence of two genotypes would be present. One genotype where $m \approx \beta - \delta m$ and one where $m = \delta m$. Vertical line represents the analytically predicted α_{max} above which branching can occur in m Eq. (3.10). The blue curve left of this line is the numerical solution to $dm/d\tau|_{\alpha=\alpha_{max}} = 0$ in Eq. (3.34) and the blue horizontal lines towards the right represent the theoretical masses of the macrogametes ($m = \beta - \delta m$) and microgametes ($m = \delta m$). Parameters are $A = 100$, $M = 1$, $T = 0.1$, $C_z = 0.3$, $C_p = 0$, $\beta = 1$, $\delta = 0.01$, $\mu = 10^{-3}$ and $f_0 = 2 \times 10^{-3}$.

for true oogamy to evolve. This is illustrated in Figure 3.5. Essentially if the costs to the macrogamete of fertilization are not sufficiently high, then there is no longer a selective pressure (as described in Section 3.3.2) for the macrogamete to avoid fertilizations by decreasing its fertilization rate. While the population does not evolve towards oogamy, its ultimate state is dependent on the maximum fertilization rate, $\alpha_{\max}$.

Suppose we initially place no limit on the maximum fertilization rate, $\alpha_{\max} \to \infty$, and that the costs C_p and C_z are such that there is no longer selection for the fertilization rate of the macrogamete to decrease (i.e. the inequality in Eq. (3.11) does not hold). While there may be a selective pressure for the fertilization rate of the macrogamete to increase, this selection pressure is weaker for the parthenogenetic macrogamete than the microgamete, as the microgamete relies more heavily on the fertilization pathway for its survival. Thus although the fertilization rates of both macrogametes and microgametes, α_{macro} and α_{micro}, evolve to increase α indefinitely the macrogamete does so at a considerably slower rate. This scenario is known as pseudooogamy [198] where the fertilization rate of the macrogamete does not

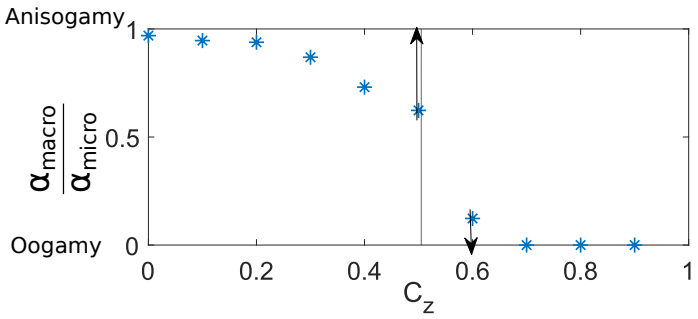

Figure 3.5: Numerical illustration of the ratio of macrogamete to microgamete fertilization rate, $\alpha_{macro}/\alpha_{micro}$. Oogamy is favoured over anisogamy above a sufficiently high fertilization cost C_z, predicted analytically by the vertical line (see Eq. (3.11)). Here $C_p = 0.5$, $(m(0), \alpha(0)) = (0.25, 0.02)$ and $\alpha_{max} = 1.3$. All other parameters are as in Figure 3.2 except $f_0 = 0.02$ and the simulation is run for 6×10^6 generations. Back arrows point in the direction towards which $\alpha_{macro}/\alpha_{micro}$ evolves in infinite time.

evolve towards zero but still evolves to a fertilisation rate lower than that of the microgamete. While this situation is only observed when $\alpha_{\max} \to \infty$, in the case of large but finite $\alpha_{\max}$, pseudooogamy is a prolonged transient state (see Figures 3.5 and 3.14). Lastly, should $\alpha_{\max}$ be smaller, such that the increasing fertilization rate of macrogametes can "catch up" with that of microgametes, the population can return to anisogamy following a transient period of oogamy.

3.3.5 In a switching environment, the population can evolve a bet-hedging strategy that stabilizes isogamy

In Figure 3.6, we see that the analysis for the evolutionary dynamics in the case of a switching environment (see Eq. (3.8)) provides a good approximation to the dynamics of the full model (which accounts for multiple traits coexisting under a mutation-selection balance) realised via numerical simulation. A mutation-selection balance refers to the balance between the rate at which deleterious mutations arise and the rate at which they are eliminated by selection. The number of coexisting traits depends on this balance, whereby increasing the mutation rate would increase the number of traits that coexist under this mutation-selection balance. We now see three broad evolutionary outcomes.

We begin by considering the intuitive limit of $\tau_1 \gg \tau_2$. In this scenario the population spends almost all of the time in environment 1, and a comparatively

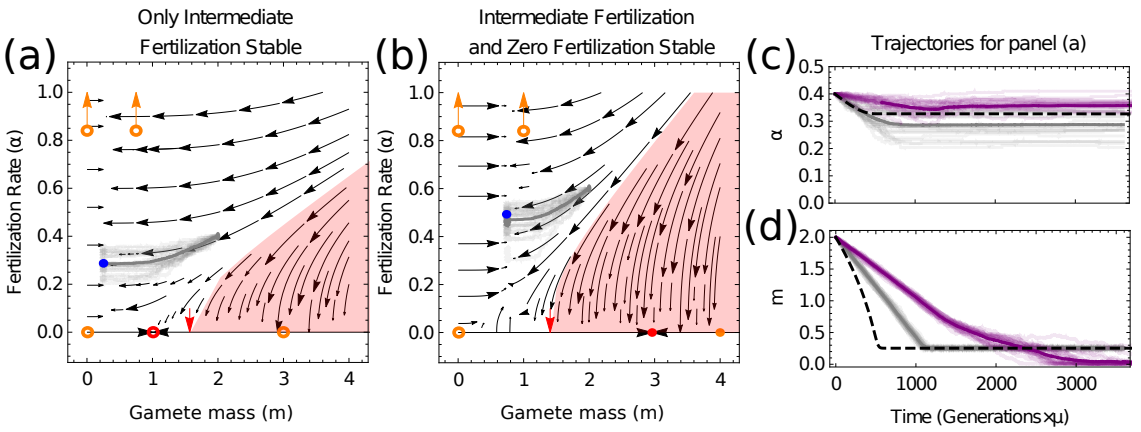

Figure 3.6: Phase portraits for the co-evolutionary dynamics in a switching environment (see Eq. (3.8)). In addition to qualitatively similar dynamics as in the fixed environment (see Figure 3.2), two new evolutionary scenarios are now possible, including populations in which stable intermediate fertilization rates (filled blue circle) are the only evolutionarily stable state (see panel (a)) and populations in which there is an additional zero fertilization stable state (filled red circle, see panel (b)). Open orange circles represent the (now unstable) states to which the population can be attracted in either environment 1 or 2 (where $\beta = \beta_1$ or $\beta = \beta_2$). Average population trait trajectories, $(\langle m \rangle(t), \langle \alpha \rangle(t))$, from simulation of the full stochastic model in the FRTI regime are overlaid in gray. In panel (c) we plot the time-series in the FRTI regime (gray, as in panel (a)) alongside those in the FRTE regime (purple) and our analytic predictions (black-dashed). In all panels, $A = 100$, $M = 10$, $T = 0.1$ and $f_0 = 2 \times 10^{-3}$. Under the FRTI regime (overlaid in gray in panels (a), (b) and (c)), $\mu = 3.5 \times 10^{-4}$, $\delta = 7 \times 10^{-3}$ and the simulation is run for 10^7 generations. Under the FRTE regime (overlaid in purple in panel (c)), $\mu = 5 \times 10^{-3}$, $\delta = 5 \times 10^{-3}$ and the simulation is run for 7×10^6 generations. In panel (a) (and panel (c) gray), $C_z = 0.35$, $\beta_1 = 3$, $P_1 = 0.335$, $(m(0), \alpha(0)) = (2, 0.4)$, $\lambda_{1\to 2} = 0.250$ and $\lambda_{2\to 1} = 0.126$. In panel (b), $C_z = 0.7$, $\beta_1 = 4$, $P_1 = 0.74$, $(m(0), \alpha(0)) = (2, 0.6)$, $\lambda_{1\to 2} = 5.86 \times 0.01$ and $\lambda_{2\to 1} = 0.167$. In panels (c) and (d), the switching rates for the FRTE regime (purple) are $\lambda_{1\to 2} = 2.93 \times 10^{-5}$ and $\lambda_{2\to 1} = 8.34 \times 10^{-5}$. Switching rates and mutation rates μ are all measured in units of (number of generations)$^{-1}$ and $C_p = 0$ in all panels.

insignificant amount of time in environment 2 (i.e. $P_1 \approx 1$ and $P_2 \approx 0$). Consequently, the population evolves approximately as if it were simply in a fixed environment with $\beta = \beta_1$, and the conditions given for the fixed environment, Eq. (3.9), can be used to infer the evolutionary outcome. An analogous argument holds for the dynamics when $\tau_2 \gg \tau_1$, but with $\beta = \beta_2$ in Eq. (3.9). Once on the $m = \beta_i/4$ manifold towards

the high-fertilization rate attractor, evolutionary branching towards anisogamy and oogamy can occur as described in Section 3.3.2. When the time spent in each environment is of the a comparable order however (e.g when τ_1, and τ_2 have not entirely dissimilar magnitudes), we find the emergence of bet-hedging strategies (see Figures 3.6, 3.7 and Eq. (3.14)).

Under a range of parameter conditions, we find analogous evolutionary attractors for the fertilization rate as in the fixed-environment case (see Eq. (3.9) and Figure 3.6), but with bet-hedging strategies for the gamete mass;

$$\begin{aligned}(m^*, \alpha^*) &\to (m^*_{\alpha=0}, 0) \\ (m^*, \alpha^*) &\to (m^*_{\alpha\to\infty}, \alpha_{\max})\end{aligned} \tag{3.12}$$

with

$$\begin{aligned}m^*_{\alpha=0} &= P_1\beta_1 + (1-P_1)\beta_2 \\ m^*_{\alpha\to\infty} &= \frac{1}{4}\left(P_1\beta_1 + (1-P_1)\beta_2\right) .\end{aligned} \tag{3.13}$$

The population thus initially evolves either to high fertilization rates or zero fertilization rates, but with a mass which is the weighted average of the optimal strategy in either environment.

Under a more restricted set of parameter conditions however we find that a bet-hedging strategy for the fertilization rate can also evolve; essentially a switching-induced fixed point can manifest, as illustrated in Figures 3.6 and 3.7. Here the tension between the evolutionary dynamics in the two environments (which can select for high fertilization rate and large gametes in one environment, and zero fertilization rate and small gametes in the other) can lead to the population being held in a state at which intermediate finite fertilization rate and isogamy form the evolutionarily stable bet-hedging strategy. In Appendix 3.11.3, we show that the switching-induced fixed point is given by

$$\begin{aligned}m^* &= \frac{1}{4}\left[P_1\beta_1 + (1-P_1)\beta_2\right] \\ \alpha^* &= \frac{(1-C_p)m^*}{AMT(1-C_z)} \\ &\times \left(\frac{(1-C_z)\left(P_1 e^{\frac{\beta_1}{2m^*}} + (1-P_1)e^{\frac{\beta_2}{2m^*}}\right) - (1-C_p)}{(1-C_p)(P_1 e^{\frac{\beta_2}{2m^*}} + (1-P_1)e^{\frac{\beta_1}{2m^*}}) - (1-C_z)e^{\frac{\beta_1+\beta_2}{2m^*}}}\right) .\end{aligned} \tag{3.14}$$

In Figure 3.6 (c-d) we see that this switching induced fixed point is observed in the evolutionary simulations in which multiple traits can be held in the population

under a mutation-selection balance. In the FRTI regime (gray lines), in which the

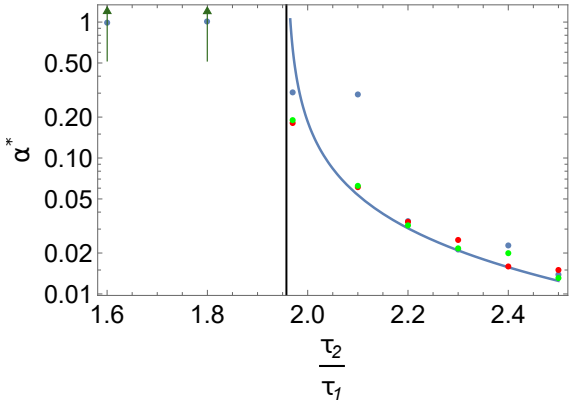

Figure 3.7: Analytic theory for the fertilization rate at the switching-induced isogamous fixed point (blue line, see Eq. (3.14)) against simulations (dots). The black vertical line is the analytic prediction for when this fixed point vanishes, with parameters left of the line corresponding to destabilized isogamy. Simulations are obtained by averaging over 25 realisations of $\langle \alpha(t) \rangle$ corresponding to stochastic evolutionary trajectories. Parameters are $\delta = 0.01$, $\mu = 5 \times 10^{-4}$, $f_0 = 2 \times 10^{-3}$, $\tau_1 = 0.25$ and simulation run for 1.2×10^7 generations. All other parameters are the same as in the FRTI regime of Figure 3.6 (a). Red markers are for $(m(0), \alpha(0)) = (0.29, 0.075)$, green for $(m(0), \alpha(0)) = (0.29, 0.15)$ and blue for $(m(0), \alpha(0)) = (0.29, 0.3)$.

environment changes more quickly, the population is held at the predicted mass m, while more variability between simulations is seen around the predicted fertilization rate α as a result of the weaker selection on this trait. In the slower-switching FRTE regime (purple lines), a greater quantitative difference between the results of simulations and the analytic predictions is observed. However in both regimes, the key prediction of finite fertilization rate (Figures 3.6 and 3.7) and isogamy is indeed captured (see Figures 3.16 and 3.17).

3.4 CONCLUSIONS

In this chapter we have extended classic results on the evolution of anisogamy [124, 173] to account for parthenogenetic development [37, 130], the co-evolution of fertilization rate and gamete cell mass, and stochastically varying environments. In doing so we have demonstrated the possibility of a continuous evolutionary trajectory from an initial state of isogamy to anisogamy followed by oogamy, an evolutionary trajectory observed throughout the eukaryotes [81, 112].

Consistent with earlier theoretical results on the evolution of anisogamy that neglected the pre-existence of self-incompatible mating types [143], we have shown that anisogamy can evolve from isogamy through evolutionary branching. However our model also shows that when the possibility of parthenogenetic development is accounted for, a subsequent branching to oogamy in the population is possible. Importantly, this transition is possible even without explicitly accounting for pheromone-receptor systems [79, 96], speed-mass relationships [38], mechanistic gamete encounter rates [50] or costly motility [199], as in previous work. Instead in our model, oogamy can naturally evolve from anisogamy via sexual conflict [29], following the evolution of anisogamy from isogamy. In particular, microgametes receive a survival advantage by increasing their fertilization rate with macrogametes, while macrogametes evolve to zero fertilization rate to overcome fertilization costs. This requirement of fertilization costs for oogamy to evolve under parthenogenesis provides a complementary result to [198], where instead high motility costs were required for oogamy.

We have shown analytically that various conditions exist that can arrest the population at different stages of the isogamy-anisogamy-oogamy evolutionary trajectory. In particular our model suggests that isogamy can be stabilized under greater energetic constraints or in highly turbulent environments. We have also shown that isogamy can be stabilized as a bet-hedging strategy under switching environmental conditions. Here, in organisms that reproduce parthenogenetically, microgametes that fail to find a partner face low survival prospects under harsh environmental conditions, and thus the transition to anisogamy is frustrated. Such dynamics may be at play in isogamous species that reproduce parthenogenetically, such as those among the *ectocarpus* [125] and *Blidingia minima* [216].

The results described above are dependent on the capacity for parthenogenetic development amongst gametes that fail to find a partner to fertilize with. Such a capacity is widespread in the brown algae [151] and can also be found in the green algae [130]. We have also shown that for the evolution of oogamy in our model, we require costs of fertilization to exceed those of parthenogenetic development. While empirically parthenosporophytes have lower survival probabilities than zygotes [130], disentangling the costs of development pathways (parthenogenesis or fertilization) from those of increased mass is challenging [37]. By invoking costs to zygote formation arising from cytoplasmic conflict [86], our work is reminiscent of alternative hypotheses for the evolution of anisogamy [100, 101], that suggest that anisogamy

reduces the potential for such conflict by limiting cytoplasmic contributions from the microgamete. However many other costs to fertilization could be present, including the possibility of failure during cell fusion [80], and the energetic and time costs of sexual reproduction [131]. Unpacking and quantifying these costs remains an interesting area of empirical research.

Our primary reason for neglecting the existence of mating types in our model was for mathematical simplicity [143]. Including such types in a model would be a natural next step, and based on the biological reasoning above, such a model should demonstrate similar qualitative behaviours. However it is worthwhile noting that many models of the evolution of sexual reproduction in early eukaryotes suppose the existence of a "unisexual" early ancestor that mated indiscriminately [89]. Thus our model, which has neglected the existence of mating types, could be seen as reflecting such a system. In this context, the evolution we have seen of incompatibility between macrogametes (that fuse with each other at rate zero in the oogamous scenario) is interesting, as the origins of self-incompatible mating types are still debated [16, 78]. This mechanism would represent an "anisogamy consequence" model for the evolution of mating types that manages to identify the conditions under which fusion between large macrogametes is disadvantageous [17]. However as such "anisogamy consequence" models are not consistent with most empirical observations (that suggest mating types preceded anisogamy), this interpretation should be treated with caution.

As with any model, there are omissions from our formulation. These include the mortality of gametes (known to stabilize isogamy [124]) the discrete nature of cell divisions leading to gametes [130], and non-local trait mutations. While these would be interesting additions to our model, the key insights derived from the PBS model have remained remarkably robust to such generalizations, and so the inclusion of these additional considerations to our model may lead primarily to quantitative, rather than qualitative, changes in results. More generally, extending our mathematical approach leveraging adaptive dynamics to switching environments [155] in other facultatively sexual populations might prove particularly fruitful [36, 42].

In this chapter, we have extended the models of [143, 171] in several ways; by allowing the fertilization rate to evolve, accounting for the possibility for unfertilized gametes to develop parthenogenetically should they fail to locate a partner, and subjecting the system to switching environments. In doing so, we have shown its capacity to parsimoniously capture continuous evolutionary trajectories from isogamy

to oogamy in parthenogens. Moreover, our models emphasise the importance of investigating the co-evolutionary dynamics for a range of evolutionary parameters and its implications for the evolution of oogamy.

3.5 APPENDIX I: INDIVIDUAL SPOROPHYTES MUST PRODUCE TWICE AS MANY SPORES AS PARTHENOSPOROPHYTES TO MAINTAIN THE SAME REPRODUCTIVE OUTPUT

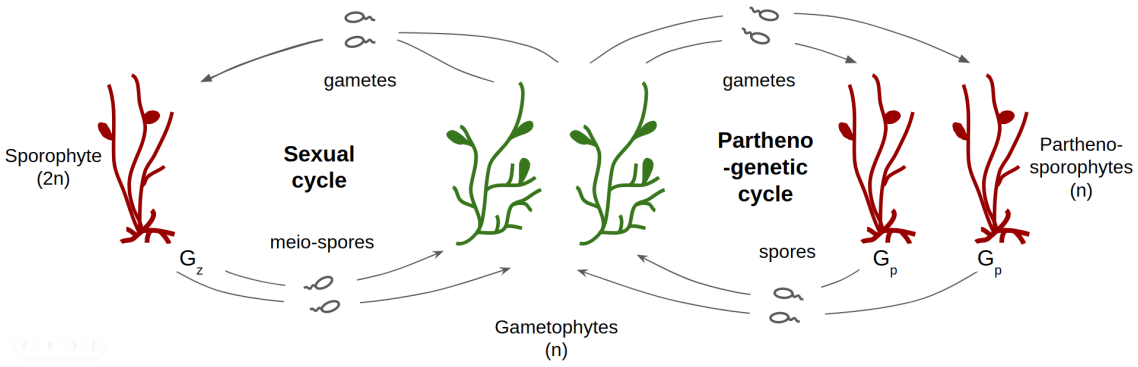

Figure 3.8: Diagram, inspired by the life-cycle of Ectocarpus, illustrating the model dynamics in terms of diploid sporophytes and haploid parthenosporopytes. The figure is adapted from [151] to account for the fact that in our model, both male and female gameteophytes can in principle develop parthenogenetically.

Here, we show that diploid sporophytes must produce twice as many meio-spores as are produced by haploid parthenosoprophytes in order to maintain a parity in reproductive output (see Figure 3.8). Further, we show that meio-spore production in sporophytes below this level constitutes an implicit cost of fertilization.

Recall that, in the main text, C_z is the survival cost incurred by taking the sexual reproductive route while C_p is the survival cost incurred by taking the parthenogenetic reproductive route. We denote $F(T)$ as the number of fertilized zygotes, $2F(T)$ as the number of haploid gametes that have formed zygotes at the end of a generation, and $N(T)$ as the number of unfertilized gametes at the end of a generation. The

functional form for the absolute fitness of a single genotype is

$$\begin{aligned} w &= 2F(T)S(\beta, m_z)(1 - C_z) + N(T)S(\beta, m_p)(1 - C_p) \\ &= (1 - C_p)\left[2F(T)S(\beta, m_z)\left(\frac{1 - C_z}{1 - C_p}\right) + N(T)S(\beta, m_p)\right] \end{aligned} \tag{3.15}$$

where m_z is the mass of fertilized zygotes, m_p is the mass of unfertilized gametes and $S(\beta, m)$ is the survival function (see Eq. (3.2)).

Alternatively, we can express the model in terms of the number of sporophytes and parthenosporophytes, and their individual rates of meio-spore/spore production, G_z and G_p respectively. We again denote the number of zygotes (destined to become diploid sporophytes) as $F(T)$ and the number of unfertilized gametes (destined to become haploid parthenosporophytes) as $N(T)$. The absolute fitness of a single genotype can now be written

$$\begin{aligned} w &= F(T)S(\beta, m_z)G_z + N(T)S(\beta, m_p)G_p \\ &= G_p\left(F(T)S(\beta, m_z)\frac{G_z}{G_p} + N(T)S(\beta, m_p)\right). \end{aligned} \tag{3.16}$$

We now note that in terms of relative fitness, the constant factors $(1 - C_p)$ and G_p preceding Eqs. (3.15-3.16) are inconsequential.

Now equating the pre-factors of $F(T)S(\beta, m_z)$ in Eqs. (3.15-3.16) allows us to evaluate the mass-independent costs to zygote formation in the model. We see that

$$2\left(\frac{1 - C_z}{1 - C_p}\right) = \frac{G_z}{G_p} \tag{3.17}$$

we can rearrange in terms of C_z to get

$$C_z = 1 - \frac{G_z}{2G_p}(1 - C_p). \tag{3.18}$$

For zero costs to parthenogenesis, $C_p = 0$, production of meio-spores by sporophytes (G_z) must be twice that of spores by parthenosporophytes (G_p) in order to achieve zero cost to zygote formation. This result has a straightforward biological interpretation. Since under Mendelian inheritance the reproductive fitness of sporophytes is shared between the gametes that contribute towards its production, the fitness of sporophytes must be at least twice that of parthenosporophytes to avoid an implicit cost to zygote formation.

3.6 APPENDIX II: WITHIN GENERATION DYNAMICS

At the start of each generation, we assume a total of A organisms of mature cell size (or energy budget) M divide to form gametes. We further assume that the frequency of mutants in this adult population is given by $1 \geq \hat{f}_i \geq 0$, with $i = m$ for mutants that change the mass of gametes and $i = \alpha$ for mutants that change the fertilization rate of gametes.

3.6.1 FERTILIZATION KINETICS: MUTANT WITH DIFFERENT MASS

If a mutation occurs changing the size of gametes produced by the mutant ($\hat{m} = m + \delta m$ for resident gamete mass m), the fertilization kinetics themselves (see Eq. (3.1)) are unaffected by the change in mass. However the number of gametes produced by the mutant will change. Denoting by N the number of resident gametes and $\hat{N}$ the number of mutant gametes, we have

$$\begin{aligned} \frac{\mathrm{d}N}{\mathrm{d}t} &= -\alpha(N^2 + N\hat{N})\,; \qquad N(0) = \frac{AM(1-\hat{f}_m)}{m} \\ \frac{\mathrm{d}\hat{N}}{\mathrm{d}t} &= -\alpha(\hat{N}^2 + N\hat{N})\,; \qquad \hat{N}(0) = \frac{AM\hat{f}_m}{\hat{m}}\,, \end{aligned} \tag{3.19}$$

which has a solution

$$\begin{aligned} N(t) &= \frac{N(0)}{1 + \Big(N(0) + \hat{N}(0)\Big)\,\alpha t} \\ \hat{N}(t) &= \frac{\hat{N}(0)}{1 + \Big(N(0) + \hat{N}(0)\Big)\,\alpha t}\,. \end{aligned} \tag{3.20}$$

This allows us to determine the number of unfertilized cells of each type at the end of the fertilization window at $U = N(T)$ and $\hat{U} = \hat{N}(T)$, where we recall that T is the length of the fertilization window.

We also need to determine the number of fertilized cells formed from the fertilization of two residents, F_N^N, two mutants, $F_{\hat{N}}^{\hat{N}}$, and a mutant and a resident, $F_N^{\hat{N}}$. For this we need to solve

$$\begin{aligned} \frac{\mathrm{d}F_N^N}{\mathrm{d}t} &= \frac{1}{2}\alpha N^2; \qquad F_N^N(0) = 0 \\ \frac{\mathrm{d}F_N^{\hat{N}}}{\mathrm{d}t} &= \alpha N\hat{N}; \qquad F_N^{\hat{N}}(0) = 0 \\ \frac{\mathrm{d}F_{\hat{N}}^{\hat{N}}}{\mathrm{d}t} &= \frac{1}{2}\alpha \hat{N}^2; \qquad F_{\hat{N}}^{\hat{N}}(0) = 0\,, \end{aligned} \tag{3.21}$$

which can be solved by substituting for N and $\hat{N}$ from Eq. (3.20) to yield

$$\begin{aligned} F_N^N(t) &= \frac{1}{2}\frac{N(0)^2 t}{1+\left(N(0)+\hat{N}(0)\right)\alpha t} \\ F_N^{\hat{N}}(t) &= \frac{N(0)\hat{N}(0)t}{1+\left(N(0)+\hat{N}(0)\right)\alpha t} \\ F_{\hat{N}}^{\hat{N}}(t) &= \frac{1}{2}\frac{\hat{N}(0)^2 t}{1+\left(N(0)+\hat{N}(0)\right)\alpha t}\,. \end{aligned} \tag{3.22}$$

3.6.2 Fertilization kinetics: mutant with different fertilization rate

If a mutation occurs changing the fertilization rate of gametes produced by the mutant ($\hat{\alpha} = \alpha + \delta\alpha$ for residents with fertilization rate α), the fertilization kinetics themselves are altered relative to Eq. (3.1). We assume for simplicity that the fertilization rate between resident-pairs is the mean of the fertilization rate between the two types in isolation, such that

$$\begin{aligned} \frac{\mathrm{d}N}{\mathrm{d}t} &= -\alpha N^2 - \left(\frac{\alpha+\hat{\alpha}}{2}\right) N\hat{N}\,; \qquad N(0) = \frac{AM(1-\hat{f}_\alpha)}{m} \\ \frac{\mathrm{d}\hat{N}}{\mathrm{d}t} &= -\hat{\alpha}\hat{N}^2 - \left(\frac{\alpha+\hat{\alpha}}{2}\right) N\hat{N}\,; \qquad \hat{N}(0) = \frac{AM\hat{f}_\alpha}{m}\,. \end{aligned} \tag{3.23}$$

Solving this equation is slightly less straightforward than solving Eq. (3.19). However we can make analytic progress by making a change of variables and applying an approximation based on small mutational step size $\delta\alpha$.

We introduce the transformed variables N_{tot} and $\hat{r} = \frac{\hat{N}}{N+\hat{N}}$, representing the total number of unfertilized cells and the frequency of unfertilized mutant cells respectively. Eq. (3.23) then becomes

$$\begin{aligned} \frac{\mathrm{d}N_{\mathrm{tot}}}{\mathrm{d}t} &= -N_{\mathrm{tot}}^2(\alpha + \delta\alpha\,\hat{r}) \\ \frac{\mathrm{d}\hat{r}}{\mathrm{d}t} &= -\frac{1}{2}\delta\alpha\, N_{\mathrm{tot}}\hat{r}(1-\hat{r})\,. \end{aligned} \tag{3.24}$$

We now see that although this equation is also intractable, the leading order dynamics of N_{tot} are governed by α. Therefore when $\alpha \gg \delta\alpha$, we make the approximation $(\mathrm{d}N_{\mathrm{tot}}/\mathrm{d}t) \approx -\alpha N_{\mathrm{tot}}^2$. We then obtain

$$N_{\mathrm{tot}}(t) \approx \frac{N_{\mathrm{tot}}(0)}{1+N_{\mathrm{tot}}(0)\alpha t}\,, \tag{3.25}$$

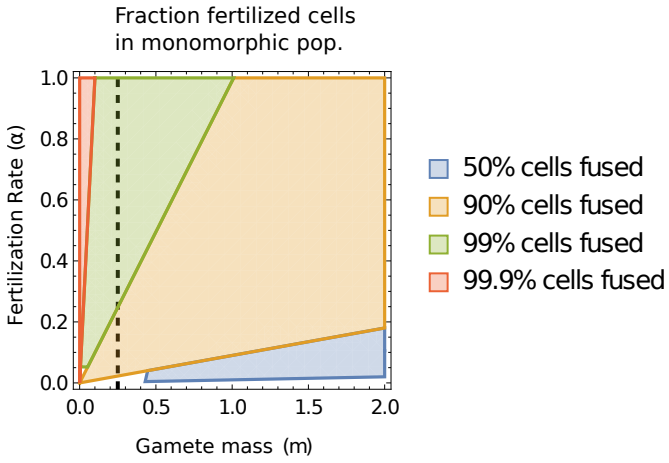

Figure 3.9: Illustration of the total proportion of cells that are fertilized at the end of a fertilization period (length $T = 1$) in a monomorphic isogamous population (without branching) as a function of trait variables m and α. Parameters used are the same as those in Figure 3.2. The vertical black dashed line gives the location of the manifold $(\beta/4, \alpha)$, along which the population is attracted to when approaching the high α attractor.

and substituting this into our equation for $\hat{r}$ in Eq. (3.24), we can solve to obtain

$$\hat{r}(t) = \frac{\hat{r}(0)}{\hat{r}(0) + (1 - \hat{r}(0))(1 + N_{\text{tot}}(0)\alpha t)^{\frac{\delta\alpha}{2\alpha}}} \,. \tag{3.26}$$

Inverting the transformation, we then arrive at

$$\begin{aligned} N(t) &= N_{\text{tot}}(t)(1 - \hat{r}(t)) \\ \hat{N}(t) &= N_{\text{tot}}(t)\hat{r}(t) \,, \end{aligned} \tag{3.27}$$

with N_{tot} and $\hat{r}(t)$ taken from Eqs. (3.25-3.26). We shall see in Appendix 3.6.4 that calculating F_N^N, $F_{\hat{N}}^{\hat{N}}$, and $F_N^{\hat{N}}$ explicitly is in fact unnecessary, and these expressions for $N(t)$ and $\hat{N}(t)$ are sufficient for analytical progress.

3.6.3 CHANGE IN MUTANT FREQUENCY OVER A GENERATION: MUTANT WITH DIFFERENT MASS

We begin by calculating the fitness of the resident and a mutant that changes the mass of gametes, w_m and $\hat{w}_m$ respectively, which are simply given by the total number of cells of each type at the end of a generation. Recalling that fertilized and unfertilized gametes both survive with a probability governed by the parameter β

and the mass of the cell m_c (see Eq. (3.2)), and that fertilized cells survive with an additional probability $(1 - C_z)$, we have

$$\begin{aligned} w_m &= (1 - C_z)\left(2F_N^N(T)S(\beta, 2m)\right. \\ &\qquad \left. + F_N^{\hat{N}}(T)S(\beta, m + \hat{m})\right) + N(T)S(\beta, m)(1 - C_p) \\ \hat{w}_m &= (1 - C_z)\left(2F_{\hat{N}}^{\hat{N}}(T)S(\beta, 2\hat{m})\right. \\ &\qquad \left. + F_N^{\hat{N}}(T)S(\beta, m + \hat{m})\right) + \hat{N}(T)S(\beta, \hat{m})(1 - C_p) \end{aligned} \tag{3.28}$$

where $F_N^N(T)$, $F_{\hat{N}}^{\hat{N}}(T)$, and $F_N^{\hat{N}}(T)$ are taken from Eq. (3.22), and $N(t)$ and $\hat{N}(t)$ are taken from Eq. (3.20). These expressions can be used to calculate the frequency at the end of the generation, $\hat{f}'_m$, of a mutant that changes the mass of gametes as

$$\hat{f}'_m = \frac{\hat{w}_m}{w_m + \hat{w}_m}\,. \tag{3.29}$$

The change is the frequency of the mutant over the course of a generation is then $\hat{f}'_m - \hat{f}_m$.

3.6.4 CHANGE IN MUTANT FREQUENCY OVER A GENERATION: MUTANT WITH DIFFERENT FERTILIZATION RATE

Taking an analogous approach to Appendix 3.6.3, we begin by calculating the fitness of the resident and a mutant that changes the fertilization rate, w_α and $\hat{w}_\alpha$ respectively. We obtain

$$\begin{aligned} w_\alpha &= (1 - C_z)\left(2F_N^N(T)S(\beta, 2m)\right. \\ &\qquad \left. + F_N^{\hat{N}}(T)S(\beta, 2m)\right) + (1 - C_p)N(T)S(\beta, m) \\ &= (1 - C_z)S(\beta, 2m)\left(N(0) - N(T)\right) \\ &+ (1 - C_p)N(T)S(\beta, m) \\ \hat{w}_\alpha &= (1 - C_z)\left(2F_{\hat{N}}^{\hat{N}}(T)S(\beta, 2m)\right. \\ &\qquad \left. + F_N^{\hat{N}}(T)S(\beta, 2m)\right) + (1 - C_p)\hat{N}(T)S(\beta, m) \\ &= (1 - C_z)S(\beta, 2m)\left(\hat{N}(0) - \hat{N}(T)\right) \\ &+ (1 - C_p)\hat{N}(T)S(\beta, m) \end{aligned} \tag{3.30}$$

where $N(t)$ and $\hat{N}(t)$ are now taken from Eq. (3.27). Here we have used the fact that since the survival function for fertilized cells, $S(\beta, 2m)$, is independent

of the composition of the fertilized cells (mutations here only affect fertilization rate) the number of cell-types contributing to the fertilized cells can be inferred under cell conservation during the fertilization period (e.g. for resident cell types $\mathrm{d}/\mathrm{d}t\left[2F_N^N(t)+F_N^{\hat{N}}(t)+N(t)\right]=0$, and similarly for mutant cell types).

These expressions can be used to calculate the frequency at the end of the generation, $\hat{f}'_\alpha$, of a mutant that changes the fertilization rate as

$$\hat{f}'_\alpha=\frac{\hat{w}_\alpha}{w_\alpha+\hat{w}_\alpha}\,. \tag{3.31}$$

The change is the frequency of the mutant over the course of a generation is then $\hat{f}'_\alpha-\hat{f}_\alpha$.

3.7 APPENDIX III: INVASION DYNAMICS

Our aim is to derive the dynamics of the frequency of a mutant with a small mutation in either of the traits, m or α, over multiple generations. We introduce t_g as a measure of the number of discrete generations.

3.7.1 INVASION ODE: MUTANT WITH DIFFERENT MASS

We begin by deriving the dynamics for the frequency of a mutant that changes the mass of gametes. We begin by assuming that the mutational step size, δm, is small. Under these conditions, the frequency of mutants changes only by a small amount over the course of one generation, and we can approximate the frequency of mutants at the end of the generation (see Appendix 3.6.3) by $\hat{f}'_m=\hat{f}_m+\delta m\,(\mathrm{d}\hat{f}'_m/\mathrm{d}\,\delta m)|_{\delta m=0}$, where $\hat{f}_m$ is the frequency of the mutants at the beginning of the generation. The dynamics of the mutant frequency over an invasion can then be approximated by

$$\begin{aligned}
\frac{\mathrm{d}\hat{f}_m}{\mathrm{d}t_g}&=\left(\frac{\hat{f}'_m-\hat{f}_m}{t_g}\right)\Bigg|_{t_g\to 1}\\
&=\delta m\left[\frac{\mathrm{d}\hat{f}'_m}{\mathrm{d}\,\delta m}\right]\Bigg|_{\delta m=0}\\
&=\delta m\left[\frac{AM\alpha T(1-C_z)e^{\frac{\beta}{2m}}(\beta-4m)-4m(m-\beta)(1-C_p)}{4m^2\left(AM\alpha T(1-C_z)e^{\frac{\beta}{2m}}+m(1-C_p)\right)}\right]\\
&\qquad\times\hat{f}_m(1-\hat{f}_m)\\
&\equiv h_m(m,\alpha,\beta,C_z,C_p)\hat{f}_m(1-\hat{f}_m)\,,
\end{aligned} \tag{3.32}$$

(see also Eq. (3.3)), where $\mathrm{d}\hat{f}'_m/\mathrm{d}\,\delta m$ is derived from Eq. (3.29). We show in Figure 3.10 that as expected, this is a good approximation for the dynamics when δm is small.

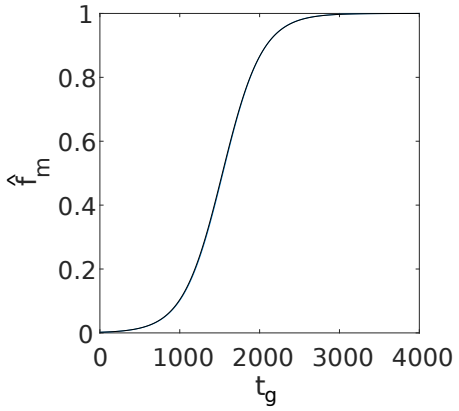

Figure 3.10: Invasion dynamics for a mutant with mass $m + \delta m$. Blue - analytical prediction using Eq. (3.32), black - numerical simulation. The initial condition is $(m(0), \alpha(0)) = (0.4, 0.1)$. Parameters are $\delta m = -0.005$, $f_0 = 0.002$, $G = 4 \times 10^3$, $A = 100$, $M = 1$, $T = 1$, $C_z = 0.6$, $C_p = 0$ and $\beta = 1$.

3.7.2 Invasion ODE: mutant with different fertilization rate

We now derive the dynamics for the frequency of a mutant that changes the fertilization rate of gametes. Taking an analogous approach to Appendix 3.7.1, we assume the mutational step size for mutation, $\delta\alpha$, is small. We can then approximate the frequency of mutants at the end of the generation (see Appendix 3.6.4) by $\hat{f}'_\alpha = \hat{f}_\alpha + \delta\alpha\,(\mathrm{d}\hat{f}'_\alpha/\mathrm{d}\,\delta\alpha)|_{\delta\alpha=0}$, where $\hat{f}_\alpha$ is the frequency of the mutants at the beginning of the generation. The dynamics of the mutant frequency over an invasion

can then be approximated by

$$\begin{aligned}\frac{\mathrm{d}\hat{f}_\alpha}{\mathrm{d}t_g} &= \left(\frac{\hat{f}'_\alpha - \hat{f}_\alpha}{t_g}\right)\Bigg|_{t_g\to 1} \\ &= \delta\alpha\left[\frac{\mathrm{d}\hat{f}'_\alpha}{\mathrm{d}\,\delta\alpha}\right]\Bigg|_{\delta\alpha=0} \\ &= \delta\alpha\left[\frac{m\left((1-C_z)e^{\frac{\beta}{2m}} - (1-C_p)\right)\ln\left(\frac{\alpha AMT}{m}+1\right)}{2\alpha\left(\alpha A(1-C_z)MTe^{\frac{\beta}{2m}} + m(1-C_p)\right)}\right] \\ &\qquad \times \hat{f}_\alpha(1-\hat{f}_\alpha) \\ &\equiv h_\alpha(m,\alpha,\beta,C_z,C_p)\hat{f}_\alpha(1-\hat{f}_\alpha)\,,\end{aligned} \tag{3.33}$$

(see also Eq. (3.4)) where $\mathrm{d}\hat{f}'_\alpha/\mathrm{d}\,\delta\alpha$ is derived from Eq. (3.31). We show in Figure 3.11 that as expected, this is a good approximation for the dynamics when $\delta\alpha$ is small.

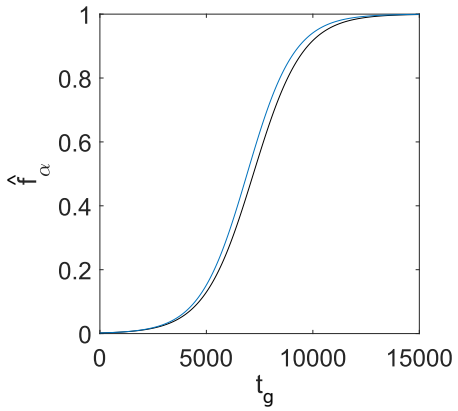

Figure 3.11: Invasion dynamics for a mutant with fertilization rate $\alpha + \delta\alpha$. Blue - analytical prediction using Eq. (3.33), black - numerical simulation. The initial conditions and parameters are the same as in Figure 3.10 except $\delta\alpha = 1/200$, $\delta m = 0$ and $G = 1.5 \times 10^4$.

3.8 APPENDIX IV: EVOLUTIONARY DYNAMICS: FIXED ENVIRONMENT

3.8.1 DERIVATION OF EVOLUTIONARY ODES

We begin by noting that the functional form of $\mathrm{d}\hat{f}_m/\mathrm{d}t_g$ (see Eq. (3.32)) and $\mathrm{d}\hat{f}_\alpha/\mathrm{d}t_g$ (see Eq. (3.33)) is $\hat{f}_i(1-\hat{f}_i)$, which implies a situation of trait substitution [150];

mutants are either driven to fixation or extinction, and polymorphic equilibria are not possible. This simplifies the subsequent analysis considerably.

Taking a classic adaptive dynamics approach [150] and define the invasion fitness of the mutants as their per-capita rate of reproduction upon arising in the population (i.e. when $\hat{f}_m \approx 0$ and $\hat{f}_\alpha \approx 0$). Under the standard assumptions of adaptive dynamics (i.e. that mutations are of small effect, $1 \gg \delta m, \delta\alpha$, and occur sufficiently rarely that each mutation can fixate before a new mutation occurs), the evolutionary dynamics are given by

$$\begin{aligned}
\frac{\mathrm{d}m}{\mathrm{d}\tau} &= \frac{\mathrm{d}}{\mathrm{d}\hat{f}_m}\frac{\mathrm{d}\hat{f}_m}{\mathrm{d}t_g}\bigg|_{\hat{f}_m=0} = h_m(m,\alpha,\beta,C_z,C_p) \\
&= H_m(m,\alpha,\beta,C_z,C_p) \\
\frac{\mathrm{d}\alpha}{\mathrm{d}\tau} &= \frac{\mathrm{d}}{\mathrm{d}\hat{f}_\alpha}\frac{\mathrm{d}\hat{f}_\alpha}{\mathrm{d}t_g}\bigg|_{\hat{f}_\alpha=0} = h_\alpha(m,\alpha,\beta,C_z,C_p) \\
&= H_\alpha(m,\alpha,\beta,C_z,C_p)\,.
\end{aligned} \tag{3.34}$$

Substituting for $\mathrm{d}\hat{f}_m/\mathrm{d}t_g$ from Eq. (3.32), and $\mathrm{d}\hat{f}_\alpha/\mathrm{d}t_g$ from Eq. (3.33), we obtain Eq. (3.5) in the main text.

3.8.2 Analysis of ODEs

In this section we aim to analytically characterise the long term evolutionary behaviour of the population in a fixed environment, with dynamics given by Eq. (3.5), as illustrated in Figure 3.2.

We begin by calculating the evolutionary behaviour of m when the fertilization rate is fixed to zero ($\alpha = 0$). Solving $H_m(m,0,\beta,C_z,C_p) = 0$, for $m = m^*_{\alpha=0}$ we obtain

$$\begin{aligned}
\delta m\frac{(\beta-m)}{m^2} &= 0 \\
&\implies m^*_{\alpha=0} = \beta\,.
\end{aligned} \tag{3.35}$$

We now turn to the evolutionary behaviour of m and α. For $(m^*_{\alpha=0}, 0)$ identified above to remain stable if evolution on α is allowed requires that $[\mathrm{d}\alpha/\mathrm{d}\tau]|_{(m=\beta,\alpha=0)} < 0$ (i.e.

that evolution selects against increases in α). Stating this condition in full, we have

$$\begin{aligned} 0 > \delta\alpha \left[\frac{m\left((1-C_z)e^{\frac{\beta}{2m}} - (1-C_p)\right)\ln\left(\frac{\alpha AMT}{m}+1\right)}{2\alpha\left(\alpha A(1-C_z)MTe^{\frac{\beta}{2m}} + m(1-C_p)\right)} \right]\Bigg|_{(m=\beta,\alpha=0)} \\ \implies 0 > (1-C_z)e^{\frac{1}{2}} - (1-C_p) \\ \implies C_z > 1 - (1-C_p)e^{-\frac{1}{2}}\,. \end{aligned} \tag{3.36}$$

If fertilization costs exceed this this value, evolutionary trajectories starting at $(m^*_{\alpha=0}, 0)$ will remain there (see Figure 3.2, panel (b)). Conversely if fertilization costs do not exceed this value, evolutionary trajectories starting at $(m^*_{\alpha=0}, 0)$ will initially experience a selective pressure for increasing α, due to the fact $[\mathrm{d}\alpha/\mathrm{d}\tau]|_{m=\beta} > 0$ under these conditions (see Figure 3.2, panel (a)); the dynamics are then pulled to another attractor, which we characterise below.

When $C_z > 1 - (1-C_p)\exp(-1/2)$ (but less than another critical value, yet to be determined), only a subset of initial conditions fall within the basin of attraction of this attractor described above (see Figure 3.2, panel (b)), with remaining initial conditions leading to an evolutionarily state at which $\alpha^* \to \infty$. Conversely, when $C_z < 1 - (1-C_p)\exp(-1/2)$, *all* initial conditions lead to this attractor at which $\alpha^* \to \infty$ (see Figure 3.2, panel (a)). We now calculate the mass to which the population evolves at this second attractor. Taking the limit $\alpha \to \infty$ in the evolutionary dynamics for m (see Eq. (3.5)) and solving for zero;

$$\begin{aligned} 0 &= H_m(m,\alpha,\beta,C_z,C_p)|_{\alpha\to\infty} \\ &= \delta m\,\frac{(\beta-4m)}{4m^2} \\ \implies m^* &= \frac{\beta}{4}\,. \end{aligned} \tag{3.37}$$

Therefore the second early evolutionary attractor is at $(m,\alpha) \to (\beta/4, \infty)$.

For even greater costs of fertilization, C_z, our mathematical analysis (which assumes monomorphic resident populations and no evolutionary branching) suggests that $(m,\alpha) \to (m^*_{\alpha=0}, 0)$ becomes the only attractor, with $(m,\alpha) \to (\beta/4, \infty)$ ceasing to be an attractor. To determine the critical cost at which this occurs, we take $[\mathrm{d}\alpha/\mathrm{d}\tau]_{m=\beta/4}$ and calculate the conditions under which this is negative when α is large (i.e. when $(m,\alpha) \to (\beta/4, \infty)$ is no longer attracting, but repelling). Expanding

$[\mathrm{d}\alpha/\mathrm{d}\tau]_{m=\beta/4}$ in small $1/\alpha$, we find that to leading order we must have;

$$\begin{aligned} 0 &> \delta\alpha \left[\frac{m\left((1-C_z)e^2-(1-C_p)\right)\ln\left(\frac{\alpha AMT}{\beta}\right)}{8\alpha^2 A(1-C_z)MTe^2} \right] \\ \implies 0 &> (1-C_z)e^2-(1-C_p) \\ \implies C_z &> 1-(1-C_p)e^{-2} \\ \implies C_z &\gtrapprox 0.86\,, \qquad \text{for } C_p = 0\,. \end{aligned} \tag{3.38}$$

However, while this accurately captures the short-term evolutionary dynamics, we see evolutionary branching (which our model Eq. (3.5) does not account for) should trajectories approach the $m \approx \beta/4$ manifold. Unless costs are exceedingly high ($C_z \approx 1$), this eventually leads to anisogamy followed by oogamy (see Figure 3.12). We discuss this more in Section 3.9.

3.8.3 Implementation of simulations

Here we detail the process of numerical simulation (see *Code availability statement*). We employ a multigenotype model whereby a mutation occurs at rate μ. This rate is the inverse of the expected number of fertilization processes (i.e. generations) until the next mutation event. Before each successive mutation event, the number of fertilization processes until the next mutation event is determined by generating a number from a $Geo(\mu)$ distribution and taking the inverse of that number. In an S genotype model, if a given genotype i has a trait value of (m_i, α_i) and has frequency f_i, then the mean population trait value is given by

$$\langle m \rangle = \sum_{i=1}^{S} m_i f_i \quad , \quad \langle \alpha \rangle = \sum_{i=1}^{S} \alpha_i f_i\,. \tag{3.39}$$

The simulations in Figure 3.2 are repeated for 5500 mutation events. Below, we detail how we simulate the dynamics on each timescale.

Fertilization Kinetics

We construct our model for fertilization kinetics assuming that each genotype, characterised by their unique (m_i, α_i), can fertilize with one another. The input parameters of this function are the trait values of each genotype $\boldsymbol{m}$ and $\boldsymbol{\alpha}$, their frequency in the preceding adult generation $\boldsymbol{f}$, and the parameters A, M and T.

The fertilization kinetics simulations are then run according to

$$\frac{\mathrm{d}N_i}{\mathrm{d}t} = -\frac{1}{2}\sum_{j=1}^{S}(\alpha_i+\alpha_j)N_iN_j \tag{3.40}$$

$$\frac{\mathrm{d}F_{ij}}{\mathrm{d}t} = \frac{1}{2}(\alpha_i+\alpha_j)N_iN_j \tag{3.41}$$

$$\frac{\mathrm{d}F_{ii}}{\mathrm{d}t} = \frac{1}{2}\alpha_iN_i^2$$

where F_{ij} is the number of fertilized cells formed from genotypes i and j, and F_{ii} is the number of fertilized cells formed from two cells of genotype i for any $i,j \in [1,S]$. The initial conditions are

$$N_i(0) = \frac{AMf_i}{m_i} \tag{3.42}$$

$$F_{ij}(0) = 0\,. \tag{3.43}$$

The fertilization process is run for a fixed time period T. At the end of this time period, the function outputs the number of cells of each type. These include the number of unfertilized cells with each trait pair, $N_i(T)$, and fertilized cells of each type $F_{ij}(T)$. The fertilization kinetics are simulated by solving Eqs. (3.40-3.41) using the ODE solver *ode45* in MATLAB.

Single Generation Dynamics

We simulate the frequency of each genotype after the end of each generation taking into account the survival probability of each genotype. Upon maturation, only a fraction of unfertilized and fertilized cells survive to adulthood. We use the outputs of the fertilization kinetic function along with the Vance survival functions Eq. (3.2) to calculate the probability that each progeny survives into adulthood. The single generation dynamics are run according to

$$f_i' = \frac{w_i}{\sum_{j=1}^{S} w_j} \tag{3.44}$$

$$\begin{aligned} w_i &= (1-C_p)N_i(T)S(\beta,m_i) \\ &+ (1-C_z)\Bigg(\sum_{j=1,j\neq i}^{S} F_{ij}(T)S(\beta,m_i+m_j) \\ &+ 2F_{ii}(T)S(\beta,2m_i)\Bigg) \end{aligned} \tag{3.45}$$

where f_i' is the frequency of genotype i in the subsequent generation and w_i is its absolute fitness. The input parameters of this function are $N_i(T)$ and $F_{ij}(T)$, $\boldsymbol{m}$,$\boldsymbol{\alpha}$,

β, C_z and C_p whilst the output is $\boldsymbol{f}$, the frequency of all genotypes at the end of the generation.

Invasion Dynamics

The invasion dynamics are run for approximately one unit of τ (i.e. until the next randomly chosen mutation event after G generations). As outlined in the opening paragraph of Appendix 3.8.3, G is generated from a $Geo(\mu)$ distribution following each mutation event. The inputs of this function are $\boldsymbol{f}$, $\boldsymbol{m}$, $\boldsymbol{\alpha}$, A, M, T, C_z, C_p, G and β and the output is the frequency of each genotype $\boldsymbol{f}$ after one unit of τ (i.e after G fertilization processes). Please note we have adopted one single notation for the frequency of all genotypes.

Evolutionary Dynamics

The evolutionary dynamics have the input parameters δ, μ, $\boldsymbol{f}$, Nmut, f_0, A, M, T, C_z and C_p where δ is the mutational stepsize and Nmut is the number of mutations that occur in our simulation. The initial frequency of a newly introduced mutant f_0 is chosen to be small (equal to 0.002 in our model). We initialise the simulation with two genotypes, where each genotype is characterised by a unique pair of trait values e.g. (m_i, α_i) for genotype i. A mutant is introduced into the population after a random number of fertilization processes G (generated from a $Geo(\mu)$ distribution) at frequency f_0. The mutation is chosen to occur in either α or m with equal probability 1/2. The mutation also acts to increase/decrease the trait value each with probability 1/2. Upon introduction of this mutant, the invasion dynamics of the population is run for G fertilization processes. In the meantime, the mean mass and fertilization rate of the population is recorded using Eq. (3.39). Next, we repeat the process of introducing a new mutant into the population. Since the population now has more than two genotypes, the genotype that mutates is chosen with probability weighted by the frequency of each genotype. There is now the possibility of back mutation to one of the existing genotypes. In this case, if an existing genotype k with frequency f_k mutates to another existing genotype l that has frequency f_l, then following mutation, the frequency of genotype k becomes $f_k - f_0$ and the frequency of genotype l becomes $f_l + f_0$. Furthermore, a genotype is thought to be extinct if its frequency falls below 10^{-3}, in which case we remove that genotype.

3.8.4 EVOLUTIONARY BRANCHING IN GAMETE MASS AND FERTILIZATION RATE

In this section we present additional numerical results investigating the evolutionary branching that occurs in simulations on the manifold $m \approx \beta/4$ (along which $\mathrm{d}\alpha/\mathrm{d}t \approx 0$).

In a fixed environment, when C_z is small relative to C_p, branching in fertilization rate still occurs but no longer gives rise to oogamy

When the cost to fertilization, C_z is small compared to C_p, we find that although branching in both gamete mass and fertilization rate still occurs, the branching in α no longer acts to decrease the fertilization rate of macrogametes. Therefore, we observe pseudooogamy but not oogamy. This behaviour is illustrated in Figures 3.5 and 3.14. We can estimate the critical value of C_z at which this occurs by comparing the survival probability of a macrogamete that does not fertilize with a microgamete, which has cost C_p with that of a macrogamete that does fertilize with a microgamete at cost C_z; if this first probability exceeds the second, there should be no evolutionary pressure for oogamy to evolve. We find

$$(1-C_p)\exp\left(-\frac{\beta}{m_{\mathrm{macro}}}\right) > (1-C_z)\exp\left(-\frac{\beta}{m_{\mathrm{macro}}+m_{\mathrm{micro}}}\right) \tag{3.46}$$

or, noting that m_{macro} approaches $\beta - \delta m$ and m_{micro} approaches δm

$$(1-C_p)\exp\left(-\frac{\beta}{\beta-\delta m}\right) > (1-C_z)\exp\left(-1\right) \tag{3.47}$$

$$\implies 1-(1-C_p)\exp\left(-\frac{\delta \mathrm{m}}{\beta-\delta m}\right) > C_z \geq 0\,. \tag{3.48}$$

In a fixed environment, when C_z is very large, branching in mass and fertilization rates can still occur, with oogamy possible (dependent on initial conditions)

Our analysis of the dynamics of Eq. (3.5) in Appendix 3.8.2 suggested that zero fertilization rates is the only evolutionary attractor when costs to fertilization are high (see Eq. (3.38)). However, our analysis in Section 3.9 along with our simulations reveal that although $\mathrm{d}\alpha/\mathrm{d}t < 0$ along the $m \approx \beta/4$ manifold, any trajectory that approaches this manifold can experience branching in gamete mass, unless $C_z = 1$. Once branching in gamete mass occurs, the smaller gametes (e.g. microgametes) again experience a strong selective pressure to increase their fertilization rate, despite the high costs imposed by fertilization; sexual conflict drives the population towards

obligate oogamy imposed by motile microgametes. Thus as for $C_z < 1$, there is always a subset of initial conditions that lead to trajectories along the $m \approx \beta/4$ manifold (most obviously initial conditions on the manifold itself) and thus oogamy remains one of the two evolutionary outcomes, albeit requiring an increasingly small and biologically unrealistic set of initial conditions (see Figure 3.12 below).

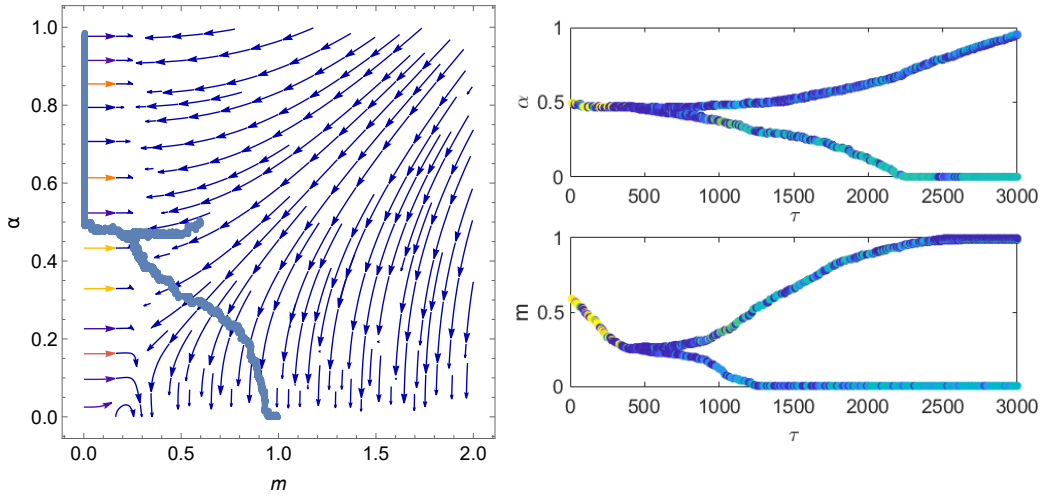

Figure 3.12: Numerical illustration of evolutionary branching for the case where $C_z = 0.9$ and $C_p = 0$. All other parameters the same as Figure 3.2, except $(m(0), \alpha(0)) = (0.6, 0.5)$ and run for $3000/\mu$ generations.

3.9 APPENDIX V: $\alpha_{\max}$ BELOW WHICH BRANCHING TO ANISOGAMY CAN BE ARRESTED

The two necessary conditions for evolutionary branching to occur is for the fitness gradient of an invading mutant to be zero (i.e. $dm/d\tau$ in Eq. (3.34) to equal zero if the mutation occurs in mass, this is also known as an evolutionary singularity) and the second derivative of Eq. (3.29) with respect to the mutational stepsize δm to be positive [46, 64] when $\delta m = 0$. Using Eq. (3.34), we see that the vertical manifold $m \approx \beta/4$ is an approximate evolutionary singularity since $[dm/d\tau]|_{m=\beta/4} \approx 0$. The condition for selection in mass to be disruptive about an evolutionary singularity is

$$\frac{d}{d\hat{f}}\left[\frac{\mathrm{d}^2 \hat{f}'}{\mathrm{d}\,\delta m^2}\right]_{\delta m=0}\Bigg|_{\hat{f}=0} > 0\,, \tag{3.49}$$

where $\hat{f}'$ is the frequency of the mutant in the subsequent generation, given by Eq. (3.29). The explicit expression for the left hand side of Eq. (3.49) is given by

$$\frac{d}{d\hat{f}}\left[\frac{\mathrm{d}^2\hat{f}'}{\mathrm{d}\,\delta m^2}\right]_{\delta m=0}\Bigg|_{\hat{f}=0} = U/V\,, \tag{3.50}$$

where

$$\begin{aligned} U &= A(-1+C_z)e^{\frac{\beta}{2m}}MT\alpha\left(32m^2-12m\beta+\beta^2\right) \\ &\quad +16(-1+C_p)m\left(2m^2-4m\beta+\beta^2\right) \\ V &= 16m^4\left((-1+C_p)m+A(-1+C_z)e^{\frac{\beta}{2m}}MT\alpha\right)\,. \end{aligned}$$

The first condition for branching to occur in mass i.e $dm/d\tau = 0$ can be calculated straightforwardly by setting Eq. (3.34) to zero and solving for α, which gives

$$\alpha = \frac{4m(C_p-1)(m-\beta)e^{-\frac{\beta}{2m}}}{AMT(C_z-1)(4m-\beta)}\,. \tag{3.51}$$

We then substitute Eq. (3.51) into Eq. (3.50) to obtain

$$\frac{\partial}{\partial\hat{f}}\left[\frac{\mathrm{d}^2\hat{f}'}{\mathrm{d}\,\delta m^2}\right]_{\delta m=0}\Bigg|_{\hat{f}=0} = \frac{(3\beta-7m)(\beta-4m)\beta}{12\beta m^4}\,, \tag{3.52}$$

which needs to be positive for selection to be disruptive. Eq. (3.52) equals 0 if $m = 3\beta/7$ or $m = \beta/4$ and is positive for $\beta/4 < m < 3\beta/7$. By substituting the boundaries of this interval into Eq. (3.51), we find that it corresponds to

$$\frac{48(1-C_p)\beta}{35AMT(1-C_z)e^{\frac{7}{6}}} < \alpha_{\max} < \infty\,, \tag{3.53}$$

and thus the value of $\alpha_{\max}$ above which branching will occur in mass is the value given on the left hand side of the inequality Eq. (3.53). In other words, Eq. (3.53) is the interval in $\alpha_{\max}$ in which branching is expected to occur. The vertical line in Figure 3.4 corresponds to the lower limit of this interval. In Figure 3.13 below, we provide a numerical example showing how isogamy can be stabilized below a sufficiently low $\alpha_{\max}$.

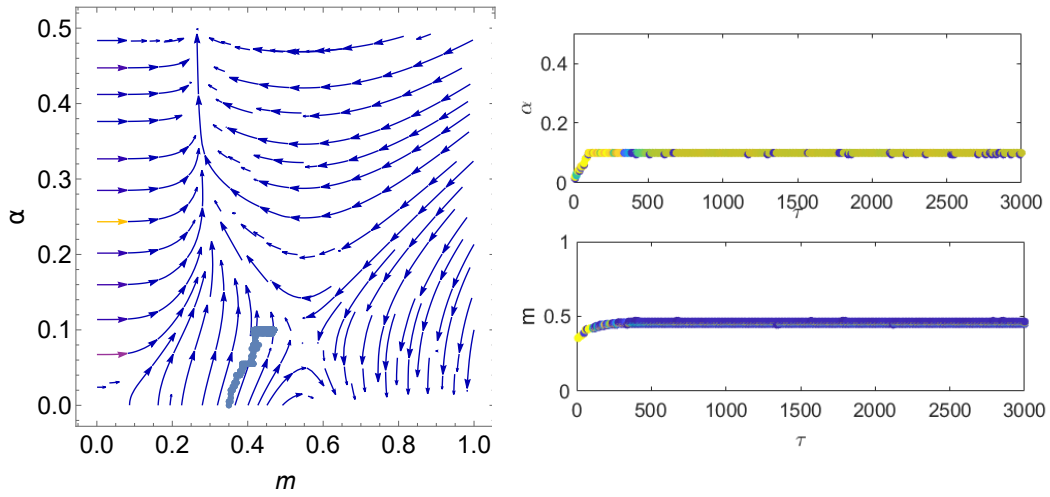

Figure 3.13: Numerical illustration of the stabilization of isogamy below a sufficiently low α_{max}. System parameters are $A = 100$, $M = 1$, $T = 0.1$, $C_z = 0.6$, $C_p = 0$, $\beta = 1$, $\alpha_{max} = 0.1$ and simulation parameters are $\delta = 5 \times 10^{-3}$, $f_0 = 2 \times 10^{-3}$ and run for 6×10^6 generations. Using Eq. (3.53) we can calculate that branching would occur if $\alpha_{max} \gtrapprox 0.1068$ for these parameter values.

3.10 APPENDIX VI: THE STABILIZATION OF ANISOGAMY UNDER HIGH COSTS OF PARTHENOGENESIS RELATIVE TO FERTILIZATION

In Figure 3.14, we provide an example of a parameter regime where pseudooogamy occurs. When the parameters are close to the boundary where we observe the transition between oogamy and anisogamy (i.e close to where the inequality in Eq. (3.48) becomes an equality), we observe pseudooogamy, where there is a considerably stronger selection pressure for microgametes to increase their fertilization rate than macrogametes. Under pseudooogamy, macrogametes are still motile but are considerably less motile than microgametes.

3.11 APPENDIX VII: EVOLUTIONARY DYNAMICS: SWITCHING ENVIRONMENTS WITH BET-HEDGING

We first tackle the derivation of the approximate dynamics in the FRTI (fast relative to invasion) switching regime in Appendix 3.11.1, before verifying the qualitative

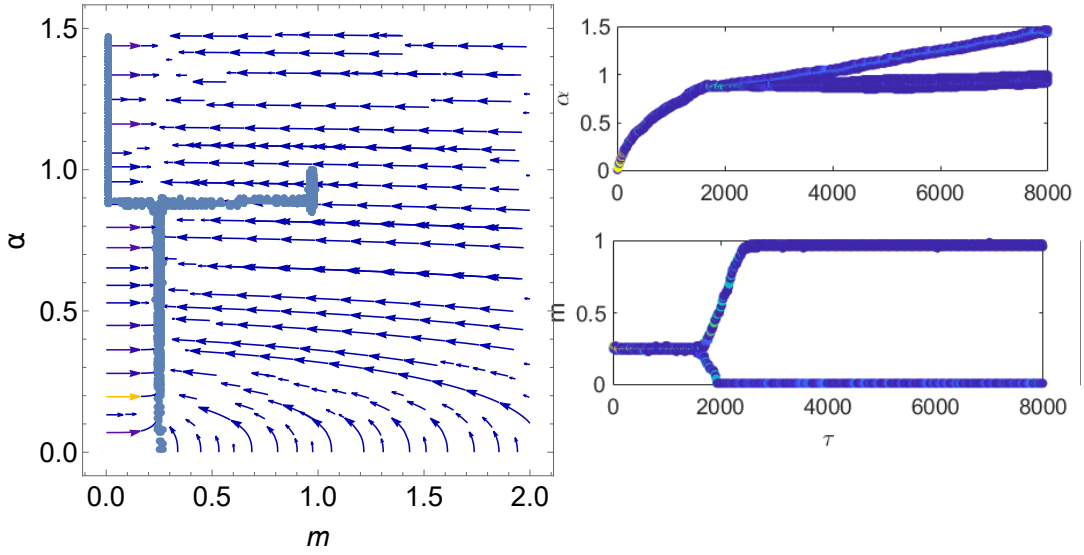

Figure 3.14: Numerical illustration of the evolution of pseudo-oogamy in fixed environment. Parameters are $C_z = 0$, $C_p = 0$, $\delta m = 0.005$, $(m(0), \alpha(0)) = (0.25, 0)$ and simulation run for 1.6×10^7 generations. Remaining parameters are as given in Appendix 3.12.

robustness of these results in the FRTE switching regime in Appendix 3.11.1.

3.11.1 DERIVATION OF EVOLUTIONARY ODES: FRTI

$$\begin{aligned}\frac{\mathrm{d}\hat{f}_m}{\mathrm{d}t_g} &\approx [P_1 h_m(m,\alpha,\beta_1,C_z,C_p) \\ &\qquad + P_2 h_m(m,\alpha,\beta_2,C_z,C_p)]\,\hat{f}_m(1-\hat{f}_m) \\ \frac{\mathrm{d}\hat{f}_\alpha}{\mathrm{d}t_g} &\approx [P_1 h_\alpha(m,\alpha,\beta_1,C_z,C_p) \\ &\qquad + P_2 h_\alpha(m,\alpha,\beta_2,C_z,C_p)]\,\hat{f}_\alpha(1-\hat{f}_\alpha)\,,\end{aligned} \tag{3.54}$$

where $h_m(m,\alpha,\beta,C_z,C_p)$ and $h_\alpha(m,\alpha,\beta,C_z,C_p)$ are given in Eq. (3.32) and Eq. (3.33) respectively. In Figure 3.15 we show that this indeed is a good approximation of the dynamics when δm and $\delta\alpha$ are small and when the residency times in each environment are small relative to the invasion time. Note that as in the case of the fixed environment, the functional form of f_i in these equations implies that trait substitution occurs for independent mutations on the gamete mass and fertilization rate.

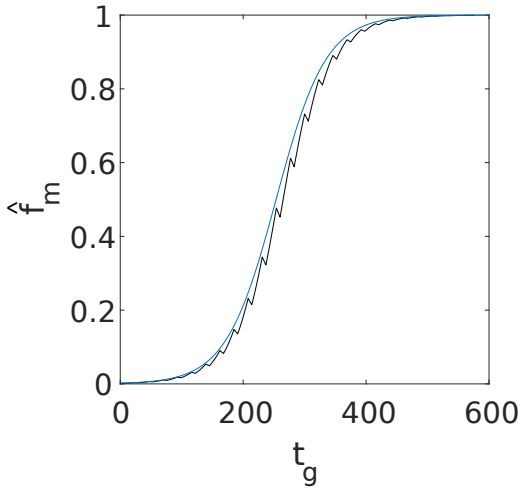

Figure 3.15: Invasion dynamics for a population undergoing bet-hedging when the environment switches FRTI. Blue curve is the analytical approximation using Eq. (3.54) and jagged curve is the numerical simulation. Mutation occurred in mass with $\delta = 0.005$, $f_0 = 0.002$, $(m(0), \alpha(0)) = (0.3, 0.1)$, $\lambda_{1\to 2} = 13/222$, $\lambda_{2\to 1} = 1/6$. All other parameters the same as Figure 3.6.

We now apply the same approach as in Appendix 3.8.1 (see Eq. (3.34)) to obtain

$$\begin{aligned}
\frac{\mathrm{d}m}{\mathrm{d}\tau} &= [P_1 h_m(m,\alpha,\beta_1,C_z,C_p) \\
&\qquad + P_2 h_m(m,\alpha,\beta_2,C_z,C_p)]\,\frac{\mathrm{d}}{\mathrm{d}\hat{f}_m}\left(\hat{f}_m(1-\hat{f}_m)\right)\Big|_{\hat{f}_m=0} = \\
&= P_1 H_m(m,\alpha,\beta_1,C_z,C_p) + P_2 H_m(m,\alpha,\beta_2,C_z,C_p) \\
\frac{\mathrm{d}\alpha}{\mathrm{d}\tau} &= [P_1 h_\alpha(m,\alpha,\beta_1,C_z,C_p) \\
&\qquad + P_2 h_\alpha(m,\alpha,\beta_2,C_z,C_p)]\,\frac{\mathrm{d}}{\mathrm{d}\hat{f}_\alpha}\left(\hat{f}_\alpha(1-\hat{f}_\alpha)\right)\Big|_{\hat{f}_\alpha=0} = \\
&= P_1 H_\alpha(m,\alpha,\beta_1,C_z,C_p) + P_2 H_\alpha(m,\alpha,\beta_2,C_z,C_p)\,.
\end{aligned} \tag{3.55}$$

We see in Figure 3.6 that these also provide a good approximation of the evolutionary dynamics.

3.11.2 Derivation of evolutionary ODEs: FRTE

In the FRTI scenario in the previous section, we supposed that switching between the environments was happening sufficiently regularly relative to the timescale of

invasion that the effective invasion dynamics could be described by a weighted mean of the invasion dynamics in both environments (see Eq. (3.54)). Applying an analogous logic, we now assume in the FRTE scenario that switching between the environments occurs regularly relative to the timescale of evolution (the arrival rate of new mutations) that the effective evolutionary dynamics can be described by a weighted mean of the evolutionary dynamics in both environments; that is

$$\begin{aligned} \frac{\mathrm{d}m}{\mathrm{d}\tau} &= P_1 H_m(m,\alpha,\beta_1,C_z,C_p) + P_2 H_m(m,\alpha,\beta_2,C_z,C_p) \\ \frac{\mathrm{d}\alpha}{\mathrm{d}\tau} &= P_1 H_\alpha(m,\alpha,\beta_1,C_z,C_p) + P_2 H_\alpha(m,\alpha,\beta_2,C_z,C_p)\,. \end{aligned} \tag{3.56}$$

We note that these are in fact exactly the same evolutionary dynamics as derived in the FRTI scenario (see Eq. (3.55)). In Figure 3.6 we show that these do indeed remain a good approximation for the dynamics in the FRTE scenario. The difference between the dynamics in both regimes is quantitative, rather than qualitative. In the very-fast switching FRTI regime, the populations follow the effective dynamics very closely. In the comparatively slower FRTE regime, although the populations no longer follow the dynamics as well, the qualitative picture of the dynamics is still captured by Eq. (3.56). In particular, we still observe an evolutionarily stable state of intermediate fertilization rate (see Figure 3.6).

3.11.3 Analysis of ODEs

We begin, as in Appendix 3.8.2, by considering the evolutionary behaviour of m when the fertilization rate is fixed to zero ($\alpha = 0$). We recall that if the population was fixed in environment 1 or 2 with $\alpha = 0$, the mass of gametes would evolve to $m = \beta_1$ and $m = \beta_2$ respectively. In the switching environment we instead find the bet-hedging strategy

$$m^*_{\alpha=0} = P_1\beta_1 + (1-P_1)\beta_2\,. \tag{3.57}$$

Meanwhile the region of the boundary $\alpha = 0$ over which reduced fertilization rates are selected for is given by $[\mathrm{d}\alpha/\mathrm{d}\tau]|_{\alpha=0} < 0$, or

$$\begin{aligned} 0 &> (1-C_z)\left[P_1 e^{\frac{\beta_1}{2m}} + (1-P_1)e^{\frac{\beta_2}{2m}}\right] - (1-C_p) \\ \implies\; C_z &> (1-C_p) - \frac{1}{P_1 e^{\frac{\beta_1}{2m}} + (1-P_1)e^{\frac{\beta_2}{2m}}}\,. \end{aligned} \tag{3.58}$$

If this condition holds for $m = m^*_{\alpha=0}$, then the switching-induced fixed point $(m,\alpha) = (m^*_{\alpha=0}, 0)$ is stable. We next turn to the high fertilization rate attractor, for which

$\alpha \to \infty$. In a similar manner to Eq. (3.37) (albeit with $H_m(m, \alpha, \beta, C_z, C_p)$ replaced with $[P_1 H_m(m, \alpha, \beta_1, C_z, C_p) + P_2 H_m(m, \alpha, \beta_2, C_z, C_p)]$), we find the bet-hedging strategy

$$m^* = \frac{1}{4}\left[P_1\beta_1 + (1-P_1)\beta_2\right] . \tag{3.59}$$

However, unlike in the fixed environment case, $(m, \alpha) = (m^*, \infty)$ is not always an evolutionary attractor.

In the switching environment, a third evolutionary attractor can emerge, brought about by a balance between selection for high fertilization rates in one environment and zero fertilization rates in the other. Solving $P_1 H_\alpha(m, \alpha, \beta_1, C_z, C_p) + P_2 H_\alpha(m, \alpha, \beta_2, C_z, C_p) = 0$ for α we obtain

$$\begin{aligned}\alpha^* = &\frac{m^*(1-C_p)}{AMT(1-C_z)} \\ &\times \left(\frac{(1-C_z)\left(P_1 e^{\frac{\beta_1}{2m^*}} + (1-P_1)e^{\frac{\beta_2}{2m^*}}\right) - (1-C_p)}{(1-C_p)(P_1 e^{\frac{\beta_2}{2m^*}} + (1-P_1)e^{\frac{\beta_1}{2m^*}}) - (1-C_z)e^{\frac{\beta_1+\beta_2}{2m^*}}} \right) .\end{aligned} \tag{3.60}$$

A good approximation for m^* in this equation can be deduced by noting that the fixed point sits on a vertical manifold of trajectories along which m is approximately held constant as $\alpha \to \infty$; thus we substitute $\approx m^*$ from Eq. (3.59) to obtain the approximate expression for the attractor (m^*, α^*) given in Eq. (3.14). In Appendix 3.11.5, we verify from simulations that evolutionary branching does not occur at this switching-induced fixed point, and that the population is instead held in a state of isogamy.

3.11.4 Implementation of simulations

In the bet-hedging scenario, we simulate the evolutionary dynamics using the same principles as in the fixed environment scenario (see Appendix 3.8.3), however we now implement a Gillespie algorithm [68] to allow us to incorporate environmental switching events as well as mutations, both of which occur stochastically. In particular, we introduce environmental switching events and mutations randomly with geometrically distributed waiting times, measured in units of number of fertilization processes t_g. To simulate the FRTI regime, we set the environmental switching rates to larger values than the mutation rate i.e. $\lambda_{2\to1}, \lambda_{1\to2} >> \mu$. Likewise, in the FRTE regime, we set $\lambda_{2\to1}$ and $\lambda_{1\to2}$ to smaller values than μ. The mutation rate in the

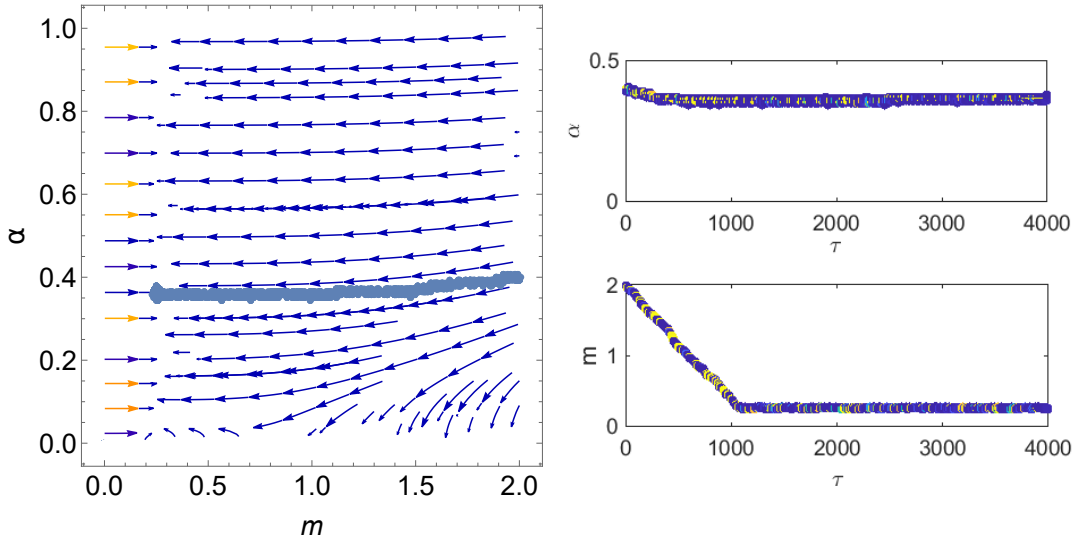

Figure 3.16: Numerical simulation showing an absence of branching for a system undergoing bet-hedging in an environment that switches FRTI. Parameters are same as Figure 3.6 (a) and system run for $4000/\mu$ generations.

numerical simulation of the FRTI regime overlaid in Figure 3.6 (a) is $\mu = 3.5 \times 10^{-4}$ and the switching rates are $\lambda_{2\to1} = 67/532$ and $\lambda_{1\to2} = 1/4$. For Figure 3.6 (b) we have $\mu = 3.5 \times 10^{-4}$ and the switching rates are $\lambda_{2\to1} = 1/6$ and $\lambda_{1\to2} = 13/222$. For the FRTE regime, $\mu = 5 \times 10^{-4}$ and the switching rates are $\lambda_{2\to1} = (67\mu)/532$ and $\lambda_{1\to2} = \mu/4$ in Figure 3.6 (a) and $\lambda_{2\to1} = \mu/6$ and $\lambda_{1\to2} = (13\mu)/222$ in Figure 3.6 (b).

3.11.5 Absence of Evolutionary Branching at Switching-Induced fixed point

In this section we present additional numerical results that confirm that evolutionary branching does not occur at the switching-induced fixed point calculated in Eq. (3.14). We see that under both FRTI and FRTE, the population is held in a state of isogamy. This behaviour is illustrated in Figures 3.16 and 3.17.

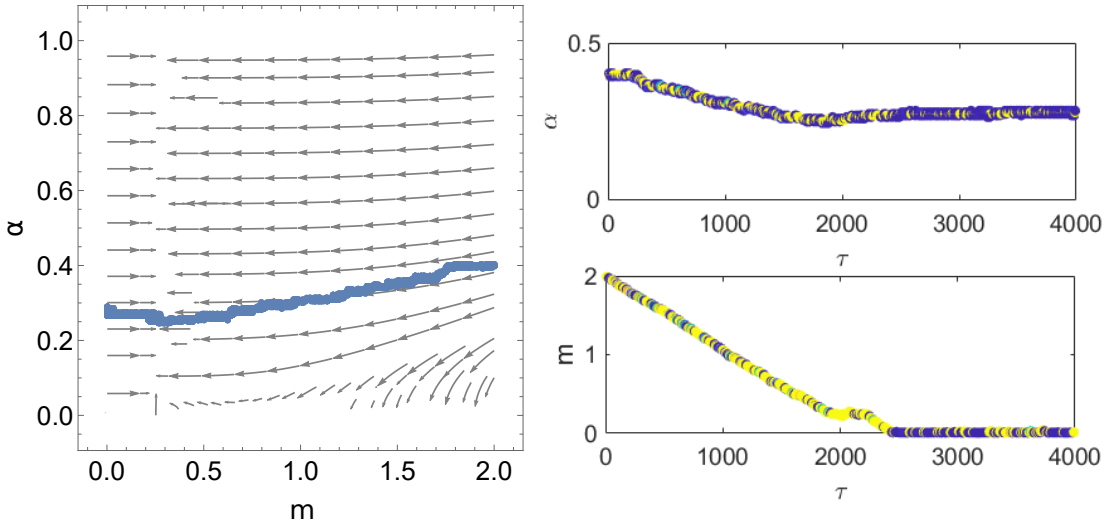

Figure 3.17: Numerical illustration showing an absence of evolutionary branching for a system undergoing bet-hedging in an environment that switches FRTE. Parameters the same as Figure 3.6 (c) and the system is run for $4000/\mu$ generations.

3.12 APPENDIX VIII: MODEL PARAMETERS IN FIGURES

3.12.1 FIGURE 3.2

The initial points of the trajectories are $(m(0), \alpha(0)) = (1.5, 0.6)$ and $(m(0), \alpha(0)) = (2, 0.1)$. Remaining model parameters are $A = 100$, $M = 1$, $T = 1$ and $\beta = 1$. Simulation parameters: Initial frequency of novel mutant genotype $f_0 = 2 \times 10^{-3}$, mutation rate $\mu = 5 \times 10^{-4}$ (number of generations)$^{-1}$, run for 1.1×10^7 generations in panel (a) and 1.24×10^7 generations in panel (b). The mutational stepsize is $\delta = 5 \times 10^{-3}$. Throughout the paper, we use δm and $\delta\alpha$ to denote mutational stepsize in m and α respectively, however in simulations involving co-evolution of m and α, the same stepsize is used for both traits. For convenience, we thus denote mutational stepsize as δ in these simulations.

3.13 APPENDIX IX: BRANCHING TO ANISOGAMY CAN OCCUR UNDER PERIODIC ENVIRONMENTAL SWITCHING

In the case where environmental switching is periodic in the FRTI regime, evolutionary branching can occur about the switching induced fixed point (m^*, α^*) Eq. (3.14), as shown in Figures 3.18 and 3.21. This can be justified as follows. In the case of periodic switching, the invasion dynamics of each genotype is deterministic, hence a gamete with mass greater than the evolutionary stable mass at the switching induced fixed point and one with a smaller mass than that evolutionary stable mass are guaranteed to coexist (see bottom panel of Figure 3.19). This coexistence is what gives rise to branching. Here, the frequencies of both genotypes oscillate between two fixed frequencies under switching environments. However when switching is stochastic, as shown in the top panel of Figure 3.19, a longer than average period spent in a given environment can cause fixation of one of the genotypes, which is likely to prevent coexistence. The increased chance of extinction due to the long period spent in a given environment has also been noted in Berríos-Caro *et al.* [15], which investigated the effect of environmental switching on the evolution of number of mating types.

Furthermore, repeating the simulations of the FRTE regime of Figure 3.6 for periodic switching, we find that branching does not occur (see Figure 3.20).

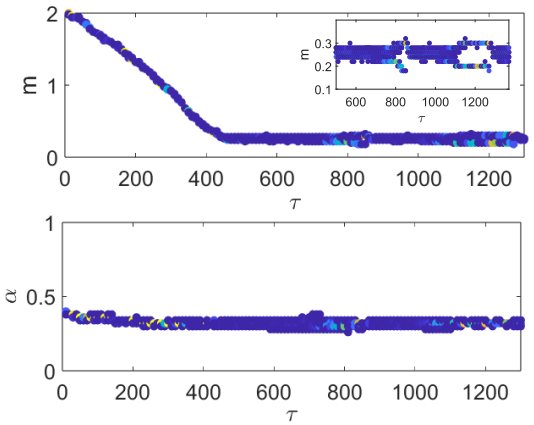

Figure 3.18: Simulation of Figure 3.6 (a) under periodic environmental switching in the FRTI regime. Evolutionary branching can occur to give rise to mild anisogamy. Following branching, the microgamete extincts, giving rise to branching-extinction cycles, as shown in the inset of the top panel. All simulation parameters are identical to the FRTI regime in Figure 3.6 except $\delta = 0.02$ and $\mu = 1/500$. This is to help speed up simulations. To allow periodic switching with a discrete number of generations, $\lambda_{1\to 2} = 1/4$ and $\lambda_{1\to 2} = 1/8$, which is a good approximation of the switching rates in Figure 3.6 (a).

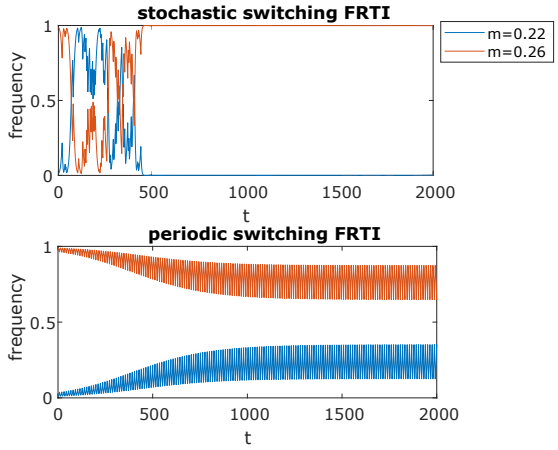

Figure 3.19: Invasion dynamics of a population with one gamete with smaller mass than that at the switching induced fixed point Eq. (3.14), and one with mass larger than at the switching induced fixed point Eq. (3.14), under the FRTI regime with stochastic environmental switching (top) and periodic switching (bottom). Under periodic switching, coexistence between the small and large gamete occurs, while under stochastic switching, one type of gamete can be driven to extinction due to a long period between two switching events. Parameters are identical to Figure 3.6 (a), with $f_0 = 0.01$, The small and large gametes have masses 0.22 and 0.26 respectively.

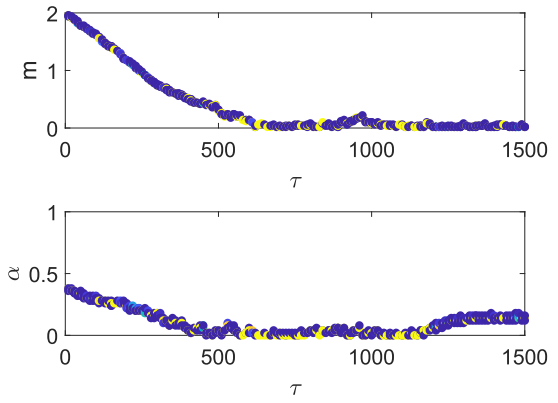

Figure 3.20: Simulation of Figure 3.6 (a) under the FRTE regime. Evolutionary branching does not occur. All simulation paraemeters are identical to the FRTE regime in Figure 3.6 except $\delta = 0.02$ and $\mu = 1/500$ to speed up simulations. Switching rates are $\lambda_{1\to 2} = 1/2000$ and $\lambda_{2\to 1} = 1/4000$ to ensure that switching occurs after a discrete number of generations, as in Figure 3.18.

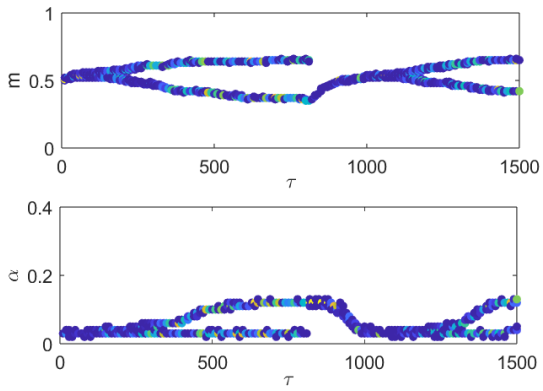

Figure 3.21: Simulation of evolutionary trajectories with periodic switching under the FRTI regime for a different set of parameters as Figure 3.18. Here, we observe branching-extinction cycles about the switching induced fixed point Eq. (3.14). System parameters are, $A = 100$, $M = 1$, $T = 1$ $\beta_1 = 4$, $\beta_2 = 0.1$, $C_z = 0.6$, $C_p = 0$, $\lambda_{1\to 2} = \lambda_{2\to 1} = 1/5$. Simulation parameters are $\delta = 0.01$, $\mu = 1/1000$, $f_0 = 0.002$ and $(m(0), \alpha(0)) = (0.5, 0.03)$ which is located in the vicinity of the switching induced fixed point.

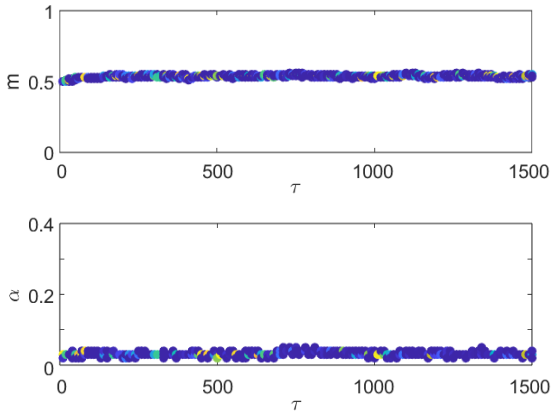

Figure 3.22: Simulation of evolutionary trajectories as in Figure 3.21 but with stochastic switching. All system and simulation parameters are identical to Figure 3.21.

— 4 —

Cell size, switching environments and selection for stress-induced binary cell fusion

Abstract

Binary cell fusion is a crucial process in sexually reproducing organisms and can form the basis for the evolution of various reproductive modes. Microorganisms exhibit a variety of reproductive modes, and these modes vary with changing environments. Adverse environments are typically associated with cell fusion, while benign environments favour cell fission. Here, we develop mathematical models which capture the co-evolution of mass and cell fusion rate to investigate the environmental conditions that select for cell fusion. Cells are shown to evolve to increase their mass in response to environmental stress. However, if the physiological machinery for cell-fusion becomes available, binary cell fusion is selected for instead of increasing mass. Furthermore, we show that in the presence of phenotypic plasticity, where genotypes can produce different phenotypes in response to different environmental conditions, evolution of binary cell fusion may occur as a response to environmental stress. These results highlight a nascent mechanistic advantage to cell fusion, due to the increase in cytoplasmic volume. Since the conditions giving rise to the evolution of facultative sex and facultative multicellularity broadly coincide, both occurring under harsh environmental conditions and involve cell fusion, our results have the potential to shed light on the coincidence between the early evolution of facultative multicellularity and sex.

4.1 INTRODUCTION

Although the details of the early evolution of sexual reproduction in the last common eukaryotic common ancestor (LECA) are shrouded in mystery, it is argued that the process began with the evolution of cell–cell fusion and meiosis [73] in an archaeal ancestor [51, 52]. This step can be further broken down into the evolution of binary cell fusion, the one spindle apparatus, homologous pairing and chiasma, and finally reduction, division and syngamy [147]. The vast majority of theoretical studies investigating the evolution of sexual reproduction have focused on later stages of this evolutionary trajectory, namely the conditions that give rise to a selective pressure for genetic recombination [87, 154, 221]. However, comparatively few studies have investigated the selective pressures that may have first given rise to binary cell fusion, which sets the stage for the evolution of homologous pairing and genetic recombination by bringing nuclei together in the same cell.

Hypotheses for the evolution of binary cell fusion generally rely on hybrid fitness advantage. It has been suggested that selection for cell–cell fusions might have initially been driven by "selfish" transposons and plasmids [93, 94, 191], or negative epistatic interactions between mitochondrial mutations [182, 217]. However, once a heterokaryotic cell has been formed (binucleate with nuclei from both parental cells), the advantage of hybrid vigor and the masking of deleterious mutations could lead to the maintenance of cell fusion [147]. Such benefits are required to alleviate costs to cell-fusion, which include selfish extra-genomic elements in the cytoplasm [86] and cytoplasmic conflict [100, 101].

In this chapter, we avoid the complexity associated with these genetic factors. Instead, we focus on how the survival advantage associated with increasing cytoplasmic volume might select for binary cell fusion; this relies on the physiological advantages conferred by cell-cell fusion and is independent of the question of the genetic advantages (and disadvantages) of sexual reproduction. This alternative perspective offers useful new insights that can be usefully compared with empirical observation.

That size-based processes could play a role in the early evolution of sexual reproduction has theoretical support. The "food hypothesis" [225] suggests that metabolic uptake could drive horizontal gene transfer in bacteria and archaea, with DNA molecules providing nutrients for the receiving cell [186, 185]. Indeed, horizontal gene transfer has been shown experimentally to be an important source of carbon

and nutrients in bacteria [156, 169]. Meanwhile amongst eukaryotes, the benefits of increasing cytoplasmic volume are understood to be strong enough to drive selection for the sexes themselves [128, 173]. However, the role of cytoplasmic volume in the evolution of binary cell fusion has yet to be explored. In suggesting a mechanistic hypothesis for the evolution of binary cell fusion, our work has interesting parallels with [144], where an advantage to cell fusion is identified in terms of shortening the cell-cycle.

The fusion of any pair of genetically distinct cells may incur conflicts of interest [215], which requires cooperation in order to be resolved [163]. This presents another cost to cell fusion. However we show here that the increased mass conferred by cell fusion presents a sheer physiological advantage, which may be sufficient to overcome such costs.

An early theory known as Muller's Ratchet states that sex evolved due to the ability for genetic recombination to prevent the buildup of deleterious mutations [154]. Another hypothesis known as the Red queen hypothesis [221] states that recombination generates genetic diversity which hastens adaptation to fluctuating environments.

4.2 MODEL

4.2.1 Insights from Simulations

We consider a computational model of a haploid population that reproduces via binary cell fission. This population dynamics proceed as follows: there is an initial growth and binary fission phase where a fixed energy budget E can be used for population growth and binary fission (see Figure 4.1 (a)), followed by an environmentally-induced mortality phase where growth and fission are suspended and survival depends on cell size. Explicitly, we assume that cells reach maturity at size M, so that in the absence of binary fission ($n = 0$ rounds of fission) the total number of mature cells at the end of a growth cycle is E/M (see Figure 4.1 (a)). If, however, the population undergoes $n > 0$ rounds of binary fission, then each resulting daughter cell has mass $m = M/2^n$ and the total number of daughter cells is $(2^n E)/M$. After this growth and fission phase each daughter cell is subject to an extrinsic mass-dependent mortality, such that larger daughter cells are more likely to survive into the next growth cycle (see Figure 4.1, panel (a)). We denote this survival function $S(m; \beta)$;

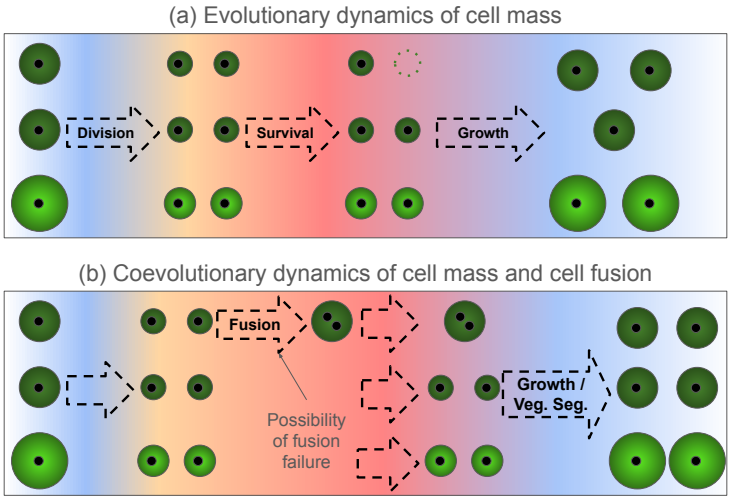

Figure 4.1: Schematic for the model dynamics within each growth cycle. Panel (a): Illustrative dynamics for the evolutionary dynamics of cell mass alone. Due to energetic constraints genotypes in the population can either produce fewer, larger mature cells or more numerous, smaller cells (see different shades of green). Daughter cells are produced following cell division. Their survival is dependent on mass, such that smaller cells are more likely to die (see Eq. (4.2)). Surviving cells seed the next growth cycle. panel (b): Illustrative dynamics for the coevolutionary dynamics of cell mass and cell fusion rate. The model is similar to that in panel (a), but now a fraction of daughter cells are given the opportunity to risk fusing to form binucleated cells; with probability C fusion fails, and both daughter cells are lost. However should a fused cell successfully form, it experiences an enhanced survival probability as a result of its larger cytoplasmic volume. Following growth and vegetative segregation, surviving cells seed the next growth cycle.

the parameter β describes the magnitude of the mortality process (i.e. the harshness of the environment). For a given value of m, an increase in β decreases the survival probability.

In an initial investigation, we view the mass of daughter cells, m, as a single trait subject to evolution. To increase the number of daughter cells they produce (E/m) mature cells can grow to smaller sizes (reduced M) or increase their number of cell-divisions (increased n) . However by increasing n, individuals also produce smaller daughter cells that are more vulnerable to extrinsic mortality. The size of daughter cells is thus subject to a quality-quantity trade-off. For simplicity, we model the mass of daughter cells m as a continuous trait, and explore its evolution using simulations.

Figure 4.2 summarises the outcome of such evolutionary dynamics. Panel (a) shows that the population evolves towards an evolutionarily stable strategy (ESS) in m for a given environment. In the context of adaptive dynamics, an ESS is a strategy where the local fitness gradient is zero and is uninvadable by any nearby mutants [150] (i.e a fixed point in the evolutionary phase plane). Should the environment suddenly become harsher (via an increase in β) the population evolves towards a new ESS, in which daughter cells are larger.

We now modify the model to allow for the possibility of binary cell fusion. Daughter cells may now fuse to form a binucleated cell (e.g. a dikaryon, in which the cytoplasm of the contributing cells are mixed but their nuclei remain distinct [122]) or remain a mononucleated cell. The rate of cell fusion is given by α, such that when $\alpha = 0$ all cells remain mononucleated, and cell survival into the next growth cycle is calculated as before. Conversely for $\alpha > 0$, some proportion of daughter cells will have fused and in the limit $\alpha \to \infty$, all cells will have fused. These fused cells will receive a survival advantage from their increased mass. However they will also pay an additional cost, C, resulting from factors such as cell-fusion failure [80], selfish extra-genomic elements in the cytoplasm [86], cytoplasmic conflict [100, 101] and maintenance of a binucleated cell [227]. Together this means that fused cells survive with a total probability $(1 - C)S(2m; \beta)$. Surviving adults divide to form a new growth cycle of mononucleated haploid daughter cells, with binucleated parental cells producing mononucleated progeny through vegetative segregation [189]. Note that although we do not account for the possibility of binucleated cells failing to form mononucleated progeny (i.e. failed segregation), this can be accounted for by their additional survival cost, C (see Figure 4.1, panel (b)).

We now explore the coevolution of daughter cell mass, m, and fusion rate, α. In Figure 4.2, panel (b), we see that in the benign environment, α remains at zero, and the population evolves towards an ESS in m as in Figure 4.2, panel (a). However now when the population is introduced to a harsher environment, the evolutionary dynamics differ from those in Figure 4.2, panel (b) (where α was held artificially at zero). Rather than cells evolving to be larger, we see a different response emerging; selection for binary cell fusion.

The result above is in some sense surprising. Despite the presence of additional survival costs associated with binary cell-fusion, selection for non-zero fusion rates (rather than increased daughter cell size) persists in the harsh environment. We explain the emergence of this behaviour mathematically in the Results section.

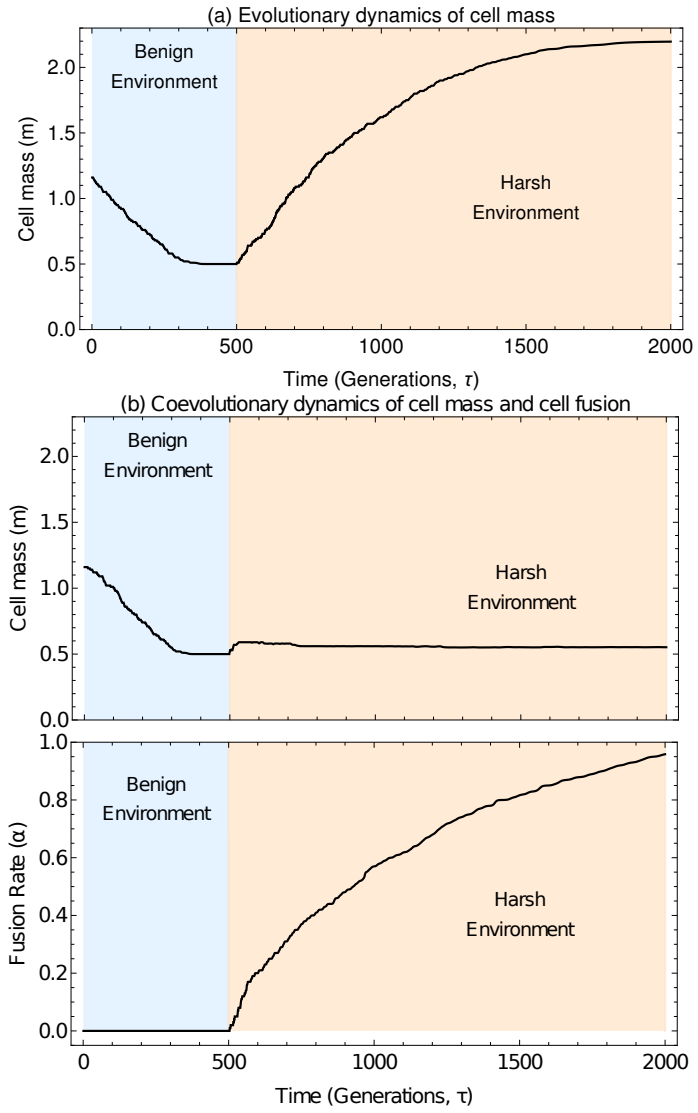

Figure 4.2: Stochastic simulations of evolutionary trajectories when the system is subject to a switch from the benign environment ($\beta_1 = 0.5$, green region) to the harsh environment ($\beta_2 = 2.2$, orange region) at growth cycle 500. Panel (a) for the case where the fusion rate is held at 0, representing the case where the physiological machinery for fusion has not evolved, and panel (b) for the case where fusion rate is subject to evolution. Remaining model and simulation parameters are given in section 4.6 and the initial condition is $(m(0), \alpha(0)) = (1.16, 0)$.

4.2.2 MATHEMATICAL MODEL

Our model takes inspiration from the classic Parker-Baker-Smith (PBS) model for the evolution of anisogamy [173] (the production of sex cells of differing size). However,

whereas such models typically consider the binary cell fusion (fertilization) rate a fixed parameter, we here treat it as a trait subject to evolution. In doing so our work builds on Chapter 3, where a very similar model with a different biological motivation was used to investigate the evolution of anisogamy with parthenogenesis. In order to analyse the dynamics of the model, we use tools from adaptive dynamics [22], assuming that traits are continuous and that mutations have small effect.

In addition we will explore the effect of switching environments, another departure from the PBS model. As such it is important to keep track of the hierarchy of timescales at play. The shortest timescale is the timescale of a growth cycle. The intermediate timescale is that over which the invasion of a rare mutant (taking place over many growth cycles) can occur and the longest timescale is the evolutionary timescale, representing the cumulative effect of multiple mutations and invasions. Finally we assume that environmental switching can take place on either intermediate or long timescales (see section 3.2.2 in Chapter 3).

4.2.3 Dynamics within each growth cycle

A total of $(2^n E)/M$ daughter cells enter a pool in which binary cell fusion can occur. After a finite time window, the resultant cells are subject to a round of mass dependent mortality, such that cells of larger mass are more likely to survive. The surviving cells form the basis of the next growth cycle, completing the growth cycle, as illustrated in Figure 4.1.

Fusion Kinetics

We assume that the attractor to which the population evolves depends on the initial conditions cells may fuse with each other, an assumption consistent with most models of the early evolution of sexual reproduction, which suppose the existence of a "unisexual" early ancestor that mated indiscriminately [89]. Initially, the population is comprised of N unfused daughter cells. Fusion between mononucleated cells occurs at a rate of α, such that the number of unfused cells, N, is given by the solution to

$$\frac{\mathrm{d}N}{\mathrm{d}t} = -\alpha N^2\,, \qquad N(0) = \frac{2^n E}{M}\,. \tag{4.1}$$

At the end of the fusion window of duration T there are then $N(T)$ unfused (mononucleated) cells remaining, and $(N(0) - N(T))/2$ fused (binucleated) cells.

Survival Probability

We assume that both unfused and fused cells are subject to the same extrinsic mass-dependent mortality function, $S(m;\beta)$, while fused cells pay an additional mass-independent cost C. Many choices for such a function are possible, so long as it is an increasing function of cell size (which we equivalently refer to as cell mass m). However here we assume that $S(m;\beta)$ is the Vance survival function [222], a common assumption in the literature [25, 124, 135]. We thus have that at the end of the fusion window, the survival probability of unfused and fused cells are given respectively by

$$\begin{aligned} S(m_i;\beta) &= \exp\left(-\frac{\beta}{m_i}\right), \\ (1-C)S(m_j+m_k;\beta) &= \exp\left(-\frac{\beta}{m_j+m_k}\right), \end{aligned} \tag{4.2}$$

where m_i is the mass of a particular unfused cell and m_j and m_k are the masses of two daughter cells that have fused. For a given cell mass, increasing β will decrease the survival probability. We therefore refer to β as the environmental harshness parameter, with high β corresponding to harsh environments in which survival is difficult, and low β corresponding to more benign environments in which even cells of modest mass have a high probability of surviving.

Having defined how the survival of a cell depends on its mass, we have the necessary tools to mathematically characterise the fitness of a rare mutant, and whether it can invade the resident population. In the following section, we provide mathematical approximations of the invasion dynamics of such a mutant.

4.2.4 Invasion Dynamics

Adopting the classical assumptions of adaptive dynamics as in Chapter 3, we mathematically characterise the invasion dynamics of a mutant (which occur over discrete growth cycles). Denoting by $\hat{f}_m$ the frequency of mutants of size $m \pm \delta m$ in the population where δm is the mutational stepsize in m, which is assumed to be small and t_g the number of growth cycles, we find

$$\frac{\mathrm{d}\hat{f}_m}{\mathrm{d}t_g} = \quad h_m(m,\alpha,\beta,C)\hat{f}_m(1-\hat{f}_m)\,, \tag{4.3}$$

where $h_m(m,\alpha,\beta,C)$ is a constant that depends on the parameters m,α,β,C. This constant provides the fitness gradient of the mutant. Similarly, denoting by $\hat{f}_\alpha$ the

frequency of mutants with fusion rate $\alpha \pm \delta\alpha$ in the population where $\delta\alpha$ is the mutational stepsize in α, which is again assumed to be small, we find

$$\frac{\mathrm{d}\hat{f}_\alpha}{\mathrm{d}t_g} = \; h_\alpha(m, \alpha, \beta, C)\hat{f}_\alpha(1 - \hat{f}_\alpha)\,, \tag{4.4}$$

where $h_\alpha(m, \alpha, \beta, C)$ is the fitness gradient of a mutant with fusion rate $\alpha + \delta\alpha$. We see that in the case of a single mutant, we have frequency-independent selection for mutants with different masses and fusion rates. We assume for the remainder of the mathematical analysis that mutants encounter a monomorphic resident population (trait substitution), which substantially simplifies the subsequent evolutionary analysis in the following section.

4.2.5 Evolutionary Dynamics

We assume that haploid daughter cells are characterised by two genetically non-recombining traits mass m and cell fusion rate α. As in Chapter 3, we assume that mutations occur in m or α independently at a fixed rate μ, where μ is measured in units of (number of growth cycles)$^{-1}$. A mutation in m represents a change in the mass of the daughter cell produced, and a mutation in α represents a change in the fraction of the population that undertakes either one of the reproductive routes (obligate fusion/no fusion).

Mutants with a different mass to their ancestor can produce either more or fewer daughter cells than their ancestor (see Eq. (4.1)), which impacts their survival (see Eq. (4.2)). When mutants have a different fusion rate to their ancestor, although the number of daughter cells produced does not differ from their ancestor, the number of fused cells at the end of a growth cycle can either increase/decrease, which impacts their survival, since fused cells have greater mass. The survival of fused cells is also influenced by the cost of fusion C.

Our mathematical analysis in the remainder of this section is based on the assumptions of adaptive dynamics [22], which assumes that mutations are of small effect, and mutants encounter a strictly monomorphic population (i.e. that mutations fixate before the introduction of a new mutant). However in our numerical simulations, mutants arise in the population stochastically at a finite rate, making it possible that a subsequent mutation occurs before the fixation of the previous mutant. We therefore allow for the coexistence of multiple traits in the population, held under a mutation-selection balance, as described in Section 3.3.5 of Chapter 3.

Fixed Environment

We first consider the evolutionary dynamics in the case where the environment is fixed (i.e. when the parameter β, which measures the harshness of the environment (see Eq. (4.2)), is constant throughout the evolution. Assuming that δm and $\delta\alpha$ are small (small mutational step size), we use techniques from adaptive dynamics [37, 138] to obtain equations for the evolutionary dynamics of m and α. These are obtained from Eqs. (4.3-4.4) and are given by

$$\begin{aligned}\frac{\mathrm{d}m}{\mathrm{d}\tau} &= H_m(m,\alpha;\beta,C)\\ &= -\frac{4m(m-\beta)+E\alpha T(1-C)e^{\frac{\beta}{2m}}(4m-\beta)}{4m^2(m+E\alpha T(1-C)e^{\frac{\beta}{2m}})}\\ \frac{\mathrm{d}\alpha}{\mathrm{d}\tau} &= H_\alpha(m,\alpha;\beta,C)\\ &= -m\frac{(1-(1-C)e^{\frac{\beta}{2m}})\ln(1+\frac{E\alpha T}{m})}{2\alpha(E\alpha T(1-C)e^{\frac{\beta}{2m}}+m)}.\end{aligned} \tag{4.5}$$

Switching environments with phenotypic plasticity

Now, we consider the case where evolution acts on the same traits as before, but the environment is subject to change. We model environmental change as switching between two environments β_1 and β_2. If $\beta_1 > \beta_2$, then β_1 is the harsher environment (see Eq. (4.2)). We also allow for phenotypic plasticity such that the population can evolve different strategies in different environments [180]. The population's evolutionary state is now described by four traits; the daughter cell mass in environments 1 and 2 (m_1 and m_2) and fusion rate in these environments (α_1 and α_2).

For simplicity we assume that any cost of phenotypic switching or environmental sensing is negligible and that this plastic switching is instantaneous upon detection of the change in environmental conditions. The evolutionary dynamics in each environment are then decoupled. However the evolutionary trajectories in each environment are coupled by the initial trait values for the population in each environment, which we assume are the same (i.e. the population begins in a phenotypically undifferentiated state). With phenotypic plasticity, the evolutionary

dynamics are then given by

$$\begin{aligned} \frac{\mathrm{d}m_1}{\mathrm{d}\tau} &= P_1 H_m(m_1, \alpha_1; \beta_1, C), \qquad \frac{\mathrm{d}m_2}{\mathrm{d}\tau} = P_2 H_m(m_2, \alpha_2; \beta_2, C) \\ \frac{\mathrm{d}\alpha_1}{\mathrm{d}\tau} &= P_1 H_\alpha(m_1, \alpha_1; \beta_1, C), \qquad \frac{\mathrm{d}\alpha_2}{\mathrm{d}\tau} = P_2 H_\alpha(m_2, \alpha_2; \beta_2, C) \end{aligned} \tag{4.6}$$

with initial conditions

$$m_1(0) = m_2(0) = m_0, \quad \text{and} \quad \alpha_1(0) = \alpha_2(0) = \alpha_0\,. \tag{4.7}$$

Here, $H_m(m, \alpha; \beta, C)$ and $H_\alpha(m, \alpha; \beta, C)$ retain the functional form in Eq. (4.5), and P_1 and P_2 represent the probability of finding the population in environment 1 and 2 respectively, as described in Chapter 3.

As Eqs. (4.6) are only coupled through their shared initial conditions, m_0 and α_0, the choice of these initial conditions is an important consideration. Since we are interested in the initial evolution of binary cell fusion, it is natural to assume that the population evolves from a state of zero fusion, $\alpha_0 = 0$. Deciding on a plausible initial daughter cell mass takes more thought. One parsimonious choice would be that the population is already adapted to either environment 1 or environment 2 and that m_0 is given by an evolutionary fixed point in one of these environments. However if the population has been exposed to *both* the environments before phenotypic placticity has evolved, it is possible that m_0 is instead given by a bet-hedging strategy. We explore what such a strategy would look like in the following section.

Switching environments without phenotypic plasticity

We now consider the case where there is switching between environments but where the population exhibits no phenotypic plasticity. As described in the previous section, we are particularly concerned with the period before the physiological machinery for cell fusion has evolved, and so focus on the case where the cell fusion rate is fixed to zero, $\alpha = 0$. Evolution then solely acts on the daughter cell mass, m.

As in Chapter 3, environmental switching is modelled as a discrete stochastic telegraph process, with the time spent in each environment distributed geometrically. The population spends an average of $\tau_1 = 1/\lambda_{1\to 2}$ in environment 1 and $\tau_1 = 1/\lambda_{2\to 1}$ in environment 2, where $\lambda_{i\to j}$ is the transition rate from environment i to j.

The two switching rates most relevant to our model are when the environment switches many times before an invasion can complete, (fast relative to invasion, FRTI) and when each switching event occurs after multiple invasions have completed,

(fast relative to evolution, FRTE). More detail of these switching rates are provided in section 3.2.2 in Chapter 3 . However in Chapter 3, we show that the evolutionary dynamics for m in both these regimes can be approximated using the same dynamical equations. Using adaptive dynamics techniques modified to account for such environmental switching [155], we obtain

$$\frac{\mathrm{d}m}{\mathrm{d}\tau} = P_1 H_m(m, 0; \beta_1, C) + P_2 H_m(m, 0; \beta_2, C) \tag{4.8}$$

where $H_m(m, \alpha; \beta, C)$ retains the functional form in Eq. (4.5) and $P_1 = \tau_1/(\tau_1 + \tau_2)$ and $P_2 = \tau_2/(\tau_1 + \tau_2)$ are the probabilities of finding the population in the two respective environments. We therefore see that in the absence of phenotypic plasticity, the evolutionary dynamics is the weighted average of the dynamics in the two environments.

Obtaining the ESS for Eq. (4.8) is relatively straightfoward. Substituting for $H_m(m, 0; \beta_1, C)$ and $H_m(m, 0; \beta_2, C)$ using the functional form given in Eq. (4.5) and setting $\mathrm{d}m/\mathrm{d}\tau = 0$ in Eq. (4.8), we obtain the ESS

$$m^*_{\mathrm{BH},\alpha=0} = P_1\beta_1 + (1 - P_1)\beta_2 \,. \tag{4.9}$$

This strategy constitutes a bet-hedging strategy in cell mass when the population has yet to evolve phenotypic plasticity nor the capacity for cell-cell fusion. In the limits $P_1 \to 1$ and $P_2 \to 1$, we can recover the ESS strategies in the two respective environments:

$$m^*_{1,\alpha=0} = \beta_1 \,, \qquad m^*_{2,\alpha=0} = \beta_2 \,, \tag{4.10}$$

which can be verified from a consideration of the equations for $\mathrm{d}m/\mathrm{d}\tau = 0$ in a fixed environment (see Eq. (4.5)). Thus far, our analyses have only considered the scenarios where the environment switches numerous times before invasion has time to complete (FRTI) and where the environment switches once every few mutation events (FRTE). An interesting area for future consideration would be the evolutionary dynamics for the regime between FRTI and FRTE i.e. if the environment switches a small number of times before invasion has time to complete. We can now proceed to analyse how binary cell fusion can be selected for when the fusion rate α is allowed to increase from zero in the following Results section.

4.2.6 Implementation of Numerical Simulations

The stochastic simulations of the evolutionary trajectories are also implemented using a Gillespie algorithm [68] where successive mutations and environmental

switching events occur randomly with geometrically distributed waiting times. The rates of mutations μ and environmental switching λ are measured in units of (number of growth cycles)$^{-1}$. In the simulations, multiple traits coexist under a mutation-selection balance (see Section 3.3.5 of Chapter 3 and [183] for more detail), which allows us to account for variations in selection strengths in simulations of our evolutionary trajectories.

4.3 RESULTS

In this section we proceed to analyse the evolutionary dynamics derived from the mathematical model and compare our results to numerical simulations of the full stochastic simulations.

4.3.1 In a fixed environment the population evolves to either no cell fusion, or to high levels of cell-fusion, dependent on the cost of cell fusion

In Figure 4.3, we see two potential evolutionary outcomes for the fusion rates in a single fixed environment; the population can either evolve to high (technically infinite) fusion rates or to zero fusion rates. To which of these ESS the population is attracted depends both on the parameters and the initial conditions.

When the costs to cell fusion are low ($C < 0.39$), the only evolutionary attractor is the high fusion rate attractor (see Eq. (4.11) and Figure 4.3, panel (a)). In this scenario, obligate fusion is the only evolutionary outcome.

For intermediate costs to cell fusion $0.39 < C < 0.86$, there are two evolutionary attractors. The attractor to which the population evolves depends on the initial conditions (see Eq. (4.11) and Figure 4.3, Panel B). If the initial mass on the $\alpha = 0$ boundary is small, selection acts to increase fusion rate and obligate fusion is the ESS. However, if the initial mass on the boundary is sufficiently large, the state of no cell fusion becomes the evolutionarily stable state.

Finally when costs to fusion, C, are extremely high (Figure 4.3, Panel C), selection for decreased fusion rate acts regardless of initial conditions and a state in which $\alpha = 0$ (zero fusion rate) is the only evolutionary outcome. Under this high cost regime, $d\alpha/dt < 0$ along the entire line $\alpha = 0$ and so fusion rate is never selected to increase given any initial conditions.

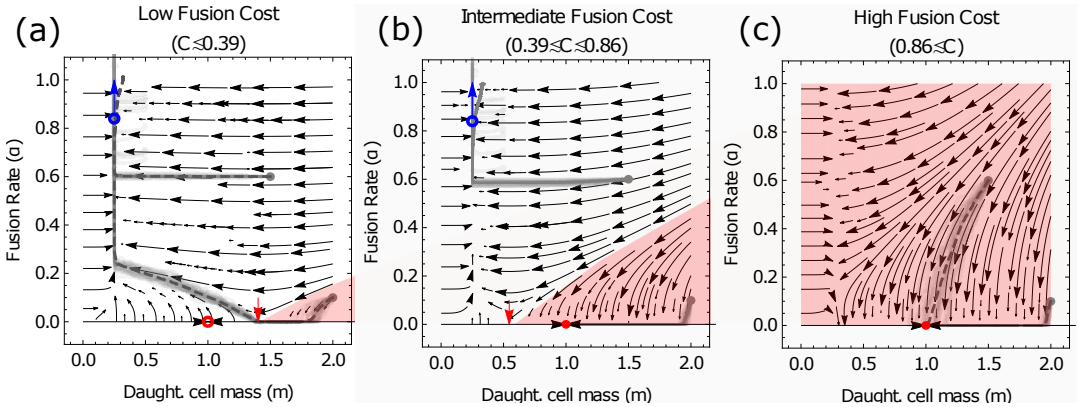

Figure 4.3: Phase portraits for the co-evolutionary dynamics in a fixed environment (see Eq. (4.5)). High fusion rates are the only evolutionary outcome when costs to cell fusion are low (panel (a)), while under intermediate costs (panel (b)), high fusion rate and zero fusion rate (obligate asex) are both evolutionary outcomes, and under high costs, zero fusion rate becomes the only evolutionary outcome (Panel C), as summarised analytically in Eq. (4.11). The red shaded region shows trajectories leading to points on the $\alpha = 0$ boundary for which evolution selects for decreasing fusion rate ($\mathrm{d}\alpha/\mathrm{d}\tau < 0$) and the critical point at which $\mathrm{d}\alpha/\mathrm{d}\tau = 0$ is marked by the red arrow (see section 4.5.2). The red circles mark a fixed point in the evolutionary dynamics of m ($m^* = \beta$, see Eq. (4.11)), which may be unstable (open circles) or stable (filled circle) under coevolution with α. The blue circles and arrows illustrate the evolutionary attractor for high fusion rates ($(m^*, \alpha^*) \to (\beta/4, \infty)$, see Eq. (4.11)). Average population trait trajectories, $(\langle m \rangle(t), \langle \alpha \rangle(t))$, from simulation of the full stochastic model are plotted in light gray, and their mean over multiple realisations are dashed. Initial conditions: $(m(0), \alpha(0)) = (1.5, 0.6)$ and $(m(0), \alpha(0)) = (2, 0.1)$. Simulation is run for 1.1×10^7 growth cycles in panel (a), 1.24×10^7 growth cycles in panel (b) and 10^7 growth cycles in panel (C). Remaining system parameters are given in section 4.6.

In section 4.5.2, we conduct a mathematical and numerical analysis to formalise the arguments above. In summary, the possible early evolutionary attractors, (m^*, α^*), are given by

$$(m^*, \alpha^*) \to \begin{cases} (\beta/4, \infty) & \text{if } 1 - \frac{1}{\sqrt{\mathrm{e}}} > C \geq 0 \\ (\beta/4, \infty) \quad \text{or} \quad (\beta, 0) & \text{if } 1 \gtrapprox C > 1 - \frac{1}{\sqrt{\mathrm{e}}} \\ (\beta, 0) & \text{if } C \approx 1 \end{cases} \tag{4.11}$$

where we note $1 - \mathrm{e}^{-1/2} \approx 0.39$ and $1 - \mathrm{e}^{-2} \approx 0.86$. While intermediate costs ($0.86 \gtrapprox C \gtrapprox 0.39$) lead to two potential evolutionary outcomes depending on the

initial conditions, it is the second of these, $(m^*, \alpha^*) = (\beta, 0)$, that is arguably the most relevant for the evolution of early cell fusion; if evolution had acted on daughter cell size, m, before the physiological machinery necessary for cell fusion had evolved, the initial condition for the co-evolutionary dynamics would be $(m(0), \alpha(0)) = (\beta, 0)$, at which the population would be subsequently held by costs to fusion.

In Figure 4.3 we also see that our mathematical analysis is a good predictor of the outcome of stochastic simulations (gray shaded lines). One minor point of departure is that at high fusion rates our simulated trajectories begin to diverge from from our analytic prediction. This discrepancy is the result of evolutionary branching in cell mass, which we explored in Chapter 3. However this branching happens at a later evolutionary stage than the focus of this study, the early emergence of binary cell fusion.

We conclude this section by addressing the key biological result that arises from this analysis; cell fusion is uniformly selected for even under moderately high costs (with a fraction of up to $C \approx 0.39$ of fused cells failing to survive) and can even be selected for under higher costs (up to $C \approx 0.86$) given necessary initial conditions. In the context of the evolution of early binary cell fusion, this provides a surprising nascent advantage to cell fusion. This advantage could even help compensate for other short-term costs arising from the later evolution of sex and recombination. The selective advantage experienced by fusing cells comes from their increased cytoplasmic volume, which leads to increased survival probabilities.

4.3.2 In a switching environment with phenotypic plasticity, binary cell fusion can evolve as a facultative stress response to harsh environments

Having considered the case of the evolutionary dynamics in a fixed environment, we now move on to consider the evolutionary dynamics of a population exhibiting phenotypic plasticity in a switching environment (see Eq. (4.6)). We recall that under the assumptions of costless and immediate phenotypic switching, the dynamics of (m_1, α_1) and (m_2, α_2) are decoupled. The evolution of the traits in the respective environments *are* coupled however through the initial conditions from which they evolve, which must be the same (i.e. a phenotypically undifferentiated state).

We consider two parsimonious choices for these initial conditions, both beginning in a state without fusion ($\alpha_1(0) = \alpha_2(0) = 0$). In the first scenario, we assume that the

population has evolved to a stable non-fusing mass adapted to a single environment (see Eq. (4.10)) such that $m_1(0) = m_2(0) = m^*_{1,\alpha=0}$ or $m_1(0) = m_2(0) = m^*_{2,\alpha=0}$. This is a situation in which the alternate environment is in some sense novel and to which the population has not adapted. In the second scenario, we instead assume that the population has evolved to a bet-hedging strategy in mass (optimising the mass of daughter cells across the two environments, see Eq. (4.9)) such that $m_1(0) = m_2(0) = m^*_{\mathrm{BH},\alpha=0}$. An illustrative phase portrait is shown in Figure 4.4.

We initially consider the scenario in which a population has initially evolved under environment 1 to reach a stable state $(m_1, \alpha_1) = (\beta_1, 0)$ (see red disk and surrounding purple circle, Figure 4.4 (a)). The population is now exposed to a second, harsher environment ($\beta_2 > \beta_1$) and allowed to evolve a phenotypically plastic response to this new environment. Starting from an initial state $(m_2(0), \alpha_2(0)) = (\beta_1, 0)$ (see purple circle, panel (b)), traits m_2 and α_2 can evolve to $(m_2, \alpha_2) = (\beta_2/4, \infty)$.

Similarly, we now consider second scenario in which the population has initially evolved under the switching environments to reach a stable state bet-hedging strategy in mass $(m_1(0), \alpha_1(0)) = (m_2(0), \alpha_2(0)) = (m^*_{\alpha=0}, 0)$ (see orange circles in Figure 4.4). Upon the evolution of phenotypic plasticity, in the benign environment we see the population traits relax to a stable state $(m_1, \alpha_1) = (\beta_1, 0)$ (no cell fusion, see Figure 4.4 (a)). However in the harsh environment, we see that the bet-hedging strategy in mass becomes unstable, and the population evolves towards traits $(m_2, \alpha_2) = (\beta_2/4, \infty)$ in environment 2 (see Figure 4.4 (b)).

In both scenarios described above, we see the emergence of facultative binary cell-fusion as a response to harsh environmental conditions that lower the survival probability of daughter cells. However we note that this is only possible if there is an appreciable increase in environmental harshness, β, between the environments. In Figure 4.5, we summarise the key results over the $\beta_1 - \beta_2$ parameter plane. Here we assume that the cost to cell fusion is intermediate ($1 - \mathrm{e}^{-2} > C > 1 - \mathrm{e}^{-1/2}$, i.e. $0.86 \gtrapprox C \gtrapprox 0.39$) such that there are regions on the boundary $\alpha = 0$ at which increased fusion rates are both selected for and against depending on the value of m (see section 4.5.2 and Eq. (4.11)); this restricts us to the more interesting parameter regime in which different evolutionary outcomes are possible in each environment.

In Figure 4.5, we see that when one of the environments is not appreciably worse than the other, binary cell fusion does not evolve in either environment. However when the difference between the environments grows more substantial, it is possible to evolve cell fusion in the harsher environment from initial condition

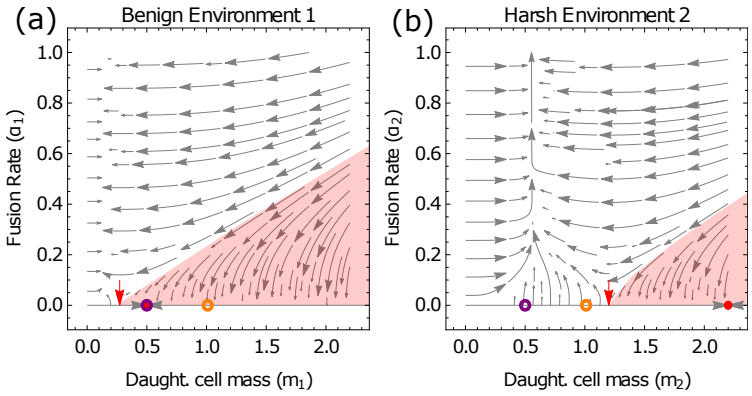

Figure 4.4: Illustrative phase portrait for co-evolutionary dynamics of $(m_1, \alpha_1, m_2, \alpha_2)$ in a switching environment with phenotypic switching that exhibits facultative binary cell fusion. In both environment 1 (panel (a)) and environment 2 (panel (b)) the cost to cell fusion is $C = 0.6$, purple circles represent the initial condition $(m_1(0), \alpha_1(0)) = (m_2(0), \alpha_2(0)) = (\beta_1, 0)$, and orange circles represent the initial condition $(m_1(0), \alpha_1(0)) = (m_2(0), \alpha_2(0)) = (m^*_{\alpha=0}, 0)$, with $m^*_{\alpha=0}$ taken from Eq. (4.9). Environmental parameters are $\beta_1 = 0.5$ and $\beta_2 = 2.2$ making environment 1 the more benign environment, in which the population typically spends a proportion $P_1 = 0.7$ of its time.

$(m_1(0), \alpha_1(0)) = (m_2(0), \alpha_2(0)) = (\beta_i, 0)$ (where the population has first evolved towards the evolutionary optimum of the more benign environment). Finally when the difference between the environments is extreme, it is also possible to evolve cell fusion in the harsher environment from initial condition $(m_1(0), \alpha_1(0)) = (m_2(0), \alpha_2(0)) = (m^*_{\alpha=0}, 0)$ (where the population has first evolved towards a bet-hedging strategy in cell mass).

4.4 DISCUSSION

The evolution of sexual reproduction and its consequences for the subsequent evolutionary trajectory of populations is of general importance to biologists [2, 87, 182]. In this chapter we have illustrated a reversal of the classic two fold cost of sex in organisms with distinct sexes [131]; in unicellular organisms, binary cell fusion can be selected for even in the presence of substantial costs due to a survival benefit that comes from increased mass. It is particularly interesting that the benefits conferred to cell-fusion through increased mass are sufficient to withstand remarkably high

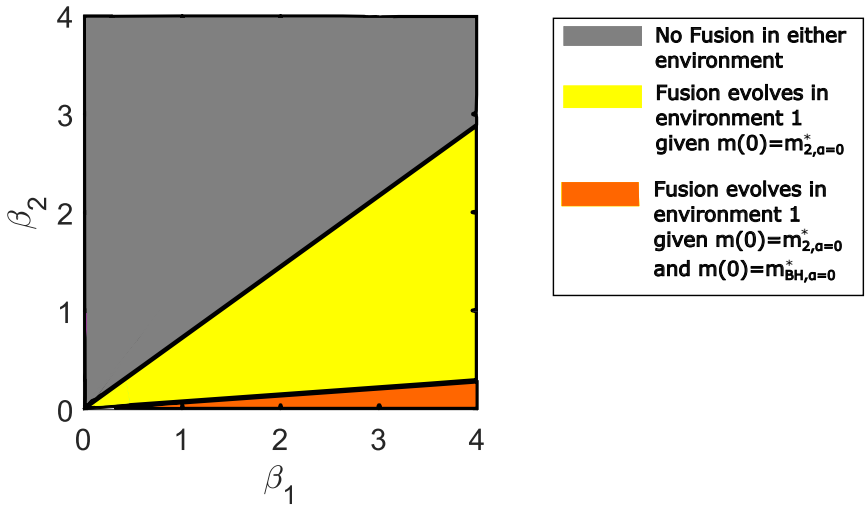

Figure 4.5: Regions in the β_1-β_2 parameter space where we evolve obligate fusion, given various initial conditions $(m(0), 0)$. The region plot is independent of E and T. Here, $C = 0.5$ and $P_1 = 0.7$. Since $C > 1 - 1/\sqrt{e}$ (see Eq. (4.11)), fusion can only evolve in at most one of the two environments. In this case it is environment 1 where fusion can evolve since $m(0) = m^*_{2,\alpha=0}$ (see Eq. (4.10)). A numerical simulation to support this regionplot is shown in Figure 4.6. This plot is obtained as described in Appendix 4.5.2.

costs, with obligate sexuality remaining the only evolutionary attractor with costs equivalent to a loss of $\sim 39\%$ cells that attempt to fuse.

Perhaps most interesting is the case of switching environments with phenotypic plasticity. Here we find under a broad set of biologically reasonable conditions (costs to cell fusion equivalent to $39\% - 86\%$ additional mortality to fused cells and at least moderate changes in environmental quality) that high fusion rates are selected for in harsh environments and zero fusion rates are maintained in benign environments. This behaviour parsimoniously recapitulates the empirically observed reproductive strategies of numerous facultatively sexual species, including *C. reinhardtii* [18], *S. pombe* [197] and *D. discoideum* [159]. This mechanism, under which cell fusion evolves to increase the survival probability of daughter cells, provides a complementary perspective on the frequent evolution of survival structures (resistant to environmental stress) that form following the formation of a zygote. These include ascospores in fungi [161] and zygote-specific stress-resistant stress wall in *C. reinhardtii* [226]. Note that such correlations between sexual reproduction and the formation of survival structures are not as easily explained under genetic explanations

for the evolution of sexual reproduction, where engaging in both behaviours at once constitutes a simultaneous (and therefore potentially costly) change in genotype and temporal dislocation in environment [62, 115].

The results above are particularly interesting in the case of the evolution of early binary cell-fusion as a first step in the evolution of sexual reproduction. While most studies focus on the genetic benefits of cell-fusion [168] (including a functionally-diploid dikaryotic cell [147]), or the genetic benefits of mixed cytoplasm [182, 217] (which can also come with costs), the mechanism at play here is purely physiological. Yet, as addressed above, it naturally captures the empirical observation of binary cell-fusion as response to challenging environmental conditions. While the mechanism does not explain the evolution of sexual reproduction and genetic recombination itself, it does provide a nascent advantage to binary cell fusion that sets the stage for the evolution of sex by bringing nuclei into contact for prolonged periods, as well as countering short-term costs associated with recombination. Conceivably, if genetic recombination is beneficial for myriad genetic reasons in the long-term [158], it would seem natural that it would be instigated when the opportunity arises (i.e. when physiological survival mechanisms bring nuclei into close contact). In this scenario recombination may not be only a direct response to environmental variability [28, 40], but also to the correlated formation of a survival structure.

One element absent from our model is the fusion of multiple cells, which is likely to be selected for under the assumptions implicit in our model. There would clearly be an upper-limit on the number of fusions selected for, arising from the likely multiplicative effect of the fusion cost C. However in this context, it is interesting to note that one of the hypotheses for the evolution of self-incompatible mating types is as a signal to prevent the formation of polyploid cells [174]. Such a mechanism could also prevent the formation of trikaryotic cells should the cost of multiple fusions be too great. Thus, the model neatly preempts the second stage in models for the evolution of eukaryotic sex, the regulation of cell–cell fusion [73].

More generally, it is interesting to note that the conditions for facultative sexuality (e.g. harsh environmental conditions) broadly coincide with those for facultative multicellularity in both bacteria and eukaryotes, with starvation triggering the formation of fruiting bodies in myxobacteria [20, 108] and flocking in yeast [45, 223]. Meanwhile in *C. reinhardtii*, the formation of multicellular palmelloids and aggregates are an alternate stress response to sexual reproduction [30], as are the formation of fruiting bodies in *D. discoideum* [211]. In this multicellular context,

the sexual behaviour of *D. discoideum* is particularly interesting, as once formed, the zygote attracts hundreds of neighboring cells that are then cannibalised for the provision of a macrocyst [48]. These various survival strategies are unified in our model as a mechanism for the evolution of binary cell fusion.

The evolutionary transition from unicellularity to facultative multicellularity was first observed in cyanobacteria [213, 215]. The life cycles of these organisms are observed to alternate between a multicellular filamentous phase and unicellularity, in response to environmental cues. Environmental stresses such as drops in salinity trigger multicellularity, while increases in population density due to filament formation causes fragmentation back into solitary cells [213]. Furthermore, eco-evolutionary models have shown that environmental fluctuations and costs of fragmentation can give rise to a diverse range of fragmentation patterns [177, 178, 179], which could potentially explain the diverse modes of facultative multicellular life cycles observed in bacteria [118]. However, most existing work has focused on the fragmentation of clusters rather than the selective pressures that drive solitary cells to come together initially. The process of forming small aggregates by single cells coming together is costly, due largely to the risk of conflict between the cells [163, 215]. By analysing the evolutionary benefits of binary cell fusion, our model may provide an alternative perspective on the mechanistic benefits of solitary cells coming together.

As addressed above, trade-offs on the evolution of cell-fusion rate, inbreeding, and the possibility of higher-order cell-fusions offer interesting avenues to extend this analysis. In addition, we have not accounted for the discrete nature of divisions leading to daughter cells, costs to phenotypic switching, non-local trait mutations, or pre-existing mating types. More generally, extending our mathematical approach leveraging adaptive dynamics to switching environments in other facultatively sexual populations might prove particularly fruitful [42, 121].

In this chapter we have extended the classic PBS model [173] in two key ways; allowing the fusion rate to evolve and subjecting the population to switching environments. In doing so, we have shown its capacity to parsimoniously capture the evolution of obligate binary cell fusion, no fusion and stressed induced binary cell fusion in unicellular organisms. These results offer particularly interesting implications for the evolution of binary cell-fusion as a precursor to sexual reproduction, as well as suggesting common mechanistic links to the evolution of binary cell fusion. Moreover, our analysis emphasises the importance of exploring the coevolutionary dynamics of a range of evolutionary parameters, and of developing computational

and mathematical approaches to elucidate facultative sexual reproduction.

4.5 APPENDIX I: EVOLUTIONARY DYNAMICS: SWITCHING ENVIRONMENTS WITH PHENOTYPIC PLASTICITY

4.5.1 Derivation of evolutionary ODEs

With phenotypic plasticity, the evolutionary trajectory of the population in one environment is independent of that in the other. A mutation would exert a selection pressure in only one environment. If a mutation occurs on traits m_i or α_i in environment i, the invasion dynamics are given respectively by

$$\begin{aligned}
\frac{d\hat{f}_{m_1}}{dt_g} &= \hat{f}_{m_1}(1-\hat{f}_{m_1})\Gamma_m(\beta_1)P_1 \\
\frac{d\hat{f}_{\alpha_1}}{dt_g} &= \hat{f}_{\alpha_1}(1-\hat{f}_{\alpha_1})\Gamma_\alpha(\beta_1)P_1 \\
\frac{d\hat{f}_{m_2}}{dt_g} &= \hat{f}_{m_2}(1-\hat{f}_{m_2})\Gamma_m(\beta_2)P_2 \\
\frac{d\hat{f}_{\alpha_2}}{dt_g} &= \hat{f}_{\alpha_2}(1-\hat{f}_{\alpha_2})\Gamma_\alpha(\beta_2)P_2\,,
\end{aligned} \tag{4.12}$$

for each mutant type, where $\Gamma_m(\beta)$ and $\Gamma_\alpha(\beta)$ are of the same functional form as $h_m(m,\alpha,\beta,C)$ and $h_\alpha(m,\alpha,\beta,C)$ respectively in Eq. (4.3). From here, we can apply the same approach as section 4.2.5 to obtain the evolutionary dynamics given by Eq. (4.6) of the main text.

4.5.2 Analysis of ODEs

By inspection of Eq. (4.6) of the main text, we note that the co-evolutionary dynamics of the mass and fusion rate in environment 1 (m_1,α_1) and environment 2 (m_2,α_2) themselves are identical to that of a fixed environment with $\beta=\beta_1$ and $\beta=\beta_2$ respectively. Since the dynamics of (m_1,α_1) and (m_2,α_2) are decoupled, we can infer the behaviour of the system in each environment by treating them as two separate fixed environments, but with shared initial conditions.

Using the analysis given in section 4.3.1, we infer that in cases where $C<1-1/\sqrt{\mathrm{e}}$, obligate fusion is the only evolutionary outcome in both environments, given any

initial condition within the phase plane. In other words, $(m_1, \alpha_1) = (\beta_1/4, \infty)$ and $(m_2, \alpha_2) = (\beta_2/4, \infty)$ is the only evolutionary attractor of the system under this regime.

In the case where $C > 1-1/\sqrt{\mathrm{e}}$, there exists a scenario whereby obligate fusion can evolve if the initial state of the population is $(m_1(0), \alpha_1(0)) = (m_2(0), \alpha_2(0)) = (\beta_1, 0)$. Since $(m_1, \alpha_1) = (\beta_1, 0)$ is a stable fixed point, obligate fusion does not evolve in environment 1. The condition for obligate fusion to evolve in environment 2 is for $\beta_1 < -\beta_2/(2\ln(1-C))$. Using the same argument, if the initial state of the population is $(m_1(0), \alpha_1(0)) = (m_2(0), \alpha_2(0)) = (\beta_2, 0)$, obligate fusion does not evolve in environment 2. Given this initial state, the condition for obligate fusion to evolve in environment 1 is $\beta_2 < -\beta_1/(2\ln(1-C))$. In other words, if $C > 1-1/\sqrt{\mathrm{e}}$, it is only possible for obligate fusion to be selected for in the alternate environment to which the population has stabilised, prior to evolving the mechanism of fusion. The criterion for obligate fusion to evolve in this alternate environment j is for $(m_j^0, \alpha_j^0) = (\beta_i, 0)$ $(i \neq j)$ to sit left of the critical point at which $\mathrm{d}\alpha_j/\mathrm{d}\tau = 0$ on the $\alpha_j = 0$ boundary (this critical point is represented by the red arrow in Figure 4.3 of main text). This critical point is situated at approximately $m_j = -\beta_j/(2\ln(1-C))$. If the criterion for obligate fusion in environment j is fulfilled, the evolutionary attractor in environment j is $(m_j, \alpha_j) = (\beta_j/4, \infty)$ as is the case for a fixed environment with $\beta = \beta_j$ (see Eq. (4.11) in the main text).

If the initial state of the population is $(m_1(0), \alpha_1(0)) = (m_2(0), \alpha_2(0)) = (m^*_{\alpha=0}, 0)$ where $m^*_{\alpha=0} = P_1\beta_1 + (1-P_1)\beta_2$, obligate fusion evolves in environment 1 if $P_1\beta_1 + (1-P_1)\beta_2 < -\beta_1/(2\ln(1-C))$, provided that $\beta_1 > \beta_2$. From the same initial state, obligate fusion evolves in environment 2 if $P_1\beta_1 + (1-P_1)\beta_2 < -\beta_2/(2\ln(1-C))$, provided that $\beta_2 > \beta_1$.

Again, using the analysis given in Eq. (4.11) of the main text, we infer that in the case where $C > 1 - 1/\mathrm{e}^2$, zero fusion is the only evolutionary outcome in both environments too.

4.5.3 Implementation of simulations

We numerically simulate the evolutionary dynamics starting from the initial states as described in Figure 4.4 of the main text. To simulate the evolutionary dynamics for a population with initial conditions $(m_1(0), \alpha_1(0)) = (m_2(0), \alpha_2(0)) = (\beta_1, 0)$, we use a fixed value of β_2, and run the evolutionary dynamics as described in section 4.2.6,

for various values of β_1 (ranging from 0 to 4 in our simulations). To simulate the evolutionary dynamics for a population with initial condition $(m_1(0), \alpha_1(0)) = (m_2(0), \alpha_2(0)) = (P_1\beta_1 + (1 - P_1)\beta_2, 0)$, we again run the evolutionary dynamics as described in the above sentence, using the same fixed value of β_2. The parameters used to run the evolutionary dynamics are $\delta = 1/50$, $\mu = 1/500$, $f_0 = 1/500$ and simulation run for $400/\mu$ growth cycles. All the other parameters are as in Figure 4.5 of main text. Running it for 2×10^5 growth cycles ($400/\mu$ growth cycles) provides sufficient time for the stochastic trajectory to reach the $m \approx \beta/4$ manifold in the alternate environment. At the 400th mutation event, we check whether the conditions $|m - \beta/4| < \delta$ and $\alpha > \delta$ are satisfied in order to determine whether the trajectory has reached the vicinity of the $m \approx \beta/4$ manifold. A numerical simulation is shown in Figure 4.6.

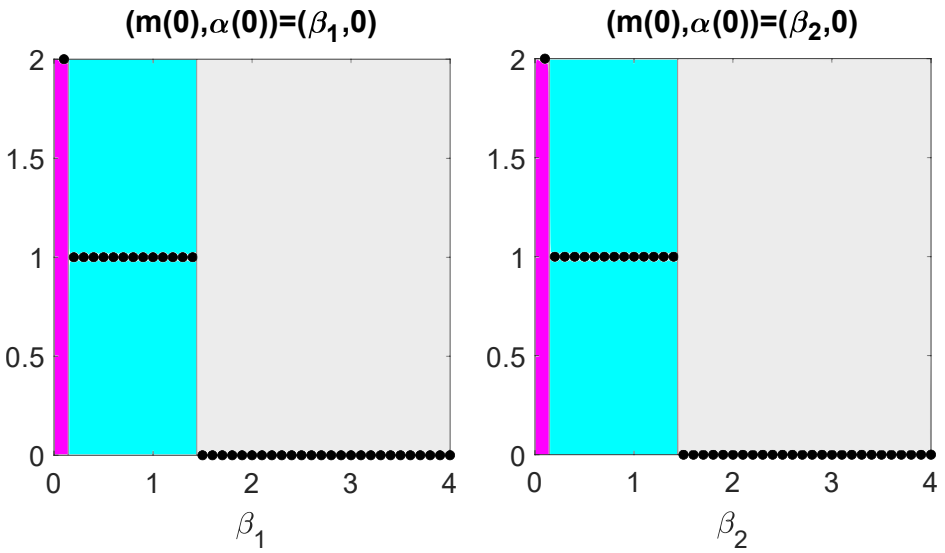

Figure 4.6: Black markers show numerical simulations of the evolutionary outcome in a switching environment with phenotypic plasticity. If the marker has a y-coordinate of 2, we evolve obligate fusion in environment j given both $(m(0), \alpha(0)) = (\beta_i, 0)$ and $(m(0), \alpha(0)) = (P_1\beta_1 + (1 - P_1)\beta_2, 0)$ for $i \in \{1, 2\}$, $i \neq j$. If the marker has a y-coordinate of 1, we evolve obligate fusion in environment j given $(m(0), \alpha(0)) = (\beta_i, 0)$ only (for $j \neq i$) and if the marker sits at 0, we evolve no fusion given either initial condition. Different coloured regions represent the behaviour predicted analytically predicted in Figure 4.5 of the main text. In panel (a) $\beta_2 = 2$ and $i = 1$ and in Panel B $\beta_1 = 2$ and $i = 2$. We see that the numerical simulations match the analytical predictions. The parameter conditions are $E = 100$, $T = 1$, $C = 0.5$, ($P_1 = 0.7$ if $\beta_1 > \beta_2$ and $P_1 = 0.3$ if $\beta_2 > \beta_1$), $\delta = 2 \times 10^{-2}$, $\mu = 2 \times 10^{-3}$ and $f_0 = 2 \times 10^{-3}$. Simulation is run for 2×10^5 growth cycles ($500/\mu$ growth cycles).

4.6 APPENDIX II: PARAMETERS FOR SIMULATIONS

In all the relevant figure panels, unless otherwise stated, the system parameters used are $E = 100$, $T = 1$ and $\beta = 1$ and the simulation parameters are $\delta = 5 \times 10^{-3}$, $f_0 = 2 \times 10^{-3}$, $\mu = 5 \times 10^{-4}$ (number of growth cycles)$^{-1}$

— 5 —

Evolution of environmentally-triggered sexual reproduction and hibernation

Abstract

Environmentally triggered facultative sex is observed in a range of organisms, including green algae, yeast and water fleas. These species switch to sexual reproduction in response to environmental adversity (e.g. nitrogen limitation) and revert to their default asexual reproductive mode when the ideal environmental condition returns. Sexual reproduction is also frequently observed to co-occur with the formation of stress-resistant survival structures that are associated to hibernation. Existing theories for the evolution of environmentally triggered facultative sex include the Red Queen hypothesis, which proposes that sex hastens adaptation in fluctuating environments due to its ability to generate genetic diversity rapidly through genetic recombination. However, the Red Queen hypothesis alone is incompatible with the theory of hibernation. Conversely, Muller's ratchet argues that sex evolved due to its ability to prevent deleterious mutations from accumulating through genetic recombination, but has not been linked to hibernation. By modifying Muller's ratchet to account for population dynamics as well as the selective advantages of genetic recombination in terms of purging deleterious mutations, we develop an eco-evolutionary model that explains how hibernation as well as sex can evolve at high population densities when opportunity costs for population growth diminish. By suitably modifying our model, we demonstrate that hibernation is subject to the same selective pressure as sex, and evolves at high population densities too, which parsimoniously explains how they both evolve to occur in tandem. In addition to our analysis, we provide simulations of evolutionary trajectories, which reveal richer dynamical behaviour beyond what is predictable by our

analytics.

5.1 INTRODUCTION

There are numerous organisms that exhibit environmentally triggered sex, such as in algae, fruitfly, slime moulds and aphids [62, 200]. In the green algae *C. reinhardtii* and the yeast *S. pombe* [197], sexual reproduction is triggered in response to nitrogen limitation, while *daphnia* water fleas have been shown to switch to sexual reproduction in response to extreme temperatures [77]. The social amoeba *D. discoideum* [159] undergoes sexual reproduction in response to darkness and moisture. Environmentally triggered sex is also present in several plant species such as *R. reptans* [114], where it was shown experimentally that germination of sexually produced seeds was more frequent when grown at higher population densities.

An early theory for the evolution of sex is Muller's Ratchet [154], which proposes that sex evolved due to its ability to prevent the build up of deleterious mutations through genetic recombination. It summarises that in a finite non-recombining population, the buildup of deleterious mutations is so rapid that it cannot be prevented by the selection that acts against them. Genetic recombination in sexual reproduction allows advantageous genes from separate individuals to be combined into the genome of one individual more often, which can help prevent the buildup of deleterious mutations. Redfield [187] was the first to use computer simulations to demonstrate this selective advantage, which showed that bacteria with a heavy deleterious load that underwent genetic transformation (equivalent to meiotic sex) evolved a higher fitness than asexual bacterial populations. Although this has been used to explain the evolution of sex, it does not explain why sex occurs in response to fluctuating environments, which is frequently observe empirically [162].

Another theory, known as the Red Queen hypothesis [221] proposes that sex evolved due to its ability to facilitate adaptation in actively changing environments [10, 31, 221], as different genotypes are favoured in different environments. Since sex involves genetic recombination, it allows well adapted genotypes to be generated more rapidly than asexual populations, where mutations are the only means of generating novel genotypes (see Figure 5.2 (a)). The Red Queen hypothesis is also the most common explanation for the occurrence of sex in hosts as a means of resisting infection by parasites [82, 106], which are themselves rapidly evolving.

Although the Red Queen hypothesis explains how sex occurs in response to changing environments, it does not explain how dormancy occurs following sex in most facultatively sexual organisms (see Figure 5.2 (b)). In fact, almost all organisms exhibiting environmentally triggered sex enter a phase of dormancy [21, 62, 197].

One key element not accounted for in the aforementioned theories for the evolution of sex is the underlying ecological factors, which may influence the cost of sex. A mechanistic cost of sex is that it reduces the population growth rate [62, 131], meaning that in a population under logistic growth, the demographic cost of sex is diminished as the population approaches carrying capacity [47, 63] (i.e upon resource limitation or reaching a sufficient population density). Experimentally, it has been shown that sex is induced in *daphnia* upon reaching high population densities [63]. Here, we detail some literature that considers how the propensity for sex in a population is influenced by the population density.

Computational simulations of the consumer-resource model by [193, 203] have shown how sexual populations can outcompete asexual populations under resource scarcity. In particular, sexual populations are favorable in conditions such as low population death rates and low resource replenishment rate. However, their model simply considered the regions in the genotype space that allow sexual populations to invade asexual populations, without accounting for any aspects of evolution, including trait mutations or population genetics.

Furthermore, Gerber *et al.* [61] investigated the evolution of environmentally triggered facultative sex in the *daphnia* reproductive system, where sex necessitates the production of dormant female eggs required for survival, while asexual reproduction promotes fast population growth. In particular, they used a genetic algorithm to investigate the time during a growing season that the population engages in sex to maximise survival across seasons. They showed that if the season length is less predictable, the population adopts a bet-hedging strategy by switching from asexual to sexual reproduction earlier in the season [206].

Taking inspiration from the work of [61], we develop an eco-evolutionary model that provides a novel analytical explanation for environmentally triggered facultative sex, based on the ability for sex to prevent the buildup of deleterious mutations through genetic recombination (referred to as "recombination" throughout the chapter). By incorporating biologically realistic population dynamics into Muller's Ratchet and accounting for the ecological costs of sex, we investigate the optimal time for a population to switch from asexual reproduction with logistic growth to sexual

reproduction. As dormancy incurs similar ecological costs as sex [62, 134], our model parsimoniously explains how dormancy can evolve in response to environmental stress.

Evolving populations typically contain genotypes with different trait values competing against each other, which may impose eco-evolutionary effects that cause the evolutionary optimal state to differ with the state that maximises the fitness of an individual genotype (individual optimum) [53, 163]. It is thus vital for us to consider both these optimal states. In Section 5.2, we provide details of our models for hibernation and genetic recombination and in Section 5.3, we include details of our results.

5.1.1 Illustration of Muller's Ratchet and the Red Queen hypothesis

Here, we use diagrams to illustrate the Red Queen hypothesis and Muller's Ratchet. In Figure 5.1, we demonstrate Muller's Ratchet, which describes how recombination can increase population fitness by reversing the depletion of genotypes free of deleterious mutations. In the absence of recombination, the only way by which genetic diversity can increase is through mutations. An asexual population would have to wait until the mutants arise by chance [33, 54, 153], which is very inefficient due to the slow timescale over which mutations accumulate. In sexual populations, recombination allows multiple well adapted genes from different individuals to be combined into one individual, which speeds up adaptation by reducing the number of mutations per individual required for adaptation. This is known as the Fisher-Muller hypothesis [54, 153].

In Figure 5.1 (b), we see that a recombination event between two genotypes that contain one deleterious mutation (single mutants) can produce genotypes that contain no deleterious mutations (non mutants), which are better adapted than both the recombining genotypes themselves. Although genotypes containing two deleterious mutations (double mutants) can be produced, those are likely to be selected out of the population due to their maladaptedness. We have hereby demonstrated the advantage conferred by genetic recombination.

In Figure 5.2 (a), we demonstrate how recombination can speed up adaptation to novel environments when the environment is actively changing (Red Queen hypothesis), and furthermore its inability to explain the evolution of dormancy

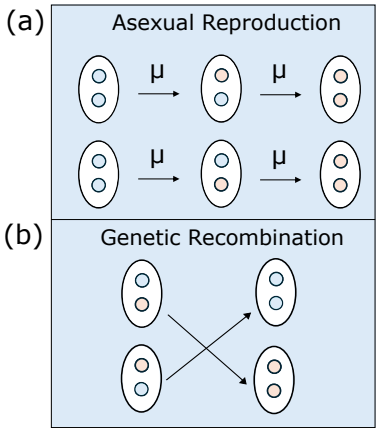

Figure 5.1: Schematic illustrating Muller's Ratchet. Each oval represents a 2-locus bialleic genotype as in Figure 5.2. Genotypes containing 2 blue alleles are optimally adapted (with no deleterious mutations) and those with 2 yellow alleles are the most maladapted (two deleterious mutations). Assuming that deleterious mutations are the only possible mutation event, staying in asexual reproduction causes buildup of maladapted genotypes and depletion of optimally adapted genotypes (Panel (a)). In panel (b), we show how genetic recombination can restore optimally adapted genotypes following deleterious mutations.

Figure 5.2 (b). Many living organisms have evolved to enter the process where they enter a dormant state in response to environmental stress e.g seasonal changes [232]. This process is known as "hibernation", which we refer to throughout this chapter.

5.2 MODELS

Our aim is to show why the onset of sexual reproduction is often correlated with hibernation or dormancy sometime after. To this end we investigate two models. We begin with our model for hibernation in Section 5.2.1 before moving onto our model for sex, as our model for hibernation is the simpler one. By accounting for the buildup of deleterious mutations, our hibernation model can be modified to one that investigates sex, which is shown in Section 5.2.2. Since we are considering sex in terms of its ability to improve the survival of the population via genetic recombination, sex is equivalent to genetic recombination in our model, and we use these terms interchangably. We refer to our model for sex as "model for genetic recombination".

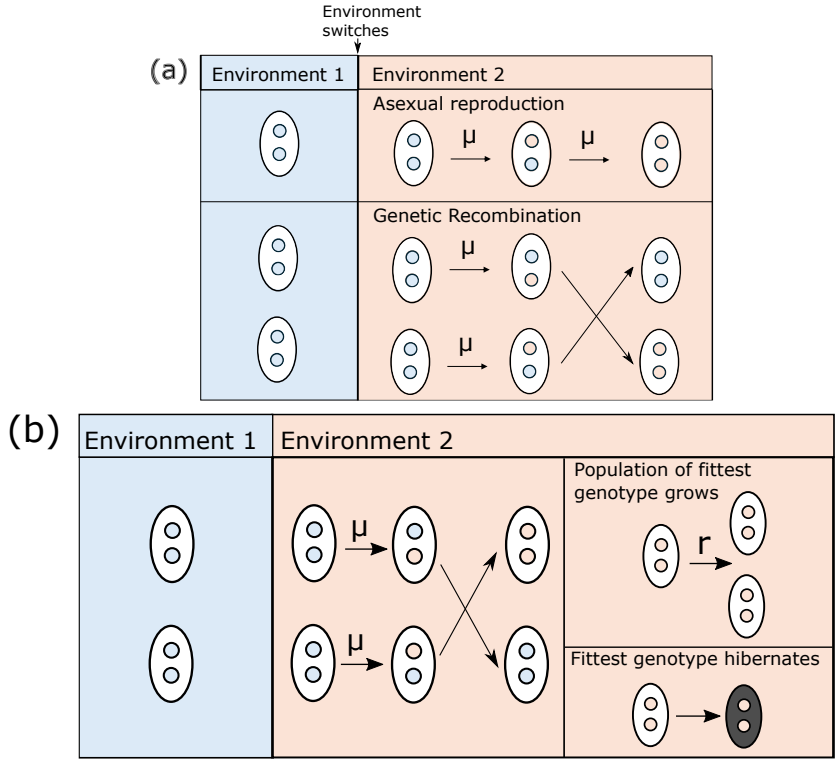

Figure 5.2: Schematic illustrating the Red Queen hypothesis. Each oval represents a 2-locus biallelic genotype with circles representing alleles. Genotypes containing 2 blue alleles are optimally adapted to environment 1 and those containing 2 orange alleles are optimally adapted to environment 2. Each allele mutates at rate μ. Panel (a): Upon switching to environment 2, an asexual population requires two or more mutations per individual to adapt to its environment, whereas a sexual population can adapt through genetic recombination following one mutation. This hastens adaptation as recombination occurs on a faster timescale than the buildup of mutations. Panel (b): If hibernation immediately follows sex, the fitness advantage conferred by recombination is not utilised, since the fittest genotype cannot grow in hibernation.

In both our models for hibernation and genetic recombination, we assume a time period T over which the population is capable of logistic growth, defined as the "growth period". The end of the growth period (time T) represents the time when either the resources fall below a threshold abundance or when the population density exceeds a threshold (see Appendix 5.5) such that population growth can no longer occur, and only sufficiently fit population members can survive. We assume that the population may begin entering hibernation (in our hibernation model) or genetic recombination (in our genetic recombination model) anytime during this period.

As population growth is naturally attenuated during hibernation and sexual reproduction [62, 63], we assume that population growth stops once it begins to engage in sex or hibernation (i.e once it enters the sex or hibernation phase). As mentioned in the introduction, there are two ways to investigate the optimal timing of hibernation or recombination. One is the "individual optimal timing" and the other is the "evolutionary optimal timing". The individual optimal timing represents the timing of hibernation t_H or recombination t_R that maximises the fitness of a population that is monomorphic in t_R or t_H (i.e a population only one trait value for t_R or t_H). The evolutionary optimal timing represents the value of t_H or t_R that the population evolves to when the population contains genotypes with different values of t_H or t_R interacting with each other. We denote the individual optimal timing of hibernation and sex as $t_H^{(I)}$ and $t_R^{(I)}$ respectively, and the evolutionary optimal timing of these processes as t_H^* and t_R^*. To investigate the evolutionary dynamics, we apply an adaptive dynamics type framework to our simulations, where we assume continuous trait spaces with small mutational stepsizes, as in Chapters 3 and 4.

5.2.1 Evolution of Hibernation Time

The growth period begins with the population under logistic growth. At some time during the growth period t_H, population members may begin entering a phase of hibernation at a fixed rate to ensure that some members survive at time T (see Figure 5.3). The value of t_H thus impacts the number of population members in hibernation at time T. We assume that non-hibernating population members cannot survive the environment at time T, so at time T, the population is subject to a round of mortality where only population members in hibernation survive (see Figure 5.3). To this end, we are interested in the optimal value of t_H that maximises the population that survives at time T. To do this, we investigate the individual optimal timing of hibernation $t_H^{(I)}$ and the evolutionary optimal timing of hibernation t_H^* (see Appendix 5.5). Both quantities require us to consider the population dynamics within a growth period, which is where we begin our analysis.

5.2.1.1 Dynamics within a growth period

At a given time t_H during the growth period where $0 < t_H < T$, members of the population begin entering hibernation at a rate of γ. We refer to the period after the population begins entering hibernation as the "hibernation phase" (i.e when

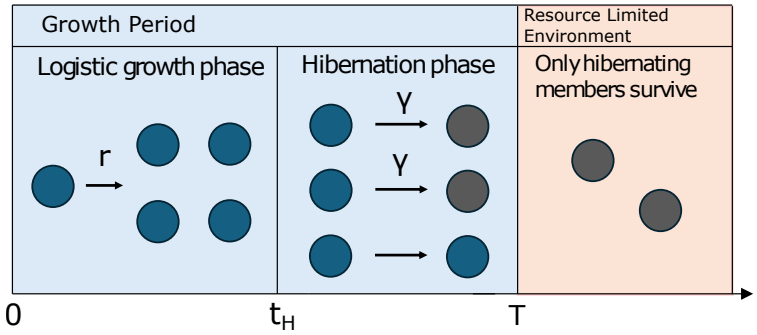

Figure 5.3: Schematic of hibernation model behaviour within a growth period. The population begins each growth period in logistic growth (at rate r). At some time t_H, members of the population begin entering hibernation (at rate γ). In the resource limited environment at the end of the growth period (time T), the population is subject to a round of mortality, where only members that have ended up in hibernation (grey) survive. Surviving members form the beginning of the next growth period, and this cycle repeats itself.

$t_H < t < T$). In addition to the survival benefits, there are costs to hibernation, such as missed opportunities for population growth [134] and diapause occurring at the wrong time [97], which are accounted for by assuming that population growth stops during hibernation phase. Due to the logistic growth that occurs before the hibernation phase, the value of t_H can influence population fitness, as population fitness is governed by the number of hibernating members at time T (see Figure 5.4). Below, we provide equations that describe the population dynamics within a growth period. Denoting $x(t)$ as the population under logistic growth and $y(t)$ as the population in hibernation at a given time, the population dynamics of the system before entering hibernation phase ($0 < t < t_H$) is given by

$$\begin{aligned} \frac{\mathrm{d}x}{\mathrm{d}t} &= rx\left(1 - \frac{x}{k}\right) \\ \frac{\mathrm{d}y}{\mathrm{d}t} &= 0 \end{aligned} \tag{5.1}$$

where r is the intrinsic growth rate and k is the carrying capacity. Once the population enters hibernation phase (i.e during $t_H < t < T$), the population dynamics are given by

$$\begin{aligned} \frac{\mathrm{d}x}{\mathrm{d}t} &= -\gamma x \\ \frac{\mathrm{d}y}{\mathrm{d}t} &= \gamma x\,. \end{aligned} \tag{5.2}$$

The population dynamics of a system that enters hibernation at various times during the growth period (various t_H) are shown in Figure 5.4. As in Chapters 3 and 4, we assume a continuous number of population members for simplicity. In Figure 5.4,

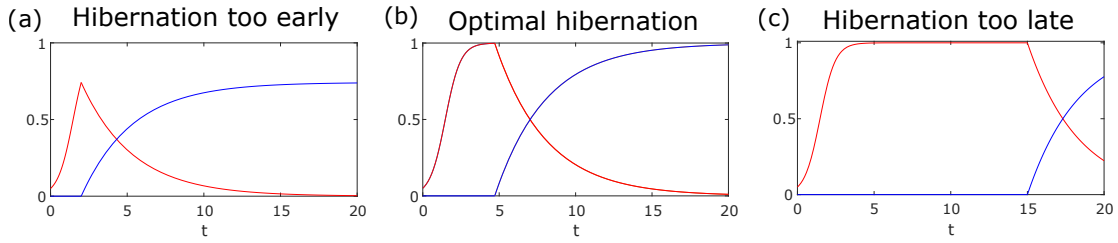

Figure 5.4: Population dynamics within a growth period when the hibernation phase begins at different times. Parameters are given in Appendix 5.9. For the parameters in this figure, the optimal timing of hibernation is $t_H = 4.7$ (as in panel (b)), which can be found using Eq. (5.3) as described in Section 5.2.1.2 below. If the hibernation phase occurs too early panel (a) the population sacrifices too much opportunity for growth, and if hibernation occurs too late panel (c), there will be less time for population members to enter hibernation. In Appendix 5.6.5 we prove that the individual and evolutionary optimal timing of hibernation are identical. Blue curves represent the population in hibernation $y(t)$ and red represents the population not in hibernation $x(t)$. In panel (a), $t_H = 2$ and in panel (b), $t_H = 15$.

we see that entering hibernation either too early or too late reduces the population in hibernation at time T. It is thus ideal for the population to enter hibernation some time during the middle of a growth period. For convenience, we refer to t_H as the "timing of hibernation" throughout this chapter. Below, we investigate the individual optimal timing of hibernation.

5.2.1.2 Individual Optimal Timing of Hibernation

To determine the individual optimal timing of hibernation $t_H^{(I)}$, we need to use the solution for $y(T)$ in Eqs. (5.1-5.2), which is given by

$$y(T) = \frac{k x_{\text{total}}(0) e^{r t_H}}{k + x_{\text{total}}(0)(e^{r t_H} - 1)} (1 - e^{-\gamma(T - t_H)}), \tag{5.3}$$

where $x_{\text{total}}(0)$ is the total population at the beginning of each growth period. Eq. (5.3) is obtained by first solving for $x(t)$ and $y(t)$ in Eq. (5.1) and evaluating them at $t = t_H$. Next we solve Eq. (5.2) for $x(t)$ and $y(t)$. As Eq. (5.2) is defined over the time interval $(t_H < t < T)$, it is necessary here to label t as t' in $x(t)$ and $y(t)$ where $t' = t - t_H$, so that the initial condition $x(t')|_{t'=0} = x(t_H)$. To calculate $t_H^{(I)}$, it

is necessary to solve for t_H in the expression $dy(T)/dt_H = 0$ (see Appendix 5.6.5 for details). Since $dy(T)/dt_H = 0$ is a transcendental equation with no explicit solution in t_H, we evaluate $t_H^{(I)}$ numerically using the `Findroot` command in Mathematica (see `hibernation.nb` in the `Evol_Env_triggered_sex` repository, where its link is provided in the Accompanying Material section).

5.2.1.3 Evolutionary Optimal Timing of Hibernation

Now, we determine the evolutionary dynamics of t_H, which requires an approximation for the invasion dynamics. In order to approximate the invasion dynamics, it is necessary to determine the population dynamics within a growth period for the case where a mutant that enters hibernation at time $\hat{t}_H = t_H + \delta t_H$ is in competition with the resident population, which enters hibernation at time t_H (see Appendix 5.6.2).

As in Chapters 3 and 4, we adopt the classical assumptions of adaptive dynamics here. The frequency of mutants in the subsequent growth period is given by the proportion of mutants in the hibernating population (see Eq. (5.23)).

5.2.1.4 Invasion Dynamics

Here, we mathematically characterise the invasion dynamics of a mutant (see Appendix 5.6.3). The invasion dynamics of the mutant with a different timing of hibernation is of the form

$$\frac{\mathrm{d}\hat{f}_{t_H}}{\mathrm{d}t} = \hat{f}_{t_H}(1 - \hat{f}_{t_H})h_{t_H}(t_H) \tag{5.4}$$

where $h_{t_H}(t_H)$ is a constant that depends on t_H and all the system parameters. We observe in Eq. (5.4) that we have frequency independent selection. Eq. (5.4) can then be used to mathematically characterise the evolutionary dynamics of t_H (see Appendix 5.6.4).

5.2.1.5 Evolutionary Dynamics

As shown in Figure 5.4, the timing of hibernation affects the number of hibernating population members at the end of a growth period. The evolutionary dynamics for the timing of hibernation t_H are derived using Eq. (5.4), and is given by

$$\frac{\mathrm{d}t_H}{\mathrm{d}\tau} = \frac{r(k - x_{\text{total}}(0))}{(k - x_{\text{total}}(0) + e^{rt_H}x_{\text{total}}(0))} - \frac{\gamma}{(e^{\gamma(T-t_H)} - 1)} \tag{5.5}$$

(see Appendix 5.6.4 for derivation). The evolutionary optimal timing of hibernation t_H^* is given by the solution for t_H to Eq. (5.5) that corresponds to the stable fixed

point in the interval $[0, T]$. As for the case of $y(T)$ in Eq. (5.3), $\mathrm{d}t_H/d\tau = 0$ is again a transcendental equation in t_H, which makes it necessary to seek a numerical solution for t_H. Here, we again use the `Findroot` command in Mathematica (see `hibernation.nb` in the `Evol_Env_triggered_sex` repository). In Appendix 5.6.5, we also prove that t_H^* is equal to $t_H^{(I)}$ i.e. the timing of hibernation that the population evolves to is equal to the timing of hibernation that maximises the number of hibernating population members at the end of a growth period (individual optimal timing). This is expected in our hibernation model, since the only form of interactions between different genotypes are competitive interactions, with no public goods or other interaction types.

5.2.2 Evolution of timing of Genetic Recombination

In this section, we demonstrate how our model for hibernation can be modified to investigate the evolution of environmentally triggered sex. As in our hibernation model, we assume a period T over which the population can undergo logistic growth and genetic recombination. The population is again subject to a round of fitness dependent mortality at time T, but this time, it is the genotypes with relatively few deleterious mutations that survive. These surviving population members form the beginning of the subsequent growth period.

At some time t_R during the growth period, population members enter a phase of recombination. In principle, it is possible to obtain an analytical approximation for the evolutionary dynamics for the timing of recombination, however due to the lengthy expression for the population of the genotypes, an analytical derivation of the evolutionary dynamics would be of little practical use. We therefore restrict our analysis to the individual optimal timing of recombination $t_R^{(I)}$. Although, we can still obtain insights into the evolutionary dynamics of t_R by simulating evolutionary trajectories. Predicting $t_R^{(I)}$ analytically requires us to consider the population dynamics within a growth period, however here, there are two regimes for genetic recombination.

In the first regime, population members can engage in sex many times during their lifetime (iteroparous), which is observed in species including *daphnia*. In the second regime, population members can only engage in sex once in their lifetime (semelparous), and this is observed in organisms such as *S. pombe* and *C. reinhardtii* [85, 197]. We abbreviate these two regimes as "iteroparous" and "semelparous"

respectively. For both regimes, we analytically investigate the individual optimal timing of recombination $t_R^{(I)}$.

Our model for recombination is inevitably more complex than our model for hibernation, due to the presence of deleterious mutations, and four different genotypes that emerge through deleterious mutations or genetic recombination events. To this end, we provide an overview of the exact process by which genotypes emerge and the precise definitions for each genotype, before providing the equations for the population dynamics within a growth period.

5.2.2.1 Model Behaviour

In our model, each growth period starts with a population of genotypes free of deleterious mutations. These genotypes are subject to deleterious mutations as the population grows logistically. We construct a three locus model where one locus is a non-recombining region that encodes the timing of recombination t_R. The other two loci determine the genetic identity (genotype) of the population members. Each of these two loci can contain two possible alleles 0 or 1, where 0 is the non mutant allele and 1 as the deleterious mutant. We assume that mutations are strictly deleterious, i.e mutations from 0 to 1 are the only possible mutation event. These mutations occur at a fixed rate of μ at each locus.

We represent the genotype of an individual using vectors $\binom{a}{b}$ where a and b correspond to the alleles in the top and bottom locus respectively. There are four possible genotypes in our model $\binom{0}{0}$, $\binom{0}{1}$, $\binom{1}{0}$ and $\binom{1}{1}$. A deleterious mutation event transforms $\binom{0}{0}$ into either $\binom{0}{1}$ or $\binom{1}{0}$ depending on the locus in which the mutation occurred, and $\binom{0}{1}$ and $\binom{1}{0}$ transforms into $\binom{1}{1}$.

We class $\binom{0}{0}$, $\binom{0}{1}$ and $\binom{1}{0}$ as viable genotypes and $\binom{1}{1}$ as the lethal genotype. The lethal genotype cannot survive in the resource limited environment, nor undergo logistic population growth (i.e $r = 0$) or recombination, all of which the viable genotypes can (see Figure 5.5). It is important to stress that $\binom{0}{1}$ and $\binom{1}{0}$ thus have the same fitness as the $\binom{0}{0}$ genotype, as they have the same growth rate and survive equally well in the resource limited environment. However as the $\binom{0}{1}$ and $\binom{1}{0}$ genotypes are one mutation away from becoming the lethal $\binom{1}{1}$ genotype, it is more favorable to have a smaller population of these types in the beginning of a growth period, to reduce the buildup of deleterious mutations. Throughout the text, we refer to the $\binom{0}{0}$ as the non mutant genotype, $\binom{0}{1}$ or $\binom{1}{0}$ as the single mutant genotype and $\binom{1}{1}$ as the double mutant genotype. During each growth period, the

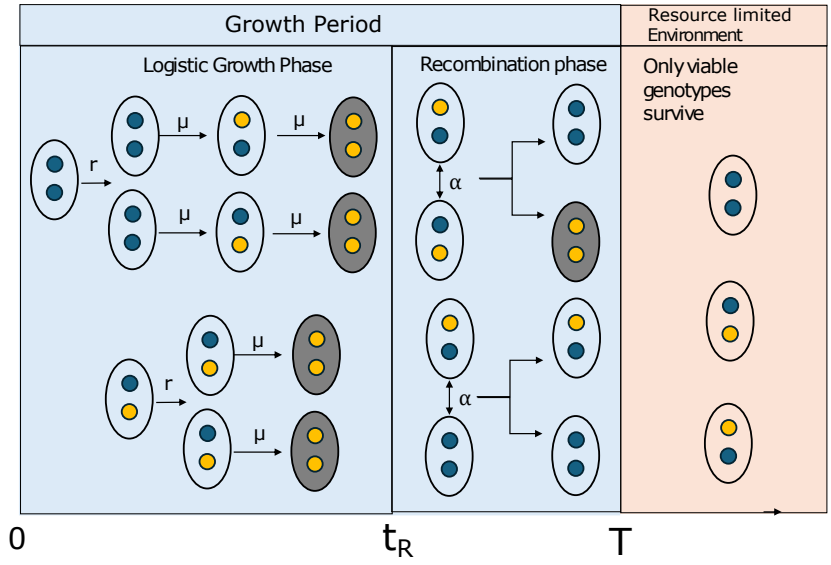

Figure 5.5: Schematic of genetic recombination model behaviour within a growth period. During the logistic growth phase, genotypes are subject to deleterious mutations (at rate μ). At time t_R, members of the population begin entering recombination, which occurs at rate α. At the end of a growth period (time T), individuals are subject to a round of mortality, where only viable genotypes survive. We also assume that recombination does not occur in the resource limited environment at time T. Blue dot represents the non mutant 0 allele and yellow represents the deleterious mutant 1 allele.

population of each genotype is subject to continuous change, due either to the buildup of deleterious mutations or the population switching from logistic growth to recombination. In the following sections, we use equations and figures to describe how the population of the genotypes change throughout each growth period in both the iteroparous and semelparous regimes.

5.2.2.2 Iteroparous regime (Multiple rounds of sex)

In this section, we analytically investigate the individual optimal timing of recombination $t_R^{(I)}$ for the iteroparous regime. In this regime, organisms can engage in multiple rounds of sex during a growth period. The equations describing the population dynamics within a growth period are given below.

Dynamics within a growth period

The population dynamics of the system during the logistic growth phase ($0 < t < t_R$) is given by

$$\begin{aligned}
\frac{\mathrm{d}x_{00}}{\mathrm{d}t} &= rx_{00}\left(1 - \frac{x_{\text{total}}}{k}\right) - 2\mu x_{00} \\
\frac{\mathrm{d}x_{01}}{\mathrm{d}t} &= rx_{01}\left(1 - \frac{x_{\text{total}}}{k}\right) + \mu x_{00} - \mu x_{01} \\
\frac{\mathrm{d}x_{10}}{\mathrm{d}t} &= rx_{10}\left(1 - \frac{x_{\text{total}}}{k}\right) + \mu x_{00} - \mu x_{10} \\
\frac{\mathrm{d}x_{11}}{\mathrm{d}t} &= \mu(x_{01} + x_{10})
\end{aligned} \tag{5.6}$$

where $x_{00}(t)$, $x_{01}(t)$, $x_{10}(t)$ and $x_{11}(t)$ are the populations of the $\binom{0}{0}$, $\binom{0}{1}$, $\binom{1}{0}$ and $\binom{1}{1}$ genotypes respectively and $x_{\text{total}}(t) = x_{00}(t) + x_{01}(t) + x_{10}(t) + x_{11}(t)$. Deleterious mutations result in the buildup of lethal genotypes $\binom{1}{1}$ throughout the logistic growth phase. We keep explicit track of $x_{11}(t)$ as resources have been used to produce such offspring, limiting population growth (see Appendix 5.5). Note in Eq. (5.6) that we assume that single mutants have the same growth rate r as the non mutants.

Now we describe the population dynamics for the recombination phase ($t_R < t < T$). Recombination involves crossover events where the alleles in one of the two loci gets swapped between the pair of population members engaging in sexual encounter (see Figure 5.6). Sexual encounters are assumed to occur at rate α, and the probability that a crossover event occurs in at least one locus during a sexual encounter is P_R. With probability $1 - P_R$, the offspring is identical to the parent.

As mentioned in the introduction, sex incurs costs which limit population growth [62]. These costs include the need to go through the time consuming process of meiosis [131] and reduced fecundity [131, 139]. *Daphnia* is an extreme example where sexual production reduces fecundity; their sexual clutches can only contain at most 2 eggs while asexual clutches can contain 110 eggs [63]. It was noted by [47, 63, 193] that such demographic costs of sex can be reduced as the population of asexuals approach carrying capacity. As in our model for hibernation, we account for these costs by assuming that population growth stops once it enters the recombination phase. The population dynamics for the recombination phase ($t_R < t < T$) is given

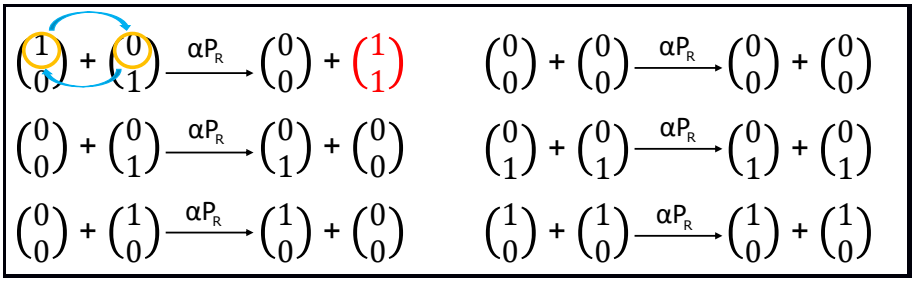

Figure 5.6: During genetic recombination, crossover events occur whereby the alleles in one of the two loci gets exchanged between the two population members undergoing sexual encounter. A successful crossover between a $\binom{1}{0}$ and $\binom{0}{1}$ genotype produces one $\binom{0}{0}$ genotype and one lethal genotype $\binom{1}{1}$ (red genotype). A successful crossover event between all other pairs of viable genotypes does not change the population composition of any genotype. Sexual encounters occur at rate α, and the probability that a crossover occurs upon each sexual encounter is P_R.

by

$$\begin{aligned} \frac{\mathrm{d}x_{00}}{\mathrm{d}t} &= \alpha P_R x_{10} x_{01} \\ \frac{\mathrm{d}x_{01}}{\mathrm{d}t} &= -\alpha P_R x_{01} x_{10} \\ \frac{\mathrm{d}x_{10}}{\mathrm{d}t} &= -\alpha P_R x_{01} x_{10} \\ \frac{\mathrm{d}x_{11}}{\mathrm{d}t} &= \alpha P_R x_{10} x_{01} \,. \end{aligned} \tag{5.7}$$

We note that although the lethal genotype $\binom{1}{1}$ cannot engage in sex, they can still be a product of recombination between the $\binom{1}{0}$ and $\binom{0}{1}$ genotypes (see Figure 5.6), which justifies how their population can change throughout the recombination phase. The population of $\binom{0}{0}$ remains unchanged following recombination with any genotype (see Figure 5.6), but since they are also a product of recombination between $\binom{1}{0}$ and $\binom{0}{1}$, their population can still increase over the recombination phase (see Eq. (5.7) and Figure 5.6).

In Appendix 5.7.1.2, we derive an analytical approximation for the population dynamics of each genotype $x_{00}(t)$, $x_{01}(t)$, $x_{10}(t)$ and $x_{11}(t)$ within a growth period (approximate solutions to Eq. (5.6-5.7)). These approximations assume that the rate of deleterious mutation μ is small (i.e of order 10^{-2} or less). In Appendix 5.7.4, we show how these analytical approximations break down for sufficiently large μ.

Below, we provide plots that show how the population of each genotype changes within a growth period. We see in Figure 5.7 (a) that if recombination does not occur, non mutant genotypes $\binom{0}{0}$ deplete as the lethal double mutants build up over the course of the growth period. If recombination occurs during the growth period (Figure 5.7 (b)), the population can reverse the depletion of non mutant genotypes, increasing its population at time T.

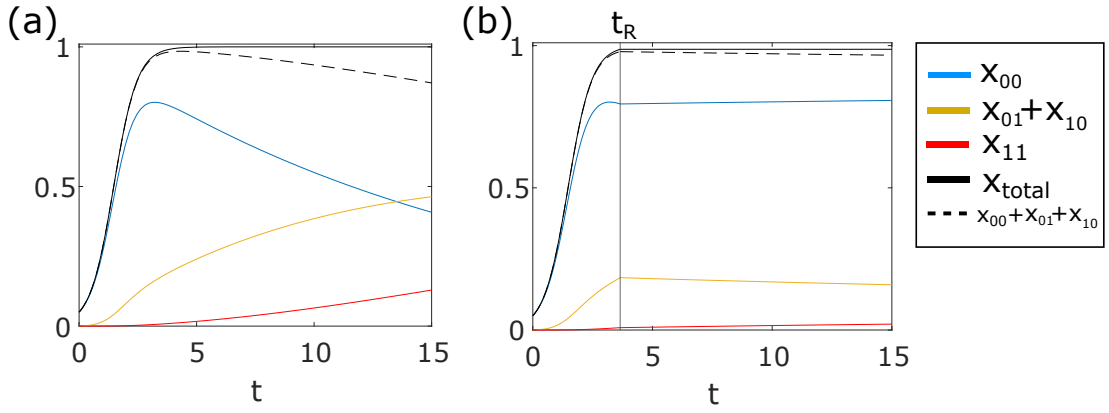

Figure 5.7: Population dynamics within a growth period. The population composition of each genotype changes within a growth period due to deleterious mutations. Panel (a) is the case where recombination does not occur and panel (b) is the case where recombination occurs at $t_R = 3.65$. These plots correspond to numerical solutions to Eqs. (5.6-5.7). Remaining parameters are given in Appendix 5.9. The total population at any given time is denoted as $x_{\text{total}}(t) = x_{00}(t) + x_{01}(t) + x_{10}(t) + x_{11}(t)$.

If the population engages in recombination too early in the growth period, it gives up too much opportunity for population growth, whereas if it engages in recombination too late in the growth period, the number of viable genotypes will have depleted by too much due to the buildup of deleterious mutations. The best time to attempt recombination is some time during the middle of the growth period, which we show analytically in Section 5.3.2.

Although the analytical approximations of $x_{00}(t)$, $x_{01}(t)$, $x_{10}(t)$ and $x_{11}(t)$ (see Eq. (5.45) in Appendix 5.7) enable us to calculate $t_R^{(I)}$, we note that due to the difference in the fitness of each genotype in Eqs. (5.6-5.7), a mutation selection balance will be reached after multiple growth periods, which we demonstrate in the following section.

Mutation Selection Balance in the Genotypes

In our model, mutation selection balance is characterised by the state where the frequency of each genotype at the beginning of each growth period stops changing over multiple growth periods (equilibriates). We demonstrate this in Figure 5.8 below.

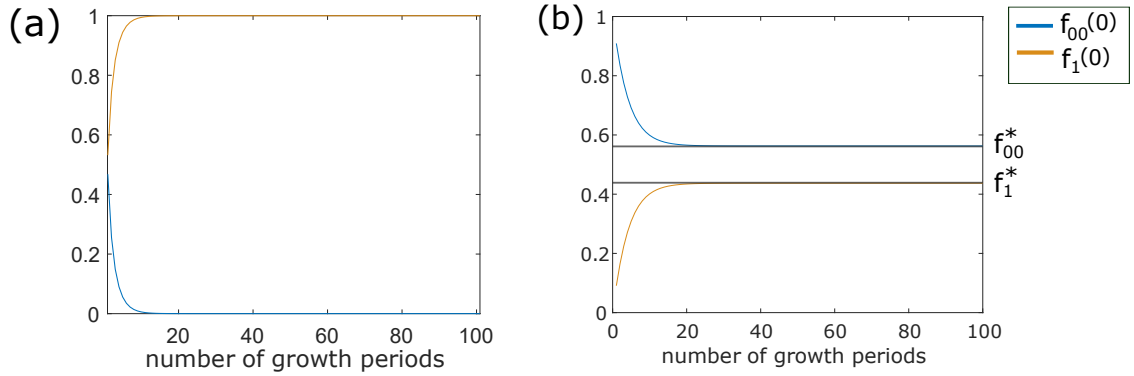

Figure 5.8: Change in the frequency of non mutant genotypes $\binom{0}{0}$ and single mutant genotypes ($\binom{0}{1}$ and $\binom{1}{0}$) at the beginning of a growth period ($f_{00}(0)$ and $f_1(0)$ respectively) over 100 growth periods for Eqs. (5.6-5.7). In panel (a) there is no recombination $t_R \geq T$ and in panel (b) there is recombination $t_R < T$. Upon reaching mutation selection balance, $f_{00}(0)$ and $f_1(0)$ tend towards an equilibrium (f^*_{00} and f^*_1 respectively). Black lines in panel (b) are analytical predictions for the equilibrium frequencies of both genotypes f^*_{00} and f^*_1, (see Eq. (5.48) in Appendix 5.7). Remaining parameters are given in Appendix 5.9.

In Figure 5.8 we see that over a large number of growth periods, the frequency of non mutant and single mutant genotypes at the beginning of each growth period ($f_{00}(0)$ and $f_1(0)$ respectively) tends to an equilibrium, which we define as the "equilibrium frequency". We denote the equilibrium frequencies of non mutants and single mutants as f^*_{00} and f^*_1 respectively.

If recombination does not occur (Figure 5.8 (a)), $f_{00}(0)$ tends to 0 and $f_1(0)$ tends to 1 after a large number of growth periods. If recombination occurs (Figure 5.8 (b)), $f_{00}(0)$ and $f_1(0)$ would both tend to a value between 0 and 1 (see Eq. (5.48) in Appendix 5.7 for analytical expression of this equilibrium frequency). Since the double mutant is lethal, their frequencies are always 0 at the beginning of each growth period, and we have the condition that $f_{00}(0) + f_1(0) = 1$. Note also that

if the growth period T or sexual encounter rate α becomes infinitely large, the f_{00}^* corresponding to Eqs. (5.6-5.7) tends to 1 (see Eq. (5.48) in Appendix 5.7 and Figure 5.21). However, this is not true for the case where recombination occurs once (semelparous regime), which we talk about in Section 5.2.2.3.

Using the analytical approximation for f_{00}^* (see Eq. (5.48) for the "iteroparous" regime and Eq. (5.62) in Appendix 5.7 for the "semelparous" regime), we obtain an analytical expression for the individual optimal timing of recombination $t_R^{(I)}$.

Given the lengthy analytical expressions of f_{00}^* (Eqs. (5.48) and (5.62) in Appendix 5.7) and the large number of genotypes in this complex model, analytical expressions for $x_{00}(t_R)$, $x_{00}(\hat{t}_R)$ (where $\hat{t}_R$ is the recombination timing of a mutant) which are necessary for analytically predicting the evolutionary dynamics would be even lengthier. It would thus be of little practical use to obtain analytical predictions for the evolutionary dynamics of t_R. However, a convenient way to acquire insights into the evolutionary dynamics of t_R is through stochastic simulations of its evolutionary trajectories (see Appendix 5.7.3). In Section 5.2.2.3 below, we consider the semelparous regime, where each individual can attempt recombination once in a growth period.

5.2.2.3 Semelparous regime (one round of sex)

Here, we consider the case where population members only have one opportunity to engage in sex (semelparous). In this case, an inherent cost of sex is manifested through the risk of unsuccessful recombination (occurring with probability $1 - P_R$), as once a population member has engaged in sex, it cannot engage in further sexual encounters. As sexual encounters occur between random members of the population, there is a chance that recombination occurs between a pair of genotypes such that even a successful recombination will not increase the population of the $\binom{0}{0}$ genotype, which further accentuates the cost of sex. An example of this is when $\binom{0}{1}$ recombines with $\binom{0}{1}$ to produce $\binom{0}{1}$ and $\binom{0}{1}$ (see Figure 5.9), which can undergo no further sexual encounters. Out of all the possible recombination events, it is only the successful recombination between $\binom{1}{0}$ and $\binom{0}{1}$ that can reverse the depletion of $\binom{0}{0}$ genotypes due to deleterious mutations.

Dynamics Within a growth period

In the semelparous regime, the dynamics of the population before it enters recombination is identical to that of the iteroparous regime (Section 5.2.2.2). In this

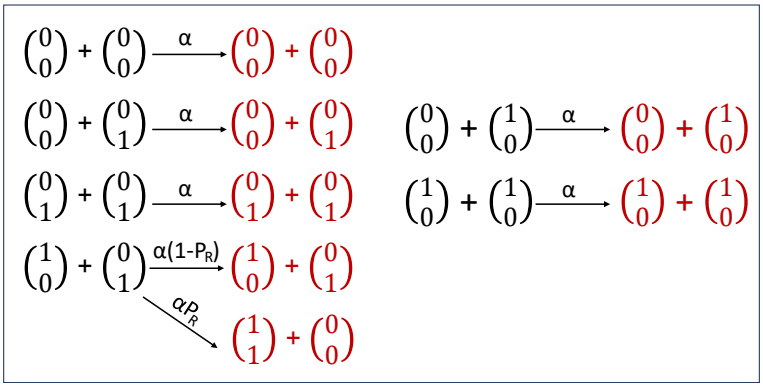

Figure 5.9: Schematic of the genetic recombination process in the semelparous regime. The difference between the iteroparous regime in Figure 5.6 is that once a sexual encounter has occurred, no further sexual encounters occur. The products of recombination shown in maroon cannot undergo further recombination.

section, we only consider the population dynamics of the recombination phase (i.e at $t_R < t < T$), which is given by Eq. (5.8) below. Here, we adopt separate notations for the population that have engaged in sex (attempted recombination) and the population that have not engaged in sex (not attempted recombination). Denoting by χ_{00}, χ_1 and χ_{11} as the population of non mutants, single mutants and double mutants that have not attempted recombination and y_{00}, y_1 and y_{11} as the corresponding populations that have attempted recombination, the population dynamics are given by

$$
\begin{aligned}
\frac{\mathrm{d}\chi_{00}}{\mathrm{d}t} &= -\alpha\chi_{00}(\chi_{01}+\chi_{10}) - \alpha\chi_{00}^2 \\
\frac{\mathrm{d}\chi_{01}}{\mathrm{d}t} &= -\alpha\chi_{01}(\chi_{10}+\chi_{01}+\chi_{00}) \\
\frac{\mathrm{d}\chi_{10}}{\mathrm{d}t} &= -\alpha\chi_{10}(\chi_{01}+\chi_{10}+\chi_{00}) \\
\frac{\mathrm{d}\chi_{11}}{\mathrm{d}t} &= 0 \\
\frac{\mathrm{d}y_{00}}{\mathrm{d}t} &= \alpha P_R\chi_{10}\chi_{01} + \alpha\chi_{00}(\chi_{10}+\chi_{01}) + \alpha\chi_{00}^2 \\
\frac{\mathrm{d}y_{01}}{\mathrm{d}t} &= \alpha\chi_{00}\chi_{01} + \alpha(1-P_R)\chi_{01}\chi_{10} + \alpha\chi_{01}^2 \\
\frac{\mathrm{d}y_{10}}{\mathrm{d}t} &= \alpha\chi_{00}\chi_{10} + \alpha(1-P_R)\chi_{01}\chi_{10} + \alpha x_{10}^2 \\
\frac{\mathrm{d}y_{11}}{\mathrm{d}t} &= \alpha P_R\chi_{10}\chi_{01}\,.
\end{aligned}
\tag{5.8}
$$

As we can observe, there is a remarkable difference between Eq. (5.8) and the population dynamics of the recombination phase in the iteroparous regime Eq. (5.7) (see Figure 5.14 in Section 5.3.2.2 for illustration). In this semelparous regime, the frequency of each genotype also reaches a mutation selection balance over many generations as in the iteroparous regime (see Figure 5.8 and Eq. (5.62) in Appendix 5.7). Due to the conceptual similarity between the iteroparous regime, a figure showing the mutation selection balance for this regime is omitted.

5.2.2.4 Simulations of evolutionary trajectories

In our analyses above, we have only accounted for the optimal timing of recombination for a monomorphic population, where all members enter the recombination phase at the same time. When the timing of recombination t_R is an evolving trait, the population will contain multiple genotypes that enter recombination at different times during the growth period (i.e with different values of t_R). In such a case, the genotypes with different values of t_R can have sexual encounters with each other and the single mutants $\binom{1}{0}$ and $\binom{0}{1}$ become public goods. For instance, a $\binom{1}{0}$ from an individual that enter recombination at $t_R = t_R^{(1)}$ can recombine with a $\binom{0}{1}$ from an individual with $t_R = t_R^{(2)}$, to reverse the depletion of non-mutants $\binom{0}{0}$ in the population of each other. This can give rise to phenomena such as cheating, whereby one genotype (the cheater) can reap the benefits of another genotype (the cooperator) without paying a reasonable cost. In our model, this can occur when a mutant with a different t_R to the resident genotype enters the population at a small frequency (see Appendix 5.8). Such phenomena can have an effect on the subsequent evolutionary dynamics for t_R. In Figure 5.10, we show different ways in which genotypes with different values of t_R can interact with each other in a two genotype population.

The stochastic simulations for the iteroparous regime are implemented according to Eqs. (5.52-5.53) in Appendix 5.7 and the corresponding simulations for the semelparous regime are implemented according to Eqs. (5.65-5.66) in Appendix 5.7).

5.3 RESULTS

In this section, we present our analytical and numerical results for the hibernation model and recombination model. We begin by considering the hibernation model. In Section 5.3.1, we provide analytical results and numerical simulations for the

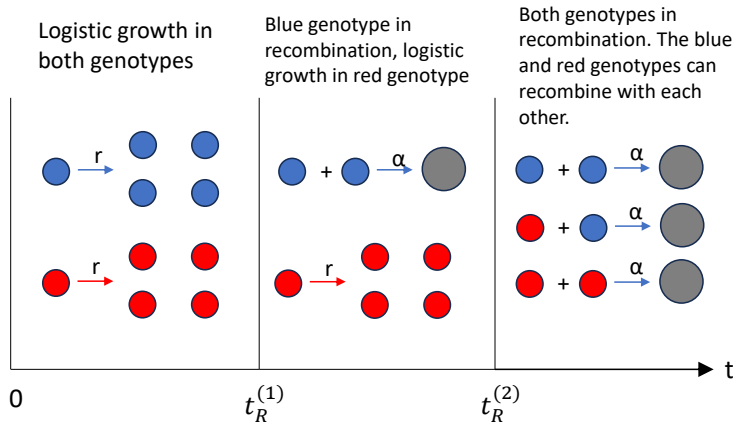

Figure 5.10: When the population contains two genotypes with different values of t_R, there are times when both genotypes are under logistic growth ($0 < t < t_R^{(1)}$), one under logistic growth and one in recombination phase ($t_R^{(1)} < t < t_R^{(2)}$), or both in recombination phase ($t_R^{(2)} < t < T$). When both genotypes are in the recombination phase, sexual encounters can occur between members of the other genotype.

evolutionary dynamics of the timing of hibernation. In Section 5.3.2, we present our analytical results for the individual optimal timing of recombination along with numerical simulations for the evolutionary dynamics of the timing of recombination for both the iteroparous and semelparous regime.

5.3.1 Environmental Stress Induces Hibernation

In Figure 5.11, we provide stochastic simulations for the evolutionary dynamics of the timing of hibernation t_H (bottom row) alongside their corresponding population dynamics within each growth period (top row). In the top panels of Figure 5.11, we see that as expected, the evolutionarily optimal timing of hibernation occurs during the middle of a growth period, where both the opportunity cost of hibernation is sufficiently low while ensuring that a sufficient population ends up in hibernation at time T. In Figure 5.11, we also see that for shorter growth periods, the population evolves to enter hibernation phase earlier to allow time for population members to enter hibernation. In Appendix 5.6.6, we show that the fixed point of Eq. (5.5) (red lines in Figure 5.11 top) is indeed an evolutionarily stable state for the parameters in Figure 5.11, which justifies why we do not observe evolutionary branching.

Having provided simulations of evolutionary trajectories for t_H, we are now interested in how the evolution of hibernation is influenced by various system

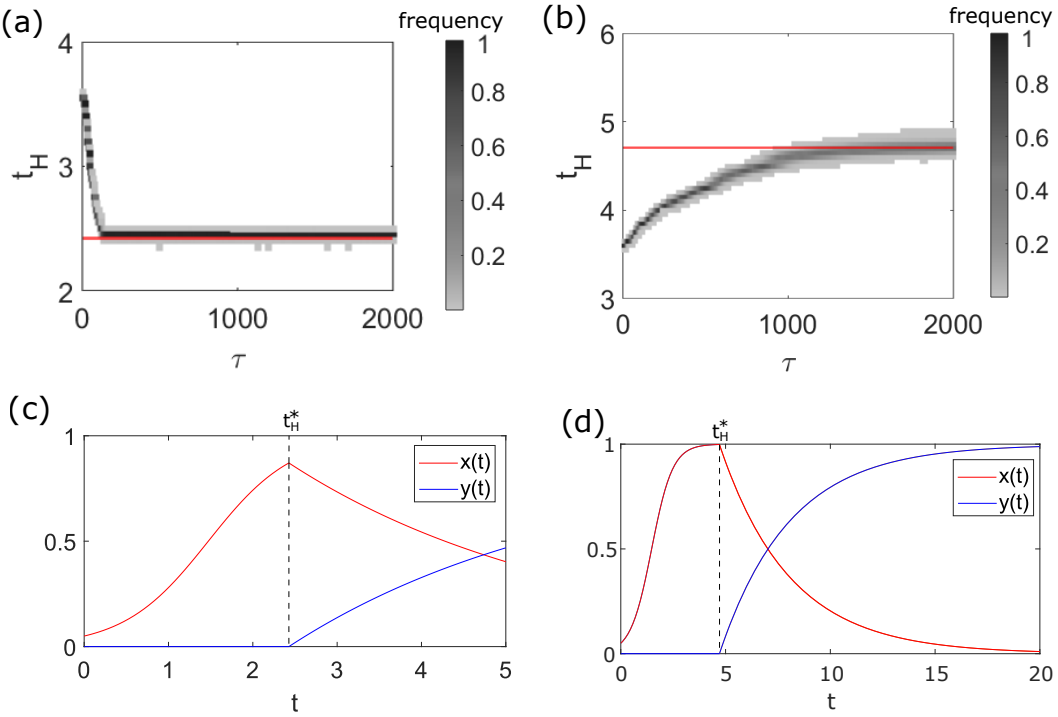

Figure 5.11: Top: Stochastic simulations of the evolutionary trajectories $t_{H(i)}(\tau)$ for each i-th genotype in the population. Darkness of the trajectories represents the frequency of a given genotype $t_{H(i)}$ in the population and red lines correspond to the analytical prediction of the stable fixed point of Eq. (5.5). Panel (a) is for the case where $T = 5$ and panel (b) for the case where $T = 20$. Bottom: Population dynamics within a growth period (Eqs. (5.1-5.2)) . Panels (c) and (d) correspond to the parameters in panels (a) and (b) respectively. Hibernation occurs at their evolutionary optimal timing of hibernation t_H^*, with $t_H^* \approx 2.42$ in panel (c) and $t_H^* \approx 4.71$ in panel (d). Remaining parameters are given in Appendix 5.9.

parameters. In Figure 5.12 below, we show how T and $x_{\text{total}}(0)$ affects the fraction of the population in hibernation at time T. We see that the fraction of the population that end up in hibernation at time T increases with both T and $x_{\text{total}}(0)$. This is because for small T, the population needs to enter hibernation phase sooner while opportunity costs are still high, in order to allow time for population members to enter hibernation.

We also note that should $x_{\text{total}}(0)$ be too large or T be sufficiently small in our model, the opportunity for population growth becomes so limited that the population would be better off residing in the hibernation phase throughout the growth period (black regions in Figure 5.12). We stress that this parameter regime is

not of significant biological relevance, since it either represents an extremely resource limited population at the beginning of a growth period, or a very high R_{crit} (see Section 5.5).

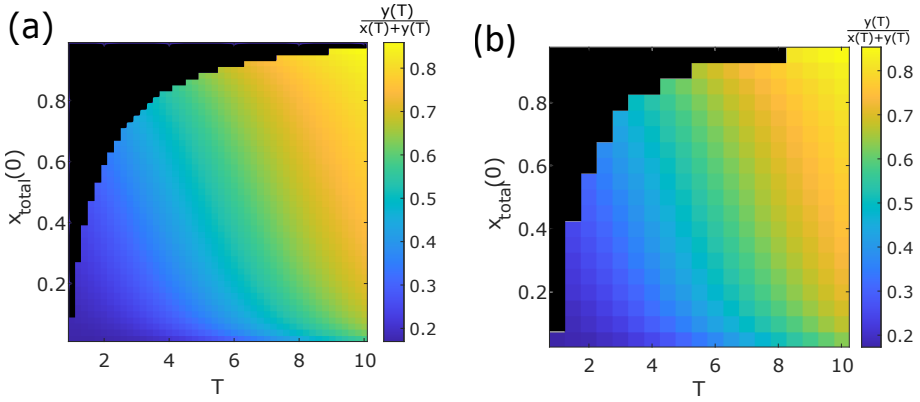

Figure 5.12: Heatmap showing the analytical prediction (panel (a)) and numerical simulation (panel (b)) of the fraction of population members in hibernation as a function of growth period T and the initial population $x_{\text{total}}(0)$. In the black region, $t_H^* \leq 0$, which represents the case where it is ideal for the population to reside in hibernation throughout the growth period. The heatmap in panel (a) are generated using Eq. (5.16) in Appendix 5.6 and in panel (b), numerical simulations are obtained by simulating evolutionary trajectories for t_H over a large number of mutation events (see Appendix 5.6.7). Remaining parameters are given in Appendix 5.9.

5.3.2 Environmental Stress Induces Genetic Recombination

In this section, we begin by presenting analytical results on the individual optimal timing of recombination for the iteroparous regime Eqs. (5.6-5.7). We then investigate the parameter regions where recombination occurs within a growth period, which relates to our original question of the conditions under which we evolve environmentally triggered sex. Afterwards, we present the corresponding results for the semelparous regime Eq. (5.8). In Section 5.3.3, we present results for our simulations of evolutionary trajectories in t_R to provide insights into the evolutionary dynamics for the timing of sex. Interestingly these simulations reveal a rich array of dynamical behaviour that depends on the rate of sexual encounter α.

5.3.2.1 Iteroparous regime (many rounds of sex)

In the iteroparous regime Eqs. (5.6-5.7), the individual optimal timing of recombination is given by

$$t_R^{(I)} = \frac{1}{r}\log\left(\frac{(r-\mu)(k-x_{\text{total}}(0))}{\mu x_{\text{total}}(0)}\right) \tag{5.9}$$

(see Appendix 5.7.1.2 for derivation) and the condition under which recombination occurs during the growth period is $t_R^{(I)} < T$. Notice that $t_R^{(I)}$ in Eq. (5.9) depends on r, μ and $x_{\text{total}}(0)$ but is independent of T, so it would be crucial to examine the values of T that result in recombination occurring during the growth period. It is also interesting to examine how varying the other parameters in Eq. (5.9) impacts the value of T that results in recombination during the growth period. In Figure 5.13, we present a heatmap to show the regions in the parameter space of T and $x_{\text{total}}(0)$ where recombination occurs.

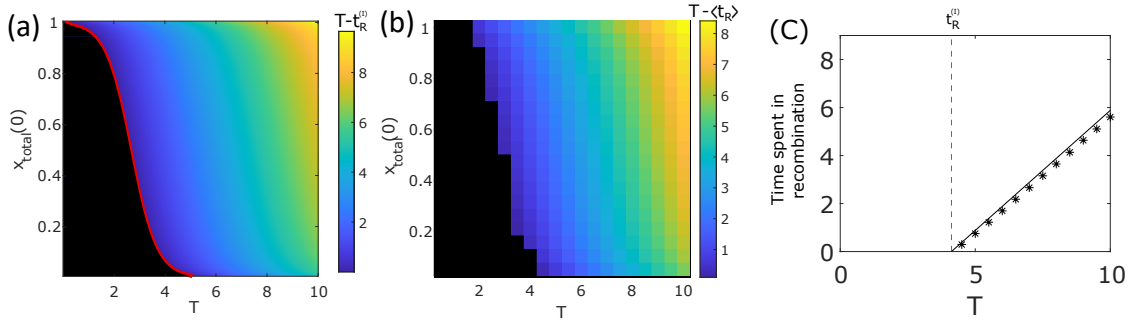

Figure 5.13: Panel (a) is a heatmap showing how the analytical approximation of $t_R^{(I)}$ varies with $x_{\text{total}}(0)$ and T. Red curve represents $t_R^{(I)}$. Its colourmap indicates the value of $T - t_R^{(I)}$. Panel (b) is the corresponding numerical result for panel (a), obtained by simulating evolutionary trajectories for t_R as described in Section 5.2.2.4 (see also Eq. (5.53) in Appendix 5.7.3). The colourmap represents the quantity $T - \langle t_R \rangle$ where $\langle t_R \rangle$ is the mean t_R in a multigenotype population (see Appendix 5.7.3). Panel (c) is the time spent in recombination as a function of T for case where $x_{\text{total}}(0) = 0.05$. Solid black line is the analytical prediction from Eq. (5.9) and markers are numerical simulations taken from panel (b). The linear increase with T of the time in recombination shows that $t_R^{(I)}$ does not vary with T for a given value of $x_{\text{total}}(0)$. Remaining parameters are given in Appendix 5.9. The parameters for this figure corresponds to the case of small α, where the evolutionary optimal t_R is well predicted by the individual optimal t_R (this is discussed more in Section 5.3.3).

From Figure 5.13, we notice that as long as the growth period is sufficiently long (i.e T sufficiently large), environmentally triggered sex is expected to occur. We also

see that the individual optimal timing of recombination can be an accurate predictor of the evolutionary optimal timing of sex, due to the close match between panels (a) and (b). However, this accuracy breaks down under certain parameter conditions which we discuss in Section 5.3.3.

5.3.2.2 Semelparous regime (one round of sex)

Surprisingly, in the semelparous regime Eq. (5.8), the individual optimal timing of recombination is given by exactly the same expression as the iteroparous regime Eq. (5.9) (see Appendix 5.7.2.2 for derivation). The parameter regions in T and $x_{\text{total}}(0)$ under which environmentally triggered sex is present is thus exactly identical to the iteroparous regime. However, the difference between these two regimes lies in the population dynamics during the recombination phase. In other words, the way in which the population composition of each genotype changes once the population enters recombination differs between the two regimes (compare Eq. (5.7) with (5.8) and see Figure 5.14). This in turn impacts the equilibrium frequencies of $\binom{0}{0}$. Here, unlike in the iteroparous regime, the frequency of $\binom{0}{0}$ does not tend to 1 as α or T tends to infinity, but tends to some value between 0 and 1 (see Eq. (5.62) in Appendix 5.7). We justify this difference using Figure 5.14 in Section 5.3.2.3 below.

5.3.2.3 Difference in population dynamics between these two regimes

As well as the population composition, there is also a qualitative difference in how the population of each genotype changes once the population enters recombination (see Figure 5.14). In the semelparous regime, the population of all genotypes eventually level off at some value above zero after every individual has attempted recombination (Figure 5.14 (b)), whereas in the iteroparous regime, the population of $\binom{0}{1}$ and $\binom{1}{0}$ genotypes will continually decay until it reaches zero (Figure 5.14 (a)).

When genotypes are at their equilibrium frequencies ($f_{00}(0) = f_{00}^*$ and $f_1(0) = f_1^*$), the population of each genotype at time T can also be evaluated analytically. This is done using the following step. For the iteroparous regime, we substitute f_{00}^* given by Eq. (5.48) into the expressions for $x_{00}(t)|_{t=T}$ and $x_1(t)|_{t=T}$ in Eq. (5.45) in Appendix 5.7, and for the semelparous regime, we substitute f_{00}^* given by Eq. (5.62) into the expressions for $(\chi_{00}(t) + y_{00}(t))\,|_{t=T}$ and $(\chi_1(t) + y_1(t))\,|_{t=T}$ in Eq. (5.60) in Appendix 5.7. This is indeed what we did to arrive at Eq. (5.9). Using these, we can determine what happens to the population of each genotype at time T in the limit as α tends to infinity (open circles in Figure 5.14). For the iteroparous regime,

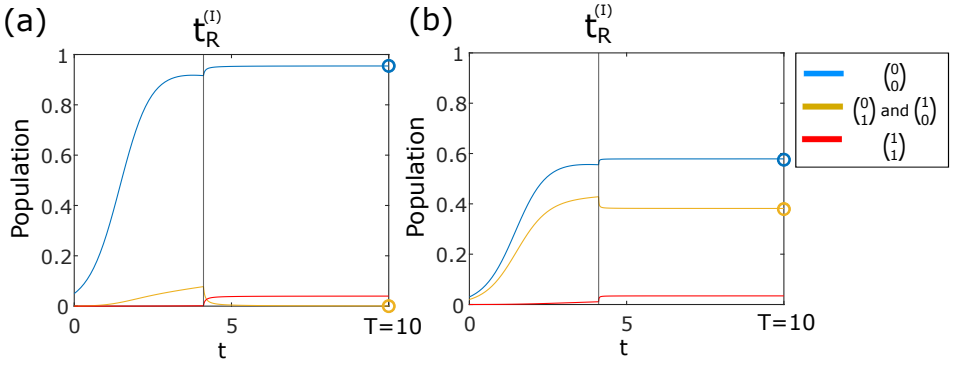

Figure 5.14: The change in population composition of each genotype following recombination differs between the iteroparous regime (panel (a)) and the semelparous regime (panel (b)). All genotypes are at their equilibrium frequencies (i.e $f_{00}(0) = f_{00}^*$ and $f_1(0) = (1 - f_{00}^*)$) in this figure. Open circles represent the populations of each genotype at time T in the limit as α tends to infinity (see Eq. (5.70) in Appendix 5.7). Remaining parameters are given in Appendix 5.9.

$x_1(T)$ tends to 0 and $x_{00}(T)$ tends to $(x_{\text{total}}(T) - x_{11}(T))$, while for the semelparous regime, $(\chi_{00}(T) + y_{00}(T))$ and $(\chi_1(T) + y_1(T))$ both tend to some value above 0.

What this result tells us is that if population members have one opportunity for recombination in a growth period, then even if all members attempt recombination instantaneously (i.e at infinite rate), the population still cannot fully reverse the depletion of the $\binom{0}{0}$ genotypes due to deleterious mutations. This illustrates the cost of sex in this regime due to either recombination failure or recombination occurring between a pair of genotypes which does not increase the population of the $\binom{0}{0}$ genotype (see Figure 5.9 in Section 5.2.2.3). In Section 5.3.3, we see that as well as the population dynamics within a growth period, the evolutionary trajectories for t_R also differs between these two regimes.

5.3.3 EVOLUTIONARY TRAJECTORIES FOR TIMING OF GENETIC RECOMBINATION REVEALS RICHER DYNAMICS

In this section, we present results for our stochastic simulations of evolutionary trajectories in the "semelparous" and "iteroparous" regimes. We show that in both regimes, evolutionary trajectories reveal a rich array of dynamical behaviour, which depends on the value of α. We begin by interpreting our simulation results for the iteroparous regime. When the encounter rate α is small (Figure 5.15 (a)), evolution

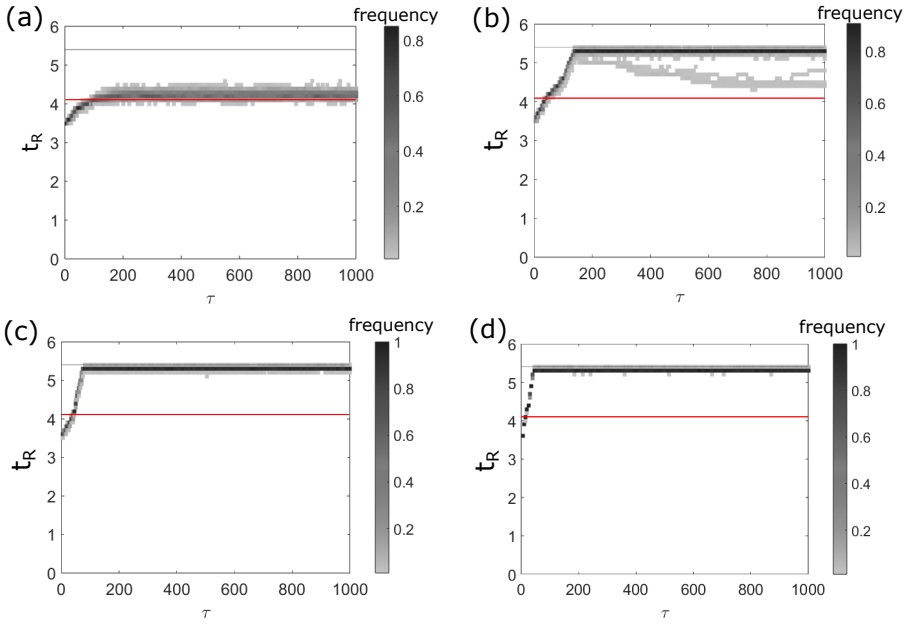

Figure 5.15: Stochastic simulations of evolutionary trajectories for t_R in the iteroparous regime (see Eq. (5.53) and Appendix 5.7.3 for implementation). In panel (a), $\alpha = 0.03$ (order 10^{-2}), panel (b) $\alpha = 3$ (order 10^0), panel (c) $\alpha = 100$ (order 10^2) and panel (d) $\alpha = 10000$ (order 10^4). Darkness of the trajectories represents the frequency of a given genotype $t_{R(i)}$ in the population. The red lines in each panel represents the individual optimal timing of recombination $t_R^{(I)}$ given by Eq. (5.9). Remaining parameters are given in Appendix 5.9.

gives rise to a population comprising one phenotypic cluster of genotypes, where the mean population trait value of t_R is well approximated by the individual optimal timing of recombination $t_R^{(I)}$ (see Eq. (5.9)). When α is of the order 10^0 (i.e a few sexual encounters occur per unit time, e.g. day), the population initially evolves to a trait value approximately equal to the individual optimal timing of recombination before undergoing evolutionary branching, with one cluster evolving to a large value of t_R (Figure 5.15 (b)). Next, when α is of the order 10^2 (i.e a hundred or so sexual encounters occur per unit time), evolution gives rise to a population comprising one cluster of genotypes that delays recombination until right before the end of the growth period (Figure 5.15 (c)). For even higher values of α (Figure 5.15 (d)), the dynamics remains the same as in (Figure 5.15 (c)).

Now we consider the evolutionary dynamics for the semelparous regime. In this regime, we see that when α is small (i.e if a small population engages in recombination during the growth period), evolution again gives rise to a population comprising one

cluster of genotypes, where the mean population trait value of t_R is well approximated by the individual optimal timing of recombination Eq. (5.9) (Figure 5.16 (a)). In the case where α is of the order 10^0, the population initially evolves to delay t_R until the end of the growth period (time T), before it undergoes branching in t_R, with one cluster of low frequency evolving to a smaller value of t_R (Figure 5.16 (b)).

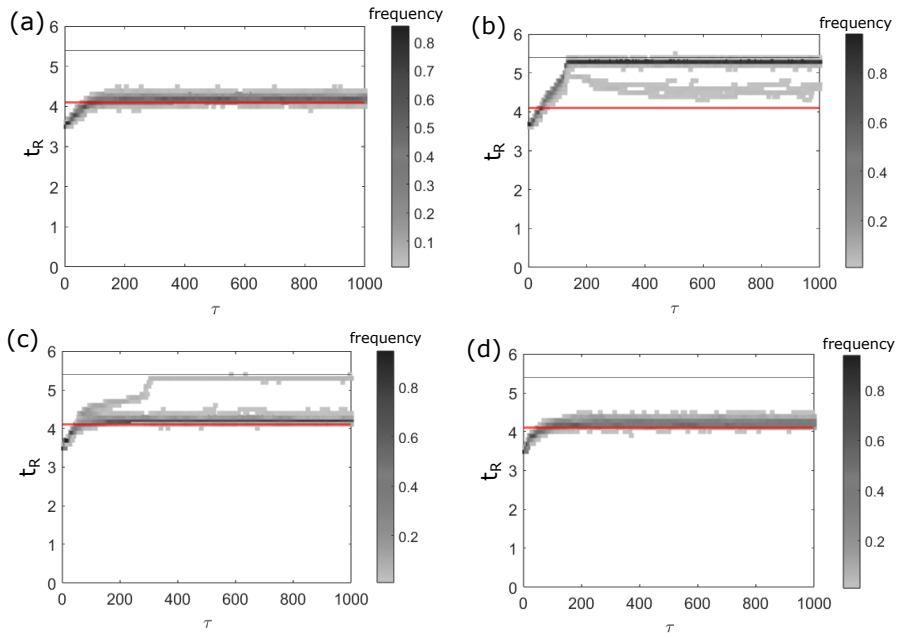

Figure 5.16: Stochastic simulations of evolutionary trajectories for t_R in the semelparous regime Eq. (5.8) (see Eq. (5.66) and Appendix 5.7.3 for implementation). In panel (a) $\alpha = 0.03$ (order 10^{-2}), panel (b) $\alpha = 7$ (order 10^0), panel (c) $\alpha = 200$ (order 10^2) and panel (d) $\alpha = 15000$ (order 10^4). All other parameters are as in Figure 5.15. Red lines in each panel represents the individual optimal timing of recombination $t_R^{(I)}$ given by Eq. (5.9). Remaining parameters are given in Appendix 5.9.

When α is of the order 10^2, the evolutionary dynamics behave differently to the iteroparous regime. Instead of delaying recombination until time T, the population initially evolves to a trait value approximately equal to the individual optimal timing of recombination before it undergoes branching, with one cluster of low frequency evolving to a large value of t_R (Figure 5.16 (c)). Interestingly, this resembles the dynamics in Figure 5.16 (b) but in a qualitatively reversed manner. In the case where α is of the order 10^4, the population evolves one cluster of genotypes which is well approximated by $t_R^{(I)}$, as in the case of small α.

As expected, the evolutionary dynamics differ between the semelparous and

iteroparous regimes for large values of α where the sexual encounter rate ceases to be a limiting factor. For the semelparous regime, the population evolves to enter recombination earlier than the iteroparous regime (compare panel (d) of Figures 5.15 and 5.16). This is likely due to the greater cost of recombination in the semelparous regime.

5.4 DISCUSSION

The main take home message of this chapter is that we have provided a novel explanation for the evolution of environmentally triggered sex using an eco-evolutionary model that incorporates a population genetics framework (Muller's Ratchet [154]), alongside biologically realistic population dynamics. In particular, we have shown that the optimal timing of hibernation broadly coincides with that of recombination within a growth period. As hibernation and recombination are both subject to the same selective pressure, and evolve when opportunities for population growth become limited, our model explains how sex and hibernation may occur in tandem as a response to the same environmental trigger. This behaviour is observed frequently in organisms such as *C. reinhardtii* and *S. pombe* [197] when subject to nitrogen limitation.

Through performing stochastic simulations of the evolutionary dynamics of the timing of recombination, we have accounted for eco-evolutionary aspects, which are key components in the evolution of sexual reproduction. These include the aspect of cheating present in sexual reproduction [230]. Accounting for these eco-evolutionary aspects enabled us to uncover a broader range of dynamical behaviour that occur on the evolutionary timescale. Our simulations reveal that when members of the population are presented with one opportunity for sex (semelparous), they may evolve to invest in sex earlier during the growth period than when they have multiple opportunities for sex (iteroparous).

Our work follows on from that of [61], by considering the selective advantages of sex explicitly in terms of its genetic benefits, whilst also providing analytical expressions that explain our results, made possible with our eco-evolutionary model. While their model was tailored specifically to the breeding cycle of *daphnia*, our model considered both semelparous and iteroparous breeding cycles, encompassing the behaviour of a range of organisms.

One limitation of our model of hibernation is that it does not explain the diverse modes of hibernation observed in different organisms. Hibernation in *C. reinhardtii* occurs through zygospore formation which is preceded by the fertilisation of gametes, and *S. pombe* [167] hibernate by forming dormant spores following the production of daughter cells in meiosis. We simply assumed that members of the population enter hibernation exponentially at a constant rate. Existing eco-evolutionary models that account for cell fusion include [37, 138], which investigated the evolution of anisogamy by assuming mass dependent survival of cells. Importantly, our results have managed to shed light on the empirical observation of why sex is correlated with the occurrence of hibernation [62].

Another feature not assumed in our models is the presence of self-incompatible mating types, which may impact reproductive success in various ways. In *C. reinhardtii* [137], gamete specific MID (minus dominant) genes are expressed in response to nitrogen limitation, which differentiates vegetative cells into mating types MT^+ and MT^- on their MT locus. Intriguingly, *S. pombe* even have the ability to switch mating types upon genetic recombination due to the genes on their MAT-1 locus [233]. It is observed in *daphina* that when there is a biased sex ratio, mate limitation can induce sexual selection [59] and reduced reproductive success. They showed experimentally that populations with biased female to male sex ratios resulted in fewer eggs per ephippum, indicating low reproductive success.

Although our results successfully explain how environmentally triggered facultative sex can occur under different parameter regions of our model, the precise details of our results differ under different modelling scenarios. For instance, if the timing of recombination is an evolving parameter, social conflict can give rise to evolutionary branching, which can delay the population mean timing of recombination to some later time than the optimal timing for an individual genotype. Crucially, both these scenarios imply that environmentally triggered facultative recombination is expected. It is only the timing of such event that differ between these scenarios. This in turn explains how sex tends to occur shortly before the environment becomes harsh in the facultatively sexual organisms discussed in this chapter.

Our work also echoes that of Gerber *et al.* [62], which used an agent based simulation to shown how sex, dormancy and dispersal can covary with each other when they are independently coevolving mechanisms. Their results showed that investment in sex can covary positively with hibernation, but negatively with dispersal in fluctuating environments. Furthermore, [207] investigated how the association

between sex and the formation of dormant survival structures can prevent facultative sexuals from being outcompeted by asexuals, which provides important implications for the maintenance of facultative sex.

As well as theories and empirical observations, environmentally triggered sex and its advantages have been demonstrated experimentally. One experiment showed that individuals of *S. cerevisiae* that underwent genetic recombination had a greater adaptation rate than their asexual counterparts [117]. Statistical analyses on experimental data [195] have shown that investment in sex in filamentous fungus is negatively correlated with the fitness of a genotype in a given environment.

5.5 APPENDIX I: RESOURCE LIMITATION AND LOGISTIC GROWTH

In this section, we describe how the one dimensional logistic growth model relates to resource limitation or saturation in population density. We begin by considering the one species MacArthur consumer-resource model [142] which is given as follows. For simplicity, we assume an absence of species death. Denoting x as the consumer species and R as the resource, the model is given by

$$\begin{aligned} \frac{\mathrm{d}x}{\mathrm{d}t} &= \theta R x \\ \frac{\mathrm{d}R}{\mathrm{d}t} &= -\theta R x \end{aligned} \tag{5.10}$$

where θ is the intrinsic growth rate of the population. Now we demonstrate how Eq. (5.10) is equivalent to the logistic growth model. First, we note that the total population of consumers and resources are constant $k = x + R$, and we define this constant as the carrying capacity. Rearranging in terms of x and substituting into Eq. (5.10), we get

$$\frac{\mathrm{d}x}{\mathrm{d}t} = \theta x(k - x) \tag{5.11}$$

which is of the same functional form as the logistic growth model. Furthermore, we assume that if the resource abundance falls below a critical abundance i.e $R(t) < R_{\text{crit}}$ or if the population density becomes sufficiently large i.e $x(t) > x_{\text{crit}}$, only consumers of sufficient fitness can survive. This is equivalent to running the logistic growth model Eq. (5.11) over a set time period T, and subjecting the population to a round of fitness dependent mortality at time T. The time T is defined as the growth period.

5.6 APPENDIX II: EVOLUTION OF HIBERNATION TIME

5.6.1 DYNAMICS WITHIN A GROWTH PERIOD

The population dynamics of logistic growth phase ($0 < t < t_H$) is given by

$$\begin{aligned} \frac{\mathrm{d}x}{\mathrm{d}t} &= rx\left(1-\frac{x}{k}\right) \\ \frac{\mathrm{d}y}{\mathrm{d}t} &= 0 \end{aligned} \tag{5.12}$$

where $x(t)$ is the non-hibernation population and $y(t)$ is the hibernating population. The population dynamics of the hibernation Phase ($t_H < t < T$) is given by

$$\begin{aligned} \frac{\mathrm{d}x}{\mathrm{d}t} &= -\gamma x \\ \frac{\mathrm{d}y}{\mathrm{d}t} &= \gamma x\,. \end{aligned} \tag{5.13}$$

The solution to Eq. (5.12) is

$$x(t) = \frac{kx_{\mathrm{total}}(0)}{x_{\mathrm{total}}(0) + (k - x_{\mathrm{total}}(0))e^{-rt}} \tag{5.14}$$

and the solution to Eq. (5.13) is

$$\begin{aligned} x(t) &= x(t_H)e^{-\gamma(t-t_H)} \\ y(t) &= x(t_H)(1-e^{-\gamma(t-t_H)}) \end{aligned} \tag{5.15}$$

where $x(t_H)$ is determined by evaluating Eq. (5.14) at $t = t_H$. The fraction of population members that have ended up in hibernation at the end of a growth period is given by

$$\frac{y(t)}{x(t)+y(t)} = 1 - e^{-\gamma(t-t_H)} \tag{5.16}$$

which is independent of $x_{\mathrm{total}}(0)$.

5.6.2 DYNAMICS WITHIN A GROWTH PERIOD IN THE PRESENCE OF A MUTANT

Here, we consider the dynamics within a generation for the case where the population contains a mutant at small frequency. When neither the mutant nor the resident has

entered hibernation phase ($t < t_H < \hat{t}_H < T$), the population dynamics are given by

$$\begin{aligned} \frac{\mathrm{d}x}{\mathrm{d}t} &= rx\Big(1 - \frac{x+\hat{x}}{k}\Big) \\ \frac{\mathrm{d}\hat{x}}{\mathrm{d}t} &= r\hat{x}\Big(1 - \frac{x+\hat{x}}{k}\Big) \\ \frac{\mathrm{d}y}{\mathrm{d}t} &= 0 \\ \frac{\mathrm{d}\hat{y}}{\mathrm{d}t} &= 0 \end{aligned} \tag{5.17}$$

where $x(t)$ and $\hat{x}(t)$ denote the population of residents and mutants not in hibernation respectively, while $y(t)$ and $\hat{y}(t)$ denotes the corresponding population in hibernation. The solutions to Eq. (5.17) are given by

$$\begin{aligned} x(t_H) &= (1 - \hat{f}_{t_H})\left(\frac{x_{\text{total}}(0)e^{rt_H}}{1 + \frac{1}{k}x_{\text{total}}(0)(e^{rt_H} - 1)}\right) \\ \hat{x}(t_H) &= \hat{f}_{t_H}\left(\frac{x_{\text{total}}(0)e^{rt_H}}{1 + \frac{1}{k}x_{\text{total}}(0)(e^{rt_H} - 1)}\right) \\ & \quad y(t) = 0 \quad , \quad \hat{y}(t) = 0 \end{aligned} \tag{5.18}$$

when evaluated at the timing of hibernation $t = t_H$, where $\hat{f}_{t_H}$ is the frequency of the mutant with a different t_H to the resident. If the resident genotype has entered hibernation phase but the mutant is not in hibernation ($t_H < t < \hat{t}_H < T$), the population dynamics become

$$\begin{aligned} \frac{\mathrm{d}x}{\mathrm{d}t} &= -\gamma x \\ \frac{\mathrm{d}\hat{x}}{\mathrm{d}t} &= r\hat{x}\Big(1 - \frac{x+\hat{x}+y}{k}\Big) \\ \frac{\mathrm{d}y}{\mathrm{d}t} &= \gamma x \\ \frac{\mathrm{d}\hat{y}}{\mathrm{d}t} &= 0\,. \end{aligned} \tag{5.19}$$

The solutions to Eq. (5.19) are given by

$$\begin{aligned} x(\hat{t}_H) &= x(t_H)e^{-\gamma(\hat{t}_H - t_H)} \\ \hat{x}(\hat{t}_H) &= \frac{\hat{x}(t_H)(\frac{x(t_H)}{k} - 1)}{e^{r(\frac{x(t_H)}{k} - 1)\delta t_H}(\frac{x(t_H)+\hat{x}(t_H)}{k} - 1) - \frac{\hat{x}(t_H)}{k}} \\ y(\hat{t}_H) &= x(t_H)(1 - e^{-\gamma(\hat{t}_H - t_H)}) \quad , \quad \hat{y}(\hat{t}_H) = 0 \end{aligned} \tag{5.20}$$

when evaluated at the mutant's timing of hibernation $t = \hat{t}_H$, where $\hat{t}_H = t_H + \delta t_H$. Finally when both the mutant and resident have entered hibernation phase ($t_H < \hat{t}_H < t < T$), the population dynamics become

$$\begin{aligned} \frac{\mathrm{d}x}{\mathrm{d}t} &= -\gamma x \\ \frac{\mathrm{d}\hat{x}}{\mathrm{d}t} &= -\gamma \hat{x} \\ \frac{\mathrm{d}y}{\mathrm{d}t} &= \gamma x \\ \frac{\mathrm{d}\hat{y}}{\mathrm{d}t} &= \gamma \hat{x} \end{aligned} \tag{5.21}$$

where the solutions are given by

$$\begin{aligned} x(t) &= x(t_H)e^{-\gamma(t-t_H)} \\ y(t) &= x(t_H)(1-e^{-\gamma(t-t_H)}) \\ \hat{x}(t) &= \hat{x}(\hat{t}_H)e^{-\gamma(t-\hat{t}_H)} \\ \hat{y}(t) &= \hat{x}(\hat{t}_H)(1-e^{-\gamma(t-\hat{t}_H)}) \end{aligned} \tag{5.22}$$

where $x(t_H)$, $x(\hat{t}_H)$ and $\hat{x}(\hat{t}_H)$ are given in Eqs. (5.18) and (5.20). We can then evaluate the expressions in Eq. (1.22) at $t = T$ to obtain expressions for $x(T)$, $\hat{x}(T)$ and $y(T)$ and $\hat{y}(T)$ and thus the population of hibernating members at the end of a growth period, which can be used to define the invasion fitness function in the following section.

5.6.3 Invasion Dynamics

The invasion dynamics is approximated using the invasion fitness function of the mutant, given by

$$\hat{f}'_{t_H} = \frac{\hat{y}(T)}{y(T) + \hat{y}(T)} \tag{5.23}$$

where $\hat{y}(T)$ is the population of mutants that have ended up in hibernation at the end of the growth period T. As in Chapters 3 and 4, we assume that the mutational stepsize parameter δt_H is small to enable us to approximate the invasion dynamics using the formula

$$\frac{\mathrm{d}\hat{f}_{t_H}}{\mathrm{d}t} = \frac{\mathrm{d}}{\mathrm{d}\delta t_H}\left[\frac{\mathrm{d}\hat{f}'_{t_H}}{\mathrm{d}t}\right]\Bigg|_{\delta t_H = 0} + O(\delta t_H^2) \tag{5.24}$$

where

$$\frac{\mathrm{d}\hat{f}'_{t_H}}{\mathrm{d}t} \approx \hat{f}'_{t_H} - \hat{f}_{t_H}. \tag{5.25}$$

Using the steps detailed in this section, we see that the invasion dynamics is given by

$$\frac{\mathrm{d}\hat{f}_{t_H}}{\mathrm{d}t} = \hat{f}_{t_H}(1-\hat{f}_{t_H})\left[\frac{r(k-x_{\text{total}}(0))}{(k-x_{\text{total}}(0)+e^{rt_H}x_{\text{total}}(0))} - \frac{\gamma}{(e^{\gamma(T-t_H)}-1)}\right]. \tag{5.26}$$

5.6.4 Evolutionary Dynamics

As in Chapters 3 and 4, the evolutionary dynamics can simply be determined using the formula

$$\frac{\mathrm{d}t_H}{\mathrm{d}\tau} = \frac{\mathrm{d}}{\mathrm{d}\hat{f}_{t_H}}\left[\frac{\mathrm{d}\hat{f}_{t_H}}{\mathrm{d}t}\right]\Bigg|_{\hat{f}_{t_H}=0}, \tag{5.27}$$

where we assume that the initial frequency of the novel mutant is small, so that we can evaluate $\mathrm{d}\hat{f}_{t_H}/\mathrm{d}t$ at the mutant free steady state $\hat{f}_{t_H}=0$. By applying Eq. (5.27) to Eq. (5.26), we get that

$$\frac{\mathrm{d}t_H}{\mathrm{d}\tau} = \frac{r(k-x_{\text{total}}(0))}{(k-x_{\text{total}}(0)+e^{rt_H}x_{\text{total}}(0))} - \frac{\gamma}{(e^{\gamma(T-t_H)}-1)}, \tag{5.28}$$

and we can find the evolutionarily stable state of Eq. (5.28) by finding the root of $\mathrm{d}t_H/d\tau = 0$ for t_H in Eq. (5.28). To show that there exists only one finite root to $\mathrm{d}t_H/d\tau = 0$ in Eq. (5.28), we check that Eq. (5.28) is monotonically decreasing in t_H for all $t_H < T$, which can be done by calculating the derivative of Eq. (5.28) with respect to t_H. This is given by

$$\frac{\mathrm{d}}{\mathrm{d}t_H}\left[\frac{\mathrm{d}t_H}{\mathrm{d}\tau}\right] = -\frac{(k-x_{\text{total}}(0))x_{\text{total}}(0)r^2e^{rt_H}}{(k+(e^{rt_H}-1)x_{\text{total}}(0))^2} - \frac{\gamma^2e^{\gamma(T-t_H)}}{(e^{\gamma(T-t_H)}-1)^2}, \tag{5.29}$$

which is strictly negative in fact for all $t_H \in \mathbb{R}$. Since $\mathrm{d}t_H/d\tau$ is a strictly decreasing function of t_H, there exists only one finite root to $\mathrm{d}t_H/d\tau = 0$ in Eq. (5.28).

5.6.5 Individual optimal timing of hibernation is equal to the evolutionary optimal timing of hibernation

To determine the individual optimal hibernation time $t_H^{(I)}$, we begin by seeking analytical solutions for $y(T)$ using Eq. (5.15). Note that Eq. (5.15) contains $x(t_H)$, which we evaluate by setting $t = t_H$ in Eq. (5.14). Denoting t_H as the timing of hibernation and $y(T)$ as the number of hibernating cells at time T, we have

$$y(T) = \frac{kx_{\text{total}}(0)e^{rt_H}}{k+x_{\text{total}}(0)(e^{rt_H}-1)}(1-e^{-\gamma(T-t_H)}) \tag{5.30}$$

and we seek the value of t_H such that $\mathrm{d}y(T)/\mathrm{d}t_H = 0$ in Eq. (5.30), i.e a t_H that maximises the number of hibernating cells at time T. We note that

$$\frac{\mathrm{d}y(T)}{\mathrm{d}t_H} = \frac{e^{-T\gamma+t_H(r+\gamma)}kx_{\text{total}}(0)\Big(A(De^{-\gamma t_H}-1)-\gamma(B+x_{\text{total}}(0)e^{rt_H})\Big)}{(k+(-1+e^{rt_H})x_{\text{total}}(0))^2} \tag{5.31}$$

where $A = r(k - x_{\text{total}}(0))$, $B = k - x_{\text{total}}(0)$ and $D = e^{\gamma T}$. As Eq. (5.31) is a transcendental in t_H, it would be necessary to obtain a numerical approximation of the solution to $dy(T)/dt_H = 0$ for t_H using `Findroot` in Mathematica (see `hibernation.nb` in the `Evol_Env_triggered_sex` repository).

Now we prove that this is equal to the evolutionary optimal timing of hibernation, which is the root of $\mathrm{d}t_H/d\tau = 0$ for t_H in Eq. (5.28). First note that Eq. (5.28) can be reexpressed as

$$\begin{aligned}\frac{\mathrm{d}t_H}{\mathrm{d}\tau} &= \frac{r(k-x_{\text{total}}(0))(e^{\gamma(T-t_H)}-1)-\gamma(k-x_{\text{total}}(0)+x_{\text{total}}(0)e^{rt_H})}{(k-x_{\text{total}}(0)+x_{\text{total}}(0)e^{rt_H})(e^{\gamma(T-t_H)}-1)} \\ &= \frac{A(De^{-\gamma t_H}-1)-\gamma(B+x_{\text{total}}(0)e^{rt_H})}{(De^{-\gamma t_H}-1)(B+x_{\text{total}}(0)e^{rt_H})}\end{aligned} \tag{5.32}$$

and here we see that the numerator of Eq. (5.31) is exactly equal to that of Eq. (5.32) multipled by a factor of $kx_{\text{total}}(0)e^{-T\gamma+t_H(r+\gamma)}$. By noting that the denominator of Eq. (5.31) as well as $kx_{\text{total}}(0)e^{-T\gamma+t_H(r+\gamma)}$ are strictly positive, the only finite positive solution for t_H such that $\mathrm{d}y(T)/\mathrm{d}t_H = 0$ is if the expression equal to the numerator of Eq. (5.32) is zero, i.e if

$$A(De^{-\gamma t_H}-1)-\gamma(B+x_{\text{total}}(0)e^{rt_H}) = 0\,. \tag{5.33}$$

This implies that the root of $\mathrm{d}y(T)/\mathrm{d}t_H = 0$ for t_H in Eq. (5.31) is equal to the root of $\mathrm{d}t_H/d\tau = 0$ for t_H in Eq. (5.32), which proves that the individual optimal timing of hibernation is equal to the evolutionary optimal timing of hibernation.

5.6.6 Is the fixed point of $\mathrm{d}t_H/d\tau$ an Evolutionarily Stable State?

Furthermore, we have the tools to calculate whether the fixed point of Eq. (5.28) for a given set of parameters is an evolutionary stable state, in which case we do not get evolutionary branching in t_H. This can be done by considering the second derivative of the mutant frequency in the subsequent growth period with respect to δt_H [46, 64]. Selection is disruptive if the condition given by

$$\frac{\mathrm{d}}{\mathrm{d}t}\left[\frac{\mathrm{d}^2\hat{f}'_{t_H}}{\mathrm{d}\delta t_H^2}\right]\Bigg|_{\delta t_H=0}\Bigg|_{\hat{f}_{t_H}=0} > 0 \tag{5.34}$$

is satisfied. To obtain an expression that allows us to check if selection is disruptive about the fixed point of Eq. (5.28), we apply Eq. (5.34) to Eq. (5.23), which is the invasion fitness of a mutant that enters hibernation at a different time. We have that

$$\begin{aligned} &\frac{\mathrm{d}}{\mathrm{d}t}\left[\frac{\mathrm{d}^2 \hat{f}'_{t_H}}{\mathrm{d}\delta t_H^2}\right]\Big|_{\delta t_H=0}\Big|_{\hat{f}_{t_H}=0} \\ = \quad &\frac{r^2(k-x_{\text{total}}(0))^2 - \frac{\gamma\left(k+(e^{rt_H}-1)x_{\text{total}}(0)\right)\left(\gamma x_{\text{total}}(0)e^{rt_H}+(k-x_{\text{total}}(0))(\gamma+2r)\right)}{e^{\gamma(t-t_H)}-1}}{\left(k+x_{\text{total}}(0)(e^{rt_H}-1)\right)^2} \end{aligned} \tag{5.35}$$

which we use to check if selection is disruptive about the fixed point of Eq. (5.28) for a given set of parameters. Since the denominator of Eq. (5.35) is strictly positive, we can simply just check the sign of the numerator of Eq. (5.35). The condition that gives rise to disruptive selection is thus equivalent to

$$e^{\gamma(t-t_H^*)}r^2(k-x_{\text{total}}(0))^2 - \left(e^{rt_H^*}\gamma x_{\text{total}}(0) + (k-x_{\text{total}}(0))(r+\gamma)\right)^2 > 0 \tag{5.36}$$

where t_H^* is the optimal timing of hibernation. As t_H^* is analytically intractable, we can only check if Eq. (5.36) is satisfied for a given set of parameters. For the parameters in Figure 5.11, we see that Eq. (5.36) is not satisfied and hence the fixed point of Eq. (5.28) are indeed evolutionary stable states, and we should not expect evolutionary branching to occur. In Figure 5.11 (a), Eq. (5.36) evaluates to -34486 and in panel (b), it evaluates to -9.0878 which suggests that selection is not disruptive.

5.6.7 Simulation of Evolutionary Dynamics

In our simulation of the evolutionary dynamics for t_H, the population will contain multiple genotypes with different values of t_H at any given time. Our dynamics within a growth period needs to account for that. If $g \subset \{1, S\}$ is the set of genotypes under logistic growth and $m \subset \{1, S\}$ is another set under recombination with $g \cap m = \emptyset$, and S is the number of genotypes in the population, the dynamics of the set g under

logistic growth is given by

$$\begin{aligned} \frac{dx^{(i)}}{dt} &= rx^{(i)}\left(1-\frac{x_{\text{total}}}{k}\right) \\ \frac{dy^{(i)}}{dt} &= 0 \\ \frac{dx^{(j)}}{dt} &= -\gamma x^{(j)} \\ \frac{dy^{(j)}}{dt} &= \gamma x^{(j)} \end{aligned} \tag{5.37}$$

for all $i \in g$ and $j \in m$. The total population of hibernating and non-hibernating members in the beginning of each generation is $x_{\text{total}}(0)$ and the x_{total} in Eq. (5.37) denotes the total population at any given time $x_{\text{total}} = \sum_{i=1}^{S} x_{\text{total}}^{(i)}$. To simulate the invasion dynamics, we calculate the change in frequency of each genotype between two successive growth periods. The frequency of genotype i with $t_H = t_H^{(i)}$ is given by

$$f_i' = \frac{y^{(i)}(T)}{\sum_{i=1}^{S} y^{(i)}(T)} \tag{5.38}$$

and the input parameters of this function are r, γ, k, $x_{total}(0)$ and T. The output of this function is $\mathbf{f}$, the frequency of all genotypes at the end of a growth period.

5.7 APPENDIX III: EVOLUTION OF TIMING OF GENETIC RECOMBINATION

5.7.1 When recombination can occur arbitrarily many times during a growth period (Iteroparous)

Here, we consider the iteroparous regime. Due to the complexity of this model, we examine the optimal timing of recombination of a genotype with a given value of t_R. This is the value of t_R that maximises the population of viable genotypes at the end of the growth period.

5.7.1.1 Dynamics within a growth period of an Individual genotype

For a genotype with a given value of t_R, the population dynamics before recombination ($0 < t < t_R$) is given by

$$\begin{aligned}
\frac{\mathrm{d}x_{00}}{\mathrm{d}t} &= rx_{00}\left(1-\frac{x_{\text{total}}}{k}\right) - 2\mu x_{00} \\
\frac{\mathrm{d}x_{01}}{\mathrm{d}t} &= rx_{01}\left(1-\frac{x_{\text{total}}}{k}\right) + \mu x_{00} - \mu x_{01} \\
\frac{\mathrm{d}x_{10}}{\mathrm{d}t} &= rx_{10}\left(1-\frac{x_{\text{total}}}{k}\right) + \mu x_{00} - \mu x_{10} \\
\frac{\mathrm{d}x_{11}}{\mathrm{d}t} &= \mu(x_{01}+x_{10})
\end{aligned} \tag{5.39}$$

where $x_{\text{total}}(t)$ is the total population at a given time. The dynamics of the population once it enters recombination phase ($t_R < t < T$) is given by

$$\begin{aligned}
\frac{\mathrm{d}x_{00}}{\mathrm{d}t} &= \alpha P_R x_{10}x_{01} \\
\frac{\mathrm{d}x_{01}}{\mathrm{d}t} &= -\alpha P_R x_{01}x_{10} \\
\frac{\mathrm{d}x_{10}}{\mathrm{d}t} &= -\alpha P_R x_{01}x_{10} \\
\frac{\mathrm{d}x_{11}}{\mathrm{d}t} &= \alpha P_R x_{10}x_{01}\,.
\end{aligned} \tag{5.40}$$

Analytical approximations of the solutions to Eqs. (5.39-5.40) are obtained as described in the section below.

5.7.1.2 Analytical Approximation of Individual Optimal timing of recombination

Using Eq. (5.39), we can approximate the frequencies of each genotype as follows.

$$\begin{aligned}
\frac{\mathrm{d}f_{00}}{\mathrm{d}t} &= r\left(1-\frac{x_{\text{total}}}{k}\right)f_{00}(1-f_{00}-f_1) - 2\mu f_{00} \\
\frac{\mathrm{d}f_1}{\mathrm{d}t} &= r\left(1-\frac{x_{\text{total}}}{k}\right)f_1(1-f_{00}-f_1) + 2\mu f_{00} - \mu f_1 \\
\frac{\mathrm{d}x_{\text{total}}}{\mathrm{d}t} &= r(f_{00}+f_1)x_{\text{total}}\left(1-\frac{x_{\text{total}}}{k}\right).
\end{aligned} \tag{5.41}$$

By assuming that μ is small (relative to r), we assume that the frequency of x_{11} in the population remains small throughout a growth period. This implies that the term $(1 - f_{00} - f_1)$ is of small magnitude. Since the term $r(1 - x_{\text{total}}/k)$ which multiplies the $(1 - f_{00} - f_1)$ is also of small magnitude, we can neglect the term

$r(1 - x_{\text{total}}/k)(1 - f_{00} - f_1)$ in Eq. (5.41), which simplifies Eq. (5.41) to

$$\begin{aligned} \frac{\mathrm{d}f_{00}}{\mathrm{d}t} &= -2\mu f_{00} \\ \frac{\mathrm{d}f_1}{\mathrm{d}t} &= 2\mu f_{00} - \mu f_1 \\ \frac{\mathrm{d}x_{\text{total}}}{\mathrm{d}t} &= r x_{\text{total}}\left(1 - \frac{x_{\text{total}}}{k}\right) \end{aligned} \tag{5.42}$$

which we can solve to obtain solutions for $f_{00}(t)$ and $f_1(t)$. The solution $f_{00}(t)$ is given simply by $f_{00}(t) = f_{00}(0)e^{-2\mu t}$, that can be substituted into the expression for $\mathrm{d}f_1/\mathrm{d}t$ in Eq. (5.42) to solve for $f_1(t)$, which is given by $f_1(t) = e^{-\mu t}(f_1(0) + 2f_{00}(0)(1 - e^{-\mu t}))$. In Figures 5.17 and 5.18, we show that this approximation yields highly accurate solutions for Eq. (5.39). The accuracy of the solutions to $f_{00}(t)$

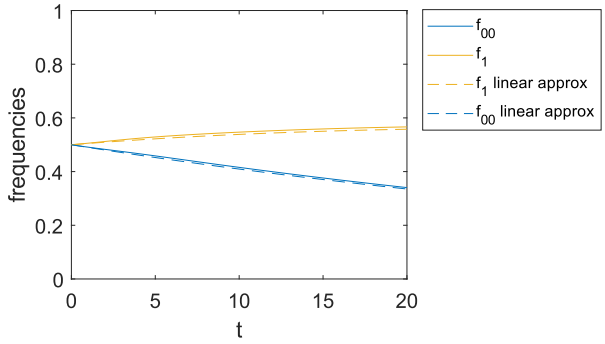

Figure 5.17: Change in the frequencies $f_{00}(t)$ and $f_1(t)$ over a growth period. Parameter values are $T = 20$, $k = 1$, $r = 1$, $x_{\text{total}}(0) = 0.05$, $f_{00}(0) = 0.5$ and $f_1(0) = 0.5$, $\mu = 0.01$. Dashed curves are analytical approximations and solid curves are actual frequencies. Analytical approximations are solutions to Eq. (5.42).

and $f_1(t)$ as shown in Figure 5.17 allows us to approximate the abundance of each genotype $x_{00}(t)$, $x_1(t)$ and $x_{11}(t)$ by multiplying those solutions by $x_{\text{total}}(t)$, which we do in Figure 5.18. In other words, $x_{00}(t) = x_{\text{total}}(t)f_{00}(t)$, $x_1(t) = x_{\text{total}}(t)f_1(t)$ and $x_{11}(t) = x_{\text{total}}(t)(1 - f_{00}(t) - f_1(t))$. The analytical approximations for the abundances we observe in Figure 5.18 are given by

$$\begin{aligned} x_{00}(t) &= \frac{k x_{\text{total}}(0) f_{00}(0) e^{(r-2\mu)t}}{k + x_{\text{total}}(0)(e^{rt} - 1)} \\ x_1(t) &= \frac{k x_{\text{total}}(0) e^{t(r-2\mu)}(-2f_{00}(0) + e^{t\mu}(f_{00}(0) + 1))}{k + x_{\text{total}}(0)(e^{rt} - 1)} \\ x_{11}(t) &= \frac{k x_{\text{total}}(0) e^{t(r-2\mu)}(e^{t\mu} - 1)(-f_{00}(0) + e^{t\mu})}{k + x_{\text{total}}(0)(e^{rt} - 1)} \end{aligned} \tag{5.43}$$

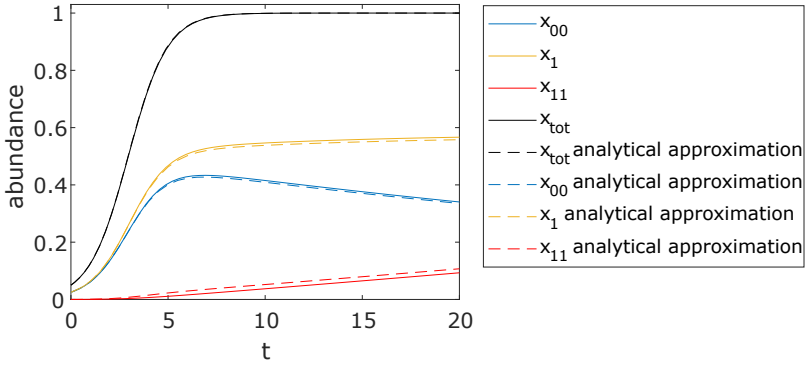

Figure 5.18: Change in the abundances $x_{00}(t)$ and $x_1(t)$ over one growth period. Parameter values same as Figure 5.17. The analytical approximations are given by multiplying the solutions to $f_{00}(t)$ and $f_1(t)$ in Eq. (5.42) with the analytical approximation of $x_{\text{total}}(t)$ derived from Eq. (5.42), just as in Figure 5.17. The expressions for these are given by Eq. (5.43).

which are obtained by multiplying $x_{\text{total}}(t)$ by the frequencies of each genotype (solutions to Eq. (5.42)). Since the $\binom{1}{1}$ genotypes die at time T, each growth period will begin with only $\binom{0}{0}$, $\binom{0}{1}$ and $\binom{1}{0}$ genotypes which means that $f_{00}(0) + f_1(0) = 1$, simplifying our analysis. In Eq. (5.43), we have substituted $f_1(0)$ with $1 - f_{00}(0)$.

5.7.1.3 Analytical approximation of the recombination phase

In order to analytically approximate the dynamics of the population once it enters the recombination phase ($t_R < t < T$), we consider Eq. (5.40). Importantly, we assume that the $\binom{1}{1}$ genotype does not attempt recombination, and its solution is given by

$$\begin{aligned} x_{00}(t) &= x_{00}(t_R) + x_{01}(t_R) - \frac{1}{x_{01}(t_R)^{-1} + \alpha P_R(t - t_R)} \\ x_1(t) &= \frac{2}{\alpha P_R(t - t_R) + x_{10}(t_R)^{-1}} \\ x_{11}(t) &= x_{11}(t_R) + x_{01}(t_R) - \frac{1}{x_{01}(t_R)^{-1} + \alpha P_R(t - t_R)}\,. \end{aligned} \tag{5.44}$$

Evaluating the expressions for Eq. (5.43) at $t = t_R$ and substituting them into Eq. (5.44), we get

$$\begin{aligned} x_{00}(t) &= \frac{kx_{\text{total}}(0)e^{t_R(r-\mu)}(f_{00}(0)+0.5(1-f_{00}(0)))}{k+(e^{rt_R}-1)x_{\text{total}}(0)} - \frac{1}{\alpha P_R(t-t_R)+C} \\ x_1(t) &= \frac{2}{\alpha P_R(t-t_R)+C} \\ x_{11}(t) &= \frac{kx_{\text{total}}(0)(e^{t_R(r-\mu)}(-f_{00}(0)-0.5(1-f_{00}(0)))+e^{rt_R})}{k+(e^{rt_R}-1)x_{\text{total}}(0)} \\ &- \frac{1}{\alpha P_R(t-t_R)+C} \end{aligned} \tag{5.45}$$

where

$$C = \frac{2e^{-t_R(r-2\mu)}(k+x_{\text{total}}(0)(e^{rt_R}-1))}{kx_{\text{total}}(0)(e^{\mu t_R}(f_{00}(0)+1)-2f_{00}(0))}\,. \tag{5.46}$$

As shown in Figure 5.8 of the main text, $f_{00}(0)$ equilibriates over a large number of growth periods, and it is necessary to calculate the equilibrium frequency of $f_{00}(0)$. We note that the frequency of each genotype at equilibrium are a function of all the system parameters $x_{\text{total}}(0)$, μ, r, k, α, P_R and T as well as t_R. The equilibrium frequency of $f_{00}(0)$ is approximated analytically by solving the equation

$$\left.\frac{x_{00}(t)}{x_{00}(t)+x_1(t)}\right|_{t=T} - f_{00}(0) = 0 \tag{5.47}$$

for $f_{00}(0)$, where $x_{00}(t)$ and $x_1(t)$ are obtained from Eq. (5.45). The solution to Eq. (5.47) for $f_{00}(0)$ is given by

$$\begin{aligned} f_{00}(0) &= \Bigg(e^{-2rt_R}(2e^{t_R(r+\mu)}(e^{t_R\mu}-1)(k+(e^{rt_R}-1)x_{\text{total}}(0)) - e^{2rt_R}kP_R(t_R-T)x_{\text{total}}(0)\alpha \\ &-\frac{1}{2}\Bigg((4e^{t_R(4r+\mu)}(e^{t_R\mu}-2)k^2P_R^2(T-t_R)^2x_{\text{total}}(0)^2\alpha^2 + 4e^{2rt_R}(2e^{t_R\mu}(k-x_{\text{total}}(0)) \\ &-2e^{2t_R\mu}(k-x_{\text{total}}(0)) + 2e^{t_R(r+\mu)}x_{\text{total}}(0) - 2e^{t_R(r+2\mu)}x_{\text{total}}(0) \\ &+e^{rt_R}kP_R(t_R-T)x_{\text{total}}(0)\alpha)^2)\Bigg)^{1/2}\Bigg)\Big/\Bigg((e^{t_R\mu}-2)kP_R(t_R-T)x_{\text{total}}(0)\alpha\Bigg) \end{aligned} \tag{5.48}$$

which is then substituted into Eq. (5.45-5.46). In the main text, we denoted the equilibrium frequency of $f_{00}(0)$ as f_{00}^*, which is literally the expression given by Eq. (5.48). In Figure 5.8, we also see that this expression is a highly accurate approximation. Having obtained an accurate approximation for the equilibrium abundances of each genotype, we have all the necessary material to calculate the

individual optimal timing of recombination, which can be calculated by solving

$$\frac{\mathrm{d}}{\mathrm{d}t_R}(x_{00}(t)+x_1(t))\Big|_{t=T,f_{00}(0)=f_{00}^*}=0 \tag{5.49}$$

for t_R, where $x_{00}(t)$ and $x_1(t)$ are obtained from Eq. (5.45). The total population of viable genotypes at time T is given by

$$x_{00}(T)+x_1(T)=\frac{kx_{\text{total}}(0)e^{t_R(r-\mu)}}{k+(e^{rt_R}-1)x_{\text{total}}(0)} \tag{5.50}$$

which we can apply Eq. (5.49) to in order to calculate the individual optimal timing of recombination $t_R^{(I)}$, given by

$$t_R^{(I)}=\frac{1}{r}\log\left(\frac{(r-\mu)(k-x_{\text{total}}(0))}{\mu x_{\text{total}}(0)}\right). \tag{5.51}$$

5.7.1.4 Dynamics within a growth period for Evolutionary Dynamics Simulation

In our simulation of the evolutionary dynamics, the population will contain multiple genotypes with different values of t_R at any given time. Our dynamics within a growth period will thus need to account for that. Given that each genotype has a different value of t_R, it is possible that at any given time, there are genotypes in recombination and genotypes that are still under logistic growth. We describe our simulation model as follows, if $g \subset \{1,...,S\}$ is the set of genotypes under logistic growth and $m \subset \{1,...,S\}$ is another set that is under recombination with $g \cap m = \emptyset$, the dynamics of the set under logistic growth are given by

$$\begin{aligned}
\frac{\mathrm{d}x_{00}^{(i)}}{\mathrm{d}t} &= rx_{00}^{(i)}\left(1-\frac{x_{\text{total}}}{k}\right)-2\mu x_{00}^{(i)} \\
\frac{\mathrm{d}x_{01}^{(i)}}{\mathrm{d}t} &= rx_{01}^{(i)}\left(1-\frac{x_{\text{total}}}{k}\right)+\mu x_{00}^{(i)}-\mu x_{01}^{(i)} \\
\frac{\mathrm{d}x_{10}^{(i)}}{\mathrm{d}t} &= rx_{10}^{(i)}\left(1-\frac{x_{\text{total}}}{k}\right)+\mu x_{00}^{(i)}-\mu x_{10}^{(i)}
\end{aligned} \tag{5.52}$$

for all $i \in g$ where the superscript i represents the i-th genotype and $x_{\text{total}} = \sum_{i=1}^{S} x_{\text{total}}^{(i)}$ is the total population. We now consider the case where a subset $g \subset \{1,...,S\}$ of genotypes are under logistic growth and another subset $m \subset \{1,...,S\}$ is in recombination with $g \cap m = \emptyset$. The timing of recombination t_R is determined by a gene in the third non-recombining locus, and therefore in our simulations, we employ a three locus model instead of a two locus model. Assuming that we have

free recombination, the dynamics are given by

$$
\begin{aligned}
\frac{\mathrm{d}x_{00}^{(i)}}{\mathrm{d}t} &= rx_{00}^{(i)}\left(1-\frac{x_{\text{total}}}{k}\right)-2\mu x_{00}^{(i)} \qquad (5.53)\\
\frac{\mathrm{d}x_{1}^{(i)}}{\mathrm{d}t} &= rx_{1}^{(i)}\left(1-\frac{x_{\text{total}}}{k}\right)+2\mu x_{00}^{(i)}-\mu x_{1}^{(i)}\\
\frac{\mathrm{d}x_{11}^{(i)}}{\mathrm{d}t} &= \mu x_{1}^{(i)}\\
\frac{\mathrm{d}x_{00}^{(j)}}{\mathrm{d}t} &= \frac{\alpha P_R}{3}\left(x_{01}^{(j)}\sum_{\substack{j'\in m\\ j'\neq j}}x_{10}^{(j')}+x_{10}^{(j)}\sum_{\substack{j'\in m\\ j'\neq j}}x_{01}^{(j')}\right)+\frac{2\alpha P_R}{3}x_{01}^{(j)}x_{10}^{(j)}\\
&\quad+\frac{2\alpha P_R}{3}\left(x_{01}^{(j)}\sum_{\substack{j'\in m\\ j'\neq j}}x_{00}^{(j')}+x_{10}^{(j)}\sum_{\substack{j'\in m\\ j'\neq j}}x_{00}^{(j')}\right)-\frac{2\alpha P_R}{3}\left(x_{00}^{(j)}\sum_{\substack{j'\in m\\ j'\neq j}}x_{01}^{(j')}+x_{00}^{(j)}\sum_{\substack{j'\in m\\ j'\neq j}}x_{10}^{(j')}\right)\\
\frac{\mathrm{d}x_{01}^{(j)}}{\mathrm{d}t} &= \frac{2\alpha P_R}{3}x_{00}^{(j)}\sum_{\substack{j'\in m\\ j'\neq j}}x_{01}^{(j')}-\frac{2\alpha P_R}{3}x_{01}^{(j)}\sum_{\substack{j'\in m\\ j'\neq j}}x_{00}^{(j')}-\alpha P_R x_{01}^{(j)}\sum_{\substack{j'\in m\\ j'\neq j}}x_{10}^{(j')}+\frac{\alpha P_R}{3}x_{10}^{(j)}\sum_{\substack{j'\in m\\ j'\neq j}}x_{01}^{(j')}\\
&\quad-\frac{2\alpha P_R}{3}x_{01}^{(j)}x_{10}^{(j)}\\
\frac{\mathrm{d}x_{10}^{(j)}}{\mathrm{d}t} &= \frac{2\alpha P_R}{3}x_{00}^{(j)}\sum_{\substack{j'\in m\\ j'\neq j}}x_{10}^{(j')}-\frac{2\alpha P_R}{3}x_{10}^{(j)}\sum_{\substack{j'\in m\\ j'\neq j}}x_{00}^{(j')}-\alpha P_R x_{10}^{(j)}\sum_{\substack{j'\in m\\ j'\neq j}}x_{01}^{(j')}+\frac{\alpha P_R}{3}x_{01}^{(j)}\sum_{\substack{j'\in m\\ j'\neq j}}x_{10}^{(j')}\\
&\quad-\frac{2\alpha P_R}{3}x_{01}^{(j)}x_{10}^{(j)}\\
\frac{\mathrm{d}x_{11}^{(j)}}{\mathrm{d}t} &= \frac{\alpha P_R}{3}\left(x_{01}^{(j)}\sum_{\substack{j'\in m\\ j'\neq j}}x_{10}^{(j')}+x_{10}^{(j)}\sum_{\substack{j'\in m\\ j'\neq j}}x_{01}^{(j')}\right)+\frac{2\alpha P_R}{3}x_{01}^{(j)}x_{10}^{(j)}
\end{aligned}
$$

for all $i \in g$ and $j, j' \in m$. Eqs. (5.52-5.53) describes the dynamics within a growth period, and is run for a time period of T. The initial population at the beginning of each growth period is $x_{\text{total}}(0)$. To simulate the subsequent evolutionary dynamics, it is necessary to simulate the invasion dynamics (see Section 5.7.3).

It is crucial to highlight that under the assumption of free recombination, the probability that a crossover event occurs in one or more loci P_R differs between the two locus and three locus model. If the probability that a crossover event occurs at a given locus is P_R, then in the two locus model (e.g Eqs. (5.39-5.40)), $P_R = P_{\text{CO}}^2+(1-P_{\text{CO}})^2$, while in a three locus model (e.g Eqs. (5.52-5.53)), $P_R = P_{\text{CO}}^3 + (1 - P_{\text{CO}})^3$. If $P_{\text{CO}} = 1/2$, then $P_R = 1/2$ in a two locus model and $P_R = 3/4$ in a three locus model.

5.7.2 When recombination occurs once during a growth period (Semelparous)

5.7.2.1 Dynamics within a growth period of Individual genotype

For this scenario, the population dynamics for the phase before recombination ($0 < t < t_R$) is also given by Eq. (5.39), as in the iteroparous regime. Thus in this section, we only include equations describing the dynamics of the population already in recombination. Denoting $\chi_{00}(t)$, $\chi_1(t)$ and $\chi_{11}(t)$ as the population that have yet to undergo recombination and $y_{00}(t)$, $y_1(t)$ and $y_{11}(t)$ as the population that have already attempted recombination, the population dynamics of the recombination phase ($t_R < t < T$) is given in Eqs. (5.54) and (5.55).

$$\begin{aligned}
\frac{\mathrm{d}\chi_{00}}{\mathrm{d}t} &= -\alpha\chi_{00}(\chi_{01}+\chi_{10}) - \alpha\chi_{00}^2 \\
\frac{\mathrm{d}\chi_{01}}{\mathrm{d}t} &= -\alpha\chi_{01}(\chi_{10}+\chi_{01}+\chi_{00}) \\
\frac{\mathrm{d}\chi_{10}}{\mathrm{d}t} &= -\alpha\chi_{10}(\chi_{01}+\chi_{10}+\chi_{00}) \\
\frac{\mathrm{d}\chi_{11}}{\mathrm{d}t} &= 0 \\
\frac{\mathrm{d}y_{00}}{\mathrm{d}t} &= \alpha P_R\chi_{10}\chi_{01} + \alpha\chi_{00}(\chi_{10}+\chi_{01}) + \alpha\chi_{00}^2 \\
\frac{\mathrm{d}y_{01}}{\mathrm{d}t} &= \alpha\chi_{00}\chi_{01} + \alpha(1-P_R)\chi_{01}\chi_{10} + \alpha\chi_{01}^2 \\
\frac{\mathrm{d}y_{10}}{\mathrm{d}t} &= \alpha\chi_{00}\chi_{10} + \alpha(1-P_R)\chi_{01}\chi_{10} + \alpha x_{10}^2 \\
\frac{\mathrm{d}y_{11}}{\mathrm{d}t} &= \alpha P_R\chi_{10}\chi_{01}
\end{aligned} \tag{5.54}$$

denoting $\chi_1(t) = \chi_{01}(t) + \chi_{10}(t)$ and assuming that $\chi_{10}(t) = \chi_{01}(t)$, Eq. (5.54) can be reexpressed as

$$\begin{aligned}
\frac{\mathrm{d}\chi_{00}}{\mathrm{d}t} &= -\alpha\chi_{00}(\chi_{00}+\chi_1) \\
\frac{\mathrm{d}\chi_1}{\mathrm{d}t} &= -\alpha\chi_1(\chi_{00}+\chi_1) \\
\frac{\mathrm{d}\chi_{11}}{\mathrm{d}t} &= 0 \\
\frac{\mathrm{d}y_{00}}{\mathrm{d}t} &= \frac{\alpha P_R\chi_1^2}{4} + \alpha\chi_{00}\chi_1 + \alpha\chi_{00}^2 \\
\frac{\mathrm{d}y_1}{\mathrm{d}t} &= \alpha\chi_{00}\chi_1 + \frac{\alpha(1-P_R)\chi_1^2}{2} + \frac{\alpha\chi_1^2}{2} \\
\frac{\mathrm{d}y_{11}}{\mathrm{d}t} &= \frac{\alpha P_R\chi_1^2}{4}
\end{aligned} \tag{5.55}$$

which can be used conveniently to obtain analytical approximations for the population dynamics, which we do below.

5.7.2.2 Analytically Approximating the Individual Optimal timing of recombination

Here, we derive the solutions to Eq. (5.55). First we can use the substitution $\chi_{\text{viable}}(t) = \chi_{00}(t) + \chi_1(t)$ and see that

$$\frac{\mathrm{d}\chi_{\text{viable}}}{\mathrm{d}t} = -\alpha\chi_{\text{viable}}^2 \tag{5.56}$$

which has solution

$$\chi_{\text{viable}}(t) = \frac{1}{\alpha\Big[(t-t_R) + (\alpha\chi_{\text{viable}}(t_R))^{-1}\Big]} \tag{5.57}$$

for $(t_R < t < T)$. Due to the symmetry of $\mathrm{d}\chi_{00}/\mathrm{d}t$ and $\mathrm{d}\chi_1/\mathrm{d}t$ in Eq. (5.55), we notice that the quantity $f_{\chi_{00}} = \chi_{00}(t)/(\chi_{00}(t)+\chi_1(t))$ is constant over time t i.e $\mathrm{d}f_{\chi_{00}}/\mathrm{d}t = 0$. As a result, we get that

$$\begin{aligned}
\chi_{00}(t) &= \chi_{\text{viable}}(t) f_{\chi_{00}} \\
\chi_1(t) &= \chi_{\text{viable}}(t)(1-f_{\chi_{00}})\,.
\end{aligned} \tag{5.58}$$

For our analysis, we set $f_{\chi_{00}}$ straightforwardly as $f_{\chi_{00}} = \chi_{00}(t_R)/(\chi_{00}(t_R)+\chi_1(t_R))$ where $\chi_{00}(t_R)$ and $\chi_1(t_R)$ are given in Eq. (5.43). From Eq. (5.55), we determine

that

$$\begin{aligned}
\frac{\mathrm{d}y_{00}}{\mathrm{d}t} &= \frac{C(t_R)}{\alpha^2\Big[(t-t_R)+(\alpha\chi_{\text{viable}}(t))^{-1}\Big]^2} \\
\frac{\mathrm{d}y_1}{\mathrm{d}t} &= \frac{D(t_R)}{\alpha^2\Big[(t-t_R)+(\alpha\chi_{\text{viable}}(t))^{-1}\Big]^2} \\
\frac{\mathrm{d}y_{11}}{\mathrm{d}t} &= \frac{E(t_R)}{\alpha^2\Big[(t-t_R)+(\alpha\chi_{\text{viable}}(t))^{-1}\Big]^2}
\end{aligned} \tag{5.59}$$

where

$$\begin{aligned}
C(t_R) &= \frac{\frac{\alpha P_R}{4}\chi_1(t_R)^2+\alpha\chi_1(t_R)\chi_{00}(t_R)+\alpha\chi_{00}(t_R)^2}{(\chi_{00}(t_R)+\chi_1(t_R))^2} \\
D(t_R) &= \frac{\frac{\alpha(1-P_R)}{2}\chi_1(t_R)^2+\alpha\chi_1(t_R)\chi_{00}(t_R)+\frac{\alpha}{2}\chi_1(t_R)^2}{(\chi_{00}(t_R)+\chi_1(t_R))^2} \\
E(t_R) &= \frac{\frac{\alpha P_R}{4}\chi_1(t_R)^2}{(\chi_{00}(t_R)+\chi_1(t_R))^2}
\end{aligned}$$

together this gives

$$\begin{aligned}
y_{00}(t) &= \frac{C(t_R)}{\alpha}\Big[\chi_{\text{viable}}(t_R)-\chi_{\text{viable}}(t)\Big] \\
y_1(t) &= \frac{D(t_R)}{\alpha}\Big[\chi_{\text{viable}}(t_R)-\chi_{\text{viable}}(t)\Big] \\
y_{11}(t) &= \frac{E(t_R)}{\alpha}\Big[\chi_{\text{viable}}(t_R)-\chi_{\text{viable}}(t)\Big].
\end{aligned} \tag{5.60}$$

As in the semelparous regime, we now approximate the equilibrium frequency of χ_{00} which is denoted as f_{00}^* in the main text, but now by solving the equation

$$\left.\frac{\chi_{00}(t)+y_{00}(t)}{\chi_{00}(t)+\chi_1(t)+y_{00}(t)+y_1(t)}\right|_{t=T} - f_{00}(0) = 0 \tag{5.61}$$

for $f_{00}(0)$, since $y_{00}(t)$ and $y_1(t)$ also forms part of the viable population now. The solution to Eq. (5.61) for $f_{00}(0)$ is given by

$$\begin{aligned} f_{00}(0) \;=\; & \Bigg(e^{-t_R(r-2\mu)}(-4(e^{rt_R}-1)(e^{t_R\mu}-1)x_{\text{total}}(0) \\ & -2k\Bigg(-2+2e^{t_R\mu}+e^{rt_R}(P_R-2)(t_R-T)x_{\text{total}}(0)\alpha \\ & -2e^{t_R(r-\mu)}(P_R-1)(t_R-T)x_{\text{total}}(0)\alpha\Bigg)+\Bigg((-4e^{2t_R(r-\mu)}k^2(e^{2t_R\mu}(P_R-4) \\ & -4e^{t_R\mu}(P_R-2)+4(P_R-1))P_R(T-t_R)^2x_{\text{total}}(0)^2\alpha^2 \\ & +4(2(e^{rt_R}-1)(e^{t_R\mu}-1)x_{\text{total}}(0) \\ & +k(-2+2e^{t_R\mu}+e^{rt_R}(P_R-2)(t_R-T)x_{\text{total}}(0)\alpha \\ & -2e^{t_R(r-\mu)}(P_R-1)(t_R-T)x_{\text{total}}(0)\alpha))^2)\Bigg)^{1/2}\Bigg)/\Bigg(2k(e^{2t_R\mu}(P_R-4) \\ & -4e^{t_R\mu}(P_R-2)+4(P_R-1))(t_R-T)x_{\text{total}}(0)\alpha\Bigg) \end{aligned} \tag{5.62}$$

where again $x_{\text{total}}(0)$ denotes the total population at the beginning of a growth period. Having obtained an accurate approximation for the abundances of each genotype, we have all the necessary material to calculate the individual optimal timing of recombination. This quantity can be calculated simply by solving

$$\frac{\mathrm{d}}{\mathrm{d}t_R}(\chi_{00}(t)+\chi_1(t)+y_{00}(t)+y_1(t))\Big|_{t=T,f_{00}(0)=f_{00}^*}=0 \tag{5.63}$$

for t_R, where $\chi_{00}(t)$ and $\chi_1(t)$ are obtained from Eq. (5.58) and $y_{00}(t)$ and $y_1(t)$ from Eq. (5.60). In this scenario, we find that the total population of viable genotypes at time T is given by the same function as that in the iteroparous regime

$$\chi_{00}(T)+\chi_1(T)+y_{00}(T)+y_1(T)=\frac{kx_{\text{total}}(0)e^{t_R(r-\mu)}}{k+(e^{rt_R}-1)x_{\text{total}}(0)} \tag{5.64}$$

which gives rise to the same optimal timing of recombination between these two regimes Eq. (5.51).

5.7.2.3 Dynamics within a growth period for Evolutionary Dynamics Simulation

As in the iteroparous regime, here we include the equations used to numerically simulate the dynamics within a growth period of a population where genotypes of different timing of recombination t_R recombine with each other. At a given time,

if $g \subset \{1, ..., S\}$ is the set of genotypes under logistic growth and $m \subset \{1, ..., S\}$ is another set that is under recombination with $g \cap m = \emptyset$, the dynamics of the set under logistic growth are given by

$$\begin{aligned}
\frac{\mathrm{d}\chi_{00}^{(i)}}{\mathrm{d}t} &= r\chi_{00}^{(i)}\left(1 - \frac{x_{\text{total}}}{k}\right) - 2\mu\chi_{00}^{(i)} \\
\frac{\mathrm{d}\chi_{01}^{(i)}}{\mathrm{d}t} &= r\chi_{01}^{(i)}\left(1 - \frac{x_{\text{total}}}{k}\right) + \mu(\chi_{00}^{(i)} - \chi_{01}^{(i)}) \\
\frac{\mathrm{d}\chi_{10}^{(i)}}{\mathrm{d}t} &= r\chi_{10}^{(i)}\left(1 - \frac{x_{\text{total}}}{k}\right) + \mu(\chi_{00}^{(i)} - \chi_{10}^{(i)}) \\
\frac{\mathrm{d}\chi_{11}^{(i)}}{\mathrm{d}t} &= \mu(\chi_{01}^{(i)} + \chi_{10}^{(i)}) \\
\frac{\mathrm{d}y_{00}^{(i)}}{\mathrm{d}t} &= 0 \\
\frac{\mathrm{d}y_{01}^{(i)}}{\mathrm{d}t} &= 0 \\
\frac{\mathrm{d}y_{10}^{(i)}}{\mathrm{d}t} &= 0 \\
\frac{\mathrm{d}y_{11}^{(i)}}{\mathrm{d}t} &= 0
\end{aligned} \tag{5.65}$$

for all $i \in g$ where the superscript i represents the i-th genotype and $x_{\text{total}} = \sum_{i=1}^{S} x_{\text{total}}^{(i)}$ is the total population that includes both the members that have engaged in sex and those that have not. The dynamics of the set of genotypes $j \in m$ in recombination

are given by

$$\frac{\mathrm{d}\chi_{00}^{(j)}}{\mathrm{d}t} = -\alpha(\chi_{00}^{(j)})^2 - \alpha\chi_{00}^{(j)} \sum_{\substack{j'\in m \\ j'\neq j}} \chi_{00}^{(j')} - \alpha\chi_{00}^{(j)}(\chi_{01}^{(j)} + \chi_{10}^{(j)}) - \alpha\chi_{00}^{(j)} \sum_{\substack{j'\in m \\ j'\neq j}} (\chi_{10}^{(j')} + \chi_{01}^{(j')}) \tag{5.66}$$

$$\frac{\mathrm{d}\chi_{01}^{(j)}}{\mathrm{d}t} = -\alpha\chi_{01}^{(j)} \sum_{\substack{j'\in m \\ j'\neq j}} \chi_{10}^{(j')} - \alpha\chi_{01}^{(j)} \sum_{\substack{j'\in m \\ j'\neq j}} \chi_{01}^{(j')} - \alpha(\chi_{01}^{(j)})^2 - \alpha\chi_{00}^{(j)}\chi_{01}^{(j)} - \alpha\chi_{10}^{(j)}\chi_{01}^{(j)} - \alpha\chi_{01}^{(j)} \sum_{\substack{j'\in m \\ j'\neq j}} \chi_{00}^{(j')}$$

$$\frac{\mathrm{d}\chi_{10}^{(j)}}{\mathrm{d}t} = -\alpha\chi_{10}^{(j)} \sum_{\substack{j'\in m \\ j'\neq j}} \chi_{01}^{(j')} - \alpha\chi_{10}^{(j)} \sum_{\substack{j'\in m \\ j'\neq j}} \chi_{10}^{(j')} - \alpha(\chi_{10}^{(j)})^2 - \alpha\chi_{00}^{(j)}\chi_{10}^{(j)} - \alpha\chi_{10}^{(j)}\chi_{01}^{(j)} - \alpha\chi_{10}^{(j)} \sum_{\substack{j'\in m \\ j'\neq j}} \chi_{00}^{(j')}$$

$$\frac{\mathrm{d}\chi_{11}^{(j)}}{\mathrm{d}t} = 0$$

$$\frac{\mathrm{d}y_{00}^{(j)}}{\mathrm{d}t} = \frac{\alpha P_R}{3}(\chi_{10}^{(j)} \sum_{\substack{j'\in m \\ j'\neq j}} \chi_{01}^{(j')} + \chi_{01}^{(j)} \sum_{\substack{j'\in m \\ j'\neq j}} \chi_{10}^{(j')}) + \alpha\left(1 - \frac{2P_R}{3}\right)\chi_{00}^{(j)} \sum_{\substack{j'\in m \\ j'\neq j}} (\chi_{10}^{(j')} + \chi_{01}^{(j')}) +$$
$$\alpha\chi_{00}^{(j)}(\chi_{01}^{(j)} + \chi_{10}^{(j)}) + \alpha(\chi_{00}^{(j)})^2 + \alpha\chi_{00}^{(j)} \sum_{\substack{j'\in m \\ j'\neq j}} \chi_{00}^{(j')} + \frac{2\alpha P_R}{3}\chi_{10}^{(j)}\chi_{01}^{(j)} +$$
$$\frac{2\alpha P_R}{3}(\chi_{01}^{(j)} + \chi_{10}^{(j)}) \sum_{\substack{j'\in m \\ j'\neq j}} \chi_{00}^{(j')}$$

$$\frac{\mathrm{d}y_{01}^{(j)}}{\mathrm{d}t} = \frac{2\alpha P_R}{3}\chi_{00}^{(j)} \sum_{\substack{j'\in m \\ j'\neq j}} \chi_{01}^{(j')} + \alpha(1 - P_R)\chi_{01}^{(j)} \sum_{\substack{j'\in m \\ j'\neq j}} \chi_{10}^{(j')} + \alpha\chi_{01}^{(j)} \sum_{\substack{j'\in m \\ j'\neq j}} \chi_{01}^{(j')} + \alpha(\chi_{01}^{(j)})^2 + \alpha\chi_{00}^{(j)}\chi_{01}^{(j)} +$$
$$\alpha\left(1 - \frac{2P_R}{3}\right)\chi_{10}^{(j)}\chi_{01}^{(j)} + \alpha\left(1 - \frac{2P_R}{3}\right)\chi_{01}^{(j)} \sum_{\substack{j'\in m \\ j'\neq j}} \chi_{00}^{(j')} + \frac{\alpha P_R}{3}\chi_{10}^{(j)} \sum_{\substack{j'\in m \\ j'\neq j}} \chi_{01}^{(j')}$$

$$\frac{\mathrm{d}y_{10}^{(j)}}{\mathrm{d}t} = \frac{2\alpha P_R}{3}\chi_{00}^{(j)} \sum_{\substack{j'\in m \\ j'\neq j}} \chi_{10}^{(j')} + \alpha(1 - P_R)\chi_{10}^{(j)} \sum_{\substack{j'\in m \\ j'\neq j}} \chi_{01}^{(j')} + \alpha\chi_{10}^{(j)} \sum_{\substack{j'\in m \\ j'\neq j}} \chi_{10}^{(j')} + \alpha(\chi_{10}^{(j)})^2 + \alpha\chi_{00}^{(j)}\chi_{10}^{(j)} +$$
$$\alpha\left(1 - \frac{2P_R}{3}\right)\chi_{10}^{(j)}\chi_{01}^{(j)} + \alpha\left(1 - \frac{2P_R}{3}\right)\chi_{10}^{(j)} \sum_{\substack{j'\in m \\ j'\neq j}} \chi_{00}^{(j')} + \frac{\alpha P_R}{3}\chi_{01}^{(j)} \sum_{\substack{j'\in m \\ j'\neq j}} \chi_{10}^{(j')}$$

$$\frac{\mathrm{d}y_{11}^{(j)}}{\mathrm{d}t} = \frac{\alpha P_R}{3}(\chi_{10}^{(j)} \sum_{\substack{j'\in m \\ j'\neq j}} \chi_{01}^{(j')} + \chi_{01}^{(j)} \sum_{\substack{j'\in m \\ j'\neq j}} \chi_{10}^{(j')}) + \frac{2\alpha P_R}{3}\chi_{01}^{(j)}\chi_{10}^{(j)}\,.$$

5.7.3 Simulation of Evolutionary Dynamics

To perform stochastic simulations of the evolutionary trajectories, it is necessary to consider the invasion dynamics. For our simulation of the evolutionary trajectories, we employ a multi-genotype model as in Chapters 3 and 4 whereby a novel mutant with a different value of t_R to the resident genotype emerges in the population at rate μ_{t_R}. This rate is the inverse of the expected number of growth periods until the next mutation event G. Before each successive mutation event, the number of growth periods until the next mutation event G is determined by generating a random number from a $Geo(\mu_{t_R})$ distribution and inverting that number. In an S genotype model, if a given genotype i has a t_R of $t_R^{(i)}$ and has frequency f_i, then the mean population trait value is given by

$$\langle t_R \rangle = \sum_{i=1}^{S} t_R^{(i)} f_i \,. \tag{5.67}$$

This value is recorded immediately before each novel mutant emerges in the population. Simulations of the evolutionary trajectories in our hibernation model (e.g in Figures 5.11 and 5.12) is performed using the exact same principles as described above, but with the rate at which a novel mutant with a different value of t_H emerges denoted as μ_{t_H}.

5.7.3.1 Dynamics within a growth period

Upon arrival of each novel mutant, the invasion dynamics is simulated. To simulate the invasion dynamics, we first calculate the change in frequency of each genotype between two successive growth periods. For the iteroparous regime, the frequency of genotype i with $t_R = t_R^{(i)}$ in the subsequent growth period is given by

$$f_i^{'} = \frac{x_{00}^{(i)}(T) + x_1^{(i)}(T)}{\sum_{j=1}^{S}(x_{00}^{(j)}(T) + x_1^{(j)}(T))} \tag{5.68}$$

and for the semelparous regime, it is given by

$$f_i^{'} = \frac{\chi_{00}^{(i)}(T) + \chi_1^{(i)}(T) + y_{00}^{(i)}(T) + y_1^{(i)}(T)}{\sum_{j=1}^{S}(\chi_{00}^{(j)}(T) + \chi_1^{(j)}(T) + y_{00}^{(i)}(T) + y_1^{(i)}(T))} \tag{5.69}$$

and the input parameters of this function are $x_{\text{total}}(0)$, r, μ, k, α, P_R and T. The output of this function is $\mathbf{f}$, the frequency of all genotypes at the end of the growth period.

5.7.3.2 Invasion Dynamics

The invasion dynamics are run for one unit of τ i.e for G growth periods. The input parameters of this function are G, $x_{\text{total}}(0)$, r, μ, k, α, P_R and T and the output is $\mathbf{f}$, the frequency of all genotypes at the end of the growth period.

5.7.3.3 Evolutionary Dynamics

The evolutionary dynamics have the input parameters δt_R, $t_R(0)$, f_0, NEVOL, μ_{t_R} where δt_R is the mutational stepsize, $t_R(0)$ is the timing of recombination of the initial genotype in the simulation and f_0 is the initial frequency of a novel mutant, which is chosen to be small. All the remaining system parameters are as given in the Invasion Dynamics. The mutation acts to increase/decrease the value of t_R each with probability 1/2. Upon introduction of the novel mutant, the invasion dynamics is run for G subsequent growth periods, where G is generated from a $Geo(\mu_{t_R})$ distribution at frequency f_0. In the meantime, the mean value of t_R is recorded using Eq. (5.67). Next, the process of introducing a new mutant into the population is repeated. Since the population now has more than two genotypes, the genotype that mutates is chosen with probability weighted by the frequency of each genotype. There is now the possibility of back mutation to one of the existing genotypes. In this case, if f_k is frequency of this genotype before the introduction of an identical mutant, then the frequency of this genotype becomes $f_k + f_0$ after its introduction. Furthermore, a genotype is thought to be extinct if its frequency falls below the extinction threshold, in which case we remove that genotype.

5.7.4 ANALYTICAL APPROXIMATIONS BREAK DOWN FOR LARGE μ

Here, we demonstrate how our analytical approximations break down above a sufficiently large value of μ. In Figure 5.19 we see that for $\mu = 0.15$, the analytical approximation for $x_{00}(t)$, $x_1(t)$ and $x_{11}(t)$ breaks down drastically. In Figure 5.20 below, we show that for sufficiently large μ, our analytical approximation of $t_R^{(I)}$ no longer matches our simulations of evolutionary trajectories either. In fact, even for $\mu = 0.03$, our analytical approximation of the individual optimal timing of recombination $t_R^{(I)}$ no longer matches our stochastic simulation, despite the small value of α and P_R.

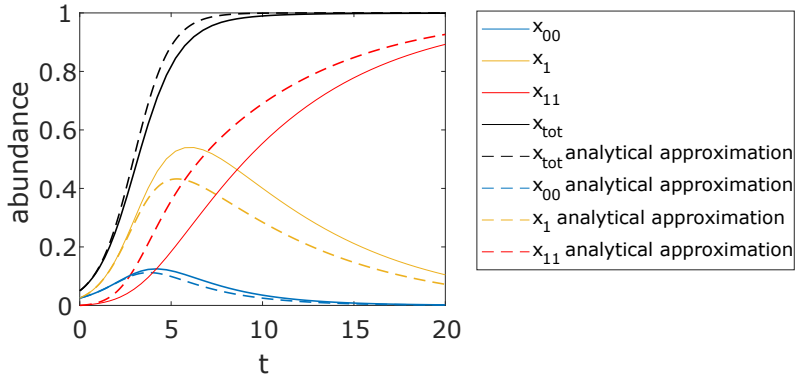

Figure 5.19: Change in the abundances $x_{00}(t)$ and $x_1(t)$ over one growth period as in Figure 5.18, but for $\mu = 0.15$ where our analytical approximation breaks down.

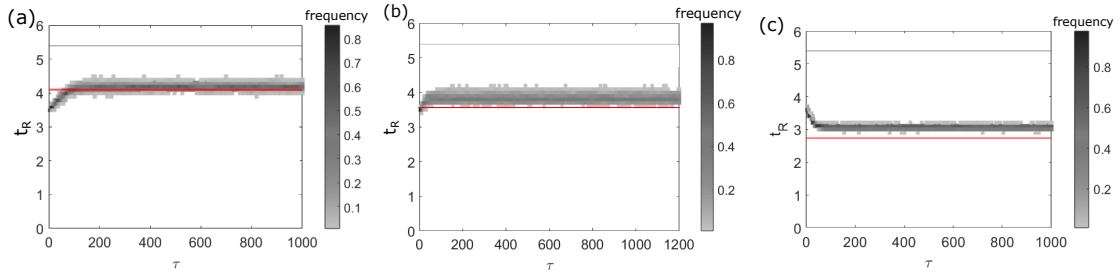

Figure 5.20: Analytical prediction of individual optimal timing of recombination $t_R^{(I)}$ (red line) breaks down when $\mu \approx 0.03$. In panel (a) $\mu = 0.01$ and in panel (b) $\mu = 0.03$ and panel (c) $\mu = 0.15$. Remaining parameters are $\alpha = 0.3$, $P_R = 0.1$, $r = 2$, $k = 1$, $x_{\text{total}}(0) = 0.05$ and $T = 5.4$.

5.7.5 EQUILIBRIUM FREQUENCIES OF NON MUTANT GENOTYPE

In the case where recombination an occur arbitrarily many times in a growth period, the equilibrium frequency of the $\binom{0}{0}$ genotype f_{00}^* tends to 1 as α or T tends to infinity. However in the semelparous regime, f_{00}^* tends to another limit smaller than 1 as α or T tends to infinity. In other words, the limit of Eq. (5.62) as α tends to infinity is

$$\begin{aligned} & e^{-t_R(r-2\mu)}\Big(- e^{rt_R}k(P_R-2)(t_R-T)x_{\text{total}}(0) + 2e^{t_R(r-\mu)}k(P_R-1)(t_R-T)x_{\text{total}}(0) \\ + \; & 2\sqrt{e^{2t_R(r-\mu)}(e^{t_R\mu}-1)k^2(-1+e^{t_R\mu}+P_R)(T-t_R)^2x_{\text{total}}(0)^2}\Big) \Big/ \Big(k(e^{2t_R\mu}(P_R-4) \\ - \; & 4e^{t_R\mu}(P_R-2)+4(P_R-1))(t_R-T)x_{\text{total}}(0)\Big) \end{aligned} \tag{5.70}$$

see Figure 5.22 for illustration, and the limit of Eq. (5.62) as T tends to infinity is

$$\begin{aligned} & e^{-t_R(r-2\mu)}\Big(- e^{rt_R}k(P_R-2)x_{\text{total}}(0)\alpha + 2e^{t_R(r-\mu)}k(P_R-1)x_{\text{total}}(0)\alpha \\ - \; & 2\sqrt{e^{2t_R(r-\mu)}(e^{t_R\mu}-1)k^2(-1+e^{t_R\mu}+P_R)x_{\text{total}}(0)^2\alpha^2}\Big)/\Big(k(e^{2t_R\mu}(P_R-4) \\ - \; & 4e^{t_R\mu}(P_R-2)+4(P_R-1))x_{\text{total}}(0)\alpha\Big) \end{aligned} \tag{5.71}$$

which is illustrated in Figure 5.23.

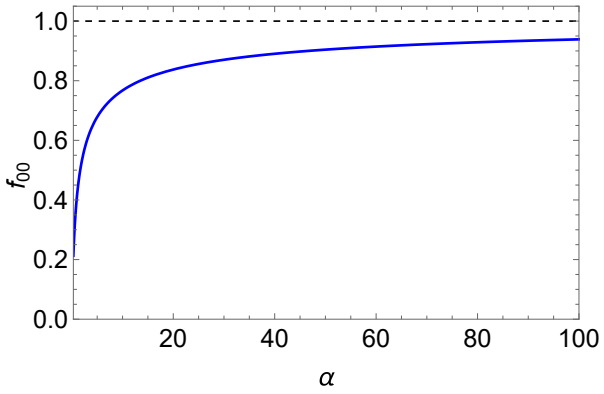

Figure 5.21: In the iteroparous regime, the equilibrium frequency of the $\binom{0}{0}$ genotype (f_{00}^*) tends to 1 as $\alpha \to \infty$ (dashed line), where f_{00}^* is given by Eq. (5.48). Here, $t_R = 4.11$, $T = 6$, $P_R = 0.1$ and all other parameters as in Figure 5.13.

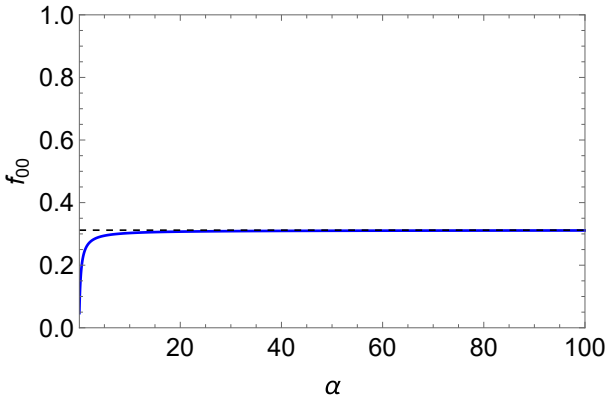

Figure 5.22: In the semelparous regime, the equilibrium frequency of the $\binom{0}{0}$ genotype (f_{00}^*) tends to a quantity smaller than 1 Eq. (5.70) as $\alpha \to \infty$ (dashed line), where f_{00}^* is given by Eq. (5.62). Here, $t_R = 4.11$, $T = 6$, $P_R = 0.1$ and all other parameters as in Figure 5.13

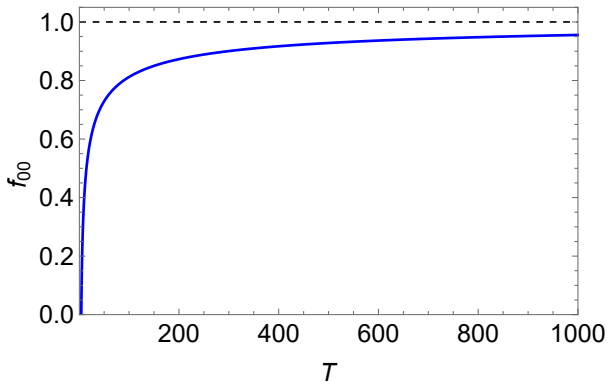

Figure 5.23: In the iteroparous regime, the equilibrium frequency of the $\binom{0}{0}$ genotype (f_{00}^*) tends to 1 as $T \to \infty$ (dashed line), where f_{00}^* is given by Eq. (5.48). Here, $t_R = 4.11$, $\alpha = 0.3$, $P_R = 0.1$ and all other parameters as in Figure 5.13.

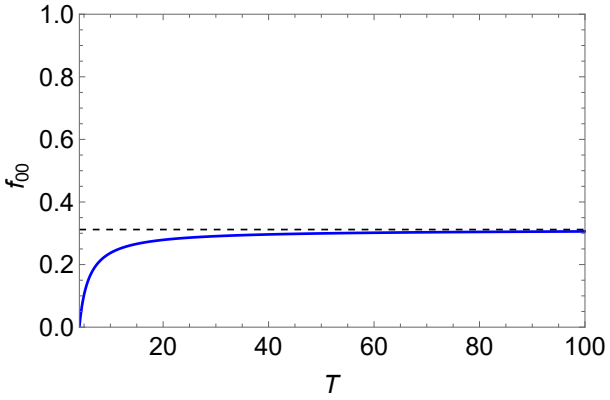

Figure 5.24: In the semelparous regime, the equilibrium frequency of the $\binom{0}{0}$ genotype (f_{00}^*) tends to a quantity smaller than 1 Eq. (5.70) as $T \to \infty$ (dashed line), where f_{00}^* is given by Eq. (5.62). Here, $t_R = 4.11$, $\alpha = 0.3$, $P_R = 0.1$ and all other parameters as in Figure 5.13

5.7.6 INDIVIDUAL OPTIMAL TIMING OF RECOMBINATION IS IDENTICAL BETWEEN ITEROPAROUS AND SEMELPAROUS REGIMES

Here, we detail why our analytical prediction for $t_R^{(I)}$ is identical in both the iteroparous and semelparous regimes. In both regimes, we note that during each genetic recombination event, the rate at which $\binom{0}{0}$ increases is exactly identical to the rate at which $\binom{1}{1}$ increases. The ability for recombination to reverse the depletion of $\binom{0}{0}$ genotypes is exactly equal to the ability to produce $\binom{1}{1}$ as byproducts. This is

known as the seesaw effect [234]. In the iteroparous regime, we have

$$\frac{\mathrm{d}x_{00}}{\mathrm{d}t} = \frac{\mathrm{d}x_{11}}{\mathrm{d}t} \tag{5.72}$$

and in the semelparous regime, we have

$$\frac{\mathrm{d}(\chi_{00}+y_{00})}{\mathrm{d}t} = \frac{\mathrm{d}(\chi_{11}+y_{11})}{\mathrm{d}t}\,. \tag{5.73}$$

We also demonstrate numerically that if the seesaw effect is violated, the value of $t_R^{(I)}$ ceases to be identical in the two regimes.

5.8 APPENDIX IV: CHEATING IN RECOMBINATION MODEL

Here we consider the case where the population contains a rare mutant with a different timing of recombination t_R to the resident, and both the resident and mutant are in recombination phase. In such case, both the mutant and resident are helping each other out in reversing the buildup of deleterious mutations, and we see cooperation. We demonstrate how this scenario involves cheating if the mutant and resident frequencies differ markedly in order of magnitude.

First, we consider the effect that the rare mutant (with frequency 10^{-2} or less as rule of thumb) has on the resident population dynamics using Eq. (5.74), which shows the resident population dynamics when both the resident and mutant are in recombination. Denoting the resident's abundance of $\binom{0}{0}$, $\binom{0}{1}$, $\binom{1}{0}$ and $\binom{1}{1}$ genotypes as x_{00}, x_{01}, x_{10} and x_{11} and the mutant's corresponding abundance as $\hat{x}_{00}$, $\hat{x}_{01}$, $\hat{x}_{10}$ and $\hat{x}_{11}$, we have

$$\begin{aligned}
\frac{\mathrm{d}x_{00}}{\mathrm{d}t} &= \frac{\alpha P_R}{3}\left(x_{01}\hat{x}_{10}+x_{10}\hat{x}_{01}\right)+\frac{2\alpha P_R}{3}x_{01}x_{10} \\
&\quad +\frac{2\alpha P_R}{3}\left(x_{01}\hat{x}_{00}+x_{10}\hat{x}_{00}\right)-\frac{2\alpha P_R}{3}\left(x_{00}\hat{x}_{01}+x_{00}\hat{x}_{01}\right) \\
\frac{\mathrm{d}x_{01}}{\mathrm{d}t} &= \frac{2\alpha P_R}{3}x_{00}\hat{x}_{01}-\frac{2\alpha P_R}{3}x_{01}\hat{x}_{00}-\alpha P_R x_{01}\hat{x}_{10}+\frac{\alpha P_R}{3}x_{10}\hat{x}_{01} \\
&\quad -\frac{2\alpha P_R}{3}x_{01}x_{10} \\
\frac{\mathrm{d}x_{10}}{\mathrm{d}t} &= \frac{2\alpha P_R}{3}x_{00}\hat{x}_{10}-\frac{2\alpha P_R}{3}x_{10}\hat{x}_{00}-\alpha P_R x_{10}\hat{x}_{01}+\frac{\alpha P_R}{3}x_{01}\hat{x}_{10} \\
&\quad -\frac{2\alpha P_R}{3}x_{01}x_{10} \\
\frac{\mathrm{d}x_{11}}{\mathrm{d}t} &= \frac{\alpha P_R}{3}\left(x_{01}\hat{x}_{10}+x_{10}\hat{x}_{01}\right)+\frac{2\alpha P_R}{3}x_{01}x_{10}\,.
\end{aligned} \tag{5.74}$$

If the frequency of the mutant is small, $\hat{x}_{00}$, $\hat{x}_{01}$, $\hat{x}_{10}$ and $\hat{x}_{11}$ would be an order of magnitude smaller than x_{00}, x_{01}, x_{10} and x_{11}, and so terms involving products of mutant abundances with resident abundance (e.g $x_{01}\hat{x}_{10}$) would be of smaller magnitude than terms involving products of resident abundance with itself (e.g $x_{01}x_{10}$), illustrating the small effect that mutants have on the resident population dynamics. In other words, the mutants have an extremely small beneficial effect on the resident here. Now we consider the mutant population dynamics

$$\begin{aligned}
\frac{\mathrm{d}\hat{x}_{00}}{\mathrm{d}t} &= \frac{\alpha P_R}{3}\left(x_{01}\hat{x}_{01} + x_{10}\hat{x}_{01}\right) + \frac{2\alpha P_R}{3}\hat{x}_{01}\hat{x}_{10} \\
&\quad + \frac{2\alpha P_R}{3}\left(x_{00}\hat{x}_{01} + x_{00}\hat{x}_{10}\right) - \frac{2\alpha P_R}{3}\left(\hat{x}_{00}x_{01} + \hat{x}_{00}x_{10}\right) \\
\frac{\mathrm{d}\hat{x}_{01}}{\mathrm{d}t} &= \frac{2\alpha P_R}{3}\hat{x}_{00}x_{01} - \frac{2\alpha P_R}{3}x_{00}\hat{x}_{01} - \alpha P_R x_{10}\hat{x}_{01} + \frac{\alpha P_R}{3}x_{01}\hat{x}_{10} \\
&\quad - \frac{2\alpha P_R}{3}\hat{x}_{01}\hat{x}_{10} \\
\frac{\mathrm{d}\hat{x}_{10}}{\mathrm{d}t} &= \frac{2\alpha P_R}{3}\hat{x}_{00}x_{10} - \frac{2\alpha P_R}{3}x_{00}\hat{x}_{10} - \alpha P_R x_{01}\hat{x}_{10} + \frac{\alpha P_R}{3}x_{10}\hat{x}_{01} \\
&\quad - \frac{2\alpha P_R}{3}\hat{x}_{01}\hat{x}_{10} \\
\frac{\mathrm{d}\hat{x}_{11}}{\mathrm{d}t} &= \frac{\alpha P_R}{3}\left(x_{01}\hat{x}_{01} + x_{10}\hat{x}_{01}\right) + \frac{2\alpha P_R}{3}\hat{x}_{01}\hat{x}_{10}\,.
\end{aligned} \tag{5.75}$$

Here, terms involving products of mutant and resident abundances (e.g $\hat{x}_{01}x_{10}$) are of greater magnitude than those involving products of mutant abundances with themselves (e.g $\hat{x}_{01}\hat{x}_{10}$), so it is the residents that have the greatest effect on the mutant population dynamics as opposed to the mutants themselves. As the benefit received by the mutant from the resident well exceeds the negligible benefit the mutant provides to the resident in return, this constitutes cheating.

5.9 APPENDIX V: PARAMETERS FOR NUMERICAL SIMULATIONS

Parameters for Figure 5.4 are $r = 2$, $k = 1$, $x_{\text{total}}(0) = 0.05$, $T = 20$, $\gamma = 0.3$.

Parameters for Figure 5.7 are $T = 15$, $\mu = 0.03$, $\alpha = 0.3$, $P_R = 0.75$, $x_{\text{total}}(0) = 0.05$, $r = 2$, $k = 1$ and $f_{00}(0) = 1$.

Parameters for Figure 5.8 are $r = 1$, $k = 1$, $\alpha = 0.3$, $P_R = 0.1$, $T = 15$, $x_{\text{total}}(0) =$

0.05, $\mu = 0.01$. In panel (b), $t_R = 5$.

Parameters in Figure 5.11 are $r = 2$, $k = 1$, $\gamma = 0.3$ and $x_{\text{total}}(0) = 0.05$ and simulation parameters are NEVOL = 2000, $\mu_{t_H} = 0.005$, $\delta = 0.05$, $f_0 = 8 \times 10^{-4}$, $t_H(0) = 3.6$ and extinction threshold is 5×10^{-4}.

Parameters in Figure 5.12 are $r = 1$, $k = 1$, $\gamma = 0.2$ and simulation parameters are NEVOL = 2000, $\delta = 0.05$, $\mu_{t_H} = 0.005$, $f_0 = 8 \times 10^{-4}$ and extinction threshold is 5×10^{-4}.

Parameters in Figure 5.13 are $r = 2$, $k = 1$, $\mu = 0.01$, $\alpha = 0.3$ and $P_R = 0.1$. Simulation parameters are NEVOL = 2000, $\delta = 0.1$, $\mu_{t_R} = 0.005$, $f_0 = 0.011$ and extinction threshold 0.01.

Parameters for Figure 5.14 are $T = 10$, $\alpha = 1000$, $P_R = 0.75$, and $t_R = t_R^{(I)}$ as defined in Eq. (5.9), with remaining parameters as in Figure 5.13.

Parameters for Figure 5.15 are $P_R = 0.75$, $T = 5.4$, $r = 2$, $k = 1$, $\mu = 0.01$ and $x_{\text{total}}(0) = 0.05$. Simulation parameters are NEVOL = 1000, $\delta t_R = 0.1$, $\mu_{t_R} = 0.005$, $f_0 = 0.011$ and extinction threshold is 0.01.

— 6 —

Discussion

In this chapter, we begin by discussing the results and wider implications of each chapter in isolation before finishing with a holistic summary of our research and its future perspectives. We begin by describing the key insights of the results for each chapter along with some future research directions based on the results.

6.1 DISCUSSING THE RESULTS OF CHAPTER 2

6.1.1 Key insights and context within research literature

The key insight of this chapter is that we have developed an analytical approximation of the feasibility probability for random ecological systems with ecologically motivated interaction structures, which combines the feasibility based approach of [210] with the RMT approach of [5].

Ecologically, a feasible (biologically meaningful) GLV system represents a community of interacting species [210]. Since feasible GLV systems in our model are almost always stable [209], the key implication of our result is on how interaction structures impact the resilience of stable communities rather than the effect of interaction structures on ecological stability itself. A biological implication of increased resilience is that communities can recover more rapidly following changes in species abundances [210], which can arise due to environmental disturbances.

As an increase in feasibility probability corresponds to a reduced likelihood of species extinctions following environmental disturbances [210] , our results demonstrate that biasing the population towards a given interaction type has a dual influence

on both the community resilience and community persistence, which considers its ability to maintain a given level of species diversity.

Although our results, along with the numerous existing results from GLV and RMT models [5, 146, 210] imply that stability decreases with diversity, it is actually observed that many of the world's largest ecological systems are in fact the most stable [88]. This contradiction is likely due to the fact that certain aspects of macroecological realism are neglected in the models for those existing results. One neglected aspect is interspecific predator interference which weakens interactions between predators due to prey sharing, and this plays a role in creating a positive diversity-stability relationship [152]. Allometrically linking metabolic rates to body mass at different trophic levels has also been shown to stabilise diverse communities [23]. Furthermore, the emergence of larger stable communities have also been shown to be promoted by the presence of stage-structure in populations [190].

Even though the aforementioned factors may all play a role in promoting a positive diversity-stability relationship, they do not explain how diversity may be the cause of stability. Recent empirical results have shown that many natural communities such as bird and fish communities exhibit sublinear population growth [88]. Under such a regime, it has been shown analytically that increased diversity may be the direct cause of stability [88].

6.1.2 FUTURE RESEARCH PRIORITIES

In contrast to macroecological systems, purely microbial ecosystems have a much simpler trophic structure and negligible size structure, and thus simpler metabolic scaling. This reduces the number of factors that play a role in creating a positive complexity-stability relationship, which makes microbial systems potentially more amenable to analysis using our model alongside existing RMT based approaches mentioned above. Indeed, the assumption of linear population growth has been observed in experimental microbial microcosms [192]. Excitingly, an empirical network reconstructed from natural human and sponge microbiome communities [235] has shown that the number of microbial species trades off with the number of interspecific interactions, thereby providing empirical evidence for May's complexity-stability criterion.

Having predicted theoretically how the interaction structure impacts ecological resilience, there is scope for future work to experimentally test such a prediction.

This has promising potential since experimental microcosms have already reached the necessary controllability to test a large number of theoretical predictions [184], and have even been able to successfully replicate several aspects of the complexity-stability relation [98]. In an experimental bacterial microcosm by Hu *et al.* [98], the fraction of surviving bacterial species was shown to decrease with interaction strength and number of species. The more generalised the ecological models become, the more novel hypotheses and model predictions we could come up with, which certainly opens up the possibility for more experimental inquiries in the future.

Throughout this chapter, our work has been centered around RMT based ecological networks which assume a random interaction profile. It is worth noting that ecological interaction networks are not random since they evolve and form piecewise. Another useful area for future consideration would be to investigate how the structure of the interaction network changes as an ecosystem evolves. It has been shown by [149] that evolution can impact the effect of dispersal network structure on the persistence of metapopulations, which has become increasingly important in the context of coral reef conservation [148]. The sheer size of macroecological networks mean that system dynamics techniques can be computationally expensive, and it would be more plausible to apply techniques such as individual based modelling.

6.2 DISCUSSING THE RESULTS OF CHAPTER 3

6.2.1 Key insights and context within research literature

One of the main insights gained from this chapter is that by co-evolving mass and fertilisation rate α in switching environments, we observe a broad range of dynamical behaviour that can explain a variety of reproductive modes. In fixed environments, obligate sexual reproduction with oogamy and obligate asexual reproduction (Figures 3.2,3.3 and Eq. (3.9)) are the only evolutionary outcomes. Our stochastic simulations of evolutionary trajectories have shown that oogamy can evolve from anisogamy through the ecological phenomena of sexual conflict, while anisogamy results from evolutionary branching in gamete mass.

Empirically we observe that the degree of anisogamy varies with environmental conditions. Gametes exhibit a greater degree of anisogamy in more turbulent gamete limited environments [132, 218]. Our model on the other hand suggests that the evolution of anisogamy from isogamy can be arrested under extreme gamete

limitation (small α). However it is worth noting that the greater anisogamy in benthic green algae can also arise due to the harsher environmental conditions in deep water environments [132]. This statement is also in line with our model, where a greater degree of anisogamy is favoured upon increasing the environmental adversity parameter β.

The smaller degree of anisogamy in shallow water green algae species may also be attributed to their photosynthetic needs, which necessitates the presence of organelles such as eyespots (*Chlamydomonas* [218]) or chlorophylls (volvacine green algae [199]), which place a lower limit on gamete size.

6.2.2 Bridge between Chapters 3 and 4

Throughout Chapter 3, we have analysed the conditions that give rise to the evolution of anisogamy and oogamy, which is the root to our understanding of the evolution of the sexes (males produce microgametes and females produce macrogametes) [125]. However, a topic that has been less emphasised in this chapter is the conditions that bring about the evolutionary transition from asexual to sexual reproduction. This can raise intriguing questions about the evolutionary origins of sex itself. A key mechanistic benefit of sexual reproduction is that it involves the fusion of cells, and fused cells have greater mass [222] which facilitates the survival of the organism in harsh environmental conditions.

Another neglected aspect in this chapter is phenotypic plasticity, where the population can adopt different strategies in different environments. Phenotypic plasticity is observed in many organisms exhibiting facultative responses in varying environments, such as *D. discoideum* [194] where cell aggregation occurs only at times of environmental stress. Furthermore, it has been suggested by [58] that the selection pressures induced by environmental switching could even be a trigger for the evolution of phenotypic plasticity in organisms.

In Chapter 4, we show how binary cell fusion can evolve as a response to environmental stress in the presence of phenotypic plasticity. Having demonstrated how switching environments can give rise to the evolution of facultative sex in the absence of phenotypic plasticity, we show that when phenotypic plasticity is present, facultative binary cell fusion can also evolve in the face of switching environments.

6.3 DISCUSSING THE RESULTS OF CHAPTER 4

6.3.1 Key insights and context within research literature

In this chapter, we have provided a novel explanation for the mechanistic advantages of sex and multicellularity due to the increased mass conferred by binary cell fusion. We have hereby demonstrated an alternative mechanism by which environmentally triggered facultative sex and multicellularity can evolve in organisms with phenotypic plasticity, without assuming any genetic aspects. Should a benign environment become sufficiently harsh (i.e if the harsh environment has a sufficiently large value of the environmental harshness parameter β), then cell fusion will be selected for in this harsh environment. Since the state of no cell fusion is an evolutionarily stable state in the benign environment, the organism is expected to revert back to that state when the environment switches back to being benign.

Our model has provided an explanation for why facultative multicellularity may evolve as an alternative stress response to facultative sex, which is observed in organisms such as *D. discodeium* [30]. The fact that binary cell fusion is favoured in harsh environments where hibernation/dormancy tends to occur in organisms also provides a hint on why sex and hibernation are frequently observed to co-occur [20, 62]. In Chapter 5 we have provided a more mathematically explicit explanation for this co-occurrence (see Section 6.4).

Our models in Chapters 3 and 4 have only considered the evolution of sexual reproduction in terms of the mechanistic benefits of cell-fusion, without accounting for any genetic factors or population dynamics. In Chapter 5, we developed an eco-evolutionary model for genetic recombination, which is another process that confers many selective advantages to sexual reproduction [154, 221].

6.3.2 Future research priorities

Capitalising on the idea of [155], we have been able to apply adaptive dynamics techniques generalised to the case of switching environments in our models (we abbreviate this generalisation as ADSE). To this end, a useful area of future work would be to incorporate environmental switching into other adaptive dynamics models alike, such as models used to study the evolution of anisogamy in populations with two mating types [124, 130]. By doing so, we are incorporating ADSE to a

more biologically realistic modelling regime since it is most commonly observed that mating types evolved before anisogamy.

Thus far, our models have been based on organisms with external fertilisation. Another area where we could incorporate ADSE would be in organisms with internal fertilisation [132] where sperms are released as discrete packets to fertilise with eggs. It would also be important to complement such models with simulations of evolutionary trajectories, as that enables us to reveal richer evolutionary dynamics that arise due to ecological interactions between multiple genotypes in the population, such as evolutionary branching.

6.4 DISCUSSING THE RESULTS OF CHAPTER 5

6.4.1 KEY INSIGHTS AND CONTEXT WITHIN RESEARCH LITERATURE

By developing an eco-evolutionary model that accounts for biological realism in the population dynamics, we have been able to demonstrate the selective advantages of both environmentally triggered genetic recombination and hibernation. In particular, we have shown how genetic recombination and hibernation both evolve to occur shortly before the end of a growth period (where resources run out or population density becomes saturated), when opportunity costs of population growth are diminished. We have hereby provided an analytical explanation for how these two traits evolved to co-occur. Although the positive covariance between these two traits has been shown by an agent based simulation by [62], our model provides a more analytically explicit explanation of this positive association by justifying how both traits can be triggered by the same environmental cue.

Even though our model justifies how hibernation and genetic recombination can be positively associated, it does not exclude the possibility that they can evolve individually as alternative stress responses, nor whether they may act antagonistically. Indeed, a small number of species such as *D. pulex* has been observed to produce dormant eggs asexually [62, 104]. Seeking an answer to this would require us to combine hibernation and recombination into one model.

The focus of our work has been on the evolutionary benefits of hibernation as opposed to the exact mechanism and molecular basis for hibernation. Traditionally, hibernation is believed to be a process whereby enzymes simply become idle in adverse environments. Recent studies have shown that hibernation is caused by

hibernation factor proteins that inhibit protein synthesis by binding to ribosomes [90]. Experimental work by [90] has identified a specific hibernation factor protein Balon which is responsible for inducing hibernation in the cold adapted bacteria *P. urativorans* during environmental stress. This together with our model highlights the need to better understand the physical mechanisms along with the evolutionary basis of hibernation.

In addition to meiosis and genetic recombination, organisms can exhibit other stress induced responses such as parasexuality, which frequently occurs in fungi [14] when challenged by the host immune system. Parasexuality is a reproductive mechanism where exchange of genetic material occurs without meiosis or fertilisation, and can involve mating between two diploids to produce tetraploids. These tetraploids then need to undergo concerned chromosomal loss to restore diploidy, however concerned chromosomal loss frequently leads to aneuploidy, where an abnormal number of chromosomes is present in a diploid set. Aneuploidy is known to enhance adaptation not only does it increase genetic diversity through copy number variation at a given locus (e.g through trisomic loci), but can increase genetic stability under chromosomal imbalances, which facilitates adaptation further. Despite its adaptive advantages, parasexuality occurs very rarely in nature due to the tremendous costs, including the need to undergo cocnerted chromosome loss [55].

6.4.2 Future research priorities

In nature, the timing of growing seasons can vary due to the unpredictability of the environment [61, 205]. This would constitute a variation in the length of the growth period. Models with varying season lengths [56, 61] have shown that organisms may adopt a bet-hedging strategy to undergo sex earlier under unpredictable season lengths. Accounting for fluctuations in growth period length in our model would thus be another fruitful area for future consideration. Given that the frequencies of each genotype in our model depends on the growth period length, incorporating these fluctuations into our model is likely to result in Red Queen dynamics, with constantly fluctuating genotype frequencies.

6.5 INTEGRATED SYNTHESIS OF CHAPTERS 3, 4 AND 5

In this thesis, we have been merely concerned with the selective forces driving the evolution of sex and its various modes as opposed to the physical origin of sex. It is worth emphasising that cell fusion is not a pre-requisite for genetic recombination to evolve, as evident in the processes of genetic transformation in bacteria [185], and parasexuality in fungi. In fact, one theory for the physical origin of meiotic sex known as "viral eukaryogenesis" posits that meiosis resulted from a symbiotic relationship of bacteria, archaea and a lysogenic DNA virus [13] in such a way that caused cells to fuse together. Crossover (genetic recombination between homologous chromosomes) occurred following cell fusion when the opportunity came. Crucially, our work provides impetus for developing further, refined mechanistic explanations for aspects concerning the evolution and the maintenance of sex [207].

Another fruitful area for future work would be to combine aspects of our recombination model in Chapter 5 with those of our model for cell fusion in Chapters 3 and 4. Doing so enables us to develop an integrative model that accounts for multiple selective advantages of sexual reproduction together. Given the complexity of the integrative model, one approach would be to utilise an agent based model, which can account for a large number of variables.

6.6 CLOSING REMARKS

One of the key insights of this thesis is that we have demonstrated how the two intrinsic mechanisms of sex; cell fusion (Chapter 4) and genetic recombination (Chapter 5) can both be triggered by environmental stress. This not only strengthens our argument for why sex often co-occurs with dormancy, but also the evolutionary benefits of facultative sex. In other words, it is more evolutionarily favorable for organisms to switch between sexual and asexual reproduction than to adopt a single reproductive strategy, due to varying environmental conditions.

In Chapter 2, we investigated the effect that a large number of interacting species can have on the dynamical behaviour of a macroecological system. In Chapters 3 and 4, we investigated how evolution can act on traits if the population contains genotypes with multiple different trait values. By doing so, we have shown explicitly how long

term evolutionary trajectories can be influenced by their underlying ecological phenomena, which is the key essence of eco-evolutionary modelling.

Finally in Chapter 5, we developed an eco-evolutionary model that accounts for population dynamics, discrete mutations in the population genetic traits as well as the ecological aspects of evolution in a population with multiple continuous trait values. We have hereby developed an integrative evolutionary model for a research problem in the evolution of sex. As a future perspective, it would be useful to apply such integrative modelling techniques to different problems in the evolution of sex literature, or other evolutionary problems alike.

Throughout the thesis, I have developed a variety of mathematical techniques, in particular techniques in ecological random matrix theory and the analytical approximation of evolutionary dynamics in multi-timescale adaptive dynamics models. I also got the opportunity to numerically simulate evolutionary trajectories by applying stochastic models.

Eco-evolutionary modelling has also found applications in a range of other areas as well including microbiome research [39] and cancer research. In cancers, enhanced cell proliferation and drug resistance [11] are traditionally known to be driven by mutations, however recent work has shown that they can be mediated by their interactions between different cellular and non-cellular components within the tumor ecosystem, known as the "tumor microenvironment" [176], complicating their eco-evolutionary feedback loop. Above all, it would be of great interest to apply the eco-evolutionary modelling techniques acquired in this thesis to these new exciting areas.

References

[1] L. Adams, K. Lyons, J. Monday, E. Larkin, and J. Wyffels. "Costs of parthenogenesis on growth and longevity in ex situ zebra sharks Stegostoma tigrinum". In: *Endangered Species Research* 50 (2023), pp. 81–91.

[2] A. F. Agrawal. "Evolution of sex: why do organisms shuffle their genotypes?" In: *Current Biology* 16.17 (2006), R696–R704.

[3] I. Akjouj and J. Najim. "Feasibility of sparse large Lotka-Volterra ecosystems". In: *Journal of Mathematical Biology* 85.6-7 (2022), p. 66.

[4] S. Allesina and S. Tang. "Stability criteria for complex ecosystems". In: *Nature* 483.7388 (2012), pp. 205–208.

[5] S. Allesina and S. Tang. "The stability–complexity relationship at age 40: a random matrix perspective". In: *Population Ecology* 57.1 (2015), pp. 63–75.

[6] T. W. Anderson, T. W. Anderson, T. W. Anderson, and T. W. Anderson. *An introduction to multivariate statistical analysis*. Vol. 2. Wiley New York, 1958.

[7] L. H. Antão, A. E. Magurran, and M. Dornelas. "The shape of species abundance distributions across spatial scales". In: *Frontiers in Ecology and Evolution* 9 (2021), p. 626730.

[8] G. Barabás, M. J. Michalska-Smith, and S. Allesina. "Self-regulation and the stability of large ecological networks". In: *Nature ecology & evolution* 1.12 (2017), pp. 1870–1875.

[9] J. W. Baron, T. J. Jewell, C. Ryder, and T. Galla. "Eigenvalues of Random Matrices with Generalized Correlations: A Path Integral Approach". In: *Physical Review Letters* 128.12 (2022), p. 120601.

[10] N. H. Barton and B. Charlesworth. "Why sex and recombination?" In: *Science* 281.5385 (1998), pp. 1986–1990.

[11] D. Basanta and A. R. Anderson. "Exploiting ecological principles to better understand cancer progression and treatment". In: *Interface focus* 3.4 (2013), p. 20130020.

[12] G. Bell. "The evolution of anisogamy". In: *Journal of Theoretical Biology* 73.2 (1978), pp. 247–270.

[13] P. J. L. Bell. "Sex and the eukaryotic cell cycle is consistent with a viral ancestry for the eukaryotic nucleus". In: *Journal of theoretical biology* 243.1 (2006), pp. 54–63.

[14] J. Berman and L. Hadany. "Does stress induce (para) sex? Implications for Candida albicans evolution". In: *Trends in Genetics* 28.5 (2012), pp. 197–203.

[15] E. Berríos-Caro, T. Galla, and G. W. A. Constable. "Switching environments, synchronous sex, and the evolution of mating types". In: *Theoretical population biology* 138 (2021), pp. 28–42.

[16] S. Billiard, M. López-Villavicencio, M. Hood, and T. Giraud. "Sex, outcrossing and mating types: unsolved questions in fungi and beyond". In: *Journal of evolutionary biology* 25.6 (2012), pp. 1020–1038.

[17] S. Billiard, M. López-Villavicencio, B. Devier, M. E. Hood, C. Fairhead, and T. Giraud. "Having sex, yes, but with whom? Inferences from fungi on the evolution of anisogamy and mating types". In: *Biological reviews* 86.2 (2011), pp. 421–442.

[18] C. Bishop. "Life cycle control of Chlamydomonas reinhardtii". In: *Genome Biology* 4.1 (2003), pp. 1–2.

[19] P. Bizeul and J. Najim. "Positive solutions for large random linear systems". In: *Proceedings of the American Mathematical Society* 149.6 (2021), pp. 2333–2348.

[20] J. T. Bonner. "The origins of multicellularity". In: *Integrative Biology: Issues, News, and Reviews: Published in Association with The Society for Integrative and Comparative Biology* 1.1 (1998), pp. 27–36.

[21] J. T. Bonner. "The relation of spore formation to recombination". In: *The American Naturalist* 92.865 (1958), pp. 193–200.

[22] Å. Brännström, J. Johansson, and N. Von Festenberg. "The hitchhiker's guide to adaptive dynamics". In: *Games* 4.3 (2013), pp. 304–328.

[23] U. Brose, R. J. Williams, and N. D. Martinez. "Allometric scaling enhances stability in complex food webs". In: *Ecology letters* 9.11 (2006), pp. 1228–1236.

[24] J. H. Brown. "Macroecology: progress and prospect". In: *Oikos* (1999), pp. 3–14.

[25] M. G. Bulmer and G. A. Parker. "The evolution of anisogamy: a game-theoretic approach". In: *Proc. R. Soc. Lond. B.* 269 (2002), pp. 2381–2388.

[26] G. Bunin. "Ecological communities with Lotka-Volterra dynamics". In: *Physical Review E* 95.4 (2017), p. 042414.

[27] M. K. Burak, J. D. Monk, and O. J. Schmitz. "Focus: ecology and evolution: eco-evolutionary dynamics: the predator-prey adaptive play and the ecological theater". In: *The Yale Journal of Biology and Medicine* 91.4 (2018), p. 481.

[28] R. Bürger. "Evolution of genetic variability and the advantage of sex and recombination in changing environments". In: *Genetics* 153.2 (1999), pp. 1055–1069.

[29] N. W. Burke and R. Bonduriansky. "Sexual conflict, facultative asexuality, and the true paradox of sex". In: *Trends in Ecology & Evolution* 32.9 (2017), pp. 646–652.

[30] F. de Carpentier, S. D. Lemaire, and A. Danon. "When unity is strength: the strategies used by Chlamydomonas to survive environmental stresses". In: *Cells* 8.11 (2019), p. 1307.

[31] B. Charlesworth. "Recombination modification in a fluctuating environment". In: *Genetics* 83.1 (1976), pp. 181–195.

[32] B. Charlesworth. "The population genetics of anisogamy". In: *Journal of theoretical biology* 73.2 (1978), pp. 347–357.

[33] F. B. Christiansen, S. P. Otto, A. Bergman, and M. W. Feldman. "Waiting with and without recombination: the time to production of a double mutant". In: *Theoretical population biology* 53.3 (1998), pp. 199–215.

[34] M. Clenet, H. El Ferchichi, and J. Najim. "Equilibrium in a large Lotka–Volterra system with pairwise correlated interactions". In: *Stochastic Processes and their Applications* 153 (2022), pp. 423–444.

[35] J. Cohen, T. Łuczak, C. Newman, and Z.-M. Zhou. "Stochastic structure and nonlinear dynamics of food webs: qualitative stability in a Lotka-Volterra cascade model". In: *Proceedings of the Royal Society of London. B. Biological Sciences* 240.1299 (1990), pp. 607–627.

[36] G. Constable and H. Kokko. "The rate of facultative sex governs the number of expected mating types in isogamous species". In: *Nat. Ecol. Evol.* 2.7 (2018), pp. 1168–1175.

[37] G. W. Constable and H. Kokko. "Parthenogenesis and the Evolution of Anisogamy". In: *Cells* 10.9 (2021), p. 2467.

[38] P. A. Cox and J. A. Sethian. "Gamete motion, search, and the evolution of anisogamy, oogamy, and chemotaxis". In: *The American Naturalist* 125.1 (1985), pp. 74–101.

[39] K. Z. Coyte, J. Schluter, and K. R. Foster. "The ecology of the microbiome: networks, competition, and stability". In: *Science* 350.6261 (2015), pp. 663–666.

[40] J. Crow. "An advantage of sexual reproduction in a rapidly changing environment". In: *Journal of Heredity* 83.3 (1992), pp. 169–173.

[41] R. N. Curnow and C. W. Dunnett. "The numerical evaluation of certain multivariate normal integrals". In: *The Annals of Mathematical Statistics* 33.2 (1962), pp. 571–579.

[42] P. Czuppon and G. W. A. Constable. "Invasion and extinction dynamics of mating types under facultative sexual reproduction". In: *Genetics* 213.2 (2019), pp. 567–580.

[43] C. Darwin. *The descent of man, and selection in relation to sex*. Vol. 2. D. Appleton, 1872.

[44] M. H. Davis. "Piecewise-deterministic Markov processes: A general class of non-diffusion stochastic models". In: *Journal of the Royal Statistical Society: Series B (Methodological)* 46.3 (1984), pp. 353–376.

[45] T. C. Day, P. Márquez-Zacarías, P. Bravo, A. R. Pokhrel, K. A. MacGillivray, W. C. Ratcliff, and P. J. Yunker. "Varied solutions to multicellularity: The biophysical and evolutionary consequences of diverse intercellular bonds". In: *Biophysics reviews* 3.2 (2022), p. 021305.

[46] F. Dercole, F. Della Rossa, and P. Landi. "The transition from evolutionary stability to branching: A catastrophic evolutionary shift". In: *Scientific reports* 6.1 (2016), pp. 1–8.

[47] C. P. Doncaster, G. E. Pound, and S. J. Cox. "The ecological cost of sex". In: *Nature* 404.6775 (2000), pp. 281–285.

[48] T. Douglas, D. C. Queller, and J. E. Strassmann. "Social amoebae mating types do not invest unequally in sexual offspring". In: *Journal of evolutionary biology* 30.5 (2017), pp. 926–937.

[49] M. Dougoud, L. Vinckenbosch, R. P. Rohr, L.-F. Bersier, and C. Mazza. "The feasibility of equilibria in large ecosystems: A primary but neglected concept in the complexity-stability debate". In: *PLoS computational biology* 14.2 (2018), e1005988.

[50] D. B. Dusenbery. "Selection for high gamete encounter rates explains the success of male and female mating types". In: *Journal of Theoretical Biology* 202.1 (2000), pp. 1–10.

[51] L. Eme, A. Spang, J. Lombard, C. W. Stairs, and T. J. Ettema. "Archaea and the origin of eukaryotes". In: *Nature Reviews Microbiology* 15.12 (2017), pp. 711–723.

[52] L. Eme, D. Tamarit, E. F. Caceres, C. Stairs, V. de Anda, M.-E. Schon, K. Seitz, N. Dombrowski, W. H. Lewis, J. H. Saw, et al. "Inference and reconstruction of the heimdallarchaeial ancestry of eukaryotes". In: *BioRxiv* (2023), pp. 2023–03.

[53] R. Ferriere and S. Legendre. "Eco-evolutionary feedbacks, adaptive dynamics and evolutionary rescue theory". In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 368.1610 (2013), p. 20120081.

[54] R. A. Fisher. *The genetical theory of natural selection. Clarendon.* 1930.

[55] A. Forche, K. Alby, D. Schaefer, A. D. Johnson, J. Berman, and R. J. Bennett. "The parasexual cycle in Candida albicans provides an alternative pathway to meiosis for the formation of recombinant strains". In: *PLoS biology* 6.5 (2008), e110.

[56] A. I. Furness, K. Lee, and D. N. Reznick. "Adaptation in a variable environment: Phenotypic plasticity and bet-hedging during egg diapause and hatching in an annual killifish". In: *Evolution* 69.6 (2015), pp. 1461–1475.

[57] G. F. Fussmann, M. Loreau, and P. A. Abrams. "Eco-evolutionary dynamics of communities and ecosystems". In: *Functional ecology* (2007), pp. 465–477.

[58] B. Gaál, J. W. Pitchford, and A. J. Wood. "Exact results for the evolution of stochastic switching in variable asymmetric environments". In: *Genetics* 184.4 (2010), pp. 1113–1119.

[59] Y. R. Galimov, A. R. Tukhbatullin, C. R. Haag, and A. V. Tchabovsky. "Sex ratio effects on reproductive success of male and female Daphnia". In: *Journal of Evolutionary Biology* 34.11 (2021), pp. 1817–1826.

[60] M. R. Gardner and W. R. Ashby. "Connectance of large dynamic (cybernetic) systems: critical values for stability". In: *Nature* 228.5273 (1970), pp. 784–784.

[61] N. Gerber, I. Booksmythe, and H. Kokko. "Sex allocation theory for facultatively sexual organisms inhabiting seasonal environments: the importance of bet hedging". In: *The American Naturalist* 192.2 (2018), pp. 155–170.

[62] N. Gerber and H. Kokko. "Abandoning the ship using sex, dispersal or dormancy: multiple escape routes from challenging conditions". In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 373.1757 (2018), p. 20170424.

[63] N. Gerber, H. Kokko, D. Ebert, and I. Booksmythe. "Daphnia invest in sexual reproduction when its relative costs are reduced". In: *Proceedings of the Royal Society B: Biological Sciences* 285.1871 (2018), p. 20172176.

[64] S. A. Geritz, E. Kisdi, G. Meszéna, and J. A. Metz. "Evolutionary singular strategies and the adaptive growth and branching of the evolutionary tree". In: *Evolutionary Ecology* (1996).

[65] D. D. Gessler and S. Xu. "Meiosis and the evolution of recombination at low mutation rates". In: *Genetics* 156.1 (2000), pp. 449–456.

[66] M. T. Ghiselin. "The economy of nature and the evolution of sex". In: *Journal of the History of Biology* 9.2 (1976).

[67] T. Gibbs, J. Grilli, T. Rogers, and S. Allesina. "Effect of population abundances on the stability of large random ecosystems". In: *Physical Review E* 98.2 (2018), p. 022410.

[68] D. T. Gillespie. "Exact stochastic simulation of coupled chemical reactions". In: *The journal of physical chemistry* 81.25 (1977), pp. 2340–2361.

[69] M. E. Gilpin. "Stability of feasible predator-prey systems". In: *Nature* 254.5496 (1975), pp. 137–139.

[70] J. Ginibre. "Statistical Ensembles of Complex, Quaternion, and Real Matrices". In: *Journal of Mathematical Physics* 6.3 (1965), pp. 440–449.

[71] V. Girko. "Elliptic law". In: *Theory of Probability & Its Applications* 30.4 (1986), pp. 677–690.

[72] B. Goh and L. Jennings. "Feasibility and stability in randomly assembled Lotka-Volterra models". In: *Ecological Modelling* 3.1 (1977), pp. 63–71.

[73] U. Goodenough and J. Heitman. "Origins of eukaryotic sexual reproduction". In: *Cold Spring Harbor perspectives in biology* 6.3 (2014), a016154.

[74] J. Grilli, M. Adorisio, S. Suweis, G. Barabás, J. R. Banavar, S. Allesina, and A. Maritan. "Feasibility and coexistence of large ecological communities". In: *Nature communications* 8.1 (2017), pp. 1–8.

[75] J. Grilli, T. Rogers, and S. Allesina. "Modularity and stability in ecological communities". In: *Nature communications* 7.1 (2016), p. 12031.

[76] R. K. Grosberg and R. R. Strathmann. "The evolution of multicellularity: a minor major transition?" In: *Annu. Rev. Ecol. Evol. Syst.* 38 (2007), pp. 621–654.

[77] L. Hadany and S. P. Otto. "The evolution of condition-dependent sex in the face of high costs". In: *Genetics* 176.3 (2007), pp. 1713–1727.

[78] Z. Hadjivasiliou, N. Lane, R. Seymour, and A. Pomiankowski. "Dynamics of mitochondrial inheritance in the evolution of binary mating types and two sexes". In: *Proc. Biol. Sci.* 280.1769 (2013), p. 20131920.

[79] Z. Hadjivasiliou, Y. Iwasa, and A. Pomiankowski. "Cell–cell signalling in sexual chemotaxis: a basis for gametic differentiation, mating types and sexes". In: *Journal of the Royal Society Interface* 12.109 (2015), p. 20150342.

[80] A. E. Hall and M. D. Rose. "Cell fusion in yeast is negatively regulated by components of the cell wall integrity pathway". In: *Molecular biology of the cell* 30.4 (2019), pp. 441–452.

[81] T. Hamaji, Y. Mogi, P. J. Ferris, T. Mori, S. Miyagishima, Y. Kabeya, Y. Nishimura, A. Toyoda, H. Noguchi, A. Fujiyama, et al. "Sequence of the Gonium pectorale mating locus reveals a complex and dynamic history of changes in volvocine algal mating haplotypes". In: *G3: Genes, Genomes, Genetics* 6.5 (2016), pp. 1179–1189.

[82] W. D. Hamilton, R. Axelrod, and R. Tanese. "Sexual reproduction as an adaptation to resist parasites (a review)." In: *Proceedings of the National Academy of Sciences* 87.9 (1990), pp. 3566–3573.

[83] B. Hannon, M. Ruth, B. Hannon, and M. Ruth. "Law of Mass Action". In: *Modeling Dynamic Biological Systems* (2014), pp. 65–68.

[84] E. R. Hanschen, M. D. Herron, J. J. Wiens, H. Nozaki, and R. E. Michod. "Multicellularity drives the evolution of sexual traits". In: *The American Naturalist* 192.3 (2018), E93–E105.

[85] E. H. Harris. "Chlamydomonas as a model organism". In: *Annual review of plant biology* 52.1 (2001), pp. 363–406.

[86] E. Harrison, R. MacLean, V. Koufopanou, and A. Burt. "Sex drives intracellular conflict in yeast". In: *Journal of Evolutionary Biology* 27.8 (2014), pp. 1757–1763.

[87] M. Hartfield and P. D. Keightley. "Current Hypotheses for the evolution of sex and recombination". In: *Integr. Zool.* 7.2 (2012), pp. 192–209.

[88] I. A. Hatton, O. Mazzarisi, A. Altieri, and M. Smerlak. "Diversity begets stability: Sublinear growth and competitive coexistence across ecosystems". In: *Science* 383.6688 (2024), eadg8488.

[89] J. Heitman. "Evolution of sexual reproduction: A view from the fungal kingdom supports an evolutionary epoch with sex before sexes". In: *Fungal Biology Reviews* 29.3 (2015). Special Issue: Fungal sex and mushrooms – A credit to Lorna Casselton, pp. 108–117. ISSN: 1749-4613.

[90] K. Helena-Bueno, M. Y. Rybak, C. L. Ekemezie, R. Sullivan, C. R. Brown, C. Dingwall, A. Baslé, C. Schneider, J. P. Connolly, J. N. Blaza, et al. "A new family of bacterial ribosome hibernation factors". In: *Nature* (2024), pp. 1–8.

[91] A. P. Hendry. *Eco-evolutionary dynamics*. Princeton university press, 2017.

[92] J. M. Henshaw, M. Bittlingmaier, and L. Schärer. "Hermaphroditic origins of anisogamy". In: *Philosophical Transactions of the Royal Society B* 378.1876 (2023), p. 20220283.

[93] D. Hickey. "Molecular symbionts and the evolution of sex". In: *Journal of Heredity* 84.5 (1993), pp. 410–414.

[94] D. A. Hickey. "Selfish DNA: a sexually-transmitted nuclear parasite". In: *Genetics* 101.3-4 (1982), pp. 519–531.

[95] R. F. Hoekstra. "The evolution of sexes". In: *The Evolution of Sex and Its Consequences*. Ed. by S. C. Stearns. Birkhaeuser Verlag, 1987, pp. 59–91.

[96] R. F. Hoekstra. "Evolution of gamete motility differences II. Interaction with the evolution of anisogamy". In: *Journal of theoretical Biology* 107.1 (1984), pp. 71–83.

[97] K. R. Hopper. "Risk-spreading and bet-hedging in insect population biology". In: *Annual review of entomology* 44.1 (1999), pp. 535–560.

[98] J. Hu, D. R. Amor, M. Barbier, G. Bunin, and J. Gore. "Emergent phases of ecological diversity and dynamics mapped in microcosms". In: *Science* 378.6615 (2022), pp. 85–89.

[99] P. G. Hufton, Y. T. Lin, T. Galla, and A. J. McKane. "Intrinsic noise in systems with switching environments". In: *Physical Review E* 93.5 (2016), p. 052119.

[100] L. D. Hurst and W. D. Hamilton. "Cytoplasmic fusion and the nature of sexes". In: *Proceedings of the Royal Society of London. Series B: Biological Sciences* 247.1320 (1992), pp. 189–194.

[101] V. Hutson and R. Law. "Four steps to two sexes". In: *Proceedings of the Royal Society of London. Series B: Biological Sciences* 253.1336 (1993), pp. 43–51.

[102] S. O. Idowu, N. Capaldi, L. Zu, and A. D. Gupta. *Encyclopedia of corporate social responsibility*. Vol. 21. Springer Berlin, 2013.

[103] R. Innami, S. Miyamura, M. Okoshi, T. Nagumo, K. Ichihara, T. Yamazaki, and S. Kawano. "Gamete dimorphism of the isogamous green alga (Chlamydomonas reinhardtii), is regulated by the mating type-determining gene, MID". In: *Communications Biology* 5.1 (2022), p. 1333.

[104] D. J. Innes, C. J. Fox, and G. L. Winsor. "Avoiding the cost of males in obligately asexual Daphnia pulex (Leydig)". In: *Proceedings of the Royal Society of London. Series B: Biological Sciences* 267.1447 (2000), pp. 991–997.

[105] P. Iyer and J. Roughgarden. "Gametic conflict versus contact in the evolution of anisogamy". In: *Theoretical population biology* 73.4 (2008), pp. 461–472.

[106] J. Jaenike. "An hypothesis to account for the maintenance of sex within populations". In: *Evol. theory* 3 (1978), pp. 191–194.

[107] A. James, M. J. Plank, A. G. Rossberg, J. Beecham, M. Emmerson, and J. W. Pitchford. "Constructing random matrices to represent real ecosystems". In: *The American Naturalist* 185.5 (2015), pp. 680–692.

[108] D. Kaiser, C. Manoil, and M. Dworkin. "Myxobacteria: cell interactions, genetics, and development". In: *Annual Reviews in Microbiology* 33.1 (1979), pp. 595–639.

[109] H. Kalmus. "Über den Erhaltungswert der phänotypischen (morphologischen) Anisogamie und die Entstehung der ersten Geschlechtsunterschiede". In: *Biol Zentralbl* 52 (1932), pp. 716–736.

[110] H. Kalmus and C. Smith. "Evolutionary origin of sexual differentiation and the sex-ratio". In: *Nature* 186 (1960), pp. 1004–1006.

[111] N. Kinoshita, C. Nagasato, and T. Motomura. "Chemotactic movement in sperm of the oogamous brown algae, Saccharina japonica and Fucus distichus". In: *Protoplasma* 254.1 (2017), pp. 547–555.

[112] D. L. Kirk. "Oogamy: inventing the sexes". In: *Current Biology* 16.24 (2006), R1028–R1030.

[113] É. Kisdi, A. Stefan, and H. Geritz. "Adaptive dynamics: a framework to model evolution in the ecological theatre". In: *Journal of mathematical biology* 61.1 (2010), p. 165.

[114] M. v. Kleunen, M. Fischer, and B. Schmid. "Effects of intraspecific competition on size variation and reproductive allocation in a clonal plant". In: *Oikos* 94.3 (2001), pp. 515–524.

[115] H. Kokko. "When synchrony makes the best of both worlds even better: How well do we really understand facultative sex?" In: *The American Naturalist* 195.2 (2020), pp. 380–392.

[116] M. Korostyshevsky, M. Schtabnoy, and V. Ratner. "On some principles of evolution viewed as a stochastic process". In: *Journal of Theoretical Biology* 48.1 (1974), pp. 85–103.

[117] K. Kosheleva and M. M. Desai. "Recombination alters the dynamics of adaptation on standing variation in laboratory yeast populations". In: *Molecular Biology and Evolution* 35.1 (2018), pp. 180–201.

[118] T. Koyama, M. Yamada, and M. Matsuhashi. "Formation of regular packets of Staphylococcus aureus cells". In: *Journal of Bacteriology* 129.3 (1977), pp. 1518–1523.

[119] R. Kress. *Linear integral equations*. 3rd ed. Vol. 82. Applied Mathematical Sciences. Springer, 2014.

[120] A. R. Kruckeberg. "The implications of ecology for plant systematics". In: *Taxon* 18.1 (1969), pp. 92–120.

[121] Y. Krumbeck, G. W. A. Constable, and T. Rogers. "Fitness differences suppress the number of mating types in evolving isogamous species". In: *Royal Society Open Science* 7 (2020), p. 192126.

[122] E. K. Kruzel and C. M. Hull. "Establishing an unusual cell type: how to make a dikaryon". In: *Current opinion in microbiology* 13.6 (2010), pp. 706–711.

[123] J. Lehtonen, M. Jennions, and H. Kokko. "The many costs of sex". In: *Trends Ecol. Evol.* 27 (2012), pp. 172–178.

[124] J. Lehtonen and H. Kokko. "Two roads to two sexes: unifying gamete competition and gamete limitation in a single model of anisogamy evolution". In: *Behav. Ecol. Sociobiol.* 65 (2011), pp. 445–459.

[125] J. Lehtonen, H. Kokko, and G. A. Parker. "What do isogamous organisms teach us about sex and the two sexes?" In: *Phil. Trans. R. Soc. B* 371 (2016), p. 20150532.

[126] J. Lehtonen. "Models of fertilization kinetics". In: *Royal Society Open Science* 2.9 (2015), p. 150175.

[127] J. Lehtonen. "Models of fertilization kinetics". In: *Royal Society Open Science* 2.9 (2015), p. 150175. eprint: `https://royalsocietypublishing.org/doi/pdf/10.1098/rsos.150175`.

[128] J. Lehtonen. "The Legacy of Parker, Baker and Smith 1972: Gamete Competition, the Evolution of Anisogamy and Model Robustness". In: *Cells* 10.3 (2021), p. 573.

[129] J. Lehtonen and L. Dardare. "Mathematical models of fertilization—an eco-evolutionary perspective". In: *The Quarterly Review of Biology* 94.2 (2019), pp. 177–208.

[130] J. Lehtonen, Y. Horinouchi, T. Togashi, and G. A. Parker. "Evolution of anisogamy in organisms with parthenogenetic gametes". In: *The American Naturalist* 198.3 (2021), pp. 360–378.

[131] J. Lehtonen, M. D. Jennions, and H. Kokko. "The many costs of sex". In: *Trends in ecology & evolution* 27.3 (2012), pp. 172–178.

[132] J. Lehtonen and G. A. Parker. "Evolution of the two sexes under internal fertilization and alternative evolutionary pathways". In: *The American Naturalist* 193.5 (2019), pp. 702–716.

[133] C. M. Lessells, R. R. Snook, and D. J. Hosken. "The evolutionary origin and maintenance of sperm: selection for a small, motile gamete mating type". In: *Sperm biology*. Elsevier, 2009, pp. 43–67.

[134] L. D. Leverett and A. K. Shaw. "Facilitation and competition interact with seed dormancy to affect population dynamics in annual plants". In: *Population Ecology* 61.4 (2019), pp. 457–468.

[135] D. R. Levitan. "Optimal egg size in marine invertebrates: theory and phylogenetic analysis of the critical relationship between egg size and development time in echinoids". In: *The American Naturalist* 156.2 (2000), pp. 175–192.

[136] J. Li, S. Pang, F. Liu, T. Shan, and S. Gao. "Spermatozoid life-span of two brown seaweeds, Saccharina japonica and Undaria pinnatifida, as measured by fertilization efficiency". In: *Chinese journal of oceanology and limnology* 31.4 (2013), pp. 774–781.

[137] H. Lin and U. W. Goodenough. "Gametogenesis in the Chlamydomonas reinhardtii minus mating type is controlled by two genes, MID and MTD1". In: *Genetics* 176.2 (2007), pp. 913–925.

[138] X. Liu, J. W. Pitchford, and G. W. Constable. "Adaptive dynamics, switching environments and the origin of the sexes". In: *bioRxiv* (2024), pp. 2024–03.

[139] C. M. Lively. "The cost of males in non-equilibrium populations". In: *Evolutionary Ecology Research* 13.1 (2011), pp. 105–111.

[140] G. Lokhorst and B. Trask. "Taxonomic studies on Urospora (Acrosiphoniales, Chlorophyceae) in western Europe". In: *Acta botanica neerlandica* 30.5/6 (1981), pp. 353–431.

[141] R. MacArthur. "Fluctuations of animal populations and a measure of community stability". In: *ecology* 36.3 (1955), pp. 533–536.

[142] R. MacArthur. "Species packing and competitive equilibrium for many species". In: *Theoretical population biology* 1.1 (1970), pp. 1–11.

[143] N. Maire, M. Ackermann1, and M. Doebeli. "Evolutionary branching and the evolution of anisogamy". In: *Selection* 2.1-2 (2002), pp. 119–131.

[144] J. Mancebo Quintana and S. Mancebo Quintana. "A Short-Term Advantage for Syngamy in the Origin of Eukaryotic Sex: Effects of Cell Fusion on Cell Cycle Duration and Other Effects Related to the Duration of the Cell Cycle—Relationship between Cell Growth Curve and the Optimal Size of the Species, and Circadian Cell Cycle in Photosynthetic Unicellular Organisms". In: *International Journal of Evolutionary Biology* 2012 (2012).

[145] H. Matsuda and P. A. Abrams. "Why are equally sized gametes so rare? The instability of isogamy and the cost of anisogamyUnopened XML element'f'closed". In: *Evolutionary Ecology Research* 1.7 (1999), pp. 769–784.

[146] R. M. May. "Will a large complex system be stable?" In: *Nature* 238.5364 (1972), pp. 413–414.

[147] J. Maynard Smith. *The Evolution of Sex*. Cambridge UK: Cambridge University Press, 1978.

[148] L. C. McManus, D. L. Forrest, E. W. Tekwa, D. E. Schindler, M. A. Colton, M. M. Webster, T. E. Essington, S. R. Palumbi, P. J. Mumby, and M. L. Pinsky. "Evolution and connectivity influence the persistence and recovery of coral reefs under climate change in the Caribbean, Southwest Pacific, and Coral Triangle". In: *Global Change Biology* 27.18 (2021), pp. 4307–4321.

[149] L. C. McManus, E. W. Tekwa, D. E. Schindler, T. E. Walsworth, M. A. Colton, M. M. Webster, T. E. Essington, D. L. Forrest, S. R. Palumbi, P. J. Mumby, et al. "Evolution reverses the effect of network structure on metapopulation persistence". In: *Ecology* 102.7 (2021), e03381.

[150] J. A. Metz, S. A. Geritz, G. Meszéna, F. J. Jacobs, and J. S. Van Heerwaarden. "Adaptive dynamics: a geometrical study of the consequences of nearly faithful reproduction". In: *International Institute for Applied Systems Analysis* (1995).

[151] L. Mignerot, K. Avia, R. Luthringer, A. P. Lipinska, A. F. Peters, J. M. Cock, and S. M. Coelho. "A key role for sex chromosomes in the regulation of parthenogenesis in the brown alga Ectocarpus". In: *PLoS genetics* 15.6 (2019), e1008211.

[152] A. Mougi. "Predator interference and complexity–stability in food webs". In: *Scientific reports* 12.1 (2022), p. 2464.

[153] H. J. Muller. "Some genetic aspects of sex". In: *The American Naturalist* 66.703 (1932), pp. 118–138.

[154] H. J. Muller. "The relation of recombination to mutational advance". In: *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis* 1.1 (1964), pp. 2–9.

[155] J. Müller, B. A. Hense, T. M. Fuchs, M. Utz, and C. Pötzsche. "Bet-hedging in stochastically switching environments". In: *Journal of theoretical biology* 336 (2013), pp. 144–157.

[156] H. P. Narra and H. Ochman. "Of what use is sex to bacteria?" In: *Current Biology* 16.17 (2006), R705–R710.

[157] M. Nei. "Stochastic theory of population genetics and evolution". In: *Vito Volterra Symposium on Mathematical Models in Biology: Proceedings of a Conference Held at the Centro Linceo Interdisciplinare, Accademia Nazionale dei Lincei, Rome December 17–21, 1979*. Springer. 1980, pp. 17–47.

[158] M. Neiman, C. M. Lively, and S. Meirmans. “Why sex? A pluralist approach revisited”. In: *Trends in ecology & evolution* 32.8 (2017), pp. 589–600.

[159] A. W. Nickerson and K. B. Raper. “Macrocysts in the life cycle of the Dictyosteliaceae. II. Germination of the macrocysts”. In: *American Journal of Botany* 60.3 (1973), pp. 247–254.

[160] B. P. Nieuwenhuis and D. Aanen. “Sexual selection in fungi”. In: *Journal of evolutionary biology* 25.12 (2012), pp. 2397–2411.

[161] B. P. Nieuwenhuis and T. Y. James. “The frequency of sex in fungi”. In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 371.1706 (2016), p. 20150540.

[162] B. B. Normark and L. R. Kirkendall. “Parthenogenesis in insects and mites”. In: *Encyclopedia of insects*. Elsevier, 2009, pp. 753–757.

[163] M. A. Nowak. “Five rules for the evolution of cooperation”. In: *science* 314.5805 (2006), pp. 1560–1563.

[164] H. Nozaki, T. K. Yamada, F. Takahashi, R. Matsuzaki, and T. Nakada. “New “missing link” genus of the colonial volvocine green algae gives insights into the evolution of oogamy”. In: *BMC Evolutionary Biology* 14.1 (2014), pp. 1–11.

[165] B. O’Neill. “The double-constant matrix, centering matrix and equicorrelation matrix: Theory and applications”. In: *arXiv preprint arXiv:2109.05814* (2021).

[166] E. P. Odum. “Fundamentals of Ecology (1953)”. In: *The Future of Nature*. Yale University Press, 2013, pp. 233–244.

[167] H. Ohtsuka, K. Imada, T. Shimasaki, and H. Aiba. “Sporulation: A response to starvation in the fission yeast Schizosaccharomyces pombe”. In: *MicrobiologyOpen* 11.3 (2022), e1303.

[168] S. P. Otto and T. Lenormand. “Resolving the paradox of sex and recombination”. In: *Nature Reviews Genetics* 3.4 (2002), pp. 252–261.

[169] V. Palchevskiy and S. E. Finkel. “Escherichia coli competence gene homologs are essential for competitive fitness and the use of DNA as a nutrient”. In: *Journal of bacteriology* 188.11 (2006), pp. 3902–3910.

[170] H. J. Park, Y. Pichugin, W. Huang, and A. Traulsen. “Population size changes and extinction risk of populations driven by mutant interactors”. In: *Physical Review E* 99.2 (2019), p. 022305.

[171] G. A. Parker, R. Baker, and V. G. Smith. "The origin and evolution of gamete dimorphism and the male-female phenomenon". In: *J. Theor. Biol.* 36.3 (1972), pp. 529–553.

[172] G. A. Parker. "How soon hath time... A history of two "seminal" publications". In: *Cells* 10.2 (2021), p. 287.

[173] G. A. Parker, R. R. Baker, and V. Smith. "The origin and evolution of gamete dimorphism and the male-female phenomenon". In: *Journal of Theoretical Biology* 36.3 (1972), pp. 529–553.

[174] N. Perrin. "What uses are mating types? The "developmental switch" model". In: *Evolution: International Journal of Organic Evolution* 66.4 (2012), pp. 947–956.

[175] S. Pettersson, V. M. Savage, and M. Nilsson Jacobi. "Predicting collapse of complex ecological systems: quantifying the stability–complexity continuum". In: *Journal of the Royal Society Interface* 17.166 (2020), p. 20190391.

[176] N. Picco, E. Sahai, P. K. Maini, and A. R. Anderson. "Integrating models to quantify environment-mediated drug resistance". In: *Cancer research* 77.19 (2017), pp. 5409–5418.

[177] Y. Pichugin, H. J. Park, and A. Traulsen. "Evolution of simple multicellular life cycles in dynamic environments". In: *Journal of the Royal Society Interface* 16.154 (2019), p. 20190054.

[178] Y. Pichugin, J. Peña, P. B. Rainey, and A. Traulsen. "Fragmentation modes and the evolution of life cycles". In: *PLoS computational biology* 13.11 (2017), e1005860.

[179] Y. Pichugin and A. Traulsen. "Evolution of multicellular life cycles under costly fragmentation". In: *PLoS computational biology* 16.11 (2020), e1008406.

[180] M. Pigliucci, C. J. Murren, and C. D. Schlichting. "Phenotypic plasticity and evolution by genetic assimilation". In: *Journal of Experimental Biology* 209.12 (2006), pp. 2362–2367.

[181] D. Pimentel. "Animal population regulation by the genetic feed-back mechanism". In: *The American Naturalist* 95.881 (1961), pp. 65–79.

[182] A. L. Radzvilavicius. "Evolutionary dynamics of cytoplasmic segregation and fusion: mitochondrial mixing facilitated the evolution of sex at the origin of eukaryotes". In: *Journal of theoretical biology* 404 (2016), pp. 160–168.

[183] C. Rafaluk-Mohr, B. Ashby, D. A. Dahan, and K. C. King. "Mutual fitness benefits arise during coevolution in a nematode-defensive microbe model". In: *Evolution Letters* 2.3 (2018), pp. 246–256.

[184] C. Ratzke, J. Barrere, and J. Gore. "Strength of species interactions determines biodiversity and stability in microbial communities". In: *Nature ecology & evolution* 4.3 (2020), pp. 376–383.

[185] R. Redfield. "Genes for breakfast: The have-your-cake and-eat-lt-too of bacterial transformation". In: *Journal of Heredity* 84.5 (1993), pp. 400–404.

[186] R. J. Redfield. "Do bacteria have sex?" In: *Nature Reviews Genetics* 2.8 (2001), pp. 634–639.

[187] R. J. Redfield. "Evolution of bacterial transformation: is sex with dead cells ever better than no sex at all?" In: *Genetics* 119.1 (1988), pp. 213–221.

[188] A. Roberts. "The stability of a feasible random ecosystem". In: *Nature* 251.5476 (1974), pp. 607–608.

[189] V. Rodilla. "Origin and evolution of binucleated cells and binucleated cells with micronuclei in cisplatin-treated CHO cultures". In: *Mutation Research/Genetic Toxicology* 300.3-4 (1993), pp. 281–291.

[190] A. M. de Roos. "Dynamic population stage structure due to juvenile–adult asymmetry stabilizes complex ecological communities". In: *Proceedings of the National Academy of Sciences* 118.21 (2021), e2023709118.

[191] M. R. Rose. "The contagion mechanism for the origin of sex". In: *Journal of Theoretical Biology* 101.1 (1983), pp. 137–146.

[192] B. Rosenbaum and E. A. Fronhofer. "Confronting population models with experimental microcosm data: from trajectory matching to state-space models". In: *Ecosphere* 14.4 (2023), e4503.

[193] S. Scheu and B. Drossel. "Sexual reproduction prevails in a world of structured resources in short supply". In: *Proceedings of the Royal Society B: Biological Sciences* 274.1614 (2007), pp. 1225–1231.

[194] C. D. Schlichting. "Origins of differentiation via phenotypic plasticity". In: *Evolution & development* 5.1 (2003), pp. 98–105.

[195] S. Schoustra, H. D. Rundle, R. Dali, and R. Kassen. "Fitness-associated sexual reproduction in a filamentous fungus". In: *Current Biology* 20.15 (2010), pp. 1350–1355.

[196] F. M. Scudo. "The adaptive value of sexual dimorphism: I, anisogamy". In: *Evolution* (1967), pp. 285–291.

[197] B. Sieber, J. M. Coronas-Serna, and S. G. Martin. "A focus on yeast mating: From pheromone signaling to cell-cell fusion". In: *Seminars in Cell & Developmental Biology*. Elsevier. 2022.

[198] J. da Silva. "The evolution of sexes: A specific test of the disruptive selection theory". In: *Ecology and evolution* 8.1 (2018), pp. 207–219.

[199] J. da Silva and V. L. Drysdale. "Isogamy in large and complex volvocine algae is consistent with the gamete competition theory of the evolution of anisogamy". In: *Proceedings of the Royal Society B* 285.1890 (2018), p. 20181954.

[200] J.-C. Simon, C. Rispe, and P. Sunnucks. "Ecology and evolution of sex in aphids". In: *Trends in Ecology & Evolution* 17.1 (2002), pp. 34–39.

[201] H. J. Sommers, A. Crisanti, H. Sompolinsky, and Y. Stein. "Spectrum of large random asymmetric matrices". In: *Physical review letters* 60.19 (1988), p. 1895.

[202] C. Song and S. Saavedra. "Will a small randomly assembled community be feasible and stable?" In: *Ecology* 99.3 (2018), pp. 743–751.

[203] Y. Song, S. Scheu, and B. Drossel. "The ecological advantage of sexual reproduction in multicellular long-lived organisms". In: *Journal of Evolutionary Biology* 25.3 (2012), pp. 556–565.

[204] C. D. Soulsbury and S. Humphries. "Biophysical Determinants and Constraints on Sperm Swimming Velocity". In: *Cells* 11.21 (2022), p. 3360.

[205] M. Spencer, N. Colegrave, and S. Schwartz. "Hatching fraction and timing of resting stage production in seasonal environments: effects of density dependence and uncertain season length". In: *Journal of Evolutionary Biology* 14.3 (2001), pp. 357–367.

[206] J. Starrfelt and H. Kokko. "Bet-hedging—a triple trade-off between means, variances and correlations". In: *Biological Reviews* 87.3 (2012), pp. 742–755.

[207] C.-P. Stelzer and J. Lehtonen. "Diapause and maintenance of facultative sexual reproductive strategies". In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 371.1706 (2016), p. 20150536.

[208] L. Stone. "The feasibility and stability of large complex biological networks: a random matrix approach". In: *Scientific reports* 8.1 (2018), pp. 1–12.

[209] L. Stone. "The Google matrix controls the stability of structured ecological and biological networks". In: *Nature communications* 7.1 (2016), pp. 1–7.

[210] L. Stone. "Some problems of community ecology: processes, patterns and species persistence in ecosystems". PhD thesis. Monash University, 1988.

[211] J. E. Strassmann and D. C. Queller. "Evolution of cooperation and control of cheating in a social microbe". In: *Proceedings of the National Academy of Sciences* 108.supplement_2 (2011), pp. 10855–10862.

[212] A. Taitelbaum, R. West, M. Assaf, and M. Mobilia. "Population dynamics in a changing environment: random versus periodic switching". In: *Physical Review Letters* 125.4 (2020), p. 048105.

[213] S. Tang, Y. Pichugin, and K. Hammerschmidt. "An environmentally induced multicellular life cycle of a unicellular cyanobacterium". In: *Current Biology* 33.4 (2023), pp. 764–769.

[214] T. Tao, V. Vu, and M. Krishnapur. "Random matrices: Universality of ESDs and the circular law". In: *The Annals of Probability* 38.5 (2010), pp. 2023–2065.

[215] C. E. Tarnita, C. H. Taubes, and M. A. Nowak. "Evolutionary construction by staying together and coming together". In: *Journal of theoretical biology* 320 (2013), pp. 10–22.

[216] M. Tatewaki and M. Lima. "Life histories of blidingia minima (chlorophyceae), especially sexual reproduction 1". In: *Journal of phycology* 20.3 (1984), pp. 368–376.

[217] A. Tilquin, J. R. Christie, and H. Kokko. "Mitochondrial complementation: a possible neglected factor behind early eukaryotic sex". In: *Journal of evolutionary biology* 31.8 (2018), pp. 1152–1164.

[218] T. Togashi, J. Bartelt, J. Yoshimura, K. Tainaka, and P. Cox. "Evolutionary trajectories explain the diversified evolution of isogamy and anisogamy in marine green algae". In: *Proc. Natl. Acad. Sci.* 109 (2012), pp. 13692–13697.

[219] J. Umen and S. Coelho. "Algal sex determination and the evolution of anisogamy". In: *Annual review of microbiology* 73 (2019), pp. 267–291.

[220] H. Van den Ende, F. Klis, and A. Musgrave. "The role of flagella in sexual reproduction of Chlamydomonas eugametos". In: *Acta botanica neerlandica* 37.3 (1988), pp. 327–350.

[221] L. Van Valen. "The red queen". In: *The American Naturalist* 111.980 (1977), pp. 809–810.

[222] R. R. Vance. "On reproductive strategies in marine benthic invertebrates". In: *The American Naturalist* 107.955 (1973), pp. 339–352.

[223] K. Verstrepen, G. Derdelinckx, H. Verachtert, and F. Delvaux. "Yeast flocculation: what brewers should know". In: *Applied microbiology and biotechnology* 61.3 (2003), pp. 197–205.

[224] H. Vogel, G. Czihak, P. Chang, and W. Wolf. "Fertilization kinetics of sea urchin eggs". In: *Mathematical Biosciences* 58.2 (1982), pp. 189–216.

[225] M. Vos and X. Didelot. "A comparison of homologous recombination rates in bacteria and archaea". In: *The ISME journal* 3.2 (2009), pp. 199–208.

[226] S. Waffenschmidt, J. P. Woessner, K. Beer, and U. W. Goodenough. "Isodityrosine cross-linking mediates insolubilization of cell walls in Chlamydomonas." In: *The Plant Cell* 5.7 (1993), pp. 809–820.

[227] J. Warr. "A mutant of Chlamydomonas reinhardii with abnormal cell division". In: *Microbiology* 52.2 (1968), pp. 243–251.

[228] A. Weismann. "The significance of sexual reproduction in the theory of natural selection". In: *Essays upon heredity and kindred biological problems* 1 (1889), pp. 255–332.

[229] E. P. Wigner. "On the Distribution of the Roots of Certain Symmetric Matrices". In: *The Annals of Mathematics* 67.2 (1958), p. 325.

[230] G. C. Williams. "Kin selection and the paradox of sexuality". In: *sociobiology: Beyond Nature/Nurture?* Routledge, 1980, pp. 371–384.

[231] R. J. Williams and N. D. Martinez. “Simple rules yield complex food webs”. In: *Nature* 404.6774 (2000), pp. 180–183.

[232] P. C. Withers and C. Cooper. “Dormancy”. In: *Encyclopedia of ecology*. Elsevier, 2008, pp. 952–957.

[233] T. Yamada-Inagawa, A. J. Klar, and J. Z. Dalgaard. “Schizosaccharomyces pombe switches mating type by the synthesis-dependent strand-annealing mechanism”. In: *Genetics* 177.1 (2007), pp. 255–265.

[234] Y. Yasui and E. Hasegawa. “The origination events of gametic sexual reproduction and anisogamy”. In: *Journal of Ethology* 40.3 (2022), pp. 273–284.

[235] Y. Yonatan, G. Amit, J. Friedman, and A. Bashan. “Complexity–stability trade-off in empirical microbial ecosystems”. In: *Nature Ecology & Evolution* 6.6 (2022), pp. 693–700.