



Universität zu Lübeck

**From the Department of Evolutionary Theory,  
Max Planck Institute for Evolutionary Biology, Plön  
of the University of Lübeck  
Director: Prof. Dr. Arne Traulsen**

**Mathematical models of competition  
between life cycles of primitive  
heterogeneous organisms**

Dissertation  
for Fulfilment of  
Requirements  
for the Doctoral Degree  
of the University of Lübeck

from the Department of Computer Sciences

Submitted by  
**Yuanxiao Gao**  
from Henan, China

Plön, 2021

*First referee:* Prof. Dr. Arne Traulsen

*Second referee:* Prof. Dr. Andreas Rößler

Date of oral examination: 23.09.2021

Approved for printing: 29.09.2021

# Abstract

Organisms on the earth have evolved into increasingly larger and more complex organisations, reflecting a series of major evolutionary transitions. Here, I focus on the major evolutionary transition from unicellular ancestors to multicellular organisms. Organisms can have different life cycles in terms of cell numbers and composition during development. The selection outcome is different when an organism undergoes different life cycles. The rule of “survival of the fittest” from Darwin determines which life cycle survives. Organism competition relies on individuals’ fitness, which depends on the traits of life cycles they undergo. Organisms mostly experience cell differentiation and include different cell types. The transformation from phenotypically homogeneous organisms into heterogeneous ones leads to the change in individuals’ fitness. Considering cellular interactions among different cell types in an organism, the question then arises: what are the effects of phenotypical heterogeneity for an organism?

In experiments, it is hard to investigate the way that cells interact and the effects that cellular interactions yield in phenotypically heterogeneous organisms. Essentially, these biological problems can be transformed into mathematical ones. We can address these problems by mathematically modelling the interactive effects of individual cells on the competitiveness of organisms. This idea leads to the thesis proposal of mathematical models of life cycle competition in heterogeneous organisms. Here, I specifically pay attention to phenotypically heterogeneous organisms, because they could include diverse cellular interaction forms. The mathematical models adopted in this thesis can depict the underlying cell interacting forms and describe their effects quantitatively. Specifically, I demonstrate the work in three chapters including the emergence of multicellular life cycles, irreversible somatic differentiation and the evolution of reproductive strategies.

Firstly, since cellular interactions could be beneficial or adverse for organisms, we ask which interaction form promotes the evolution of multicellularity? To answer this question, I present a mathematical model considering stage-structured populations. Populations have different but unique reproductive strategies. I capture the effects of cellular

interactions via evolutionary game theory by a payoff matrix. This payoff matrix determines population growth rates, which further determine the performance of populations. By comparing the growth rates between populations with unicellular organisms and populations with multicellular organisms, I found the three most important characteristics determining the emergence of multicellularity: the average performance of phenotypically homogeneous groups, heterogeneous groups, and solitary cells.

Secondly, cellular interactions can lead to cell differentiation, resulting in multiple cell phenotypes in organisms. Essentially, there are only two types in terms of reproduction: germ cells (for fertility) and somatic cells (for viability). Here, somatic cells perform extreme altruistic behaviour in terms of viability and lose their fertility entirely. I refer to this extreme altruistic behaviour as irreversible somatic differentiation (ISD). To understand the evolution of ISD, I simulate the stochastic development of organisms, which includes the one for ISD. Considering different conditions in organism development, I seek stochastic development that has evolutionary growth advantages for ISD. Our results show that ISD emerges under the conditions of both somatic cells' benefits and cell differentiation costs in larger multicellular organisms.

The forms of frequency-dependent cellular interactions are used in the first two models. However, cellular interactions are frequently observed in a threshold form in nature. These threshold effects depend on the minimum number of a certain cell type in an organism. Thus, I extend the cellular interaction form from frequency-dependent to threshold-dependent. An organism grows faster if the cells of a cell type meet a given threshold. Meanwhile, the organism size effects have also been incorporated, which are assumed as being neutral in the first model. Our results show that distinct reproductive strategies could perform uniquely or equally best for populations under the effects of sizes and thresholds. Among the unique optimal reproductive strategies, only binary-splitting ones can be optimal.

In summary, I build mathematical models to address the biological problems of cellular interaction effects on multicellular organisms. The problems include the emergence



of multicellular life cycles, irreversible somatic differentiation and the evolution of reproductive strategies. The cellular interactions and organism size both determine the growth rate of a population. The population growth rates are calculated by the characteristic equations of different structured-population models. Besides, to build the stochastic trajectories for the development of multicellularity, stochastic sampling (Monte Carlo) methods are used to capture the potential cellular interaction forms.

# Kurzfassung

Organismen auf der Erde haben sich zu immer größeren und komplexeren Organisationen entwickelt, die eine Reihe wichtiger evolutionärer Übergänge widerspiegeln. Hier konzentrieren wir uns auf den großen evolutionären Übergang von einzelligen Vorfahren zu mehrzelligen Organismen. Organismen können während der Entwicklung unterschiedliche Lebenszyklen in Bezug auf Zellzahl und Zusammensetzung haben. Das Selektionsergebnis ist unterschiedlich, wenn ein Organismus unterschiedliche Lebenszyklen durchläuft. Die Regel des "Überlebens der Stärksten" von Darwin bestimmt, welcher Lebenszyklus überlebt. Der Organismenwettbewerb hängt von der Fitness des Einzelnen ab, die von den Merkmalen der Lebenszyklen abhängt, die er durchläuft. Organismen erfahren meist eine Zelldifferenzierung und umfassen verschiedene Zelltypen. Die Umwandlung von phänotypisch homogenen in heterogene Organismen führt zu einer Veränderung der Fitness des Einzelnen. Unter Berücksichtigung der zellulären Wechselwirkungen zwischen verschiedenen Zelltypen in einem Organismus stellt sich dann die Frage: Welche Auswirkungen hat eine phänotypische Heterogenität auf einen Organismus?

In Experimenten ist es schwierig zu untersuchen, wie Zellen interagieren und welche Auswirkungen zelluläre Interaktionen auf phänotypisch heterogene Organismen haben. Im Wesentlichen können diese biologischen Probleme in mathematische umgewandelt werden, womit die Auswirkung zellulärer Effekte auf die Konkurrenzfähigkeit von Organismen modelliert werden kann. Diese Idee führt zu dieser Arbeit zu mathematischen Modellen des Lebenszykluswettbewerbs in heterogenen Organismen. Hier betrachte ich speziell phänotypisch heterogene Organismen, da diese verschiedene zelluläre Interaktionsformen aufzeigen können. Die in dieser Arbeit verwendeten mathematischen Modelle können die zugrunde liegenden zellwechselwirkenden Formen darstellen und ihre Auswirkungen quantitativ beschreiben. Im Folgenden demonstriere ich in drei Kapiteln die Entstehung mehrzelliger Lebenszyklen, der irreversiblen somatischen Differenzierung und die Entwicklung von Fortpflanzungsstrategien.

Da zelluläre Interaktionen für Organismen vorteilhaft oder nachteilig sein könnten, fragen wir uns zuerst, welche Interaktionsform die Entwicklung der Mehrzelligkeit fördert. Um diese Frage zu beantworten, stelle ich ein mathematisches Modell vor, das stadienstrukturierte Populationen berücksichtigt. Populationen haben unterschiedliche, aber einzigartige Fortpflanzungsstrategien. Ich betrachte die Auswirkungen zellulärer Interaktionen im Rahmen der evolutionären Spieltheorie mit einer Auszahlungsmatrix. Diese Auszahlungsmatrix bestimmt die Populationswachstumsraten, die die Leistung der Population weiter bestimmen. Durch Vergleich der Wachstumsraten zwischen Populationen mit einzelligen Organismen und Populationen mit mehrzelligen Organismen fand ich die drei wichtigsten Merkmale, die das Auftreten von Mehrzelligkeit bestimmen: die durchschnittliche Leistung von phänotypisch homogenen Gruppen, heterogenen Gruppen und Einzelzellen.

Weiterhin können zelluläre Wechselwirkungen zur Zelldifferenzierung führen, was zu mehreren Zellphänotypen in Organismen führt. Im Wesentlichen gibt es nur zwei Arten der Reproduktion: Keimzellen (für die Fruchtbarkeit) und somatische Zellen (für die Lebensfähigkeit). Hier zeigen somatische Zellen ein extrem altruistisches Verhalten in Bezug auf die Lebensfähigkeit und verlieren ihre Fruchtbarkeit vollständig. Ich bezeichne dieses extrem altruistische Verhalten als irreversible somatische Differenzierung (ISD). Um die Entwicklung der ISD zu verstehen, simuliere ich im dritten Kapitel die stochastische Entwicklung von Organismen, einschließlich der für ISD. In Anbetracht der unterschiedlichen Bedingungen in der Organismusentwicklung streben wir eine stochastische Entwicklung an, die evolutionäre Wachstumsvorteile für ISD bietet. Unsere Ergebnisse zeigen, dass ISD unter den Bedingungen sowohl des Nutzens somatischer Zellen als auch der Kosten für die Zelldifferenzierung in größeren mehrzelligen Organismen auftritt.

In den ersten beiden Modellen verwenden wir frequenzabhängige zelluläre Interaktionen. Zellinteraktionen werden jedoch häufig in einer Schwellenform in der Natur beobachtet. Diese Schwelleneffekte hängen von der Mindestanzahl eines bestimmten Zell-

typs in einem Organismus ab. Daher erweitern wir die zelluläre Interaktionsform von frequenzabhängig zu schwellenwertabhängig. Ein Organismus wächst schneller, wenn die Zahl der Zellen eines Zelltyps einen bestimmten Schwellenwert erreichen. Weiterhin, werden hier auch die Größeneffekte des Organismus berücksichtigt, die im ersten Modell als neutral angenommen wurden. Unsere Ergebnisse zeigen, dass bestimmte Fortpflanzungsstrategien für Populationen unter den Auswirkungen von Größen und Schwellenwerten alleinig oder gleich gut abschneiden können. Unter den alleinig optimalen Reproduktionsstrategien können nur die binär aufteilenden optimal sein.

Zusammenfassend bauen wir mathematische Modelle auf, um die biologischen Probleme zellulärer Interaktionseffekte auf mehrzellige Organismen zu beschreiben. Zu diesen Problemen gehören die Entstehung mehrzelliger Lebenszyklen, die irreversible somatische Differenzierung und die Entwicklung von Fortpflanzungsstrategien. Die zellulären Wechselwirkungen und die Größe des Organismus bestimmen beide die Wachstumsrate einer Population. Die Bevölkerungswachstumsraten werden durch die charakteristischen Gleichungen verschiedener strukturierter Bevölkerungsmodelle berechnet. Um die stochastischen Trajektorien für die Entwicklung der Mehrzelligkeit zu erstellen, werden außerdem stochastische Probenahmemethoden (Monte Carlo) verwendet, um die möglichen zellulären Interaktionsformen zu erfassen.

# Contents

---

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Motivation	1
1.2	Approaches	4
1.2.1	Exponential growth rate in stage-structured population	5
1.2.2	Evolutionary game theory	10
1.2.3	Stochastic sampling	13
1.3	Structure of this thesis	15
<b>2</b>	<b>Interacting cells driving the evolution of multicellular life cycles</b>	<b>18</b>
2.1	Introduction	19
2.2	Methods	22
2.3	Results	27
2.3.1	Games promoting a given life cycle	30
2.3.2	Life cycles promoted by prominent games	34
2.4	Supporting information	41
2.4.1	Population growth rate in the case of stochastic developmental programs	41
2.4.2	Existence of the neutral fitness landscape in the case of homogeneous groups	44
2.4.3	Life cycles of homogeneous groups	49
2.4.4	Calculation of growth rates $\lambda$ for life cycles of heterogeneous groups	53
2.4.5	Profiles of growth rates of the life cycles	60
2.4.6	Optimal life cycles landscape under the self-interaction game	63
<b>3</b>	<b>Evolution of irreversible somatic differentiation</b>	<b>65</b>
3.1	Introduction	66
3.2	Model	68

3.3	Results	72
3.3.1	For irreversible somatic differentiation to evolve, cell differentiation must be costly.	72
3.3.2	Evolution of irreversible somatic differentiation is promoted when even a small number of somatic cells provide benefits to the organism.	74
3.3.3	For irreversible somatic differentiation to evolve, the organism size must be large enough.	76
3.4	Discussion	78
3.5	Appendix	83
3.5.1	Search for the evolutionarily optimal developmental program	83
3.5.2	Under costless cell differentiation, irreversible soma strategy cannot be evolutionarily optimal	85
3.5.3	Conditions promoting the evolution of ISD, RSD, and NSD strategies	88
3.5.4	Parameters of composition effect profiles promoting ISD, RSD, and NSD strategies	89
3.5.5	Evolution of irreversible somatic differentiation under various maturity sizes and unequal cell differentiation costs	90
<b>4</b>	<b>Evolution of reproductive strategies in incipient multicellularity</b>	<b>91</b>
4.1	Introduction	92
4.2	Model	95
4.3	Results	99
4.3.1	The effects of organism sizes on reproductive strategies	99
4.3.2	The effects of thresholds on reproductive strategies	102
4.3.3	The effects of organism sizes and thresholds on reproductive strategies	105
4.4	Discussion	108

4.5	Appendix	112
4.5.1	The probability distribution of offspring	112
4.5.2	Population growth rate	113
4.5.3	Analytical prove of the general size effects on populations with maturity size $N \leq 3$	114
4.5.4	Only the binary-splitting reproductive strategies can be the optimal one under size effects	117
4.5.5	Newborn organisms distribution at the stationary growth stage of populations	119
<b>5</b>	<b>Summary and outlook</b>	<b>121</b>
5.1	Summary	121
5.2	Open questions	124
	<b>Bibliography</b>	<b>129</b>





# Chapter 1

## Introduction

---

### 1.1 Motivation

Organisms on earth have evolved into increasingly larger and more complex organisations, reflecting on a series of major evolutionary transitions [Maynard Smith and Szathmary, 1995, Szathmary and Smith, 1995]. The major evolutionary transitions have been recorded in the development of living organisms from the emergence of the first cells to the evolution of multicellular organisms to the establishment of social groups [Bonner, 1998, Grosberg and Strathmann, 2007, Claessen et al., 2014, Sebe-Pedros et al., 2017, Brunet and King, 2017]. Following these transitions, selection usually favours the hierarchically increased level of biological complexity [Carroll, 2001], rather than the fitness interests of the lower replicating units in biological organisms. Cells in organisms are referred to as cooperators or defectors depending on whether they increase the fitness of organisms or not. Cooperation implies cells follow developmental control of higher levels of organisms under natural selection, which may decrease the fitness of individual cells. Since the higher levels of organisms arise from the lower level replicating units, then how do organisms moderate the conflicts between the higher levels and lower levels of biological replication by developmental control? In other words, under which conditions can the cooperation of cells in organisms emerge on the higher level, even in the presence of defectors on the lower level? Moreover, do different modes of internal developmental control on the lower level shape different characteristics of organisms on a higher level? For example, modes can rely on the frequency of cooperators or can rely on the number of cooperators in organisms. If so, which characteristics on the higher level can be shaped under different developmental control?

The developmental control of organisms depends on natural selection acting on the

## Chapter 1. Introduction

---

fitness of the higher level. But the fitness of the higher level is impacted by the lower level of an organism. In practice, it is hard to measure and estimate the interactions between units at the lower level and their effects at the higher level in an organism. Thus, mathematical models provide a useful tool to address this question. Suppose a hierarchical organism contains cell types of cooperator and defector, and natural selection favours cooperators on the higher level but defectors on the lower level. If we treat the fraction or number of cooperators as a trait, a more general mathematical method to describe the cooperator dynamics is the Price's equation [Price, 1972, Gardner, 2008]. For a given organism,  $z$  is the value of a trait of interest, such as the frequency or the number of cooperating cells forming that organism, and  $w$  is the number of offspring of this organism.  $\Delta z$  shows the discrepancy between the trait values of the organism itself and its offspring. Price's equation shows how the average of  $z$  in the population changes from one generation to next,

$$\Delta \bar{z} = Cov\left(\frac{w}{\bar{w}}, z\right) + E\left(\frac{w}{\bar{w}} \Delta z\right), \quad (1.1)$$

where bars are average values over the population.  $Cov$  is the covariance and  $E$  is the expected value. Price's covariance approach provides a method for representing selection in hierarchically structured population [Michod and Roze, 1999, Okasha, 2006]. The first item of equation (1.1) represents the selection on the organism (higher) level, and the second item represents the selection on the cell (lower) level. Price's equation provides a general mathematical description of trait dynamics in populations, while the detailed population structures and cell interactions in hierarchical organisms need to be established further. However, it has been argued that it is not a good tool to build concrete models [van Veelen, 2005].

Evolutionary game theory has been used to estimate the effect of multilevel selection on the evolution of cooperation [Nowak, 2006a, Traulsen and Nowak, 2006]. In the model of Traulsen and Nowak [2006], organisms consist of cells of cooperators and defectors. The fitness of an organism is determined by the fitness of its consisting cells,

which further is determined by a payoff matrix. Moran process is applied both on the lower level (cell level) and the higher level (organism level). Defectors grow faster than cooperators within an organism, while an organism containing more cooperators grows faster than one containing fewer cooperators. Therefore, selection favours defectors on the population of cell level (lower level), but cooperators on the organism level (higher level). On the higher level, selection acts on the fecundity of organisms. [Traulsen and Nowak \[2006\]](#) identified the conditions favouring cooperation by comparing the fixation probabilities of cooperators  $\rho_C$  and defectors  $\rho_D$ .  $\rho_C$  is the probability that cooperators takes over a population by introducing a single cooperator in a population consisting of defectors and similarly for  $\rho_D$ . Selection at different levels in multicellularity has also been considered in infinite populations [[Roze et al., 2001](#)]. Here the generations are non-overlapping and a large number of offspring are produced at the same time and then die. Selection on the higher level acts on the average fitness of populations. The average fitness of a population is measured by its growth rate via matrix projection models, see section (1.2.1).

Many questions related to hierarchical entities have been solved, such as the cooperation problem [[Nowak et al., 2004](#), [Nowak, 2006a](#)] and the single-cell reproduction problem [[Roze et al., 2001](#)], but there remain many open questions related to multicellular organisms. In multicellular organisms, the organism and the cells represent the higher and the lower replicating units, respectively. The interactions between cells in an organism affect the characteristics of the organism. The first intriguing question would be the formation of multicellular organisms in the face of selfish cells (defectors). An ensuing question is the reproduction problem after the formation of multicellular organisms. [Roze et al. \[2001\]](#) has compared vegetative reproduction and single-cell reproduction in multicellularity. Vegetative reproduction is an asexual reproduction, which produces smaller organisms as offspring for an organism. [Pichugin et al. \[2017, 2019\]](#) investigated reproductive strategies in phenotypically homogeneous organisms. Then how do cell interactions in heterogeneous organisms affect reproductive strategies? Furthermore, organisms evolve toward the direction possessing increased cell number, thus the

reproductive ability for each cell is challenging. Cell differentiation is inevitable for larger organisms. The question then arises: to what extent can cell differentiation be maintained by an organism? Moreover, under which conditions can extreme altruistic behaviour of somatic cell differentiation emerge? The extreme altruistic behaviour of somatic cell differentiation is the situation that the somatic cells only invest in the special task viability and leave other tasks to other cells. As increases in complexity and diversity always accompany increased organism sizes in evolution, how large can organisms be to maintain their internal developmental control? All the above questions lead us to think about the competition between the lower replicating units and the higher replicating units in phenotypically heterogeneous organisms.

## 1.2 Approaches

To investigate the questions mentioned above, I adopt three approaches, namely, the exponential growth of stage-structured populations, evolutionary game theory and stochastic sampling. Here, I briefly introduce the conceptual frameworks and development of these three approaches. Since organism (high entity) growth is considered in a density-independent and rich environment, organisms follow exponential growth. Therefore, in structured populations, I mainly focus on the exponential growth of populations. Under this setting, a population's performance can be quantitatively evaluated by its growth rate, an approach I will present in the first subsection. To capture individual interactions, I follow game-theoretical approaches developed in previous work, for example, the work of [Traulsen and Nowak \[2006\]](#). Evolutionary game theory has been frequently used to investigate problems related to the emergence of cooperation, and the problems of the robustness and the stability of a strategy. In this thesis, I am interested in the payoff matrix which estimates the effects of individuals' interactions. Thus, the game-theoretic approach is conceptually used in this thesis and I will introduce it in the second subsection. Finally, in the stochastic sampling section, I focus on the Monte Carlo method to

obtain samples in high dimensional spaces.

### 1.2.1 Exponential growth rate in stage-structured population

In a stage-structured population, individuals are classified into different stages (for example by age or size). The dynamic of individuals, i.e. the number and distribution of individuals at the different stages is studied over time. Many individual-based models have been introduced to describe the dynamics of a population in terms of its individual's behaviour. Among them, the matrix projection model is a commonly used tool to estimate the asymptotical growth rate in a structured population [Caswell, 2001, Tuljapurkar and Caswell, 1997]. It is conceptually the simplest method to describe population structure in discrete time and stages. Essentially, the matrix projection model records the individual's change at different stages over a fixed time (projection interval). The stages can describe a life cycle by containing the transitions that an individual can make. For example, in a population with three stages, the transitions between individuals at different stages can be expressed in a life-cycle graph, see Fig. 1.1.

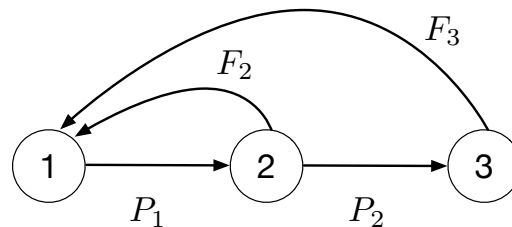


Figure 1.1: **Life-cycle graph of a  $N$ -structured model with three age stages.** The variable  $P_i$  are the survival probabilities for individuals growing from state  $i$  to  $i + 1$ , where  $i \in \{1, 2\}$ . The variable  $F_i$  are the fertilities for individuals in state  $i$ , where  $i \in \{2, 3\}$ .

The variable  $P_i$  are the survival probabilities to the next age class and the variable  $F_j$  are the fertilities producing offspring in class 1. The transitions in a population with the

## Chapter 1. Introduction

---

life cycle in Fig. 1.1 can be described by the following equations

$$\begin{aligned}n_1(t+1) &= F_2 n_2(t) + F_3 n_3(t) \\n_2(t+1) &= P_1 n_1(t) \\n_3(t+1) &= P_2 n_2(t),\end{aligned}\tag{1.2}$$

where  $n_i(t)$  is the number of individuals being at state  $i$  at time  $t$ , and the unit of time is the projection interval.

These difference equations can be written in a matrix form

$$\mathbf{n}(t+1) = \mathbf{A}\mathbf{n}(t),\tag{1.3}$$

where

$$A = \begin{pmatrix} 0 & F_2 & F_3 \\ P_1 & 0 & 0 \\ 0 & P_2 & 0 \end{pmatrix}$$

and

$$\mathbf{n}(t) = \begin{bmatrix} n_1(t) \\ n_2(t) \\ n_3(t) \end{bmatrix}.$$

$\mathbf{A}$  is a population-projection matrix and  $n(t)$  is a stage-distribution vector. In matrix  $\mathbf{A}$ ,  $a_{ij}$  is the transition probability from the stage  $j$  to stage  $i$ , where  $i, j \in [1, 3]$ . Depending on the population itself,  $a_{ij}$  could depend on size or time, which further classifies the matrix model into three types: linear (constant-coefficient models  $\mathbf{A}$ ), nonlinear models ( $\mathbf{A}_n$ ) and time-varying models ( $\mathbf{A}_t$ ).

To get the growth rate of a population, one needs to solve equation (1.3). In the long run, the population grows exponentially. The population growth rate can be solved by

equation (1.4) with the ansatz,

$$\mathbf{n}(t) = \lambda^t \mathbf{w}. \quad (1.4)$$

Together with equation (1.3), it leads to

$$\lambda^{t+1} \mathbf{w} = \lambda^t \mathbf{A} \mathbf{w}, \quad (1.5)$$

that is

$$(\mathbf{A} - \lambda \mathbf{I}) \mathbf{w} = 0. \quad (1.6)$$

To satisfy the existence of a nonzero solution for  $\mathbf{w}$ , the following characteristic equation must be satisfied

$$\det(\mathbf{A} - \lambda \mathbf{I}) = 0. \quad (1.7)$$

Thus,  $\lambda$  and  $\mathbf{w}$  are the eigenvalues and eigenvectors of  $\mathbf{A}$ , respectively. Eventually, the leading eigenvalue describes the population growth rate. The corresponding eigenvector of the leading eigenvalue gives the stable stage distribution of the population in the long term run [Tuljapurkar and Caswell, 1997]. If the stage is equal to age, then the eigenvector of the largest eigenvalue represents a population pyramid, a graphical illustration commonly used in demography.

Matrix projection models study population structures using a discrete set of stages. Thus, one has to artificially choose stages when applying matrix projection models to populations with continuous stage variables, such as mass or length. This artificial subdivision in a population inevitably yields discretisation errors. To extend the trait variable from a discrete one to a continuous one, Easterling et al. [2000] introduced the integral projection model. In doing so, they use the continuous function for stage variables. In the model,  $n(y,t)$  is the number of stage- $y$  individual at time  $t$ , which is a continuous function with respect to  $y$ .  $p(x,y)$  is the survival probability that an individual in stage  $x$

## Chapter 1. Introduction

---

at time  $t$  is alive and in the stage interval  $(y, y + dy)$  at time  $t + 1$ .  $f(x, y)dy$  is the number of newborns at time  $t + 1$  in the stage interval  $(y, y + dy)$  per stage  $x$  individual alive at time  $t$ . Instead of a matrix, they use a continuous function describing a surface to represent the survivorship ( $p(x, y)$ ) and fecundities ( $f(x, y)$ ) in a *kernel*, which is denoted in the following equation

$$\begin{aligned} n(y, t + 1) &= \int_{\Omega} [p(x, y) + f(x, y)]n(x, t)dx \\ &= \int_{\Omega} k(y, x)n(x, t)dx, \end{aligned} \tag{1.8}$$

where the  $k(y, x) = p(x, y) + f(x, y)$  is called *kernel*. [Ellner and Rees \[2006\]](#) extended the integral projection model by incorporating complex demographic attributes (including dormant and active life stages), where the population structure depends on several stages (e.g., size and age). The demographic attributes and stages impact the function of  $p(x, y)$  and  $f(x, y)$ . Thus, equation (1.8) can be written in a more general form

$$n_i(y, t + 1) = \sum_{j=1}^N \int_{\Omega_j} k_{ij}(y, x)n_j(x, t)dx, \tag{1.9}$$

where  $N$  is the total number of stage sets, possibly including discrete points and continuous domains, and  $\Omega_j$  is the  $j$ -th of the  $N$  sets. The integral projection model is similar to the matrix projection model in terms of calculating the growth rate and the stable stage distribution of a population. The estimation of population growth rates of these two models is similar, though they are different in evaluating the reproductive values and the sensitivities of a population, see the details in [\[Easterling et al., 2000\]](#).

Finally, physiologically structured population models (PSP models) treat both states and time as continuous variables. PSP models depict the changes of continuous population distribution on a continuous-time basis using partial differential equations [\[Tuljapurkar and Caswell, 1997, Chapter 6\]](#). Next, I will introduce the basic ideas of PSP models by partial differential equations. I will especially mention the following extension: allowing a population with multiple newborn stages, as it will be the case in our



Table 1.1: The notations of equation (1.10)

Parameter	Description
$x$	Stage of individuals in a population
$x_b$ or $x_m$	The stage of an individual at birth or maturation
$\tau(E_c, x, x_b)$	Growth time of an individual from $x_b$ to $x$
$g(E_c, x)$	Growth rate with respect to the stage $x$ in environment $E_c$
$b(E_c, x)$	Reproduction rate with respect to the stage $x$ in environment $E_c$
$d(E_c, x)$	Death rate with respect to the stage $x$ in environment $E_c$
$E_c$	The environmental parameter, here is a constant vector

models.

In a constant environment and density-independent interactions among individuals, a population either grows or declines exponentially. The population growth rate  $r$  is captured by the following characteristic equation by using ansatz:

$$\Pi(E_c, r) = \int_{x_b}^{x_m} e^{-r\tau(E_c, x, x_b)} \frac{b(E_c, x)}{g(E_c, x)} \exp\left(-\int_{x_b}^x \frac{d(E_c, \xi)}{g(E_c, \xi)} d\xi\right) dx = 1, \quad (1.10)$$

where parameters are explained in Table (1.1), and only one newborn stage  $x_b$  in this model, see the detailed proof in [Tuljapurkar and Caswell, 1997, Chapter 6].

In age-structured models, since age  $a$  is both growth time  $\tau(E_c, x, x_b)$  and stage  $x$ , therefore  $x_b = 0$ ,  $x_m = \infty$ ,  $g(E_c, x) = 1$  and  $\tau(E_c, x, x_b) = a$ . Equation (1.10) can be simplified into

$$\int_0^{\infty} e^{-ra} b(E_c, a) S(E_c, a) da = 1, \quad (1.11)$$

which is the Lotka-Euler equation, where  $S(E_c, a) = \exp(-\int_0^a d(E_c, \zeta) d\zeta)$  is the survival function. De Roos [2008] provided a similar computational approach as PSP models

## Chapter 1. Introduction

---

to calculate population growth rates. He extended the approach to allow individuals having multiple newborn sizes, where the population can be calculated in the following equation:

$$\det(\mathbf{L}(A_m, r) - \mathbf{I}) = 1, \quad (1.12)$$

where

$$\mathbf{L}(A_m, r) = \begin{pmatrix} L_{11}(A_m, r) & L_{12}(A_m, r) & \cdots & L_{1N}(A_m, r) \\ L_{21}(A_m, r) & L_{22}(A_m, r) & \cdots & L_{2N}(A_m, r) \\ \vdots & \vdots & \ddots & \vdots \\ L_{N1}(A_m, r) & L_{N2}(A_m, r) & \cdots & L_{NN}(A_m, r) \end{pmatrix}$$

and  $L_{ij}(A_m, r) = \int_0^{A_m} e^{r\alpha} b(\alpha) S(\alpha) d\alpha$  is the cumulative number of offspring with size  $i$  at birth produced by an individual that was born with size  $j$ . Here  $A_m$  is the maximum age of individuals and  $b(\alpha)$  and  $S(\alpha)$  are the birth rate and the survival probability at age  $\alpha$ , respectively.

### 1.2.2 Evolutionary game theory

Evolutionary game theory is a mathematical approach, which uses game theory to solve population problems in biology. It was originally introduced from the field of economy and society [von Neumann and Morgenstern, 1944]. Evolutionary game theory was first formally introduced by Maynard Smith and Price [1973] to measure the outcomes of conflicts between animals via different behavioural strategies in biology. Evolutionary games use a matrix to estimate the outcome of interactions. The variable  $a_{ij}$  of the matrix determines the interaction outcome between the player  $i$  and the player  $j$ . The variable  $a_{ij}$  is referred to as a payoff. Initially, evolutionary games were confined to a pairwise game with two players and two pure strategies. In a population, two individuals meet each other at random and the consequences of their interaction are captured by their payoffs.

Their payoffs determine their later reproductive success, called fitness. Individuals can only take one strategy at a time, but they could have many rounds of interactions. The outcome of each interaction is evaluated by a  $2 \times 2$  payoff matrix. In a  $2 \times 2$  payoff matrix, the rows represent the different strategies of a focal player and the columns represent the different strategies of a competitive player. The elements of the matrix represent the payoffs of the focal player in different situations. For example, in a population with two pure strategies;  $S_1$  and  $S_2$ , each player's payoff is captured by the matrix

$$\begin{array}{cc} & S_1 & S_2 \\ \begin{array}{c} S_1 \\ S_2 \end{array} & \begin{pmatrix} a & b \\ c & d \end{pmatrix}, \end{array}$$

where a player with strategy  $S_1$  gets a payoff  $a$  or  $b$  when interacting with another player with strategy  $S_1$  and  $S_2$ , respectively. Likewise, a player with strategy  $S_2$  gets a payoff  $c$  or  $d$  when interacting with another player with strategy  $S_1$  and  $S_2$ , respectively.

Many biological interactions can be incorporated in payoff matrices [Nowak and Sigmund, 2004]. For example, when  $b < d < a < c$  and  $a > \frac{b+c}{2}$ , the above payoff matrix describes the prisoner's dilemma game. In this scenario, the strategies are referred to as cooperator (C) and defector (D). The payoff matrix is commonly written by the matrix

$$\begin{array}{cc} & C & D \\ \begin{array}{c} C \\ D \end{array} & \begin{pmatrix} R & S \\ T & P \end{pmatrix}. \end{array}$$

If the pairs of interacting individuals are random and the same pair is not repeated, then the defection is always the best strategy for an individual. But at the level of the population, cooperation is the optimal strategy, which leads to higher payoffs. Under this situation, a population will end up with defectors [Axelrod and Hamilton, 1981]. But in the iterated prisoner's dilemma game, the outcome could be different depending on model assumptions. Another example is the snowdrift game, where each player can get a benefit  $b'$  after digging out of a snowdrift. The digger will bear a cost  $c'$ . Digging snow

## Chapter 1. Introduction

---

is referred to as cooperative behaviour. The corresponding payoff matrix of the snowdrift game is

$$\begin{array}{cc} & \begin{array}{cc} C & D \end{array} \\ \begin{array}{c} C \\ D \end{array} & \left( \begin{array}{cc} b' - \frac{c'}{2} & b' - c' \\ b' & 0 \end{array} \right).$$

With the development of the evolutionary game theory, multiple players and strategies have been considered [Diekmann, 1985] [Gokhale and Traulsen, 2014], such as the volunteer's dilemma game and the "rock-scissors-paper" game. In a volunteer's dilemma game, each player chooses between cooperation (C) and defection (D). Cooperators can voluntarily benefit all other players by paying a cost. Defectors wait for benefits from other players. All players get a 0 payoff if they all defect or "free ride". Here, I use  $b$  and  $c$  to show the benefit and the cost, respectively. In a  $N$ -player volunteer's dilemma game, the payoff matrix is

$$\begin{array}{ccccc} & 0 & 1 & 2 & \dots & N-1 \\ \begin{array}{c} C \\ D \end{array} & \left( \begin{array}{ccccc} b-c & b-c & b-c & \dots & b-c \\ 0 & b & b & \dots & b \end{array} \right),\end{array}$$

where  $b - c > 0$  and  $N \geq 2$ . The rows represent the strategies of a focal player and the columns represent the number of volunteers among the rest of the players. The elements of the matrix describe the payoffs of the focal player. If each cooperator gets the same payoff, the game is referred to as a symmetric volunteer's dilemma game. Later the situation of asymmetries in benefits or costs has also been considered in volunteer's dilemma games [Weesie, 1993]. There, cooperators have different payoffs depending on different situations, such as the waiting time of cooperation.

Initially, evolutionary game theory focused on "unbeatable strategies" or "evolutionarily stable strategies" [Hamilton, 1967] [Maynard Smith, 1982]. Unbeatable strategies are the strategies, which will dominate all other invading strategies. Evolutionarily stable strategies are the strategies that cannot be invaded or replaced by other strategies by

natural selection in a population. ESS provides a useful method to evaluate biological interactions. To seek the ESS at equilibrium states and examine their stabilities, [Taylor and Jonker \[1978\]](#) have introduced dynamics to evolutionary game theory by applying differential equations. Subsequently, there was a growing number of research on the dynamics of evolutionary game theory [[Zeeman, 1980](#)] [[Hofbauer and Sigmund, 1998](#)]. The replicator dynamics describes the changes in strategy frequencies over time in a population. Suppose that a population has  $n$  strategies. The interaction outcomes between players with different strategies are depicted in a  $n \times n$  matrix  $A = [a_{ij}]$ , where the  $a_{ij}$  is the payoff of the focal player when he plays  $i$  strategy and meets a player playing  $j$  strategy. If  $x_i$  is the frequency of the players playing strategy  $i$  in the population, then the expected payoff for the strategy  $i$  is  $f_i = \sum_{j=1}^n x_j a_{ij}$ . The average payoff of all strategies is  $\phi = \sum_{i=1}^n x_i f_i$ . Then the replicator equation is

$$\dot{x}_i = x_i(f_i - \phi). \quad (1.13)$$

We should notice that  $\sum_{i=1}^n x_i = 1$ , where  $i \in [1, n]$ . The replicator dynamics describes deterministic but frequency-dependent selection dynamics in infinite populations [[Nowak and Sigmund, 2004](#)]. The evolution game has also been investigated in finite populations [[Thomas and Pohley, 1981](#)] [[Ficici and Pollack, 2000](#)] [[Nowak et al., 2004](#)]. In this scenario, people are interested in the fixation probability and fixation time of a mutant strategy in a population of players with a resident strategy.

### 1.2.3 Stochastic sampling

When handling realistic problems, several variables implying a parameter space of high dimensions often arise in complex mathematical models. An analytical analysis often becomes less realistic in presence of high-dimensional models. Thus, obtaining a numerical solution becomes a central issue. Here the Monte Carlo method is introduced, and it is used in this thesis to sample possible functional forms of models in high dimensional spaces.

## Chapter 1. Introduction

---

Monte Carlo methods are computational algorithms that allow obtaining numerical results by repeated random sampling. In a random process, the exact outcome cannot be determined due to the intervention of random variables. Monte Carlo methods approximate the expected value of random variables by the sample mean based on the law of large numbers. It is largely used in three main problems: sampling, estimation and optimisation [Kroese et al., 2014].

To illustrate the usage of Monte Carlo methods, I show an example for approximating the constant  $\pi$ , which is defined as the ratio of a circle's circumference to its diameter. This example is related to Buffon's Needle Problem. Buffon's Needle Problem was first introduced by *Georges-Louis Leclerc, Comte de Buffon*: "Suppose we have a floor made of parallel strips of wood, each the same width, and we drop a needle onto the floor. What is the probability that the needle will lie across a line between two strips?"

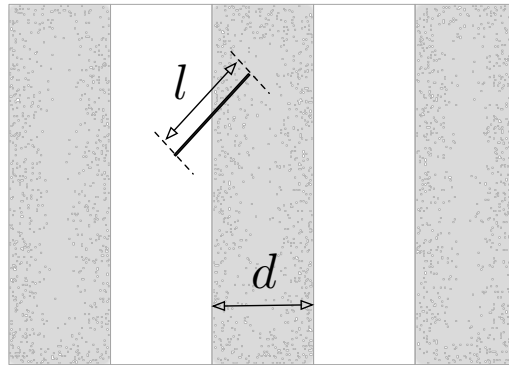


Figure 1.2: **Schematic of Buffon's Needle Problem.**  $l$  is the length of a needle.  $d$  is the width of strips, which are evenly spaced.

The probability is calculated in equation (1.14)

$$p = \int_0^\pi \frac{l \sin \theta d \theta}{\pi d} = \left(\frac{l}{\pi d}\right) \int_0^\pi \sin \theta d \theta = \frac{2l}{\pi d}, \quad (1.14)$$

where  $l$  is the length of the needle and  $d$  is the width of the strips [Ramaley, 1969], see Fig. 1.2. The probability  $p$  can be obtained by the Monte Carlo method, by dropping a

needle on the floor many times. Then we count the number of the needles crossing a strip and compare with the total number of trials. Suppose  $n'$  out of total  $n$  needles are crossing lines, then  $P \approx \frac{n'}{n}$ . Then  $\pi \approx \frac{2ln}{dn'}$  according to equation (1.14).

## 1.3 Structure of this thesis

In Chapter 2, to investigate the effects of cellular interactions on multicellular organisms, I consider organisms growing in stage-structured and infinite populations. The organisms of each population contain two cell types. The cellular interactions between cell types determine the growth rate of an organism. I apply evolutionary game theory to describe the outcomes of cellular interactions. The two cell types are corresponding to two strategies. Each daughter cells can switch to another cell type after cell divisions. The cellular interactions are captured by a pairwise game. A payoff matrix describes the outcomes of cellular interactions, which further determines the growth rate of organisms. Our model is general enough to capture cellular interactions under a wide scope of games. I consider that the payoff of a cell type depends linearly on its frequency in an organism. I further assume that selection is weak. The weak selection assumption allows us to linearise the characteristic equation of the transition matrix of a population. Furthermore, it allows us to obtain analytical results of population growth rates. Populations differ in the life cycles of their organisms. Each life cycle of organisms corresponds to a reproductive strategy. I seek the optimal population, which has the largest growth rate. The life cycle of organisms in the optimal population is the optimal life cycle. Our results show that the formation of multicellular life cycles depends on the payoff of phenotypically homogeneous groups, the payoff of the phenotypically heterogeneous groups and the payoff of the solitary cells (loners).

In chapter 3, I investigate the conditions favouring extreme altruism in multicellularity. This extreme altruism shows that somatic cells only invest in viability. Two cell types are considered: germ cells (defectors) and somatic cells (cooperators). The somatic

## Chapter 1. Introduction

---

cells are viewed as cooperators as they contribute to organism growth but lost their reproductive abilities compared with germ cells. I build a stochastic model to capture the development of a multicellular organism. The cell types can switch to each other randomly after cell divisions. Each development trajectory is a development under a sequence of fixed probabilities. The probabilities describe the cell-type switching probabilities between germ cells and somatic cells. Populations are different in their development trajectories. Populations compete for growth, which is determined by the frequency of somatic cells and the cell switching probabilities. The somatic cells provide benefits for organism growth, which is proportional to its frequency. The cell switching probabilities determine the differentiation costs indirectly. I assume that a general and flexible somatic cells' contribution from, which is embedded in a non-increasing function of growth time depending on the frequency of somatic cells. To obtain the potential forms of somatic cells' contribution, I use Monte Carlo methods to sample many possible non-increasing functions. Selection acts on population growth rates, that is the growth rate of a random development trajectory. I calculate the numerical results by the expected growth rate of a population with a unique trajectory, which is based on a large number of single simulation runs. Our results show that irreversible somatic differentiation emerges under the conditions of both somatic cells' benefits and cell differentiation costs in larger multicellular organisms.

In chapter 4, I investigate the evolution of reproductive strategies in multicellular organisms. I take the same modelling framework as in Chapter 2, but extend the model in terms of organism sizes and cellular interaction forms. I investigate the cellular interactions in a non-linear and threshold-dependent form depending on cooperators in an organism. I choose the functional forms of size effects in a general way. Similar to the work in chapter 3, I use the Monte Carlo methods to sample all possible function forms. For cellular interactions, organism growth depends on the minimum threshold of the cooperator cell type. When the number of cooperators meets a given threshold, an organism grows fast. I choose a volunteer dilemma game to demonstrate threshold effects. Compared with the first work, I remove the weak selection constraint but explore



a small cell-type switching rate. I extend multicellular life cycles to a larger set (with maturity size 8 and totally 58 life cycles) compared with the first work (37 life cycles in Chapter 2). The populations with different reproductive strategies compete for growth. The reproductive strategy is optimal if its population has the largest growth rate. Our results show that reproductive strategies can co-exist or dominate others under different size and threshold conditions. Among the dominated reproductive strategies, only the binary-splitting reproductive strategy can be the optimal and unique reproductive strategy under either condition.

In chapter 5, I summarize the problems that are addressed in this thesis based on Chapter 2, 3 and 4. I briefly reintroduce the models and methods that are used in this thesis. And then, I discuss the significant results in each chapter and compare them with previous work. For Chapter 2, I discuss the relationship between games and the life cycles with a single-cell bottleneck. For Chapter 3, I highlight the connection between the small percentage in the emergence of irreversible somatic differentiation (ISD) and the observed cell differentiation of different species in reality. For Chapter 4, I discuss the binary-splitting reproductive strategies and the cell-type switching probability. Furthermore, alternative scenarios in terms of building models are considered, such as the poor environment for organism growth, density-dependent growth, spatial structures and special fragmentation preference. I give an outlook on how other methods and models can be considered under different model assumptions to answer the questions addressed in this thesis, which may inspire future research directions or studies.

# Interacting cells driving the evolution of multicellular life cycles

---

This the work in this chapter has undergone peer-review and has been published in the journal of *PLOS Computational Biology* as: Gao, Traulsen, and Pichugin [2019]. Interacting cells driving the evolution of multicellular life cycles. <https://doi.org/10.1371/journal.pcbi.1006987>.

## Abstract

Evolution of complex multicellular life began from the emergence of a life cycle involving the formation of cell clusters. The opportunity for cells to interact within clusters provided them with an advantage over unicellular life forms. However, what kind of interactions may lead to the evolution of multicellular life cycles? Here, we combine evolutionary game theory with a model for the emergence of multicellular groups to investigate how cell interactions can influence reproduction modes during the early stages of the evolution of multicellularity. In our model, the presence of both cell types is maintained by stochastic phenotype switching during cell division. We identify evolutionary optimal life cycles as those which maximize the population growth rate. Among all interactions captured by two-player games, the vast majority promotes two classes of life cycles: (i) splitting into unicellular propagules or (ii) fragmentation into two offspring clusters of equal (or almost equal) size. Our findings indicate that the three most important characteristics, determining whether multicellular life cycles will evolve, are the average performance of homogeneous groups, heterogeneous groups, and solitary cells.

### Author summary

Multicellular organisms are ubiquitous. But how did the first multicellular organisms arise? It is typically argued that this occurred due to benefits coming from interactions between cells. One example of such interactions is the division of labour. For instance, colonial cyanobacteria delegate photosynthesis and nitrogen fixation to different cells within the colony. In this way, the colony gains a growth advantage over unicellular cyanobacteria. However, not all cell interactions favour multicellular life. Cheater cells residing in a colony without any contribution will outgrow other cells. Then, the growing burden of cheaters may eventually destroy the colony. Here, we ask what kinds of interactions promote the evolution of multicellularity? We investigated all interactions captured by pairwise games and for each of them, we look for the evolutionarily optimal life cycle: How big should the colony grow and how should it split into offspring cells or colonies? We found that multicellularity can evolve with interactions far beyond cooperation or division of labour scenarios. More surprisingly, most of the life cycles found fall into either of two categories: A parent colony splits into two multicellular parts, or it splits into multiple independent cells.

### 2.1 Introduction

The evolution of multicellular life cycles is one of the most challenging questions of modern evolutionary biology. In the history of life, multicellular organisms have independently originated at least 25 times from unicellular ancestors [Grosberg and Strathmann, 2007]. From the very beginning, multicellular life has been shaped by interactions between different cells within heterogeneous groups [Okasha, 2006, Godfrey-Smith, 2009]. The role of these interactions in the emergence (or prevention) of multicellularity is an open question. Recently, there has been a rising interest in the evolution of life cycles including multicellular stages from both experimentalists [Rossetti et al., 2011,

## Chapter 2. Interacting cells driving the evolution of multicellular life cycles

---

Ratcliff et al., 2012, 2013b,a, Hammerschmidt et al., 2014] and theoreticians [Rainey and Kerr, 2010, Tarnita et al., 2013, Libby et al., 2014, De Monte and Rainey, 2014, Rashidi et al., 2015, van Gestel and Tarnita, 2017, Pichugin et al., 2017]. Multicellular clusters can emerge either as the result of clonal development (staying together in terms of [Tarnita et al., 2013]) or by aggregation of cells and smaller clusters (coming together). In the present study, we focus on competition between various “staying together” life cycles. The life cycle that leads to the fastest population growth would eventually dominate the population. We address how interactions between different cells within heterogeneous groups affect the growth competition between unicellular and multicellular life cycles. When interactions between different types of individuals within one group accelerate growth, more complex forms of multicellularity are expected to evolve in the long run.

We design our study with two specific scenarios of interacting cells in mind: the threat of free-riders in groups relying on cooperation and division of labour between cells. The very first multicellular organisms are commonly suggested to be composed of similar cells as suggested by fossils [Knoll, 1992, Tomitani et al., 2006] and experimental studies [Ratcliff et al., 2012, 2013b,a]. Cooperation between cells in these early organisms provided them benefits unavailable to solitary cells. However, free-riders gaining the cooperation benefits without paying any costs have an evolutionary advantage over cooperators, which in turn may violate the integrity of an organism [Hardin, 1968, Bonner, 1959, Rainey and Rainey, 2003a]. One of the most efficient ways of policing free-riders is reproduction via single cell bottleneck, where an organism grows from a single cell. This suggests that interactions between cooperators and free-riders promote group reproduction with unicellular propagules.

The second scenario where cell interactions might play a significant role emerges once undifferentiated multicellularity has been established and cells begin to specialize on various tasks. For example, consider filamentous cyanobacteria. During nitrogen depletion, cells in the filaments occasionally differentiate into nitrogen-fixating heterocysts that obtain sugars from neighbouring photosynthetic cells and, in turn, provide these

cells with nitrogen. These heterocysts suffer a significant penalty to their own fitness, but are essential to the survival of the colony as a whole [Flores and Herrero, 2010]. A group reproduction mode preserving the necessary association between photosynthetic and rare nitrogen fixating cells would contribute a lot to the sustainable growth of this species. Naturally, the reproduction of cyanobacteria occurs by fragmenting the parental filament into shorter multicellular chains through programmed cell death [Claessen et al., 2014], so newly emerged multicellular colonies likely contain heterocysts and benefit from the division of labour from the very beginning. This suggests that the division of labour promotes group reproduction modes with multicellular offspring groups.

While there is no clear experimental evidence that the evolution of reproduction modes can be influenced by the interaction between cells of different types, such a hypothesis deserves close attention. There is a range of previous models investigating the evolution of the division of labour [Gavrilets, 2010, Ispolatov et al., 2012, Rodrigues et al., 2012, Cooper and West, 2018]. However, these models incorporate a single predetermined reproduction mode, or a small hand-picked set of these. The evolution of cooperation in early multicellularity gained more attention [Michod, 1997, Nowak, 2006a, van Veelen, 2009]. Given that reproduction via single cell bottlenecks is a natural policing mechanism, some aspects of the evolution of reproduction modes have been considered before. Examples are the evolution of propagule size [Roze et al., 2001, Michod and Roze, 2001], as well as the comparison between the formation of cell clusters and unicellular life [Kaveh et al., 2016]. However, the spectrum of possible interactions between cells goes way beyond specific scenarios of cooperation and the division of labour, so this topic remains largely unexplored.

In our study, we utilize the framework developed in [Pichugin et al., 2017], in which a reproduction mode is considered as a way to partition the cells comprising the parent group into two or more offspring groups. Since there is always a finite number of cells in a reproducing group, there is a finite number of possibilities for group fragmentation. However, our previous study assumed homogeneous groups composed of a single

cell type. Here, we investigate heterogeneous groups consisting of cells of two different types. Groups grow in size by means of cell division (clonal development). Upon each cell division, the cell type of newborn cells can stochastically change, so no phenotype can go completely extinct. To represent the wide spectrum of possible interactions between two types, we use a game theory approach and focus on  $2 \times 2$  games, i.e. games in which two players with two strategies interact. The result of cell interactions are given by payoff values derived from the payoff matrix of a given game. The payoff values affect both the growth rate of the whole group as well as the different growth rates of cells within the group. The combination of the game played in a group and the fragmentation mode determines the population growth rate. By screening a wide range of fragmentation modes, we find the one providing the largest growth rate, which we consider to be the evolutionarily optimal reproductive strategy for the given game, as it leads to the fastest growth of biomass. Interestingly, when group growth is independent of the group size, our model suggests that only eight life cycles can be evolutionarily optimal among all possible  $2 \times 2$  games.

## 2.2 Methods

We consider a group-structured population, where individuals of two phenotypes *A* and *B* are nested into groups. These groups incrementally grow by one cell at a time and fragment into smaller offspring groups upon reaching a critical size of  $M$  cells. For a given group, the time between cell divisions depends only on the size of this group and its cell composition. Thus, the growth of the group is independent of other groups and therefore at the level of groups, the population growth is density independent. Therefore, in the long run, the population converges to a stationary regime, characterized by exponential growth at a rate we call  $\lambda$ . As populations employing different life cycles (different critical size and/or fragmentation mode) have different growth rates, the life cycle with the largest growth rate  $\lambda$  will eventually take over the population.

### Cell payoff and cell division

Interactions among cells in a group are captured by a pairwise game. The game is determined by a  $2 \times 2$  payoff matrix

$$\begin{array}{cc} & \begin{array}{cc} A & B \end{array} \\ \begin{array}{c} A \\ B \end{array} & \begin{pmatrix} a & b \\ c & d \end{pmatrix}, \end{array}$$

where  $A$  gets payoff  $a$  or  $b$  from interacting with  $A$  or  $B$  respectively, whereas  $B$  gets  $c$  or  $d$  from  $A$  or  $B$ , respectively. The average payoffs are given by

$$\begin{aligned} \alpha_{[i,j]} &= \frac{(i-1)a + jb}{i+j-1}, \\ \beta_{[i,j]} &= \frac{ic + (j-1)d}{i+j-1}, \end{aligned} \quad (2.1)$$

where  $\alpha_{[i,j]}$  and  $\beta_{[i,j]}$  are the average payoff of  $A$  type cells and  $B$  type cells in a group of  $i$   $A$ -cells and  $j$   $B$ -cells, respectively (the  $-1$  arises from the exclusion of self-interactions, but such self interactions have only a minor influence on our results, see S6 Appendix). Solitary cells do not play the game and their payoff is zero, so  $\alpha_{[1,0]} = \beta_{[0,1]} = 0$ .

Once a cell division occurs, the probability of a cell to be chosen to divide increases linearly with its fitness,  $P \sim 1 + w\alpha$  if the cell is of type  $A$ , and  $P \sim 1 + w\beta$  if the cell is of type  $B$ , where  $w \ll 1$  is the selection strength and the 1 measures the background fitness identical for all cells. Therefore, the probabilities that the dividing cell will be of type  $A$  or  $B$  under weak selection,  $w \ll 1$ , are

$$\begin{aligned} P_{[i,j]}^A &= \frac{i(1 + w\alpha_{[i,j]})}{i(1 + w\alpha_{[i,j]}) + j(1 + w\beta_{[i,j]})} \approx \frac{i}{i+j} + w \frac{ij}{(i+j)^2} (\alpha_{[i,j]} - \beta_{[i,j]}), \\ P_{[i,j]}^B &= \frac{j(1 + w\beta_{[i,j]})}{i(1 + w\alpha_{[i,j]}) + j(1 + w\beta_{[i,j]})} \approx \frac{j}{i+j} - w \frac{ij}{(i+j)^2} (\alpha_{[i,j]} - \beta_{[i,j]}), \end{aligned} \quad (2.2)$$

where  $P_{[i,j]}^A$  is the probability that some cell of type  $A$  will be chosen to divide in a group of  $i$   $A$ -cells and  $j$   $B$ -cells, and  $P_{[i,j]}^B$  is the same for type  $B$ , so  $P_{[i,j]}^A + P_{[i,j]}^B = 1$ .

## Chapter 2. Interacting cells driving the evolution of multicellular life cycles

---

Similarly, the time between two consecutive cell divisions depends linearly on the average payoff in a group

$$t_{[i,j]} = T_{i+j} \left( 1 - w \frac{i\alpha_{[i,j]} + j\beta_{[i,j]}}{i+j} \right), \quad (2.3)$$

where  $T_{i+j}$  is the size dependent component of growth, and  $\frac{i\alpha_{[i,j]} + j\beta_{[i,j]}}{i+j}$  is an average payoff of cells in a group.

In our model, both  $P_{[i,j]}^A$ ,  $P_{[i,j]}^B$  and  $t_{[i,j]}$  are dependent on cell's payoff. Cells with larger payoff have a higher chance ( $P_{[i,j]}$ ) to reproduce, when the group grows incrementally. Thus, also groups with larger average payoff grow faster. Otherwise, under payoff-independent growth times ( $t_{[i,j]} = T_{i+j}$ ), the group composition would have no effect on the group growth. Consequently, in such a case the evolution of life cycles is driven by group size alone, a scenario which we investigated in previous work [Pichugin et al., 2017]. In other words, selection acts on the cell level via  $P_{ij}$  and selection acts on the group level via  $t_{[i,j]}$ .

When a cell divides, each of the two daughter cells may independently change their type with probability  $m$ . Thus, the daughter cells of an  $A$  cell are either two  $A$ -cells with probability  $(1 - m)^2$ , or one  $A$  and one  $B$  cell with probability  $2m(1 - m)$ , or two  $B$ -cells with probability  $m^2$ .

Once the group reaches the critical group size  $M$ , it immediately fragments into smaller pieces and all cells are randomly assigned to offspring groups. The life cycle is determined by the critical size  $M$  and the sizes of offspring groups. For instance, at  $M = 3$ , there are two possible life cycles: either split into three solitary cells (life cycle 1+1+1), or into a solitary cell and bi-cellular group (life cycle 2+1). For  $M = 4$ , there are four possible life cycles: 3+1, 2+2, 2+1+1 and 1+1+1+1. Below, we refer to different life cycles using partitions of integer numbers.



## Population growth rate

We assume that the population can grow without any bounds. For our model, the density of groups follows a linear differential equation and growth is exponential [Pichugin et al., 2017]. Our goal here is to find the overall population growth rate  $\lambda$ .

To do so, we need to take into account the stochastic nature of group development in our model. There are three sources of stochasticity: (i) the choice of the cell to divide, (ii) the phenotype of daughter cells after cell division, and (iii) the assignment of cells to offspring groups at group fragmentation. As a consequence, groups are born different: a newborn bi-cellular group may consist of two  $A$ -cells, one  $A$  cell and one  $B$  cell, or two  $B$ -cells. Also, due to the randomness in outcomes of individual cell divisions, initially identical groups may follow different developmental trajectories during their growth, where by “developmental trajectory”, we mean the record of all choices made among possible alternatives during the group growth.

Fortunately, the number of newborn states and the cell composition after each division is finite, see Fig 2.1. Therefore, for any life cycle, we take all possible developmental trajectories into account. For an arbitrary life cycle, each group is born as one of  $S$  initial types, which we enumerate as  $(1, 2, \dots, S)$ . For each available developmental trajectory  $\tau$ , we designate the initial state of the trajectory as  $i(\tau)$ , the probability that a group born at initial state  $k$  will follow the trajectory as  $p_k(\tau)$  (such that  $p_k(\tau) = 0$ , if  $k \neq i(\tau)$ ), the time necessary to complete the trajectory as  $T(\tau)$ , and the vector of numbers of each offspring type produced at the end of the trajectory as  $\mathbf{N}(\tau) = (N_1, N_2, \dots, N_S)$ .

The growth rate of population  $\lambda$  is given by the solution of the equation (Appendix 2.4.1)

$$\det(Q - I) = 0, \tag{2.4}$$

where  $I$  is the identity matrix and  $Q$  is a matrix in which

$$Q_{i,j}(\lambda) = \sum_{\tau} p_i(\tau) N_j(\tau) e^{-\lambda T(\tau)} \tag{2.5}$$

## Chapter 2. Interacting cells driving the evolution of multicellular life cycles

is the contribution of groups born as type  $i$  to the production of newborn groups of type  $j$ , see also [De Roos, 2008].

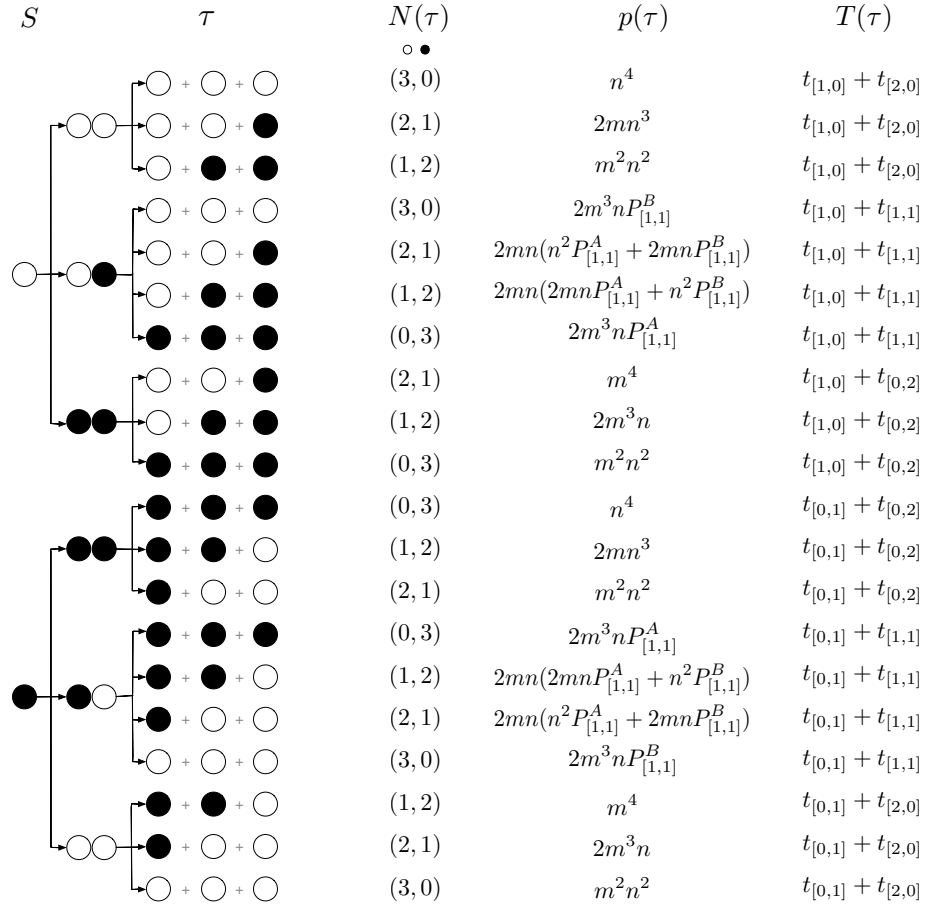


Figure 2.1: The number of developmental trajectories is finite. Here we show the full set of all 20 developmental trajectories ( $\tau$ ) in the life cycle 1+1+1, where groups are born unicellular, then grow up to size three and immediately split into independent cells. This life cycle features only two initial states  $S$ : solitary A-cell (open circles) and solitary B-cell (black circles). Stochastic phenotype switching creates 10 possible developmental trajectories for each initial state. To shorten the notation, we use  $n = 1 - m$  for the probability of a daughter cell to have the same phenotype as the mother cell.

## 2.3 Results

Our model allows us to calculate the growth rate of any given life cycle provided the elements of the payoff matrix  $(a, b, c, d)$ , the phenotype switching probability  $m$ , and the profile of the size-dependent component of development time  $(T_{i+j})$ . Here, we focus on life cycles having the largest  $\lambda$ , as these will be the winners of evolutionary growth competition.

In our study, we assume that in the absence of interactions ( $w = 0$ ), all life cycles share the same population growth rate i.e. all cells divide independently at the same rate. This assumption ensures that growth rates are exclusively determined by cell interactions alone. For this, the time of doubling of the group size must be the same for groups of any size, i.e. we need to satisfy  $T_i = \ln\left(\frac{i+1}{i}\right)$ . Only then, the time to grow from size  $k$  to  $2k$  is independent of  $k$ :  $\sum_{i=k}^{2k-1} T_i = \ln\left(\frac{(k+1)\cdots(2k)}{k\cdots(2k-1)}\right) = \ln(2)$ . As we show in Appendix 2.4.2, at  $w = 0$  this leads to the same population growth rate  $\lambda = 1$  for all life cycles. We also considered other developmental time profiles at  $w = 0$  and the results of our model are similar to our previous investigation of life cycles of homogeneous groups [Pichugin et al., 2017], see Appendix 2.4.3.

Under weak selection, the growth rate of the population with an arbitrary life cycle  $\kappa$  can be approximated by  $\lambda \approx 1 + w\lambda'_\kappa$ . The expression for  $\lambda'_\kappa$  can be obtained from a linearisation of Eq. (2.4) with respect to  $w$ . Since the payoffs  $a, b, c, d$  always come into play with a factor  $w$  (see Eqs. (2.2), (2.3)),  $\lambda'_\kappa$  is also linear in these payoffs. The dynamics of the population as a whole does not change if we exchange the two cell types  $A \leftrightarrow B$  and the corresponding payoff values  $a \leftrightarrow d, b \leftrightarrow c$ . Thus,  $a$  and  $d$  contribute to  $\lambda'_\kappa$  with the same weight and the same is true for  $b$  and  $c$ . Therefore,  $\lambda'_\kappa$  can be presented as a function of only three parameters:  $m$ ,  $\psi = a + d$  and  $\phi = \frac{b+c}{|a+d|}$ . The parameter  $\psi$  can be interpreted as whether the formation of a homogeneous group is beneficial to the cell ( $\psi > 0$ ) or not ( $\psi < 0$ ), compared to a solitary cell. The value of  $\phi$  is the benefit of interactions between cells of different types compared to interactions between cells of the

## Chapter 2. Interacting cells driving the evolution of multicellular life cycles

---

same type. The parameter  $\phi$  can also be interpreted as the benefit from the formation of a heterogeneous group compared to the formation of a homogeneous group. In a broad sense,  $\phi$  captures how well groups of mixed composition perform against pure groups. The details of the calculation of  $\lambda'_k$  can be found in Appendix 2.4.4.

M=2	M=3	M=4	M=5	M=6	M=7
1+1	1+1+1	1+1+1+1	1+1+1+1+1	1+1+1+1+1+1	1+1+1+1+1+1+1
	2+1	2+1+1	2+1+1+1	2+1+1+1+1	2+1+1+1+1+1
		3+1	2+2+1	2+2+1+1	2+2+1+1+1
		2+2	3+1+1	3+1+1+1	3+1+1+1+1
			4+1	3+2+1	2+2+2+1
			3+2	4+1+1	3+2+1+1
				2+2+2	4+1+1+1
				5+1	3+2+2
				3+3	4+2+1
				4+2	5+1+1
					3+3+1
					6+1
					5+2
					4+3

Figure 2.2: The list of all life cycles with critical sizes  $M \leq 7$ . The coloured life cycles are those found to be evolutionarily optimal for some combination of the control parameters  $m$ ,  $\psi$  and  $\phi$ . Most life cycles were never found to be optimal. Among 24 life cycles corresponding to the two largest critical sizes  $M = 6$  and  $M = 7$ , only one is found to be evolutionary optimal – 4+3.

We numerically investigated the optimality of life cycles with fragmentation size  $M$  up to 7. In total, there are 37 such life cycles, see Fig 2.2. To illustrate the results of our approach, we begin from presentation of evolutionary optimal life cycles for the specific case of a Prisoner's dilemma. Consider a game with payoff matrix

$$\begin{array}{c}
 A \quad B \\
 A \begin{pmatrix} 1 & -3 \\ c & 0 \end{pmatrix}.
 \end{array}$$

For  $c$  between 1 and 5, this game is a Prisoner's dilemma. For this payoff matrix,  $\psi = 1$  and  $\phi = c - 3$ . Additionally, we set the phenotype switching probability to  $m = 0.5$ . Among 37 considered life cycles, only three life cycles 1+1, 2+2, and 4+3 are found to be optimal in this case, see Fig 2.3. When the temptation to defect  $c$  is low, the best life cycle is unicellular, as the main outcome of the emergence of defectors is merely the harm to cooperators. With an increase of the temptation to defect  $c$ , the payoff of a heterogeneous group increases. Then, the benefits of occurring in a homogeneous group compensate the risks of occurring in a heterogeneous one. Consequently, the life cycle 2+2 becomes optimal starting from  $c = 2$  ( $\phi = -1$ ). Finally, at large  $c$ , heterogeneous groups gain a larger average payoff than homogeneous groups, so the life cycle 4+3 becomes optimal at  $c > 4$  ( $\phi > 1$ ).

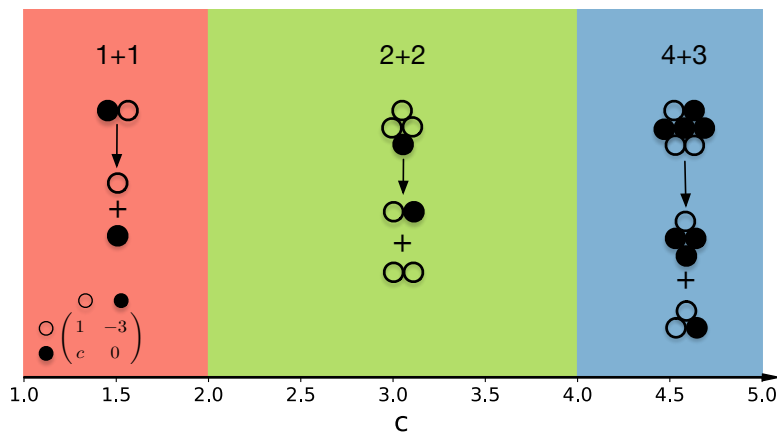


Figure 2.3: **Life cycles driven by a Prisoner's Dilemma game.** For  $c < 2$ , unicellular life cycles are optimal, as they avoid the fitness costs of heterogeneous groups. For  $2 < c < 4$ , the benefits of occurring in a homogeneous group compensate the risks of occurring in a heterogeneous one and the life cycle 2+2 becomes optimal. Finally, for  $c > 4$ , heterogeneous groups gain a larger average payoff than homogeneous groups, so the life cycle 4+3 becomes optimal. Note that the sketches of life cycles are only examples, as any distribution into black and white cells is possible (parameter values  $m = 0.5$ ,  $a = 1$ ,  $b = -3$ ,  $d = 0$ , such that  $\psi > 0$  and  $\phi = c - 3$ ).

Next, we proceed to the general game with arbitrary choice of each of the three control parameters  $\psi$ ,  $\phi$  and  $m$ . We now search for the optimal life cycle among all 37 life cycles with  $M \leq 7$ . Only eight of these life cycles were found to be evolutionarily optimal for any combination of control parameters, see Fig 2.4. These life cycles fall into one of three categories: fission into multiple unicellular offspring (1+1, 1+1+1, 1+1+1+1, and 1+1+1+1+1); binary fragmentation with group propagules (2+2 and 4+3); and the rarely observed transition between the previous two classes (2+1 and 2+1+1).

The observed set of life cycles is affected by the limit of the maximal group size, which is only 7 cells. However, among these eight, only the life cycle 4+3 reflects this limit. While for our methods it remains a challenge to investigate life cycles of larger groups, a clear pattern appears when we decrease the group size limit. If the group size is limited to  $M \leq 5$ , the life cycle 4+3 is unavailable but the life cycle 3+2 is evolutionary optimal, instead. Extending the size limit to  $M \leq 6$ , that life cycle is replaced by 3+3, and finally at  $M \leq 7$ , the life cycle 4+3 takes this place. These life cycles are likely the manifestation of the more general rule “grow as large as possible and divide into two equal or almost equal parts”. Thus, for any maximal group size, we suspect that there will be only eight evolutionary optimal life cycles, seven of which fragment at sizes five or smaller, and the eighth life cycle is the equal split at the maximal size.

We break the remaining analysis of our results into two parts. First, we consider specific life cycles and outline the conditions which promote their evolution. Then, we take the opposite direction and focus on specific games to investigate which life cycles are promoted by them.

### 2.3.1 Games promoting a given life cycle

First, we examine the optimal life cycles for negative  $a + d$  ( $\psi < 0$ ), in which homogeneous groups are in adverse conditions in the first place, see Fig 2.4D. Consequently, one of two life cycles found here is 1+1 – unicellularity, at which groups are not formed at all. Still, if  $\phi$  is sufficiently large, the highest growth rate is obtained by heterogeneous

groups. Then, evolutionary growth competition favours life cycles minimizing the fraction of homogeneous groups in the population. Due to the random partitioning of cells into offspring groups, smaller offspring have larger chances to accumulate cells of only one type during fragmentation. Thus, growth competition would likely promote larger offspring size to avoid such outcomes. If so, the optimal life cycle must be the fragmentation into two equal-sized (or nearly equal-sized) offspring groups at the maximal available size (4+3 in our case). Next, we focus on the more complex case of  $\psi > 0$ , see Fig 2.4A.

When  $\phi > 1$ , the life cycle 4+3 is evolutionarily optimal. At these values of  $\phi$ , all groups have an advantage over solitary cells, but heterogeneous groups profit more than homogeneous ones. Therefore, growth competition favours life cycles avoiding production of independent cells and minimizing the fraction of homogeneous groups in the population, i.e. an equal binary split at the maximal size. Note that at  $\phi = 1$ , where  $a + d = b + c$  (equal gains from switching), there are no benefit differences between homogeneous and heterogeneous groups. As a consequence, all life cycles with multicellular offspring have the same growth rate there, see Fig 2.4B. The popular case of constant selection, where one type is always better off than or at least equally good as the other one, can be modelled by  $a = b$  and  $c = d$ . This implies  $\phi = 1$ , and leads to the same life cycles as the more general case of equal gains from switching.

For  $0 < \phi < 1$ , 2+2 is the optimal life cycle. Here, all groups have an advantage over solitary cells, but homogeneous groups benefit more than heterogeneous ones. Therefore, growth competition would likely promote life cycles maximising the fraction of homogeneous groups in a population. First, this means producing the smallest multicellular offspring (bicellular groups) to eliminate parental heterogeneity in offspring. Second, the fragmentation has to be performed at the smallest size to minimize the risk of gaining heterogeneity in groups due to a spontaneous phenotype switch during growth. For the bi-cellular offspring, the smallest fragmentation size is four cells, therefore, the best life cycle must be 2+2. Interestingly, if  $m$  is small enough, the 2+2 life cycle can be optimal un-

## Chapter 2. Interacting cells driving the evolution of multicellular life cycles

---

der arbitrary large negative  $\phi$ , see Fig 2.4A, C. There, while heterogeneous groups have a strong disadvantage, chances of the phenotype switch to occur are low and homogeneity of groups is generally preserved.

At  $\phi < 0$ , the emergence of another cell type in homogeneous groups incurs a penalty on the group growth. To avoid the production of heterogeneous groups, growth competition is likely to promote life cycles involving dispersal into independent cells, such that each newborn group starts in a homogeneous state.

When  $\phi < 0$  and  $m$  is high enough, heterogeneous groups are likely to form after the very first cell division. In this case, 1+1 is favoured as it does not involve any group formation at all. However, once  $m$  approaches zero, the first few cell divisions performed by initially solitary cell will likely produce a homogeneous group. Thus, multicellular life cycles with fission into independent cells are favoured: 1+1+1, 1+1+1+1, and 1+1+1+1+1, see Fig 2.4C and Fig 2.7 in Appendix 2.4.5. Larger fragmentation sizes, first 3, 4, and then 5, become optimal with decreasing  $m$ . However, fission at size 6 was never found to be optimal, because at this stage, the production of multicellular offspring becomes beneficial, despite the risk of transferring parent heterogeneity into the next generation.

Transitional life cycles 2+1 and 2+1+1 are found to be optimal between areas of optimality of multiple fission life cycles (1+...+1) and multicellular offspring life cycles (2+2 and 4+3), see Fig 2.4C. These two life cycles mix unicellular and multicellular offspring. This may be a result of a compromise between producing multicellular offspring to fully utilize benefits of interactions in homogeneous groups, and the necessity to fragment into independent cells to purge emerging heterogeneous groups.



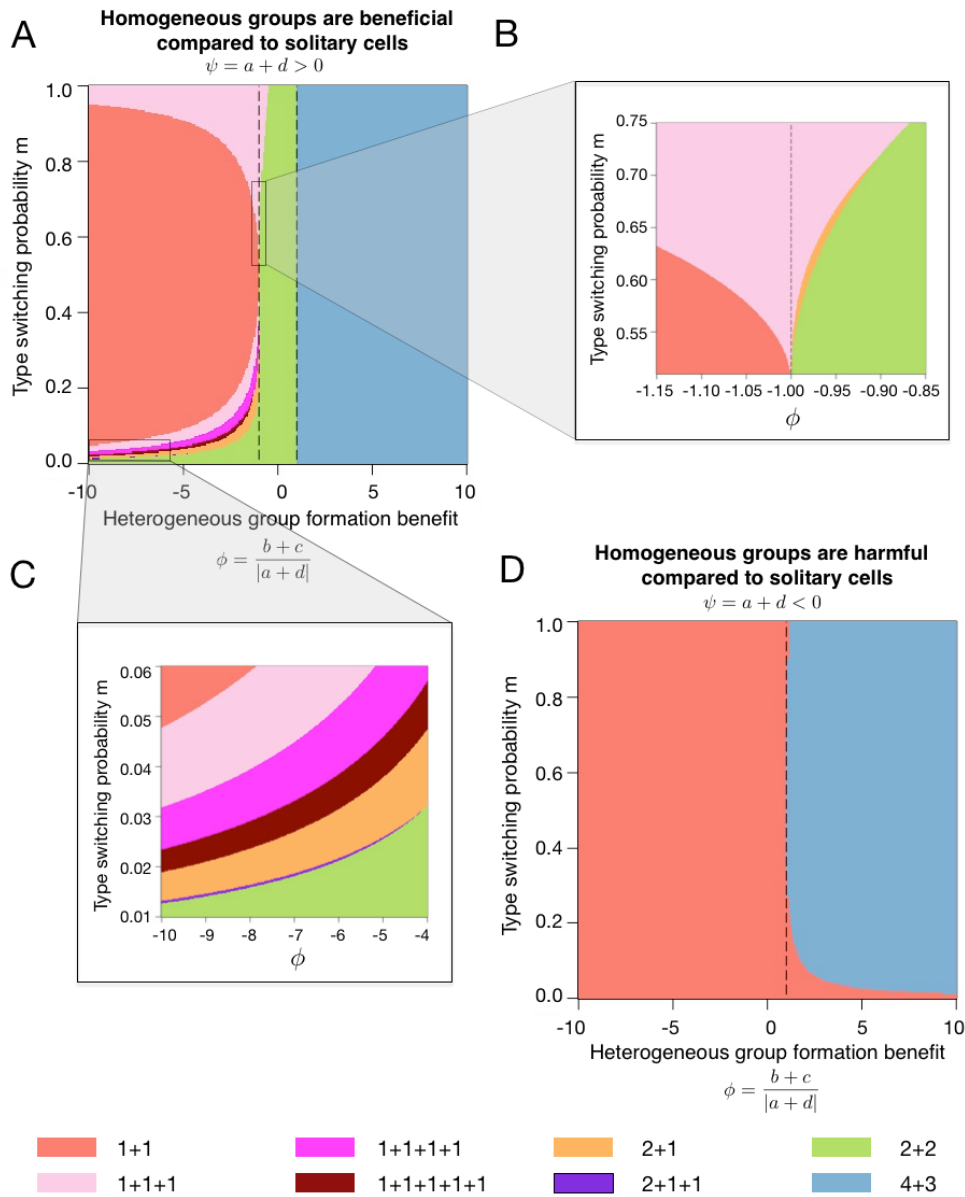


Figure 2.4: Only eight life cycles are evolutionarily optimal under weak selection for all  $2 \times 2$  games. Panel A: Optimal life cycles for  $\psi > 0$ . Dashed lines are  $\phi = -1$  and  $\phi = 1$ . Panel B: Enlargement of the area of large phenotype switching rate  $m > 0.5$  and  $\phi \approx 1$ . In a small region within this area, the life cycle 2+1 emerges. Panel C: Enlargement of the area of small phenotype switching rate  $m \ll 1$ , where a large diversity of life cycles is observed, including the rare life cycle 2+1+1. Panel D: For  $\psi < 0$ , only two life cycles are optimal. The dashed line is  $\phi = 1$ .

### 2.3.2 Life cycles promoted by prominent games

The most prominent game in the context of evolutionary game theory is the Prisoner's dilemma [Weibull, 1995, Nowak, 2006a, Pacheco et al., 2009, Hilbe et al., 2013]. In the simplest form of the Prisoner's dilemma, the donation game, each player may pay some cost  $\tilde{c}$ , so that the opposing player will receive a benefit  $\tilde{b}$  (larger than the cost). The cooperating strategy is to pay the cost, while the defecting strategy is to abstain from paying this cost (but still receive incoming benefits). The largest combined payoff is achieved by both players cooperating, while the individual's payoff resulting from defecting behaviour is always larger than payoff from mutual cooperation. The conflict between an individual's and group's interests makes this game a social dilemma.

The payoff matrix of the simplest Prisoner's dilemma is given by

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} \tilde{b} - \tilde{c} & -\tilde{c} \\ \tilde{b} & 0 \end{pmatrix}.$$

With these payoffs,  $\psi = \tilde{b} - \tilde{c} > 0$  and  $\phi = \frac{\tilde{b} - \tilde{c}}{\tilde{b} - \tilde{c}} = 1$ . Surprisingly, this game exhibits a special behaviour in our model: any life cycle which does not pass through the unicellular stage (e.g. 3+2+2) is evolutionarily optimal, independently of the phenotype switch probability  $m$  (i.e. risk of defector emergence). Contrary to our intuition, cooperative cell interactions described by the Prisoner's dilemma promote everything *except* reproduction via the single cell bottleneck. This is due to the fact that in a group with at least one cooperator, some benefit is already produced and shared across the group. Thus, preserving group living is more advantageous for the population than producing single cell propagules.

Other notable social dilemmas are the snowdrift game and the stag hunt game. In the snowdrift game, a combined cost  $\tilde{c}$  must be paid for the benefit  $\tilde{b}$  to be received by each player. Cooperators readily pay their share of the costs, while defectors abstain from paying it. The payoff matrix of the snowdrift game is

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} \tilde{b} - \tilde{c}/2 & \tilde{b} - \tilde{c} \\ \tilde{b} & 0 \end{pmatrix}. \quad (2.6)$$

This results in  $\psi = \tilde{b} - \tilde{c}/2 > 0$  and  $\phi = \frac{2\tilde{b}-\tilde{c}}{\tilde{b}-\tilde{c}/2} = 2$ . According to our findings, these parameters promote the life cycle 4+3, or, more generally, equal binary fragmentation at the maximal possible size, which ensure heterogeneous groups that maximize the combined payoff.

In the stag hunt game, players may pursue a hare – small prey providing payoff  $h$ , or a stag – large prey giving payoff  $s > h$ . A hare hunt is always successful, but only both hunters together can hunt down a stag. The payoff matrix of the stag hunt game is

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} s & 0 \\ h & h \end{pmatrix}. \quad (2.7)$$

This results in  $\psi = s + h > 0$  and  $\phi = \frac{h}{s+h} < \frac{1}{2}$ . These parameters promote the life cycle 2+2. In contrast to the Prisoner's dilemma and snowdrift games, in the stag hunt game a group of mixed composition has the smallest combined payoff (which is still larger than zero payoff for solitary cells for our choice of parameters). Therefore, the stag hunt game strongly favours a life cycle preserving homogeneity of groups, i.e. 2+2.

Many other evolutionary games have been studied and applied in a wide variety of biological situations [Maynard Smith, 1982, Hofbauer and Sigmund, 1998, Nowak and Sigmund, 2004, Broom and Rychtář, 2013]. For the case of  $2 \times 2$  games, in a large well-mixed population of players, three classes of evolutionary dynamics are possible: dominance of one strategy ( $a > c, b > d$  or  $a < c, b < d$ ), bistability ( $a > c, b < d$ ) or coexistence ( $a < c, b > d$ ) [Weibull, 1995, Nowak, 2006b].

All games experiencing a bistability (such as the stag hunt game) have  $\phi < \text{sign}(\psi)$ . According to our results, for positive  $\psi$ , bistability games can promote 7 out of 8 found life cycles: equal binary split at the maximal size (4+3) never leads to the fastest growth

rate. For negative  $\psi$ , bistability games only lead to a unicellular life cycle (1+1). Games featuring coexistence dynamics (such as the snowdrift game) satisfy  $\phi > \text{sign}(\psi)$ , which restricts the optimal life cycle to 4+3 under  $\psi > 0$  but allows both 1+1 and 4+3 under  $\psi < 0$ . Dominance games (such as the Prisoner's dilemma) may have any value  $\phi$ , so they can promote any of the 8 found life cycles.

## Discussion

In our study we performed an extensive investigation of the competition of life cycles driven by interactions between cells within a group. Key to this study is the consideration of all possible reproduction modes and all possible interactions captured by game theoretic  $2 \times 2$  payoff matrices. Among the huge variety of reproduction modes, only eight were found to be evolutionarily optimal, see Fig 2.2 and Fig 2.4. Moreover, the vast majority of games promotes either of two very specific classes of life cycles: fragmentation into strictly unicellular offspring (1+...+1) or production of exactly two strictly multicellular daughter groups of identical (or almost identical) size. Intuitively, life cycles with unicellular offspring should be promoted when the cells grow fastest in a homogeneous group, as the single cell bottleneck eliminates heterogeneity in the most effective way. Similarly, when the cells grow fastest in a heterogeneous group, life cycles with multicellular offspring should be promoted as they are best at preserving heterogeneity. Our results, in general, support this intuition. However, the current work reveals a much broader picture and we observed a number of less intuitive features of life cycle evolution driven by cell interactions. First, we observed the transition between these two major life cycles classes. This occurs via transitional life cycles mixing unicellular and multicellular offspring (such as 2+1 and 2+1+1), see Fig 2.4C. Second, we found that if being in a heterogeneous groups incurs a moderate penalty onto the cell, growth competition may still promote the life cycle with only multicellular offspring (2+2), even at high rates of phenotype switching ( $m$ ), see Fig 2.4A. Third, an arbitrary strong penalty to heterogeneous

groups ( $\phi < 0$ ), may still lead to the evolution of life cycles with multicellular offspring (2+2) given small enough  $m$ , see Fig 2.4A, C. Altogether, even with only eight life cycles observed, our model exhibits a rich behaviour and gives insights into factors shaping the evolution of life cycles.

We found that social dilemma games may not promote the evolution of single cell bottlenecks. A naive intuition suggest life cycles with unicellular offspring to be favoured by all social dilemmas, as a single cell bottleneck is an effective way to police defectors. However, social dilemmas may lead to the evolution of any of the eight life cycles. What would be the reason for such a counter-intuitive outcome?

A key difference between our approach and the most of studies utilizing evolutionary game theory is that while we allow the competition between different cell types (by means of different division probabilities  $P^A$  and  $P^B$ ), winning in such a competition is not in the focus of our attention. We consider both cell types as essential components of group development. This is in line with the previous idea of [Rainey and Kerr, 2010] that cheaters may play a significant role in the evolution of life cycles in early multicellularity. Embracing this approach, we acknowledge that life cycles showing the largest population growth rates are not necessarily the best in keeping cheaters out. Our results show that for evolution to favour single cell bottlenecks, a group mixing cooperators and defectors should have lower average fitness than an equivalent pack of independent cells. Otherwise, life cycles with multicellular offspring will be promoted.

This leads to a second key feature of our model: the role of solitary cells. Independent cells stand out as they have no other cell to interact with and, thus, do not play a game. As such, they serve as a benchmark of the cell behaviour, against which all other group compositions are compared. Our results indicate that optimality of life cycles strongly depends on whether a (homogeneous) group formation is beneficial or deleterious compared to a solitary cell, see Fig 2.4 A and D, respectively. For the Prisoner's dilemma game, a combination of a single cooperator and single defector, indeed harm the cooperator the most. However, the overall payoff to the group ( $\tilde{b} - \tilde{c} > 0$ ) is still

## Chapter 2. Interacting cells driving the evolution of multicellular life cycles

---

larger than the zero cumulative payoff these cells would obtain if separated. Thus, the Prisoner's dilemma promotes the production of the multicellular offspring in our model. The opportunity to abstain from the game (loner strategy) [Hauert et al., 2002, Fowler, 2005, Brandt et al., 2006, Hauert et al., 2007, Traulsen et al., 2009, García and Traulsen, 2012] is often viewed as a component of the secondary importance in evolutionary game theory models, despite its potential impact on microbial dynamics [Garcia et al., 2014, 2015]. For the evolution of life cycles, such an opportunity plays a central role. For any life cycle producing unicellular offspring, each member of the population passes through a developmental stage without any interaction. Also, an ultimate loner strategy, where no game is ever played, is implemented by the unicellular life cycle, which is the most basic and one of the most important reproduction modes. If we allow self-interactions, the optimality of life cycles changes insignificantly (see Appendix 2.4.6) and even fewer life cycles, only five, can be optimal in this case.

The interplay between cell interactions and life cycles has been considered in previous studies. [Roze et al., 2001] compared the growth rate of two reproductive modes: a spore reproducer (multiple fission life cycles in our terms) and the fragmentation into same sized offspring groups. Based on the fitness effects from the colony size, they investigated the question which life cycle is good at eliminating mutations deleterious at the colony level.

An explicit connection between fragmentation modes and games played within the group was first made by [Kaveh et al., 2016]. There, authors focused on fragmentation modes in a form  $x + 1$ , and explicitly considered the 2+1 life cycle. Being focused on cooperation rather than evolution of life cycles, they discussed conditions promoting the evolution of cooperation.

The results of our model can be directly compared with our previous findings in [Pichugin et al., 2017] and [Pichugin and Traulsen, 2020], which considered the evolution of life cycles in homogeneous groups. There, for costless fragmentation (as in the present study), only binary fragmentation modes (i.e. in a form  $x + y$ ) can be evolutionarily op-

timal. Once reproduction incurs a cost, fragmentation into multiple parts may evolve, but still some fragmentation modes remain “forbidden”, i.e. they cannot evolve under any fitness landscape ( $T_i$  in our terms). The set of evolutionarily optimal life cycles found in the current study is significantly different from the sets described above. Fragmentation in our model is costless, and yet we found that fragmentation into multiple parts may evolve due to the impact of cell interactions. Also, the life cycle 2+1+1, which may evolve in our model, belongs to the class of “forbidden” life cycles under costly reproduction, so it cannot evolve among homogeneous groups at all. Thus, the introduction of heterogeneity and interactions between different cell types make it possible for previously unattainable life cycles to evolve.

In our work, we adopted the minimal setup of the heterogeneous groups - colonies with two cell types. The model can be extended by considering a larger number of cell types to model more developed organisms. In such a hypothetical model, the payoff matrix is larger than 2 by 2. Consequently, the set of control parameters is larger than just  $(\phi, \psi)$  as in the current study, so the complete analysis will be significantly more complex. Additionally, more types will require more sophisticated methods of phenotype switching than the single phenotype switching probability  $m$ . Naturally, in complex multicellular organisms, the phenotypes of cells are determined by developmental programs of the organism, which might be very complex.

In our model, we consider groups as a well mixed collection of cells, where an interaction between any two cells are equally likely. However, natural and experimental multicellular clusters generally have a specific geometry. For example, the multicellularity formed by *Saccharomyces cerevisiae* after selection has a roughly spherical snowflake-like shape, in which the central cells have a 76% frequency of death compared to random cell death with a probability of 6% [Ratcliff et al., 2012]. In this snowflake-like group, central cells have more neighbours and they may have a stronger influence on cell interactions than other cells within a group. As we have shown that the interactions between cells have an impact on life cycle evolution, so must have the geometry of the group as well.

## Chapter 2. Interacting cells driving the evolution of multicellular life cycles

---

However, these geometric considerations will lead to models far more complex than ours.

It is a challenging question how the interactions between different cells within an organism shape its reproduction mode. The present study demonstrates that this topic can be addressed systematically. To do so, we combine evolutionary game theory with the theory of life cycles in simple multicellular organisms. Game theory is able to capture arbitrary pairwise interactions by a payoff matrix. At the same time, the theory of life cycles represents an arbitrary reproduction mode by the partition of an integer number. These two general frameworks naturally complement each other and allow holistic investigation of life cycles of organisms with heterogeneous composition, where it is impossible to evaluate the evolution of one factor neglecting another.



## 2.4 Supporting information

### 2.4.1 Population growth rate in the case of stochastic developmental programs

Consider a population in which each group emerges as one of  $S$  initial types. These types could be the newborn groups of different size and/or composition. With time passing, a group grows from its initial size to maturity and subsequent fragmentation. The set of growth events (cells divisions, mutations, etc) may vary from group to group. We call such an event chain “developmental trajectory” and designate it as  $\tau$ . Any two groups of the same initial type may adopt different developmental trajectories for a number of reasons, such as mutations, stochastic developmental programs, or different environmental conditions. We use the following parameters of the developmental trajectory:  $i(\tau)$  – the initial state of the group leading to the given developmental trajectory,  $p_k(\tau)$  – the probability that a group that emerged as initial type  $k$  will follow the trajectory  $\tau$ , so  $p_k(\tau) = 0$ , if  $k \neq i(\tau)$ ,  $T(\tau)$  – the time necessary to the newborn group to complete the trajectory  $\tau$  and  $\mathbf{N}(\tau) = (N_1, N_2, \dots, N_S)$  – the vector of numbers of each offspring type produced during the fragmentation at the end of the trajectory  $\tau$ .

The population features an explicit maturation component: a newborn group does not reproduce until time  $T(\tau)$  has passed. Thus, to describe the population dynamics and find the population growth rate  $\lambda$ , it is necessary to consider the population demography. To do so, we characterize each group at each moment of time by the age parameter  $\eta$ . We define the age in a way that the newborn group has  $\eta = 0$ , while the group that reached the end of the developmental trajectory and is about to fragment has  $\eta = 1$ . Along the trajectory, the age increases at a constant rate equal to  $\frac{1}{T(\tau)}$ , i.e. the rate of ageing differs between different trajectories.

From the perspective of the population dynamics, any two groups sharing the same developmental trajectory  $\tau$  and age  $\eta$  are identical. Thus, the state of the whole popula-

## Chapter 2. Interacting cells driving the evolution of multicellular life cycles

---

tion can be described by the density function  $\zeta(\tau, \eta, t)$ , which shows how many groups on the developmental trajectory  $\tau$  have age  $\eta$  at the given time  $t$ . In the stationary regime, where the fraction of groups of each type stays constant, the density function grows exponentially,

$$\zeta(\tau, \eta, t) = \rho(\tau, \eta)e^{\lambda t}, \quad (2.8)$$

where  $\rho(\tau, \eta)$  is the stationary density distribution of groups in a population.

Within a given developmental trajectory, ageing occurs at the same rate for all groups. Therefore, the dynamics of the density function at a given age  $\eta$  is determined by the balance between influx of maturing younger groups and the outflux of groups becoming too old. Both processes occur with the same rate  $\frac{1}{T(\tau)}$ , thus the density function must satisfy the transport equation

$$\frac{\partial \zeta}{\partial t} = -\frac{1}{T(\tau)} \frac{\partial \zeta}{\partial \eta}. \quad (2.9)$$

Combining Eqs. (2.8) and (2.9) we get

$$\lambda \rho = -\frac{1}{T(\tau)} \frac{\partial \rho}{\partial \eta}$$

The solution of this equation is

$$\rho(\tau, \eta) = \rho_0(\tau)e^{-\lambda T(\tau)\eta}, \quad (2.10)$$

where  $\rho_0(\tau)$  is the stationary density distribution of newborn groups with  $\eta = 0$ .

To find  $\rho_0(\tau)$ , we use the fact that each newborn organism is produced as a result of the fragmentation of some mature organism. Thus, the rate of emergence of newborn organisms in the population ( $j_0$ ) is the same as the rate of production of offspring in the course of reproduction of mature organisms ( $j_1$ ).

For any developmental trajectory  $\tau$ , the rate of entering into the newborn state per

time unit is equal to

$$j_0(\tau) = \frac{\zeta(\tau, 0, t)}{T(\tau)}, \quad (2.11)$$

where the right hand side of the equation is the product of the number of newborn groups and the rate of ageing. The number of offspring with developmental trajectory  $\tau$  is equal to the product of the total number of offspring of type  $i(\tau)$  produced by all mature organisms and the probability of the offspring to adopt this developmental trajectory ( $p_{i(\tau)}(\tau)$ )

$$j_1(\tau) = p_{i(\tau)}(\tau) \sum_{\tau'} \frac{N_{i(\tau)}(\tau')}{T(\tau')} \zeta(\tau', 1, t), \quad (2.12)$$

where summation is performed over all possible developmental trajectories of parent groups.

Since each produced propagule is a newborn organism,  $j_0(\tau) = j_1(\tau)$ . Therefore,

$$\frac{\rho_0(\tau)}{T(\tau)} = p_{i(\tau)}(\tau) \sum_{\tau'} \frac{N_{i(\tau)}(\tau')}{T(\tau')} \rho_0(\tau') e^{-\lambda T(\tau')}. \quad (2.13)$$

To obtain the expression connecting the population growth rate  $\lambda$  with parameters of developmental trajectories  $\tau$ , we multiply both parts by  $N_j(\tau) e^{-\lambda T(\tau)}$  (note that in general  $j \neq i(\tau)$ ) and sum over all possible developmental trajectories

$$\sum_{\tau} \frac{N_j(\tau)}{T(\tau)} \rho_0(\tau) e^{-\lambda T(\tau)} = \sum_{\tau} p_{i(\tau)}(\tau) N_j(\tau) e^{-\lambda T(\tau)} \left( \sum_{\tau'} \frac{N_{i(\tau)}(\tau')}{T(\tau')} \rho_0(\tau') e^{-\lambda T(\tau')} \right). \quad (2.14)$$

We define

$$X_i = \sum_{\tau} \frac{N_i(\tau)}{T(\tau)} \rho_0(\tau) e^{-\lambda T(\tau)} \quad (2.15)$$

$$Q_{i,j} = \sum_{\tau} p_i(\tau) N_j(\tau) e^{-\lambda T(\tau)}, \quad (2.16)$$

Note that  $p_j(\tau) = 0$  if  $j \neq i(\tau)$ .

Taking into account that  $p_j(\tau) = 0$  if  $j \neq i(\tau)$ , Eq (2.14) becomes

$$X_j = \sum_i Q_{i,j} X_i. \quad (2.17)$$

Also in the definition of  $Q_{i,j}$ , the result of summation over all trajectories  $\tau$  is the same as over only developmental trajectories starting from the initial state of type  $j$ , since  $p_j(\tau) = 0$ , if  $j \neq i(\tau)$ , because an organism emerged as one type has no access to developmental trajectories originated from other types.

Eq (2.17) can be satisfied only if

$$\det(Q - I) = 0, \quad (2.18)$$

where elements of matrix  $Q$  are defined by Eq (2.16) and  $I$  is identity matrix. This equation allows to infer the population growth rate  $\lambda$  if the parameters of each trajectory are known ( $i(\tau)$ ,  $p_i(\tau)$ ,  $\mathbf{N}(\tau)$  and  $T(\tau)$ ). In most interesting cases, this has to be done numerically.

### **2.4.2 Existence of the neutral fitness landscape in the case of homogeneous groups**

Consider the situation, where  $w = 0$  and, therefore, the group properties depend only on the group size. A group of size  $i$  grows in size to  $i + 1$  within time  $T_i$ . Here we show that if  $T_i = \ln\left(\frac{i+1}{i}\right)$ , all life cycles have the same growth rate  $\lambda = 1$ . We prove this by induction:

- The base of induction is given by Eq (4), which states that if  $T_1 = \ln\left(\frac{2}{1}\right)$  and  $T_2 = \ln\left(\frac{3}{2}\right)$ , then  $\lambda = 1$  for any life cycles fragmenting at size 3 or smaller.
- The step of induction must show that if the assumption of induction holds true for maximal size  $M$ , then under adding  $T_M = \ln\left(\frac{M+1}{M}\right)$ , the assumption also holds true for maximal size  $M + 1$ . To prove the step of induction, we only need to consider life cycles fragmenting exactly at the size  $M + 1$  because life cycles fragmenting at sizes smaller than  $M + 1$  have  $\lambda = 1$  according to the assumption of induction.

## 2.4. Supporting information

To construct the matrix  $Q$  and find the growth rate of considered life cycles, we need to characterize the set of offspring and developmental trajectories. In an arbitrary life cycle, the fragmentation of a homogeneous group of size  $M$  results in production of offspring groups of sizes ranging from 1 to  $M$ . In total,  $M$  different types of offspring can be produced, so the size of the matrix  $Q$  is  $M$  by  $M$ . Each of the offspring will grow up to size  $M + 1$  and then fragment, thus there is only one developmental trajectory for each type of offspring with  $p_i(\tau) = 1$ . The developmental time of the trajectory  $\tilde{T}(\tau)$  is given as the sum of incremental growth time

$$\tilde{T}_k(\tau) = \sum_{j=k}^M T_j = \ln \left( \frac{M+1}{k} \right), \quad (2.19)$$

where  $k$  denotes the size of the newborn offspring.

An arbitrary life cycle can be characterized by the distribution of offspring sizes produced upon fragmentation  $N_i$ , where  $i$  denotes the size of offspring. By the conservation of cell number during reproduction  $\sum_{i=1}^M iN_i = M + 1$ . Therefore, according to Eq (2.16), for an arbitrary life cycle, the elements of matrix  $Q_{ij}$  are given by

$$Q_{ij} = N_i e^{-\lambda \ln \left( \frac{M+1}{j} \right)} \quad (2.20)$$

To prove the step of induction, we verify whether  $\lambda = 1$  is the solution of Eq (2.18), with matrix  $Q$  given by Eq (2.20). Plugging  $\lambda = 1$  into Eq (2.20), we have  $Q_{ij} = N_i \frac{j}{M+1}$ , so the Eq (2.18) becomes

$$\begin{vmatrix} N_1 \frac{1}{M+1} - 1 & N_1 \frac{2}{M+1} & \cdots & N_1 \frac{M}{M+1} \\ N_2 \frac{1}{M+1} & N_2 \frac{2}{M+1} - 1 & \cdots & N_2 \frac{M}{M+1} \\ \vdots & \vdots & \ddots & \vdots \\ N_M \frac{1}{M+1} & N_M \frac{2}{M+1} & \cdots & N_M \frac{M}{M+1} - 1 \end{vmatrix} = 0. \quad (2.21)$$

Based on the properties of determinant, we can take out the coefficients of each row and

## Chapter 2. Interacting cells driving the evolution of multicellular life cycles

---

each column, then the left hand side of Eq (2.21) becomes

$$\frac{\prod_{i=1}^M iN_i}{(M+1)^M} \cdot \begin{vmatrix} 1 - \frac{M+1}{iN_1} & 1 & \cdots & 1 \\ 1 & 1 - \frac{M+1}{2N_2} & \cdots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \cdots & 1 - \frac{M+1}{MN_M} \end{vmatrix}. \quad (2.22)$$

For convenience, we neglect the coefficient and denote  $\frac{M+1}{iN_i}$  as  $K_i$ . Thus, the determinant is

$$\begin{vmatrix} 1 - K_1 & 1 & \cdots & 1 \\ 1 & 1 - K_2 & \cdots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \cdots & 1 - K_M \end{vmatrix}. \quad (2.23)$$

Next we calculate the determinant by splitting the first row,

$$\begin{vmatrix} 1 - K_1 & 1 & 1 & \cdots & 1 \\ 1 & 1 - K_2 & 1 & \cdots & 1 \\ 1 & 1 & 1 - K_3 & \cdots & 1 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & 1 & \cdots & 1 - K_M \end{vmatrix} = \begin{vmatrix} -K_1 & 0 & 0 & \cdots & 0 \\ 1 & 1 - K_2 & 1 & \cdots & 1 \\ 1 & 1 & 1 - K_3 & \cdots & 1 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & 1 & \cdots & 1 - K_M \end{vmatrix} + \begin{vmatrix} 1 & 1 & 1 & \cdots & 1 \\ 1 & 1 - K_2 & 1 & \cdots & 1 \\ 1 & 1 & 1 - K_3 & \cdots & 1 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & 1 & \cdots & 1 - K_M \end{vmatrix}. \quad (2.24)$$

For the second part, splitting the second row, we can get

$$\begin{aligned}
 \begin{vmatrix} 1 & 1 & 1 & \cdots & 1 \\ 1 & 1-K_2 & 1 & \cdots & 1 \\ 1 & 1 & 1-K_3 & \cdots & 1 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & 1 & \cdots & 1-K_M \end{vmatrix} &= \begin{vmatrix} 1 & 1 & 1 & \cdots & 1 \\ 0 & -K_2 & 0 & \cdots & 0 \\ 1 & 1 & 1-K_3 & \cdots & 1 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & 1 & \cdots & 1-K_M \end{vmatrix} \\
 &+ \begin{vmatrix} 1 & 1 & 1 & \cdots & 1 \\ 1 & 1 & 1 & \cdots & 1 \\ 1 & 1 & 1-K_3 & \cdots & 1 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & 1 & \cdots & 1-K_M \end{vmatrix}, \tag{2.25}
 \end{aligned}$$

The second term in Eq (2.25) is zero because the determinant has two identical columns, therefore only the first term remains. Continuing splitting the remaining rows of the first term of Eq (2.25), we finally obtain

$$\begin{aligned}
 \begin{vmatrix} 1 & 1 & 1 & \cdots & 1 \\ 1 & 1-K_2 & 1 & \cdots & 1 \\ 1 & 1 & 1-K_3 & \cdots & 1 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & 1 & \cdots & 1-K_M \end{vmatrix} &= \begin{vmatrix} 1 & 1 & 1 & \cdots & 1 \\ 0 & -K_2 & 0 & \cdots & 0 \\ 1 & 1 & 1-K_3 & \cdots & 1 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & 1 & \cdots & 1-K_M \end{vmatrix} \\
 &= \begin{vmatrix} 1 & 1 & 1 & \cdots & 1 \\ 0 & -K_2 & 0 & \cdots & 0 \\ 0 & 0 & -K_3 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & 1 & \cdots & 1-K_M \end{vmatrix} \\
 &= \dots \tag{2.26}
 \end{aligned}$$

$$\begin{aligned}
 &= \begin{vmatrix} 1 & 1 & 1 & \cdots & 1 \\ 0 & -K_2 & 0 & \cdots & 0 \\ 0 & 0 & -K_3 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & -K_M \end{vmatrix} \\
 &= (-1)^{K-1} \prod_{i \neq 1}^M K_i.
 \end{aligned}$$

Now, we look back at the first term in Eq (2.24), we split the second row

$$\begin{aligned}
 \begin{vmatrix} -K_1 & 0 & 0 & \cdots & 0 \\ 1 & 1-K_2 & 1 & \cdots & 1 \\ 1 & 1 & 1-K_3 & \cdots & 1 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & 1 & \cdots & 1-K_M \end{vmatrix} &= \begin{vmatrix} -K_1 & 0 & 0 & \cdots & 0 \\ 0 & -K_2 & 0 & \cdots & 0 \\ 1 & 1 & 1-K_3 & \cdots & 1 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & 1 & \cdots & 1-K_M \end{vmatrix} \\
 &+ \begin{vmatrix} -K_1 & 1 & 1 & \cdots & 1 \\ 1 & 1 & 1 & \cdots & 1 \\ 1 & 1 & 1-K_3 & \cdots & 1 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & 1 & \cdots & 1-K_M \end{vmatrix}.
 \end{aligned} \tag{2.27}$$

For the second term at the right hand side of Eq (2.27), similar to Eq (2.25) in the last step, we can work out that it equals  $(-1)^{M-1} \prod_{i \neq 2}^M K_i$ . That means we can get  $(-1)^{M-1} \prod_{i \neq j}^M K_i$  when split the  $j$ -th row. So we keep the same procedure to split the remaining rows of



the first term in Eq (2.27). After that, the initial determinant changes to

$$\begin{aligned}
 \begin{vmatrix} 1-K_1 & 1 & 1 & \cdots & 1 \\ 1 & 1-K_2 & 1 & \cdots & 1 \\ 1 & 1 & 1-K_3 & \cdots & 1 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & 1 & \cdots & 1-K_M \end{vmatrix} &= \begin{vmatrix} -K_1 & 0 & 0 & \cdots & 0 \\ 0 & -K_2 & 0 & \cdots & 0 \\ 0 & 0 & -K_3 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & -K_M \end{vmatrix} \\
 &+ (-1)^{M-1} \sum_{j=1}^M \prod_{i \neq j}^M K_i \\
 &= (-1)^M \prod_{i=1}^M K_i + (-1)^{M-1} \sum_{j=1}^M \prod_{i \neq j}^M K_i \\
 &= (-1)^M \left( \prod_{i=1}^M K_i - \sum_{j=1}^M \prod_{i \neq j}^M K_i \right) \\
 &= (-1)^M \left( \frac{(M+1)^M}{\prod_{i=1}^M iN_i} - \frac{(M+1)^{M-1} \sum_{i=1}^M iN_i}{\prod_{i=1}^M iN_i} \right) \\
 &= 0,
 \end{aligned} \tag{2.28}$$

where we used  $K_i = \frac{M+1}{iN_i}$  and  $\sum_{i=1}^M iN_i = M+1$  in the last two steps.

This proves that an arbitrary life cycle fragmenting at size  $M+1$  has the growth rate  $\lambda = 1$ , if  $T_i = \ln\left(\frac{i+1}{i}\right)$  for any  $i \leq M$ . This means that  $T_i = \ln\left(\frac{i+1}{i}\right)$  is a neutral fitness landscape for the scenario of homogeneous groups.

### 2.4.3 Life cycles of homogeneous groups

In the absence of cells' interactions, all cells are identical i.e. the cell type has no influence on groups. Essentially, all groups can be treated as homogeneous groups, in which only group sizes affect growth rate. In this case, groups have fixed developmental trajectories. For instance, the life cycle 1+1+1 has to go through the unique developmental trajectory: two successive divisions and then producing three single cells (see Fig 2.5). In this unique developmental trajectory, only one initial type exist – independent cell, so  $p(\tau) = 1$  and  $\mathbf{N}(\tau) = 3$ .

## Chapter 2. Interacting cells driving the evolution of multicellular life cycles



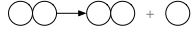


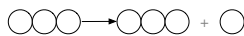




$LC$	$\tau$	$N(\tau)$ $\circ \infty \infty$	$p(\tau)$	$T(\tau)$
1 + 1		(2, 0, 0)	1	$t_{[1,0]}$
2 + 1		(1, 1, 0)	1	$t_{[1,0]} + t_{[2,0]}$
		(1, 1, 0)	1	$t_{[2,0]}$
1 + 1 + 1		(3, 0, 0)	1	$t_{[1,0]} + t_{[2,0]}$
3 + 1		(1, 0, 1)	1	$t_{[1,0]} + t_{[2,0]} + t_{[3,0]}$
		(1, 0, 1)	1	$t_{[3,0]}$
2 + 2		(0, 2, 0)	1	$t_{[2,0]} + t_{[3,0]}$
2 + 1 + 1		(2, 1, 0)	1	$t_{[1,0]} + t_{[2,0]} + t_{[3,0]}$
		(2, 1, 0)	1	$t_{[2,0]} + t_{[3,0]}$
1 + 1 + 1 + 1		(4, 0, 0)	1	$t_{[1,0]} + t_{[2,0]} + t_{[3,0]}$

Figure 2.5: Homogeneous groups have deterministic developmental trajectories for each type offspring group, i.e.  $p(\tau) = 1$ .

First, we investigate the simplest scenario, where the maximal size of the group was limited to two cells. There are three life cycles in total in this case: 1+1, 2+1 and 1+1+1. The matrices  $Q$  corresponding to these life cycles are

$$\begin{aligned}
 Q_{1+1} &= \left( 2e^{-\lambda T_1} \right), \\
 Q_{2+1} &= \begin{pmatrix} e^{-\lambda(T_1+T_2)} & e^{-\lambda(T_1+T_2)} \\ e^{-\lambda T_2} & e^{-\lambda T_2} \end{pmatrix}, \\
 Q_{1+1+1} &= \left( 3e^{-\lambda(T_1+T_2)} \right).
 \end{aligned} \tag{2.29}$$

According to Eq (2.4), the growth rate of each life cycle are given by the solutions of

$$2e^{-\lambda_{1+1}T_1} - 1 = 0, \tag{2.30}$$

$$e^{-\lambda_{2+1}(T_1+T_2)} + e^{-\lambda_{2+1}T_2} - 1 = 0, \tag{2.31}$$

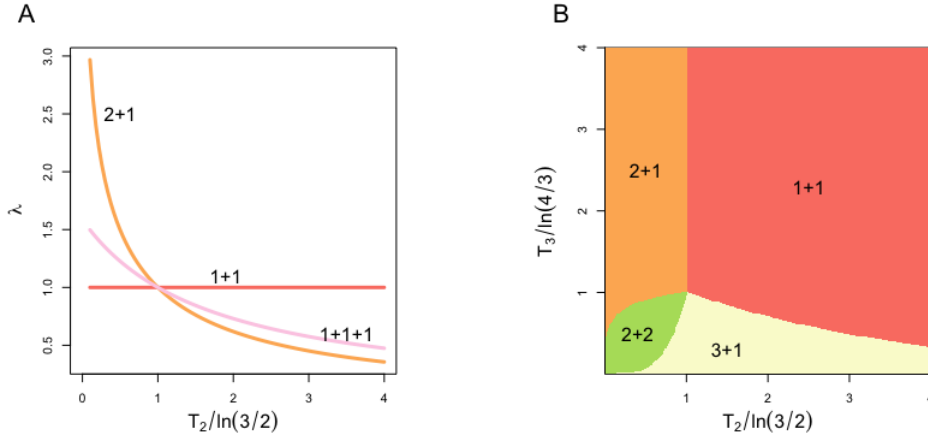


Figure 2.6: Growth rates and optimal life cycles in homogeneous groups on the condition of  $n \leq 3$  and  $n \leq 4$ , respectively. **A**) describes the growth rates of life cycles when  $n \leq 3$  i.e. 1+1, 1+1+1 and 2+1. **B**) shows the optimal life cycle when  $n \leq 4$  with respect to  $T_2$  and  $T_3$ . In both situations,  $T_i$  is the size increment time and we set  $T_1 = \ln(2)$  for convenience.

$$3e^{-\lambda_{1+1+1}(T_1+T_2)} - 1 = 0, \quad (2.32)$$

where  $\lambda_{1+1}$ ,  $\lambda_{2+1}$  and  $\lambda_{1+1+1}$  are the growth rate of 1+1, 2+1 and 1+1+1, respectively, see Fig 2.6A. For small  $T_2$ , the largest growth rate is achieved by 2+1 life cycle. In this case, bi-cellular groups produce offspring cells faster than independent cells. Consequently, the life cycle 2+1, which allows production of bi-cellular groups (unlike unicellular life cycle 1+1) and preserving one offspring group in the most productive bi-cellular state (unlike 1+1+1) is most successful in growth competition. In the opposite limit of large  $T_2$ , the life cycle 1+1 leads to the largest population growth rate. In this case, independent cells are better off than bi-cellular groups. Thus, the best reproductive strategy is to avoid the growth to bi-cellular state, which can only be achieved with a single life cycle 1+1. In both situations of  $T_2$ , the growth rate of 1+1+1 is always between that of 1+1 and 2+1.

For the next scenario, we increase the maximal size of the group to three cells. This allows four new life cycles: 3+1, 2+2, 2+1+1 and 1+1+1+1. Their growth rates are given

## Chapter 2. Interacting cells driving the evolution of multicellular life cycles

---

by the solutions of

$$e^{-\lambda_{3+1}(T_1+T_2+T_3)} + e^{-\lambda_{3+1}T_3} - 1 = 0 \quad (2.33)$$

$$2e^{-\lambda_{2+2}(T_2+T_3)} - 1 = 0 \quad (2.34)$$

$$2e^{-\lambda_{2+1+1}(T_1+T_2+T_3)} + e^{-\lambda_{2+1+1}(T_2+T_3)} - 1 = 0 \quad (2.35)$$

$$4e^{-\lambda_{1+1+1+1}(T_1+T_2+T_3)} - 1 = 0, \quad (2.36)$$

For large  $T_3$ , the life cycles which do not produce slow-growing three-cellular groups have the highest growth rates. Therefore, for large  $T_3$ , the optimal life cycles are the same as ones presented in the previous paragraph. For small  $T_3$ , the life cycles capable of producing three-cellular groups gain an evolutionary advantage. Specifically, 2+2 achieves the maximum growth rate when both  $T_2$  and  $T_3$  are comparatively small. In this case, an independent cell is the least productive state, whereas 2+2 is the only life cycle not producing independent cells. Life cycle 3+1 leads to the largest growth rate if  $T_3$  is small but  $T_2$  is large. There, the three-cellular group stands out as the most productive state, and 3+1 is the only life cycle keeping it as one of its offspring groups. Similarly to the previous scenario, life cycles with more than two offspring: 1+1+1, 2+1+1, 1+1+1+1, are never optimal. An important exception to this is the point  $T_1 = \ln(2)$ ,  $T_2 = \ln(\frac{3}{2})$  and  $T_3 = \ln(\frac{4}{3})$ , where all seven life cycles lead to the same growth rate ( $\lambda = 1$ ).

Previously, we considered another model of life cycles evolution [Pichugin et al., 2017]. There, the growth of groups from size  $i$  to size  $i + 1$  occurs spontaneously with rate  $ib_i$ . Therefore, in that model, the time between cell divisions varies between groups of the same size, in contrast to the scenario considered here, where this time is always equal to  $T_i$ . Despite the differences between two models, they both share a number of findings: existence of the neutral point, only binary fragmentation is evolutionarily optimal, same optimal life cycles in the limit cases. Therefore, these features, are independent from the model design.

**2.4.4 Calculation of growth rates  $\lambda$  for life cycles of heterogeneous groups**

To show how our approach can be used in the case of heterogeneous groups, consider the simplest unicellular life cycle 1+1. There are two types of offspring possible: independent  $A$  and  $B$  cells, so the matrix  $Q$  has dimensions 2 by 2. When a cell divides into two, three outcomes are possible: no cell, one cell, or both daughter cells change the phenotype. Since the developmental trajectory ends after the first division, there are only six developmental trajectories possible for this life cycle, see Fig 2.7.

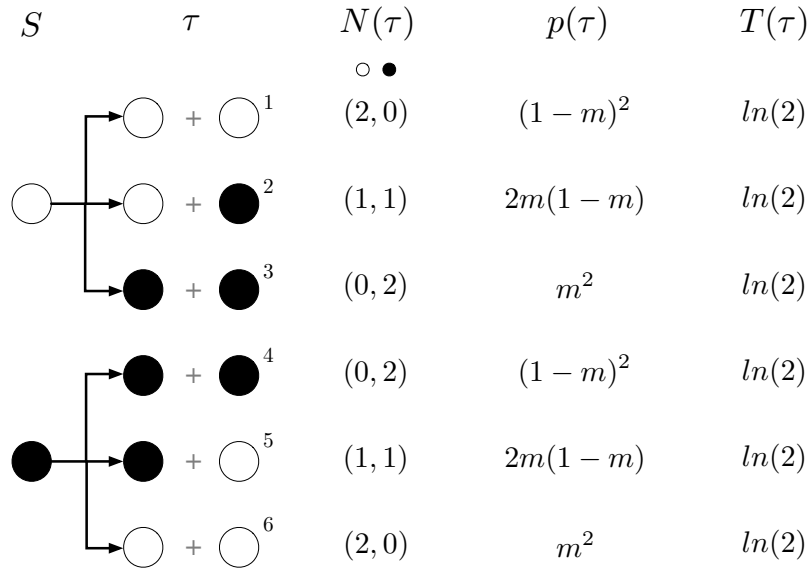


Figure 2.7: The full set of developmental trajectories in the life cycle 1+1. Here, the white and black circles denote  $A$  type cell and  $B$  type cell respectively.

To construct the matrix  $Q$ , we need to obtain the distribution of offspring ( $N_i$ ), the probability of realization ( $p$ ) and total developmental time ( $T$ ) for each trajectory. Offspring distributions are apparent from Fig 2.7. The probability of each trajectory can be directly computed from the phenotype switch probability  $m$ . The developmental time is  $T = T_1 = \ln(2)$  for each trajectory here. Therefore, the elements of  $Q$  are given by

$$Q_{11} = 2(1 - m)^2 e^{-\lambda \ln 2} \quad \leftarrow \tau_1$$

$$\begin{aligned}
 & + 2m(1-m)e^{-\lambda ln2} \quad \leftarrow \tau_2, \\
 Q_{12} = & 2m(1-m)^2 e^{-\lambda ln2} \quad \leftarrow \tau_2 \\
 & + 2me^{-\lambda ln2} \quad \leftarrow \tau_3, \\
 Q_{21} = & 2(1-m)e^{-\lambda ln2} \quad \leftarrow \tau_5 \\
 & + 2m^2 e^{-\lambda ln2} \quad \leftarrow \tau_6, \\
 Q_{22} = & 2(1-m)^2 e^{-\lambda ln2} \quad \leftarrow \tau_4 \\
 & + 2m(1-m)e^{-\lambda ln2} \quad \leftarrow \tau_5,
 \end{aligned}$$

where arrows indicate the index of the developmental trajectory contributing a given term. Solution of the Eq (18) leads to  $\lambda_{1+1} = 1$  in the life cycle 1+1.

Next, consider the life cycle 1+1+1. There are still two types of offspring possible: independent *A* and *B* cells, such that the matrix *Q* has dimensions 2 by 2. However, the life cycle requires two divisions to complete, so the number of possible developmental trajectories is increased to 20. Also, interactions play a role during the second division, so the probabilities *p* and developmental times *T* are more complicated, see Fig 2.8.

The elements of matrix *Q* are given by

$$\begin{aligned}
 Q_{11} = & 3n^4 e^{-\lambda(t_{[1,0]}+t_{[2,0]})} \quad \leftarrow \tau_1 \\
 & + 4mn^3 e^{-\lambda(t_{[1,0]}+t_{[2,0]})} \quad \leftarrow \tau_2 \\
 & + m^2 n^2 e^{-\lambda(t_{[1,0]}+t_{[2,0]})} \quad \leftarrow \tau_3 \\
 & + 6m^3 n P_{1+1}^B e^{-\lambda(t_{[1,0]}+t_{[1,1]})} \quad \leftarrow \tau_4 \\
 & + 4mn^2 (nP_{1+1}^A + 2mP_{1+1}^B) e^{-\lambda(t_{[1,0]}+t_{[1,1]})} \quad \leftarrow \tau_5 \\
 & + 2mn^2 (2mP_{1+1}^A + nP_{1+1}^B) e^{-\lambda(t_{[1,0]}+t_{[1,1]})} \quad \leftarrow \tau_6 \\
 & + 2m^4 e^{-\lambda(t_{[1,0]}+t_{[0,2]})} \quad \leftarrow \tau_8 \\
 & + 2m^3 n e^{-\lambda(t_{[1,0]}+t_{[0,2]})} \quad \leftarrow \tau_9 \\
 Q_{12} = & 2mn^3 e^{-\lambda(t_{[1,0]}+t_{[2,0]})} \quad \leftarrow \tau_2
 \end{aligned}$$

## 2.4. Supporting information

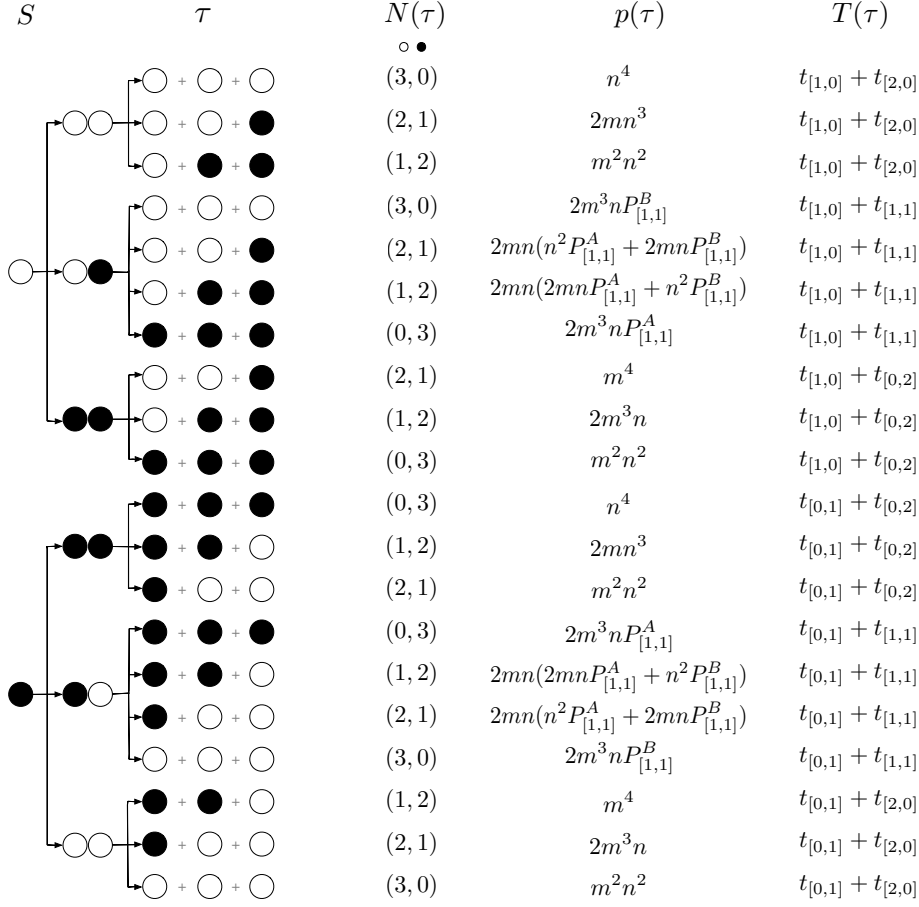


Figure 2.8: The full set of developmental trajectories in the life cycle 1+1+1. White circles represent  $A$  type cells and black circles represent  $B$  type cells. For simplicity of notation, we use  $n = 1 - m$ ,  $t_{[i,j]}$  is the time before the next cell division for a complex with  $i$   $A$  type cells and  $j$   $B$  type cells to divide; and  $P_{[1,1]}^A = \frac{1}{2} + w\frac{b-c}{4}$  and  $P_{[1,1]}^B = \frac{1}{2} - w\frac{b-c}{4}$ , see details in the model section of the main text.

$$\begin{aligned}
 &+ 2m^2n^2e^{-\lambda(t_{[1,0]}+t_{[2,0]})} \quad \leftarrow \tau_3 \\
 &+ 2mn^2(nP_{1+1}^A + 2mP_{1+1}^B)e^{-\lambda(t_{[1,0]}+t_{[1,1]})} \quad \leftarrow \tau_5 \\
 &+ 4mn^2(2mP_{1+1}^A + nP_{1+1}^B)e^{-\lambda(t_{[1,0]}+t_{[1,1]})} \quad \leftarrow \tau_6 \\
 &+ 6m^3nP_{1+1}^Ae^{-\lambda(t_{[1,0]}+t_{[1,1]})} \quad \leftarrow \tau_7
 \end{aligned}$$

## Chapter 2. Interacting cells driving the evolution of multicellular life cycles

---

$$\begin{aligned}
& + m^4 e^{-\lambda(t_{[1,0]}+t_{[0,2]})} && \longleftarrow \tau_8 \\
& + 4m^3 n e^{-\lambda(t_{[1,0]}+t_{[0,2]})} && \longleftarrow \tau_9 \\
& + 3m^2 n^2 e^{-\lambda(t_{[1,0]}+t_{[0,2]})} && \longleftarrow \tau_{10} \\
Q_{21} = & 2mn^3 e^{-\lambda(t_{[0,1]}+t_{[0,2]})} && \longleftarrow \tau_{12} \\
& + 2m^2 n^2 e^{-\lambda(t_{[0,1]}+t_{[0,2]})} && \longleftarrow \tau_{13} \\
& + 2mn^2 (2mP_{1+1}^A + nP_{1+1}^B) e^{-\lambda(t_{[0,1]}+t_{[1,1]})} && \longleftarrow \tau_{15} \\
& + 4mn^2 (nP_{1+1}^A + 2mP_{1+1}^B) e^{-\lambda(t_{[0,1]}+t_{[1,1]})} && \longleftarrow \tau_{16} \\
& + 6m^3 nP_{1+1}^B e^{-\lambda(t_{[0,1]}+t_{[1,1]})} && \longleftarrow \tau_{17} \\
& + m^4 e^{-\lambda(t_{[0,1]}+t_{[2,0]})} && \longleftarrow \tau_{18} \\
& + 4m^3 n e^{-\lambda(t_{[0,1]}+t_{[2,0]})} && \longleftarrow \tau_{19} \\
& + 3m^2 n^2 e^{-\lambda(t_{[0,1]}+t_{[2,0]})} && \longleftarrow \tau_{20} \\
Q_{22} = & 3n^4 e^{-\lambda(t_{[0,1]}+t_{[0,2]})} && \longleftarrow \tau_{11} \\
& + 4mn^3 e^{-\lambda(t_{[0,1]}+t_{[0,2]})} && \longleftarrow \tau_{12} \\
& + m^2 n^2 e^{-\lambda(t_{[0,1]}+t_{[0,2]})} && \longleftarrow \tau_{13} \\
& + 6m^3 nP_{1+1}^A e^{-\lambda(t_{[0,1]}+t_{[1,1]})} && \longleftarrow \tau_{14} \\
& + 4mn^2 (2mP_{1+1}^A + nP_{1+1}^B) e^{-\lambda(t_{[0,1]}+t_{[1,1]})} && \longleftarrow \tau_{15} \\
& + 2mn^2 (nP_{1+1}^A + 2mP_{1+1}^B) e^{-\lambda(t_{[0,1]}+t_{[1,1]})} && \longleftarrow \tau_{16} \\
& + 2m^4 e^{-\lambda(t_{[0,1]}+t_{[2,0]})} && \longleftarrow \tau_{18} \\
& + 2m^3 n e^{-\lambda(t_{[0,1]}+t_{[2,0]})} && \longleftarrow \tau_{19}
\end{aligned}$$

Here,  $P_{1+1}^A = \frac{1}{2} + w\frac{b-c}{4}$ ,  $P_{1+1}^B = \frac{1}{2} - w\frac{b-c}{4}$ ,  $t_{[1,0]} = t_{[1,0]} = \ln 2$ ,  $t_{[2,0]} = \ln \frac{3}{2}(1-wa)$ ,  $t_{[1,1]} = \ln 2 + \ln \frac{3}{2}(1-w\frac{b+c}{2})$  and  $t_{[2,0]} = \ln \frac{3}{2}(1-wd)$ . Arrows indicate the contributions of each developmental trajectory to  $Q_{ij}$ .



## 2.4. Supporting information

$S$	$\tau$	$N(\tau)$	$p(\tau)$	$T(\tau)$
		$\circ \bullet \infty \circ \bullet \bullet$		

The solution of Eq (2.18) for life cycle 1+1+1 yields

$$\lambda_{1+1+1} = 1 + w \frac{\ln(3/2)}{\ln(9)} (a+d) \left( (1-2m+2m^2) + 2m(1-m) \frac{b+c}{a+d} \right). \quad (2.37)$$

Our final example is the life cycle 2+1, where groups grow to three cells and fragment into a bi-cellular group and an independent cell. Here, five offspring types are possible: independent cells could be either *A* or *B* type and the bi-cellular group could have composition *AA*, *AB*, or *BB*, see Figs. 2.9, 2.10. Therefore,  $Q$  is  $5 \times 5$  matrix. There are 48 developmental programs possible and we refrain from showing here how elements of matrix  $Q$  are constructed in this case. The solution of the Eq (2.18) for life cycle 2+1 yields

$$\lambda_{2+1} = 1 + w \frac{3 \ln(3/2)}{2(5+8m) \ln(27/4)} (a+d) \left( (5-6m+10m^2) + 2m(7-5m) \frac{b+c}{a+d} \right). \quad (2.38)$$

For the life cycles 1+1+1+1 and 2+2, we just list the growth rate  $\lambda$

$$\lambda_{1+1+1+1} = 1 + w \frac{1}{12 \ln(2)} (a+d) \left( 3 \ln(2) + \left( 1 - \frac{b+c}{a+d} \right) \right. \\ \left. (-2m(5 \ln(2) - \ln(3)) + 2m^2(11 \ln(2) - 4 \ln(3)) - 8m^3(2 \ln(2) - \ln(3))) \right), \quad (2.39)$$

$$\lambda_{2+2} = 1 + w \frac{1}{4(8m^2 - 9m - 2) \ln(2)} (a+d) \left( -4 \ln(2) + m(\ln(3) - \ln(2)) - 4m^2 \ln(3) \right. \\ \left. + 4m^3 \ln(3) + \frac{b+c}{a+d} (-m(17 \ln(2) + \ln(3)) + 4m^2(4 \ln(2) - \ln(3)) - 4m^3 \ln(3)) \right). \quad (2.40)$$

For more complex life cycles, the analytical expressions are too large to be meaningful by naked eye analysis. Therefore, we used a combination of analytical and numerical approaches. After the linearisation with respect to  $w$ , the growth rate for any life cycle has the form

$$\lambda_{x_1+\dots+x_n} = 1 + w \frac{1}{P_1(m)} (a+d) \left( P_2(m) + \frac{b+c}{a+d} P_3(m) \right), \quad (2.41)$$

where  $P_1(m), P_2(m), P_3(m)$  are some polynomials of  $m$  of the finite power. We obtained ex-

## 2.4. Supporting information

$S$	$\tau$	$N(\tau)$ ○ ● ∞ ∞ ∞ ●●	$p(\tau)$	$T(\tau)$
	○○ + ○ <sup>33</sup>	(1, 0, 1, 0, 0)	$n^2$	$t_{[2,0]}$
	○○ + ● <sup>34</sup>	(0, 1, 1, 0, 0)	$2/3 mn$	$t_{[2,0]}$
	○● + ○ <sup>35</sup>	(1, 0, 0, 1, 0)	$4/3 mn$	$t_{[2,0]}$
	●● + ○ <sup>36</sup>	(1, 0, 0, 0, 1)	$1/3 m^2$	$t_{[2,0]}$
	○● + ● <sup>37</sup>	(0, 1, 0, 1, 0)	$2/3 m^2$	$t_{[2,0]}$
	○○ + ● <sup>38</sup>	(0, 1, 1, 0, 0)	$1/3 n^2 P_{[1,1]}^A + 2/3 mn P_{[1,1]}^B$	$t_{[1,1]}$
	○● + ○ <sup>39</sup>	(1, 0, 0, 1, 0)	$2/3 n^2 P_{[1,1]}^A + 4/3 mn P_{[1,1]}^B$	$t_{[1,1]}$
	●● + ○ <sup>40</sup>	(1, 0, 0, 0, 1)	$2/3 mn P_{[1,1]}^A + 1/3 n^2 P_{[1,1]}^B$	$t_{[1,1]}$
	○● + ● <sup>41</sup>	(0, 1, 0, 1, 0)	$4/3 mn P_{[1,1]}^A + 2/3 n^2 P_{[1,1]}^B$	$t_{[1,1]}$
	●● + ● <sup>42</sup>	(0, 1, 0, 0, 1)	$m^2 P_{[1,1]}^A$	$t_{[1,1]}$
	○○ + ○ <sup>43</sup>	(1, 0, 1, 0, 0)	$m^2 P_{[1,1]}^B$	$t_{[1,1]}$
	●● + ● <sup>44</sup>	(0, 1, 0, 0, 1)	$n^2$	$t_{[0,2]}$
	●● + ○ <sup>45</sup>	(1, 0, 0, 0, 1)	$2/3 mn$	$t_{[0,2]}$
	●● + ● <sup>46</sup>	(0, 1, 0, 1, 0)	$4/3 mn$	$t_{[0,2]}$
	○○ + ● <sup>47</sup>	(0, 1, 1, 0, 0)	$1/3 m^2$	$t_{[0,2]}$
	○● + ○ <sup>48</sup>	(1, 0, 0, 1, 0)	$2/3 m^2$	$t_{[0,2]}$

Figure 2.10: The developmental programs of bicellular newborn groups cells in the life cycle 2+1. White circles represent  $A$  type cells and black circles represent  $B$  type cells. For simplicity of notation, we use:  $n = 1 - m$ ;  $t_{[i,j]}$  is the time before the next cell division for a complex with  $i$   $A$  type cells and  $j$   $B$  type cells to divide; and  $P_{[1,1]}^A = \frac{1}{2} + w \frac{b-c}{4}$  and  $P_{[1,1]}^B = \frac{1}{2} - w \frac{b-c}{4}$ , see details in the model section of the main text.

act expressions for these polynomials using the symbolic algebra software. However, the tracking of all developmental programs means that the computation load grows exponentially with the maximal group size  $M$ . For life cycles, such as 3+2+1+1, computation of the polynomials required an extraordinary amount of RAM ( $> 70$  Gb) and the outcome is neither human-tractable nor even printable. This memory constraints is the factor, limiting the maximal group size considered to  $M = 7$ . Therefore, in our study, we only stored the numerical values of the coefficients of  $m$  in  $P_1(m), P_2(m), P_3(m)$  and used them to compute  $\lambda$ . With this approach, we are able to compute numerical values of  $\lambda$  with very high

accuracy, even if traditional closed form solutions are unavailable.

### 2.4.5 Profiles of growth rates of the life cycles

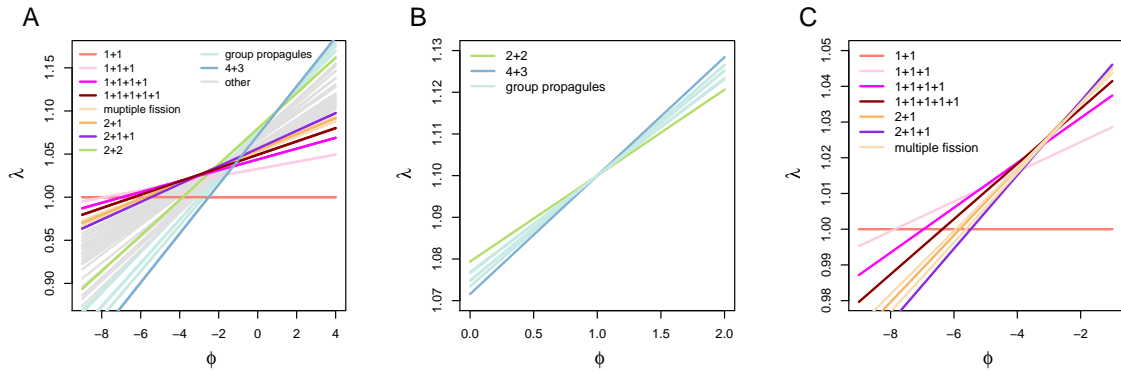


Figure 2.11: The growth rates of the considered life cycles as a function of  $\phi$  for  $\psi > 0$ .

Panel A: according to the weak selection approximation, growth rates  $\lambda$  are linear functions of  $\phi$ . For all life cycles, the slope of the line is non-negative, thus, life cycles with smaller slope dominate at  $\phi \ll -1$  (1+1 has slope zero) and life cycles with larger slope dominate at  $\phi \gg 1$  (4+3 has the largest slope for  $M \leq 7$ ). Panel B: all life cycles with multicellular offspring share the same growth rate at  $\phi = 1$  ( $\phi = -1$  under  $\psi < 0$ ). Panel C: a sequence of multiple fission life cycles is optimal at the negative  $\phi$ . At all panels  $m = 0.06$ . In all panels, multiple fission includes 1+1+1+1+1+1, 1+1+1+1+1+1+1; group fission includes 3+2, 3+3, 2+2+2, 4+2, 5+2.

In this appendix, we present profiles of growth rates at different conditions. The growth rate is determined by three parameters:  $\psi$ ,  $\phi$  and  $m$ . The greatest diversity of evolutionarily optimal life cycles is observed at  $\psi > 0$  and small  $m$ , see Fig 2.11. In this case, we observed two clusters of life cycles, where life cycles behave quite similar. One cluster contain the multiple fission life cycles such as 1+1, 1+1+1. The second cluster is the group propagules life cycles such as 3+2, 4+3. The slope of multiple fission life cycles are increasing with colony size, see Fig 2.11C. More similar growth rate patterns are ob-

## 2.4. Supporting information

served for the group propagules life cycles, which have identical growth rates at  $\phi = 1$ , see Fig 2.11B. Most other life cycles are between the area of multiple fission life cycles and group propagules life cycles, which can never be optimal. For  $\psi > 0$  and large  $m$ , only one multiple fission life cycle,  $1+1+1$ , is evolutionary optimal, see Fig 2.12A. Its area of optimality is located between unicellularity ( $1+1$ ) at large negative  $\phi$  and binary fragmentation with multicellular propagules ( $2+2$  and  $4+3$ ) at large positive  $\phi$ . Considering the dependence of growth rate from the phenotype switching probability  $m$ , we found that at  $\phi \gg 1$ , growth rate profiles are concave functions of  $m$ , see Fig 2.12B. Growth rates of most life cycles are generally bound between binary fragmentation with multicellular propagules (such as  $2+2$  and  $4+3$ ) and multiple fragmentation with unicellular propagules (such as  $1+1+1$  and  $1+1+1+1$ ). For  $\phi \ll -1$ , the pattern is very similar, with an exception, that growth rate profiles are convex, instead of concave, and the hierarchy of life cycles is reversed, see Fig 2.12C. This leads to the great diversity of evolutionary optimal life cycles at  $\phi < 0$  and small  $m$  (including also transitional life cycles  $2+1$  and  $2+1+1$ , as well as binary fragmentation  $2+2$ ), see Fig 2.12D.

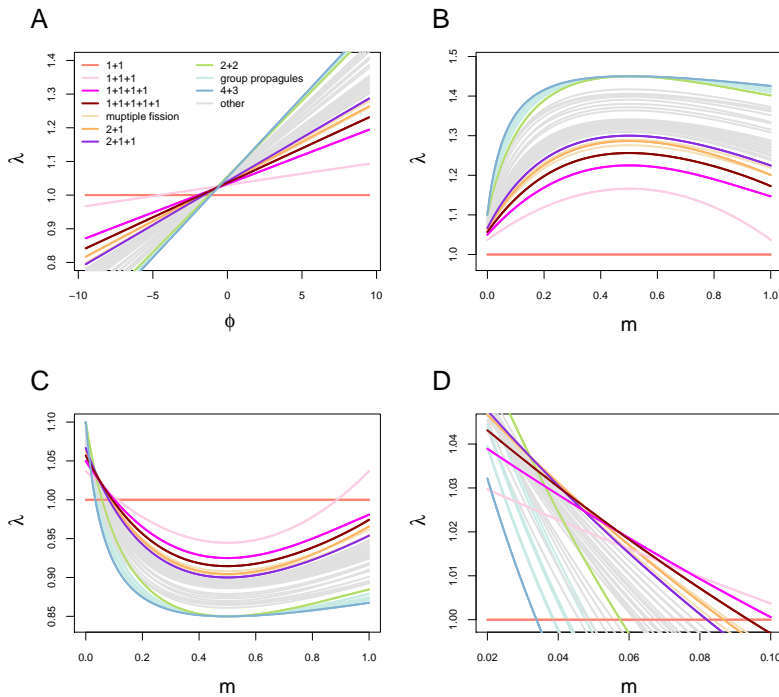


Figure 2.12: Multiple life cycles are optimal for  $\psi > 0$ . **A** Growth rates of all considered life cycles as function of  $\phi$  at  $m = 0.9$  (cf. Fig 4A for  $m = 0.06$ ). **B** Growth rates of all considered life cycles as function of  $m$  at  $\phi = 8$ . **C** Growth rates of all considered life cycles as function of  $m$  at  $\phi = -4$ . **D** Detailed view of the panel C in the range of small  $m$  showing that large number of evolutionary optimal life cycles at different  $m$ .

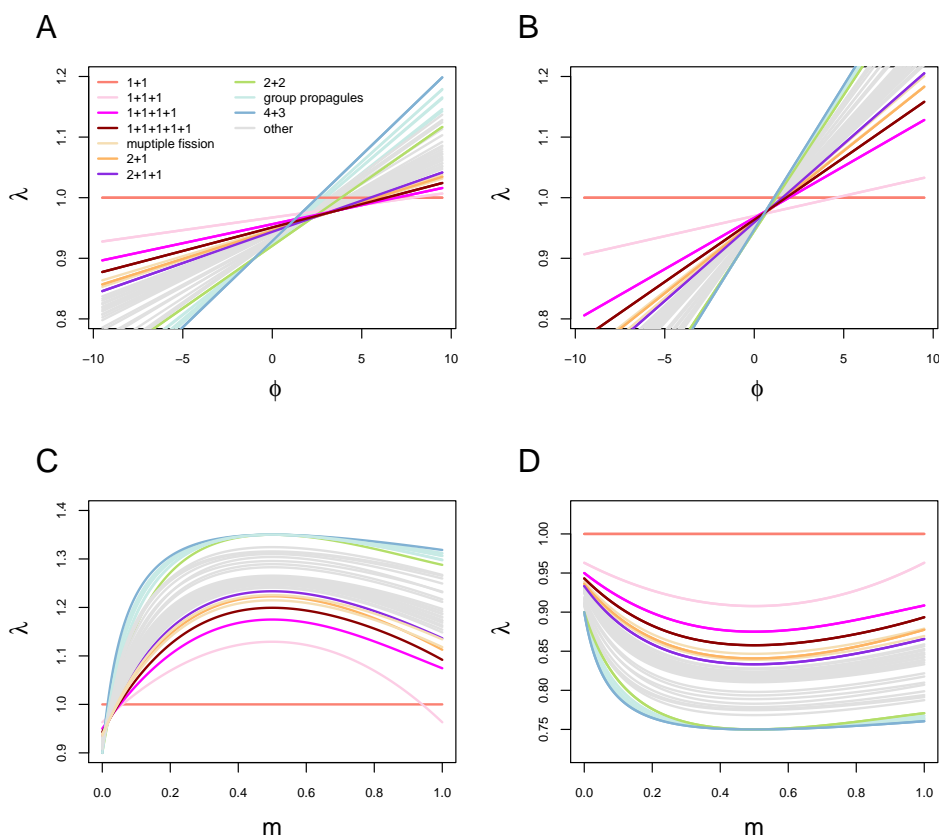


Figure 2.13: Only two life cycles are optimal for  $\psi < 0$ . **A** Growth rates of all considered life cycles as function of  $\phi$  at  $m = 0.06$ . **B** Growth rates of all considered life cycles as function of  $\phi$  at  $m = 0.9$ . **C** Growth rates of all considered life cycles as function of  $m$  at  $\phi = 8$ . **D** Growth rates of all considered life cycles as function of  $m$  at  $\phi = -4$ .

## 2.4. Supporting information

At the negative  $\psi$ , only two life cycles were found to be optimal, see Fig 2.13. The shape of individual growth rate profiles remain similar to the case of positive  $\phi$  but the relative position changes significantly. Thus, the spectrum of observed life cycles is much less diverse.

### 2.4.6 Optimal life cycles landscape under the self-interaction game

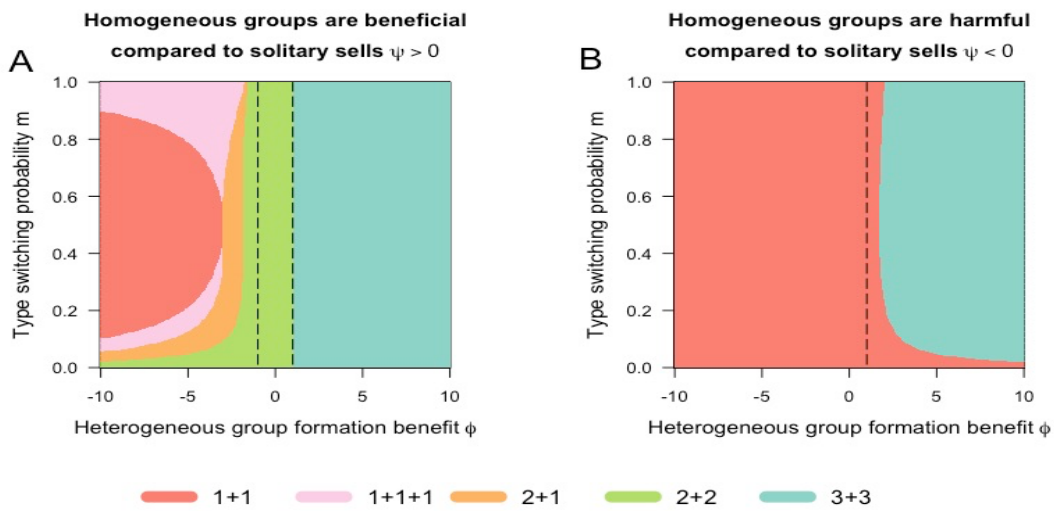


Figure 2.14: Similar optimal life cycles under the self-interaction game compared with the no self-interaction game. While we have only 5 optimal life cycles in this case, in general the results are very similar, with a large number of life cycles emerging only for  $\psi > 0$  and  $\phi < -1$ . **A** Optimal life cycles for  $\psi > 0$  under the self-interaction game. Dashed lines are  $\phi = -1$  and  $\phi = 1$ , respectively. **B** Optimal life cycles for  $\psi < 0$  under the self-interaction game. Dashed line is  $\phi = 1$ .

In our model setting, we set the payoff of single cells to zero based on the assumption that no other cells can impact their strategies. While, theoretically single cells can also

## Chapter 2. Interacting cells driving the evolution of multicellular life cycles

---

play self-interaction games to get payoff based on their cell types. Intuitively, the self-interaction game would produce the same results as the non self-interaction game, as in which only the final synergistic or antagonistic effects can really impact the outcome. To check this idea, we set the payoff of cells in a cluster to

$$\begin{aligned}\alpha_{[i,j]} &= \frac{ia + jb}{i + j}, \\ \beta_{[i,j]} &= \frac{ic + jd}{i + j},\end{aligned}\tag{2.42}$$

where  $\alpha_{[i,j]}$  and  $\beta_{[i,j]}$  are the average payoff of  $A$  type cells and  $B$  type cells in a group of  $i$   $A$ -cells and  $j$   $B$ -cells, respectively. This payoff definition allows the single cells also have non zero payoff values i.e. payoff  $a$  for  $A$  cell type and  $d$  for  $B$  cell type. Meanwhile, all other settings in the model are unchanged. Then, we investigate the optimal life cycles for population with colony size  $M$  less than seven. The results are pretty similar between the non self-interaction game (see Fig 4) and the self-interaction game (see Fig 2.14).



## Chapter 3

# Evolution of irreversible somatic differentiation

---

The work in this chapter has been submitted (March 2021) and is currently under review: Yuanxiao Gao, Hye Jin Park, Arne Traulsen and Yuriy Pichugin, “Evolution of irreversible somatic differentiation”. An updated version is available at biorxiv <https://doi.org/10.1101/2021.01.18.427219>.

### Abstract

A key innovation emerging in complex animals is irreversible somatic differentiation: daughters of a vegetative cell perform a vegetative function as well, thus, forming a somatic lineage that can no longer be directly involved in reproduction. Primitive species use a different strategy: vegetative and reproductive tasks are separated in time rather than in space. Starting from such a strategy, how is it possible to evolve life forms which use some of their cells exclusively for vegetative functions? Here, we developed an evolutionary model of development of a simple multicellular organism and found that three components are necessary for the evolution of irreversible somatic differentiation: (i) costly cell differentiation, (ii) vegetative cells that significantly improve the organism’s performance even if present in small numbers, and (iii) large enough organism size. Our findings demonstrate how an egalitarian development typical for loose cell colonies can evolve into germ-soma differentiation dominating metazoans.

### 3.1 Introduction

In complex multicellular organisms, different cells specialise to execute different functions. These functions can be generally classified into two kinds: reproductive and vegetative. Cells performing reproductive functions contribute to the next generation of organisms, while cells performing vegetative function contribute to sustaining the organism itself. In unicellular species and simple multicellular colonies, these two kinds of functions are performed at different times by the same cells – specialization is temporal. In more complex multicellular organisms, specialization transforms from temporal to spatial [Mikhailov et al., 2009], where groups of cells focused on different tasks emerge in the course of organism development.

Typically, cell functions are changed via differentiation, such that a daughter cell performs a different function than the maternal cell. The vast majority of metazoans feature a very specific and extreme pattern of cell differentiation: any cell performing vegetative functions forms a somatic line, i.e. producing cells performing the same vegetative function – somatic differentiation is irreversible. Since such somatic cells cannot give rise to reproductive cells, somatic cells do not have a chance to pass their offspring to the next generation of organisms. Such a mode of organism development opened a way for deeper specialization of somatic cells and consequently to the astonishing complexity of multicellular metazoans. In *Volvocales* – a group of green algae, serving as a model species for evolution of multicellularity, the emergence of irreversibly differentiated somatic cells is the hallmark innovation marking the transition from colonial life forms to multicellular species [Kirk, 2005].

While the production of individual cells specialized in vegetative functions comes with a number of benefits [Grosberg and Strathmann, 2007], the development of a dedicated vegetative cell lineage that is lost for organism reproduction is not obviously a beneficial adaptation. From the perspective of a cell in an organism, the guaranteed termination of its lineage seems the worst possible evolutionary outcome itself. From the

perspective of entire organism, the death of somatic cell at the end of the life cycle is a waste of resources, as these cells could in principle become parts of the next generation of organisms. For example, exceptions from irreversible somatic differentiation are widespread in plants [Lanfear, 2018] and are even known in simpler metazoans among cnidarians [DuBuc et al., 2020] for which differentiation from vegetative to reproductive functions has been reported. Therefore, the irreversibility of somatic differentiation cannot be taken for granted in the course of the evolution of complex multicellularity.

The majority of the theoretical models addressing the evolution of somatic cells focuses on the evolution of cell specialization, overlooking the developmental process how germ (reproductive specialists) and soma are produced in the course of the organism growth. For example, a huge amount of work focuses on the optimal distribution of reproductive and vegetative functions in the adult organism [Michod, 2007, Willensdorfer, 2009, Rossetti et al., 2010, Rueffler et al., 2012, Ispolatov et al., 2012, Goldsby et al., 2012, Solari et al., 2013, Goldsby et al., 2014, Amado et al., 2018, Tverskoi et al., 2018]. However, these models do not consider the process of organism development. Other work takes the development of an organism into account to some extent. In [Gavrilets, 2010], the organism development is considered, but the fraction of cells capable to become somatic is fixed and does not evolve. In [Erten and Kokko, 2020], the strategy of germ-to-soma differentiation is an evolvable trait, but the irreversibility of somatic differentiation is taken for granted. In [Rodrigues et al., 2012], irreversible differentiation was found, but both considered cell types pass to the next generation of organisms, such that the irreversible specialists are not truly somatic cells in the sense of evolutionary dead ends. Finally, in [Cooper and West, 2018] all model ingredients are present: the strategy of cell differentiation is explicitly considered and it is an evolvable trait, also soma and germ cells are considered. However, irreversible somatic differentiation was not observed in that study. Hence, the theoretical understanding of the evolution of irreversibly differentiated somatic cell lines is limited so far.

We developed a theoretical model to investigate conditions for evolution of irrevers-

ible somatic differentiation, in which vegetative soma-role cells are, in principle, capable to re-differentiate and produce reproductive germ-role cells. In our model, we incorporate factors including (i) costs of cell differentiation, (ii) benefits provided by presence of soma-role cells, (iii) maturity size of the organism. We ask under which circumstances irreversible somatic differentiation is a strategy that can maximize the population growth rate compared to strategies in which differentiation does not occur or somatic differentiation is reversible.

### 3.2 Model

We consider a large population of clonally developing organisms composed of two types of cells: germ-role and soma-role. Each organism is initiated as a single germ-role cell. In the course of the organism growth, germ-role cells may differentiate to give rise to soma-role cells and vice versa, see Fig. 3.1A,B. We assume that somatic cells accelerate growth: an organism containing more somatic cells grows faster. After  $n$  rounds of synchronous cell divisions, the organism reaches its maturity size of  $2^n$  cells. Immediately upon reaching maturity, the organism reproduces: germ-role cells disperse and each becomes a newborn organism, while all soma-role cells die and are thus lost, see Fig. 3.1A.

To investigate the evolution of irreversible somatic differentiation, we consider organisms in which the functional role of the cell (germ-role or soma-role) is not necessarily inherited. When a cell divides, the two daughter cells can change their role, leading to three possible combinations: two germ-role cells, one germ-role cell plus one soma-role cell, or two soma-role cells. We allow all these outcomes to occur with different probabilities, which also depend on the parental type, see Fig 3.1B. If the parental cell had the germ-role, the probabilities of each outcome are denoted by  $g_{gg}$ ,  $g_{gs}$ , and  $g_{ss}$  respectively. If the parental cell had the soma-role, these probabilities are  $s_{gg}$ ,  $s_{gs}$ , and  $s_{ss}$ . Altogether, six probabilities define a stochastic developmental strategy  $D = (g_{gg}, g_{gs}, g_{ss}; s_{gg}, s_{gs}, s_{ss})$ . In our model, it is the stochastic developmental strategy that is inherited by offspring cells

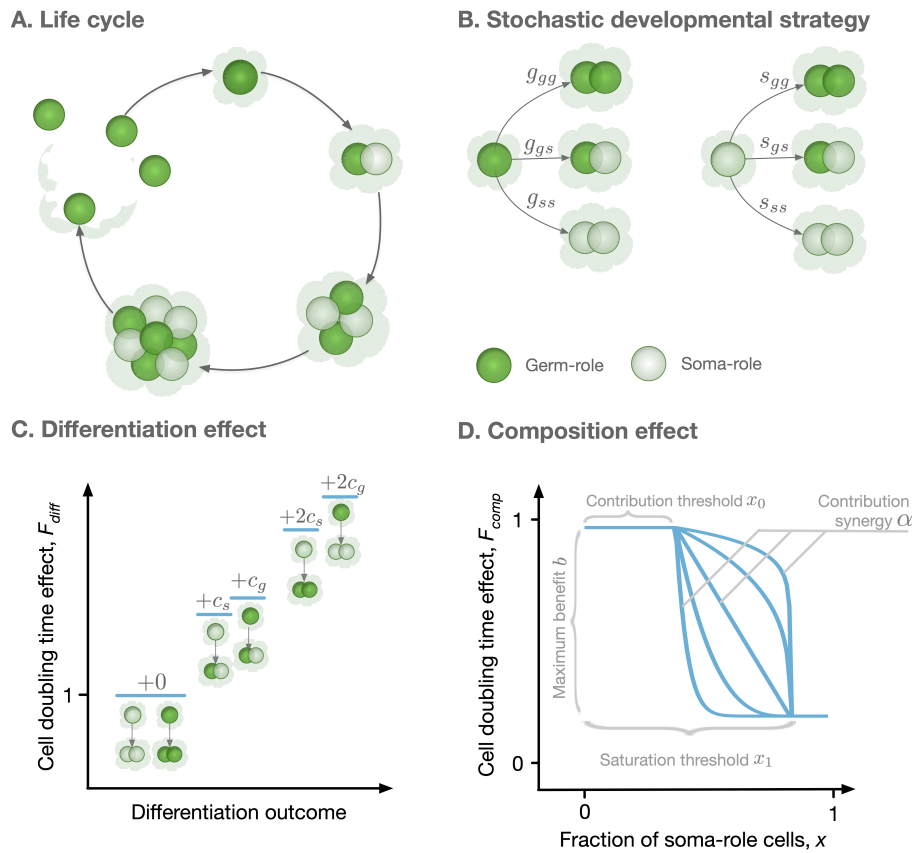


Figure 3.1: **Model overview.** **A.** The life cycle of an organism starts with a single germ-role cell. In each round, all cells divide and daughter cells can differentiate into a role different from the maternal cell's role. When the organism reaches maturity, it reproduces: each germ-role cell becomes a newborn organism and each soma-role cell dies. **B.** Change of cell roles is controlled by a stochastic developmental strategy defined by probabilities of each possible outcomes of a cell division. **C.** Differentiation of cells requires an investment of resources and, thus, slows down the organism growth. Each cell differentiation event incurs a cost ( $c_s$  or  $c_g$ ). The average cost of differentiation contributes increases the cell doubling time in a multiplicative way. **D.** The growth contribution of somatic cells is controlled by a function that decreases the doubling time with the fraction of somatic cells. The form of this function is controlled by four parameters,  $x_0$ ,  $x_1$ ,  $\alpha$ , and  $b$ .

### Chapter 3. Evolution of irreversible somatic differentiation

---

rather than the functional role of the parental cell.

To feature irreversible somatic differentiation (ISD in the following), the developmental strategy must allow germ-role cells to give rise to soma-role cells ( $g_{gg} < 1$ ) and must forbid soma-role cells to give rise to germ-role cells ( $s_{ss} = 1$ ). All other developmental strategies can be broadly classified into two classes. Reversible somatic differentiation (RSD) describes strategies where cells of both roles can give rise to each other:  $g_{gg} < 1$  and  $s_{ss} < 1$ . In the strategy with no somatic differentiation (NSD), soma-role cells are not produced in the first place:  $g_{gg} = 1$ , see Table 3.1.

Table 3.1: Classification of developmental strategies

<b>Class</b>	<b>Label</b>	$g_{gg}$	$s_{ss}$
Irreversible somatic differentiation	ISD	$< 1$	$= 1$
Reversible somatic differentiation	RSD	$< 1$	$< 1$
No somatic differentiation	NSD	$= 1$	irrelevant

In our model, evolution is driven by the growth competition between populations executing different developmental strategies. Growth competition will favour developmental strategies that lead to faster growth [Pichugin et al., 2017, Gao et al., 2019]. The rate of population growth is determined by the number of offspring produced by an organism (equal to the number of germ-role cells at the end of life cycle) and the time needed for an organism to develop from a single cell to maturity (improved with the number of soma-role cells during the life cycle). The development consists of  $n$  rounds of simultaneous cell divisions. Consequently, the total development time is a sum of  $n$  time intervals between cell doubling events. Each cell doubling time ( $t$ ) is determined by two independent effects: the differentiation effect ( $F_{\text{diff}}$ ) representing costs of changing cell roles [Gallon, 1992] and the organism composition effect ( $F_{\text{comp}}$ ) representing benefits from having soma-role cells [Grosberg and Strathmann, 1998, 2007, Shelton et al., 2012,

Matt and Umen, 2016],

$$t = F_{\text{diff}} \times F_{\text{comp}}. \quad (3.1)$$

The cell differentiation effect  $F_{\text{diff}}$  represents the costs of cell differentiation. The differentiation of a cell requires efforts to modify epigenetic marks in the genome, recalibration of regulatory networks, synthesis of additional and utilization of no longer necessary proteins. This requires an investment of resources and therefore an additional time to perform cell division. Hence, any cell, which is about to give rise to a cell of a different role, incurs a differentiation costs  $c_g$  for germ-to-soma and  $c_s$  for soma-to-germ transitions, see Fig. 3.1C. Here, we additionally assume that resource distribution among cells is coordinated at the level of the organism: Cells which need more resources will get more, such that cell division is synchronous. The resulting effect of differentiation costs is determined as  $F_{\text{diff}} = 1 + \langle c \rangle$ , where  $\langle c \rangle$  is the average differentiation cost among all cells in an organism.

The composition effect  $F_{\text{comp}}$  captures how the cell division time depends on the proportion of soma-role cells  $x$  present in an organism. In this study, we use a functional form illustrated in Fig. 3.1D and given by

$$F_{\text{comp}}(x) = \begin{cases} 1 & \text{for } 0 \leq x \leq x_0 \\ 1 - b + b \left( \frac{x_1 - x}{x_1 - x_0} \right)^\alpha & \text{for } x_0 < x < x_1 \\ 1 - b & \text{for } x_1 \leq x \leq 1 \end{cases} \quad (3.2)$$

With the functional form Eq. (3.2), soma-role cells can benefit to the organism growth, only if their proportion in the organism exceeds the contribution threshold  $x_0$ . Interactions between soma-role cells may lead to the synergistic (soma-role cells work better together than alone), or discounting benefits (soma-role cells work better alone than together) to the organism growth, controlled by the contribution synergy parameter  $\alpha$ . The maximal achievable reduction in the cell division time is given by the maximal benefit  $b$ , realized beyond the saturation threshold  $x_1$  of the soma-role cell proportion. A further in-

crease in the proportion of soma-role cells does not provide any additional benefits. With the right combination of parameters, Eq. (3.2) is able to recover various characters of soma-role cells contribution to the organism growth: linear ( $x_0 = 0, x_1 = 1, \alpha = 1$ ), power-law ( $x_0 = 0, x_1 = 1, \alpha \neq 1$ ), step-functions ( $x_0 = x_1$ ), and a huge range of other scenarios.

For a given combination of differentiation costs ( $c_g, c_s$ ) and a composition effect profile (determined by four parameters:  $x_0, x_1, b$ , and  $\alpha$ ), we screen through a number of stochastic developmental strategies  $D$  and identify the one providing the largest growth rate to the population. In this study, we searched for those parameters under which ISD strategies lead to the fastest growth and are thus evolutionary optimal, see model details in Appendix 3.5.1.

### 3.3 Results

#### 3.3.1 For irreversible somatic differentiation to evolve, cell differentiation must be costly.

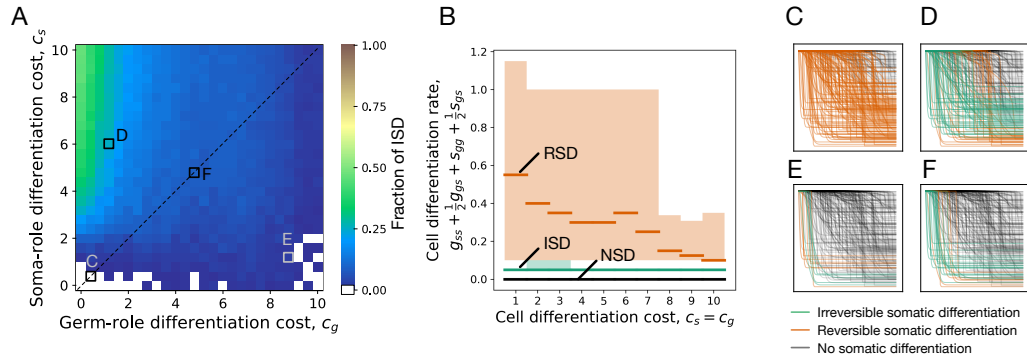
We found that irreversible somatic differentiation (ISD) does not evolve when cell differentiation is not associated with any costs ( $c_s = c_g = 0$ ), see Fig 3.2A. This finding comes from the fact that when somatic differentiation is irreversible, the fraction of germ-role cells can only decrease in the course of life cycle. As a result, ISD strategies deal with the tradeoff between producing more soma-role cells at the beginning of the life cycle, and having more germ-role cells by the end of it. On the one hand, ISD strategies which produce a lot of soma-role cells early on, complete the life cycle quickly but preserve only a few germ-role cells by the time of reproduction. On the other hand, ISD strategies which generate a lot of offspring, can deploy only a few soma-role cells at the beginning of it and thus their developmental time is inevitably longer. By contrast, reversible somatic differentiation strategies (RSD) do not experience a similar tradeoff, as germ-role cells can be generated from soma-role cells. As a result, RSD allows higher differentiation rates and



can develop a high soma-role cell fraction in the course of the organism growth and at the same time have a large number of germ-role cells by the moment of reproduction. Under costless cell differentiation, for any ISD strategy, we can find an RSD counterpart, which leads to faster growth: the development proceeds faster, while the expected number of produced offspring is the same, see Appendix 3.5.2 for details. As a result, costless cell differentiation cannot lead to irreversible somatic differentiation.

To confirm the reasoning that RSD strategies gain an edge over ISD by having larger differentiation rates, we asked which ISD and RSD strategies become optimal at various cell differentiation costs ( $c_s = c_g$ ). At each value of costs, we found evolutionarily optimal developmental strategy for 3000 different randomly sampled composition effect profiles  $F_{\text{comp}}(x)$ . We found that evolutionarily optimal RSD strategies feature much larger rates of cell differentiation than evolutionarily optimal ISD strategies, see Fig. 3.2B. Even at large costs, where frequent differentiation is heavily penalized, the distinction between differentiation rates of ISD and RSD strategies remains apparent.

We screened through a spectrum of germ-to-soma ( $c_g$ ) and soma-to-germ ( $c_s$ ) differentiation costs, see Fig 3.2A. Both differentiation costs punish RSD strategies severely due to their high differentiation rates. By contrast, strategies with irreversible somatic differentiation are insensitive to changes in soma-to-germ differentiation costs  $c_s$ , because soma-role cells never give rise to germ-role cells in ISD. Consequently, we observed that ISD is most likely to evolve, when the transition from germ-role to soma-role is cheap ( $c_g$  is small) and the reverse transition is expensive ( $c_s$  is large), see Fig 3.2A. In a similar manner, an increase in germ-to-soma differentiation costs ( $c_g$ ) punishes both RSD and ISD strategies. However, RSD strategies tend to have larger rates of germ-to-soma transitions. Thus, they are punished more than ISD, which leads to the evolution of ISD at small  $c_s$  and large  $c_g$ . Finally, the NSD strategy does not pay any costs at all, as no cell differentiation occurs. Hence, at very large germ-to-soma differentiation costs ( $c_g \approx 10$  at Fig. 3.2A), the NSD strategy outcompetes both reversible and irreversible somatic differentiation, see Appendix 3.5.3 for details. For simplicity, hereafter we focus on the case of



**Figure 3.2: Irreversible soma evolves when cell differentiation is costly.** **A.** The fraction of composition effect profiles, Eq. (3.2), promoting ISD as a function of the differentiation costs  $c_g$  and  $c_s$ . We randomly draw the parameters in Eq. (3.2) to construct 200 random profiles (see Appendix for details). The absence of costs ( $c_g = c_s = 0$ ) as well as large costs of germ differentiation (large  $c_g$ ) suppresses the evolution of ISD. Irreversible somatic differentiation is promoted the most when the cell differentiation cost is large for soma-role cells ( $c_s$ ) and small for germ-role cells ( $c_g$ ). The maturity size used in the calculation is  $2^{10}$  cells. Black dashed lines at panel B indicates the line of equal costs  $c_s = c_g$  and squares indicate the costs shown in panels C-F. **B.** Cumulative cell differentiation rate ( $g_{ss} + \frac{1}{2}g_{gs} + s_{gg} + \frac{1}{2}s_{gs}$ ) in developmental strategies evolutionarily optimal at various differentiation costs ( $c_s = c_g$ ), separated by class (ISD, RSD, or NSD). Thick lines represent median values within each class, shaded areas show 90% confidence intervals. For each cost value, 3000 random profiles are used. Evolutionary optimal RSD strategies (orange) have much higher rates of cell differentiation than ISD (green). Consequently, RSD is penalized more under costly differentiation. **C - F.** Shapes of composition effect profiles (compare Fig. 3.1D) promoting ISD (green lines), RSD (orange lines), and NSD (black lines) developmental strategies at four parameter sets indicated in panel A.

the equal differentiation costs  $c_s = c_g = c$  (a black dashed line on Fig 3.2A).

### 3.3.2 Evolution of irreversible somatic differentiation is promoted when even a small number of somatic cells provide benefits to the organism.

The composition effect profiles  $F_{\text{comp}}(x)$  that promote the evolution of irreversible somatic differentiation have certain characteristic shapes, see 3.2C-F. We investigated what

kind of composition effect profiles can make irreversible somatic differentiation become an evolutionary optimum. We sampled a number of random composition effect profiles with independently drawn parameter values and found optimal developmental strategies for each profile for a number of differentiation costs ( $c$ ) and maturity size ( $2^n$ ) values. We took a closer look at the instances of  $F_{\text{comp}}(x)$  which resulted in irreversible somatic differentiation being evolutionarily optimal.

We found that ISD is only able to evolve when the soma-role cells contribute to the organism cell doubling time even if present in small proportions, see Fig. 3.3A,B. Analysing parameters of the composition factors promoting ISD, we found that this effect manifests in two patterns. First, the contribution threshold value ( $x_0$ ) has to be small, see Fig 3.3D – ISD is promoted when soma-role cells begin to contribute to the organism growth even in low numbers. Second, the contribution synergy was found to be large ( $\alpha > 1$ ) or, alternatively, the saturation threshold ( $x_1$ ) was small, see Fig 3.3C.

Both the contribution threshold  $x_0$  and the contribution synergy  $\alpha$  control the shape of the composition effect profile at intermediary abundances of soma-role cells. If the contribution synergy  $\alpha$  exceeds 1, the profile is convex, so the contribution of soma-role cells quickly becomes close to maximum benefit ( $b$ ). A small saturation threshold ( $x_1$ ) means that the maximal benefit of soma is achieved already at low concentrations of soma-role cells (and then the shape of composition effect profile between two close thresholds has no significance). Together, these patterns give an evidence that the most crucial factor promoting irreversible somatic differentiation is the effectiveness of soma-role cells at small numbers, see Appendix 3.5.4 for more detailed data presentation.

The reason behind these patterns is a slower accumulation of soma-role cells under irreversible somatic differentiation, comparing to RSD strategies, see Appendix 3.5.2. Thus, with ISD, an organism spends a significant amount of time having only a few soma role cells. Hence, ISD strategy can only be evolutionarily successful, if these few soma-role cells have a notable contribution to the organism growth time.

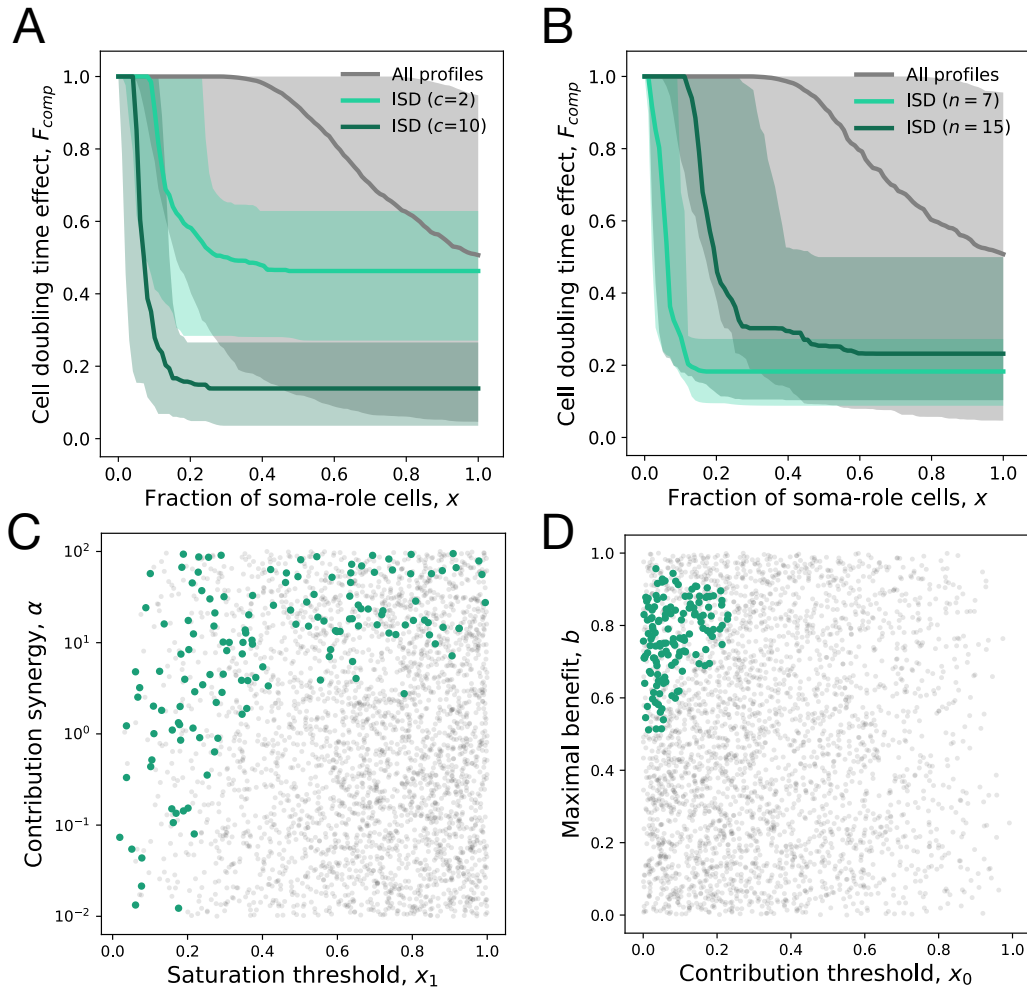
We also found that profiles featuring ISD do not possess neither extremely large, nor

extremely small maximal benefit values  $b$ , see Fig. 3.3D. When the maximal benefit is too small, the cell differentiation just does not provide enough benefits to be selected for and the evolutionarily optimal strategy is NSD. In the opposite case, when the maximal benefit is very close to one, the cell doubling time approaches zero, see Eq. (3.2). Then, the benefits of having many soma-role cells outweighs the costs of differentiation and the optimal strategy is RSD, see Appendix 3.5.4.

### 3.3.3 For irreversible somatic differentiation to evolve, the organism size must be large enough.

By screening through the maturity size ( $2^n$ ) and differentiation costs ( $c$ ), we found that the evolution of irreversible somatic differentiation is heavily suppressed at small maturity sizes, Fig 3.4A. For  $c_s = c_g$ , the minimal maturity size allowing irreversible somatic differentiation to evolve is  $2^n = 64$  cells. At the same time, organisms performing just a few more rounds of cell divisions are able to evolve ISD at a wide range of cell differentiation costs, see also Appendix 3.5.5. This indicates that the evolution of irreversible somatic differentiation is strongly tied to the size of the organism.

Evolution of ISD at sizes smaller than 64 cells is possible for  $c_s > c_g$ . For instance, at  $c_s = 2c_g$  some ISD strategies were found to be optimal at the maturity size  $2^5 = 32$  cells, Fig 3.4B. However, ISD strategies were found in a narrow range of cell differentiation costs and the fraction of composition effect profiles that allow evolution of ISD there was quite low – about 1%. The evolution of ISD at such small maturity sizes becomes likely only at extremely unequal costs of transition between germ and some roles  $c_s \gg c_g$ , see Fig 3.4C. Hence, for irreversible somatic differentiation to evolve, the organism size should exceed a threshold of roughly 64 cells.



**Figure 3.3: Irreversible soma evolves when substantial benefits arise at small concentrations of soma-role cells.** In all panels, the data representing the entire set of composition effect profiles is presented in grey, while the subset promoting ISD is coloured. **A**, **B**. Median and 90% confidence intervals of composition effect profiles at different differentiation costs (**A**, maturity size  $n = 10$ ) and maturity sizes (**B**, differentiation costs  $c = 5$ ). **C**, **D**. The set of composition effect profiles in the parameter space. Each point represents a single profile ( $c = 5$  and  $n = 10$ ). **C**. The co-distribution of the saturation threshold ( $x_1$ ) and the contribution synergy ( $\alpha$ ) reveals that either  $x_1$  must be small or  $\alpha$  must be large. **D**. Co-distribution of the contribution threshold ( $x_0$ ) and the maximal benefit ( $b$ ) shows that  $x_0$  must be small, while  $b$  must be large to promote ISD. 3000 profiles are used for panels A, C, D and 1000 profiles for panel B.

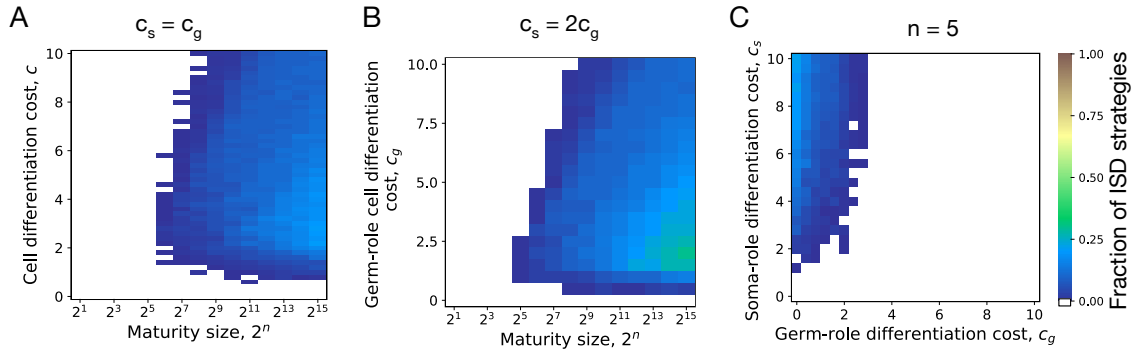


Figure 3.4: **Irreversible soma can evolve if organism grows to a large enough size in the course of its life cycle.** **A.** The fraction of composition effect profiles promoting ISD at various cell differentiation costs ( $c = c_s = c_g$ ) and maturity sizes ( $2^n$ ). ISD strategies were only found for maturity size  $2^6 = 64$  cells and larger. **B.** The fraction of composition effect profiles promoting ISD at unequal differentiation costs  $c_s = 2c_g$ . A rare occurrences of ISD ( $\sim 1\%$ ) was detected at the maturity size  $2^5 = 32$  cells in a narrow range of cell differentiation costs but not at the smaller sizes. **C.** The range of cell differentiation costs promoting ISD at at the maturity size  $2^5 = 32$  cells. For ISD strategies to evolve at such a small size, the differentiation from soma-role to germ-role must be much more costly than the opposite transition ( $c_s \gg c_g$ ).

### 3.4 Discussion

The vast majority of cells in a body of any multicellular being contains enough genetic information to build an entire new organism. However, in a typical metazoan species, very few cells actually participate in the organism reproduction – only a limited number of germ cells are capable to do it. The other cells, called somatic cells, perform vegetative functions but do not try to form an offspring organism – somatic differentiation is irreversible. We asked for the reason for the success of such a specific mode of an organism development. We theoretically investigated the evolution of irreversible somatic differentiation with a model of clonal developing organisms taking into account benefits provided by soma-role cells, costs coming from cell differentiation, and the effect of the

raw organism size.

One of the most significant assumptions we took is the synchronicity of cell divisions even if division outcomes are different. This is only possible if cell actions are coordinated at the level of organism – otherwise, cells that do not differentiate may complete their divisions before differentiating cells. When in the history of multicellularity such a coordination emerges is an open question. However, in a number of rather simple species, a synchronicity of cell divisions paired with cell differentiation is observed. One example is the green algae *Eudorina illinoensis* – one of the simplest species demonstrating the first signs of reproductive division of labour, in which four out of thirty-two cells are differentiated [Sambamurty, 2005]. Another example is 128-celled algae *Pleodorina californica*, half of the cells are differentiated. And still, the cell divisions are synchronous [Kikuchi, 1978]. Even the size of the mature organism being a power of two indicates that cells do not divide independently, but their actions are controlled at the level of the organism.

While our model can capture some key features of biological systems, it remains of course an abstraction. We assumed that populations go into an exponential growth phase – competition for space or nutrients could lead to selection of other strategies instead. Additional features such as trade-offs in growth at different colony sizes lead to further complications. Nevertheless, our model allows to start to look into the basic features of nascent life cycles at the edge of the division of labour in multicellular colonies.

Our key findings are:

- The evolution of irreversible somatic differentiation is inseparable from cell differentiation being costly.
- For irreversible somatic differentiation to evolve, somatic cells should be able to contribute to the organism performance already when their numbers are small.
- Only large enough organisms tend to develop irreversible somatic differentiation.

According to our results, cell differentiation costs are essential for the emergence of

### Chapter 3. Evolution of irreversible somatic differentiation

---

irreversible somatic differentiation, see Fig. 3.2A. For cells in a multicellular organism, differentiation costs arise from the material needs, energy, and time it takes to produce components necessary for the performance of the differentiated cell, which were absent in the parent cell. For instance, in filamentous cyanobacteria nitrogen-fixating heterocysts develop much thicker cell wall than parent photosynthetic cells had. Also, reports indicate between 23% [Ow et al., 2008] and 74% [Sandh et al., 2014] of the proteome changes its abundance in heterocysts compared against photosynthetic cells. Similarly, the changes in the protein composition in the course of cell differentiation was found during the development of stalk and fruiting bodies of *Dictyostelium discoideum* [Bakthavatsalam and Gomer, 2010, Czarna et al., 2010]

Our model demonstrates that irreversible somatic differentiation is more likely to evolve when a few soma-role cells are able to provide a substantial benefit to the organism, see Fig. 3.3. Several patterns of how the benefit provided by somatic cells changes with their numbers have been previously considered in the literature. However, the range of studied examples was restricted to concave or convex shapes [Michod, 2007, Willensdorfer, 2009, Rossetti et al., 2010, Cooper and West, 2018]. In this paper, we went beyond these shapes and additionally considered lower ( $x_0$ ) and upper ( $x_1$ ) thresholds for the somatic cells contribution (our model recover the previous approaches for  $x_0 = 0$  and  $x_1 = 1$ ). While our findings are in a qualitative agreement with past results – the profiles promoting irreversible somatic differentiation look convex-like, see Fig. 3.3A,B, our model indicates that the crucial component here is the large benefits provided by small numbers of soma-role cells, rather than overall convexity of the profile. For example, with sufficiently small  $x_1$ , the non-constant section of the composition effect profile (where the fraction of soma-role cells is between  $x_0$  and  $x_1$ , see Fig. 3.1D) can easily be concave ( $\alpha < 1$ , see Fig. 3.3C) and still promote irreversible somatic differentiation. *Volvocales* algae demonstrate that a significant contribution by small numbers of somatic cells might indeed be found in a natural population: In *E. illinoensis*, only four out of thirty two cells are vegetative [Sambamurty, 2005] (soma-role in our terms). This species has developed some reproductive division of labour and a fraction of only 1/8 of veget-



ative cells is sufficient for colony success. Thus, it seems possible that highly-efficient soma-role cells open the way to the evolution of irreversible somatic differentiation.

Our model shows that irreversible somatic differentiation can only emerge in relatively large organisms, see Fig 3.4A. The maturity size plays an important role in a organism's life cycle [Amado et al., 2018, Erten and Kokko, 2020]: Large organisms have potential advantages on the optimisation of the organism in multiple ways, such as to improve growth efficiency [Waters et al., 2010], to avoid predators [Fisher et al., 2016, Kapse-taki and West, 2019], to increase problem-solving efficiency [Morand-Ferron and Quinn, 2011], and to exploit the division of labour in organisms [Carroll, 2001, Matt and Umen, 2016]. Moreover, the maximum size has been related to the organism reproduction from the beginning of the evolution of multicellularity [Ratcliff et al., 2012]. Our results suggest that the smallest organism able to evolve irreversible somatic differentiation should typically be about 32 – 64 cells (unless the cost of soma-to-germ differentiation is extremely large and the cost of the reverse is low). This is in line with the pattern of development observed in *Volvocales* green algae. In *Volvocales*, cells are unable to move (vegetative function) and divide (reproductive function) simultaneously, as a unique set of centrioles are involved in both tasks [Wynne and Bold, 1985, Koufopanou, 1994]. *Chlamydomonas reinhardtii* (unicellular) and *Gonium pectorale* (small colonies up to 16 cells) perform these tasks at different times. They move towards the top layers of water during the day to get more sunlight. At night, however, these species perform cell division and/or colony reproduction, slowly sinking down in the process. However, among larger *Volvocales*, a division of labour begins to develop. In *Eudorina elegans* colonies, containing 16 - 32 cells, a few cells at the pole have their chances to give rise to an offspring colony reduced [Marchant, 1977, Hallmann, 2011]. In *P. californica*, half of the 128-celled colony is formed of smaller cells, which are totally dedicated to the colony movement and die at the end of colony life cycle [Kikuchi, 1978, Hallmann, 2011]. In *Volvox carteri*, most of a 10000-cell colony is formed by somatic cells, which die upon the release of offspring groups [Hallmann, 2011].

### Chapter 3. Evolution of irreversible somatic differentiation

---

Our study originated from the curiosity: what did drive the evolution of irreversible somatic differentiation? Why does green algae *Volvox* shed the most of its biomass in a single act of reproduction? And why in totally unrelated to it kingdom of Metazoa, in most of the species the majority of body cells is outright forbidden to contribute to the next generation? Our results shown, which factors makes a difference between the evolution of an irreversible somatic differentiation and other strategies of development. One of these factors, the maturity size is known in the context of the evolution of reproductive division of labour [Kirk, 2005]. Another, the costs of cell differentiation, is, in general, known in a greater biological scope but is hardly acknowledged as a factor contributing to the evolution of organism development. Finally, the early contribution of soma-role cells to the organism growth, even if they are in small numbers, is an unexpected outcome of our investigation, overlooked in the body of literature as well. Despite the simplistic nature of our model (we did not aim to model any specific organism), all our results find a confirmation among the *Volvocales* clade. Hence, we expect that the findings of this study reveal the general picture of the evolution of irreversible somatic differentiation, independently on the clade where it evolves.

## 3.5 Appendix

### 3.5.1 Search for the evolutionarily optimal developmental program

#### Finding the population growth rate for a given developmental program.

In [Gao et al., 2019] we have shown that a population of organisms, which begin their life cycle from the same state but have a stochastic development, eventually grows exponentially with the rate  $\lambda$  given by the solution of

$$\sum_i e^{-\lambda T_i} G_i P_i = 1. \quad (3.3)$$

Here,  $i$  is the developmental trajectory – in our case, the specific combination of all cell division outcomes;  $P_i$  is the probability that an organism development will follow the trajectory  $i$ ;  $T_i$  is the time necessary to complete the trajectory  $i$  – from a single cell to the maturity size of  $2^n$  cells;  $G_i$  is the number of offspring organisms produced at the end of developmental trajectory  $i$ , equal to the number of germ-role cells at the moment of maturity.

In order to find the population growth rate, we need to know  $G_i$ ,  $T_i$ , and  $P_i$  (how many offspring are produced, how long did it take to mature, and how likely is this developmental trajectory, respectively). The complete set of developmental trajectories is huge as it scales exponentially with the number of divisions  $n$ .

In our study, for each developmental strategy, we sampled  $M = 300$  developmental trajectories at random. To get each trajectory, we simulated the growth of the single organism according to the rules of our model. For each trajectory, the developmental time  $T_i$  was computed as a sum of cell doubling times at each of  $n$  cell divisions, the number of offspring  $G_i$  was given by the count of germ-role cells at the end of development. The resulting ensemble of trajectories (with  $P_i = 1/M$ ) was plugged into Eq. (3.3) to compute the population growth rate  $\lambda$ .

#### Finding the developmental program with the largest population growth rate

We assume that evolution occurs by growth competition between populations executing different developmental strategies. These strategies, which provide larger population growth rate will outgrow others. To find evolutionarily optimal strategies under given conditions, we screened through a large set of developmental strategies and identified the one with the maximal population growth rate  $\lambda$ . Since the probabilities of cell division outcomes sum into one ( $g_{gg} + g_{gs} + g_{ss} = 1$  and  $s_{gg} + s_{gs} + s_{ss} = 1$ ), these probabilities can be represented as a point on one of two simplexes, one for the division of germ-role cells, and one for the division of soma-role cells. Consequently, we choose the set of developmental strategies as a Cartesian product of two triangular lattices - one for division probabilities of germ-role cells ( $g_{gg}, g_{gs}, g_{ss}$ ) and one for soma-role cells ( $s_{gg}, s_{gs}, s_{ss}$ ). The lattice space was set to 0.1, so each of two independent lattices contained  $11 \times 12/2 = 66$  nodes, and the whole set of developmental strategies comprised  $66 \times 66 = 4356$  different strategies. For each of these strategies, the population growth rate  $\lambda$  was calculated and the strategy with the largest growth rate was identified as evolutionarily optimal.

In our investigation, parameters such as differentiation costs ( $c_s, c_g$ ) and maturity size ( $2^n$ ) were used as control parameters. In other words, we either fix them at the specific values, or screened through a range of values to obtain a map (see Figs. 3.2, 3.3). However, the parameters controlled the shape of composition effect profile ( $x_0, x_1, \alpha$ , and  $b$ ) were treated differently. For each combination of control parameters, we randomly sampled a number (between 200 and 3000) of combinations of these parameters. The thresholds ( $0 \leq x_0 \leq x_1 \leq 1$ ) were sampled as a pair of independent distributed random values from the uniform distribution  $U(0, 1)$ . The contribution threshold  $x_0$  was set to the minimum of the pair, and the saturation threshold  $x_1$  was set to the maximum. The contribution synergy ( $\alpha > 0$ ) corresponds to the concave shape of the profile at  $\alpha < 1$  and to the convex shape at  $\alpha > 1$ . Therefore,  $\log_{10}(\alpha)$  was sampled from the uniform distribution  $U(-2, 2)$ , so the profile has an equal probability to demonstrate concave and convex shape. Finally, the maximum benefit ( $0 \leq b < 1$ ) was sampled from a uniform

distribution,  $U(0,1)$ . For each tested combination of control parameters, we found the optimal developmental strategy for every sampled profile. We then classified these as irreversible somatic differentiation (ISD), reversible somatic differentiation (RSD), or no somatic differentiation (NSD).

### 3.5.2 Under costless cell differentiation, irreversible soma strategy cannot be evolutionarily optimal

In this section, we will show that an ISD strategy can never be an evolutionary optimum without cell differentiation being costly. To do that, we first consider the deterministic dynamics of the expected composition of the organism. Then, for an arbitrary ISD strategy, we identify a more advantageous RSD strategy which gives the same organism composition at the end of life cycle but higher number of soma-role cells during the life cycle.

In our model, the composition of the organism is governed by the stochastic developmental strategy and differs between different organisms. Here, as a proxy for this complex stochastic dynamics, we consider the mathematical expectation of the composition. Assume that after  $t \geq 0$  cell divisions the fraction of soma-role cells is  $s(t)$  and the fraction of germ-role cells is  $g(t) = 1 - s(t)$ . Then, the expected fractions of cells of the two types after the next cell division is

$$\begin{aligned} s(t+1) &= \left(s_{ss} + \frac{s_{gs}}{2}\right)s(t) + \left(\frac{g_{gs}}{2} + g_{ss}\right)g(t) = (1 - m_s)s(t) + m_g g(t), \\ g(t+1) &= \left(g_{gg} + \frac{g_{gs}}{2}\right)g(t) + \left(\frac{s_{gs}}{2} + s_{gg}\right)s(t) = (1 - m_g)g(t) + m_s s(t), \end{aligned} \quad (3.4)$$

where we introduced  $m_s = s_{gg} + \frac{s_{gs}}{2}$  and  $m_g = g_{ss} + \frac{g_{gs}}{2}$  – the probabilities that the offspring of a cell will have a different role. Naturally, for irreversible somatic differentiation (ISD)  $m_s = 0$  and  $m_g > 0$ , for NSD strategies  $m_g = 0$  and  $m_s$  being irrelevant, while the reversible differentiation (RSD) class covers the rest. Eq. (3.4) can be written in matrix form

$$\begin{pmatrix} s(t+1) \\ g(t+1) \end{pmatrix} = \begin{pmatrix} 1 - m_s & m_g \\ m_s & 1 - m_g \end{pmatrix} \cdot \begin{pmatrix} s(t) \\ g(t) \end{pmatrix} \quad (3.5)$$

### Chapter 3. Evolution of irreversible somatic differentiation

---

A newborn organism contains a single germ-role cell ( $s(0) = 0, g(0) = 1$ ), therefore, the expected composition of an organism after  $i$  divisions is

$$\begin{pmatrix} s(t) \\ g(t) \end{pmatrix} = \begin{pmatrix} 1 - m_s & m_g \\ m_s & 1 - m_g \end{pmatrix}^t \cdot \begin{pmatrix} 0 \\ 1 \end{pmatrix} \quad (3.6)$$

The matrix has two eigenvalues: 1 and  $1 - m_g - m_s$ , with associated right eigenvectors  $(m_g, m_s)^T$  and  $(1, -1)^T$ , respectively. Hence, the expected composition after  $t$  divisions can be obtained in the explicit form

$$\begin{aligned} s(t) &= \frac{1}{m_g + m_s} [m_g - m_g(1 - m_g - m_s)^t], \\ g(t) &= \frac{1}{m_g + m_s} [m_s + m_g(1 - m_g - m_s)^t]. \end{aligned} \quad (3.7)$$

For an arbitrary irreversible somatic differentiation strategy  $D$ ,  $m_s = 0$ , the expected number of soma-role cells changes as

$$s_D(t) = 1 - (1 - m_g)^t, \quad (3.8)$$

which is a monotonically increasing function of the number of cell divisions  $t$ , see the green line in Fig. 3.5. In the life cycle involving  $n$  cell divisions, the fraction of soma-role cells at the end of life cycle is  $s_D(n) = 1 - (1 - m_g)^n$ .

Now, we consider another developmental strategy  $D'$  with reversible somatic differentiation in which  $m'_g = s_D(n)$  and  $m'_s = 1 - s_D(n)$ . Using  $m'_g + m'_s = 1$  in Eq. (3.7), it can be shown that the expected fraction of soma-role cells in  $D'$  after the very first cell division is exactly  $s_D(n)$  and stays constant thereafter, see the orange line in Fig. 3.5. Thus, the number of offspring produced is the same for both development strategies.

If cell differentiation is costless ( $d_s = d_g = 0$ ), then the cell doubling time depends only on the fraction of soma-role cells. As all soma-role cells are then present already after the first cell division, organisms following the RSD strategy  $D'$  will grow faster than organisms using the ISD strategy  $D$  at any stage of organism development, independ-

ently of the choice of the composition effect profile ( $F_{\text{comp}}$ ). At the end of the life cycle, both strategies have the same expected number of offspring. Therefore, under costless cell differentiation, for any ISD strategy, we can find a RSD strategy that leads to a larger population growth rate.

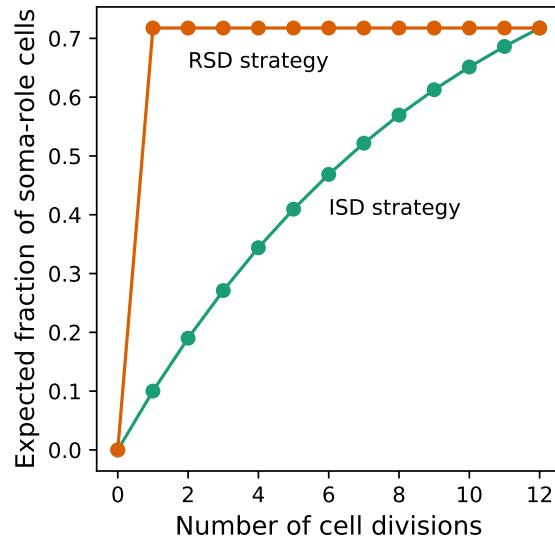


Figure 3.5: **Under costless differentiation, for any irreversible somatic differentiation strategy, exists a reversible somatic differentiation strategy dominating it.** The green curve shows the dynamics of the expected fraction of soma-role cells in an organism using an ISD developmental strategy ( $m_g = 0.1$ ,  $m_s = 0.0$ ,  $n = 12$ ). The orange curve shows the dynamics of the expected fraction of soma-role cells in an organism using the specific RSD developmental strategy [ $m'_g = 1 - (1 - m_g)^{12} \approx 0.72$ ,  $m'_s = 1 - m'_g \approx 0.28$ ]. In this strategy, the number of offspring produced at the end of the life cycle is the same as in the considered ISD strategy. At the same time, the fraction of soma-role cells during the life cycle is larger. Therefore, under costless differentiation, the presented RSD strategy is more effective than the considered ISD strategy.

### 3.5.3 Conditions promoting the evolution of ISD, RSD, and NSD strategies

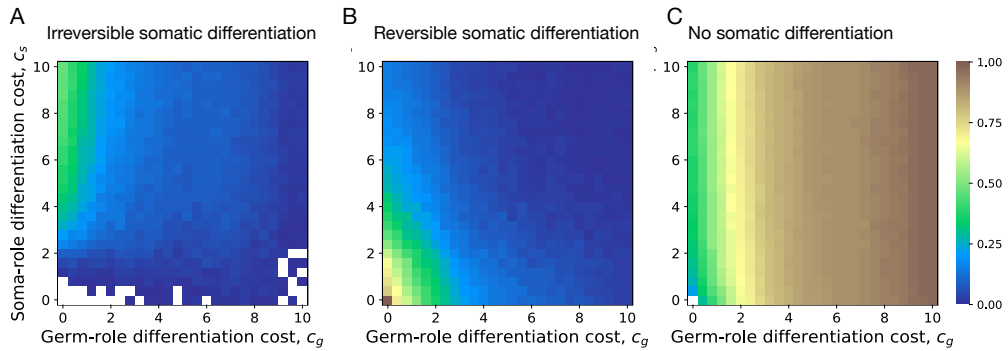


Figure 3.6: **Impact of cell differentiation costs on the evolution of development strategies.**

The fractions of 200 random composition effect profiles promoting ISD (A), RSD (B), and NSD (C) strategies at various cell differentiation costs ( $c_s, c_g$ ). In the absence of costs ( $c_g = c_s = 0$ ), only RSD strategies were observed. RSD strategies are prevalent at smaller cell differentiation costs. NSD strategies are the most abundant at large costs for germ-role cells ( $c_g$ ). ISD strategies are the most abundant at large costs for soma-role cells ( $c_s$ ). The maturity size used in the calculation is  $2^{10}$  cells.

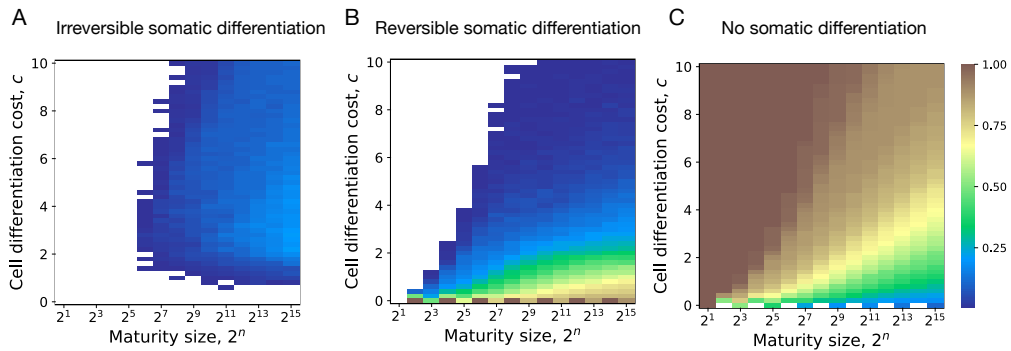


Figure 3.7: **Impact of maturity size on the evolution of development strategies.** The fractions of 200 random composition effect profiles promoting ISD (A), RSD (B), and NSD (C) strategies at various cell differentiation costs ( $c = c_s = c_g$ ) and maturity size  $2^n$ . ISD strategies are most abundant at large maturity sizes and intermediary cell differentiation costs. RSD strategies are most abundant at small cell differentiation costs. NSD strategies are most abundant at small maturity sizes and cell differentiation costs.



### 3.5.4 Parameters of composition effect profiles promoting ISD, RSD, and NSD strategies

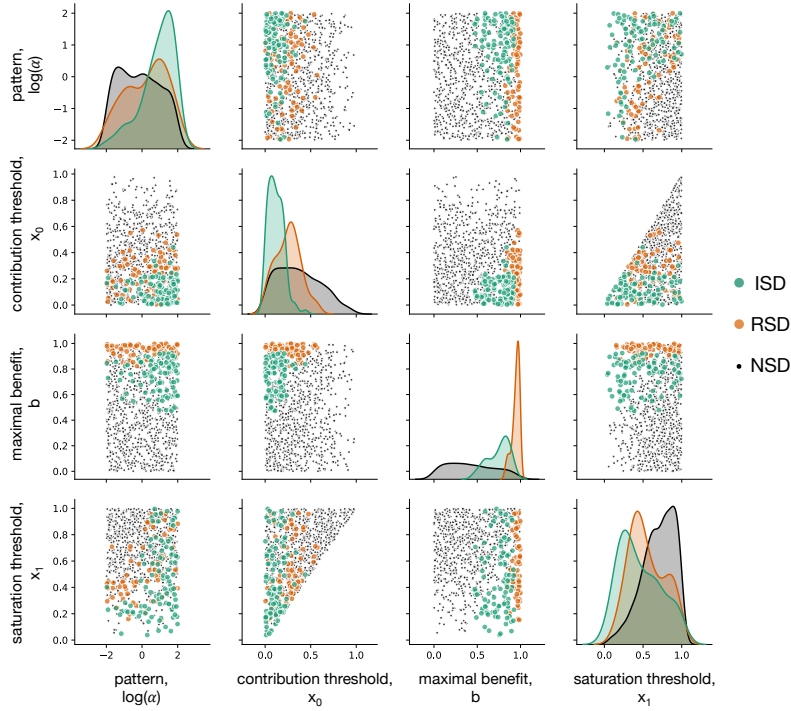


Figure 3.8: **Impact of composition effect parameters on the evolution of development strategies.** Each diagonal panel represents individual distribution of each of four parameters among composition effect profiles promoting ISD (green), RSD (orange), and NSD (black) strategies. Each non-diagonal panel represents a pairwise co-distribution of these parameters. ISD strategies are promoted at small contribution thresholds  $x_0$  and for large maximal benefit  $b$ . Also, either the contribution synergy  $\alpha$  must be large, or the saturation threshold  $x_1$  should be small - see main text for detailed discussion. RSD strategies require very large  $b$  - there the benefits of having a large number of soma-role cells outweighs costs paid by frequent differentiation. Due to the fast accumulation of soma-role cells, RSD strategies tolerate larger  $x_0$  than ISD. RSD exhibit the same restrictions with respect to  $x_1$  as ISD and are insensitive to  $\alpha$ . For this figure, 3000 composition effect profiles were investigated with costs  $c = c_s = c_g = 5$  and  $n = 10$ .

### 3.5.5 Evolution of irreversible somatic differentiation under various maturity sizes and unequal cell differentiation costs

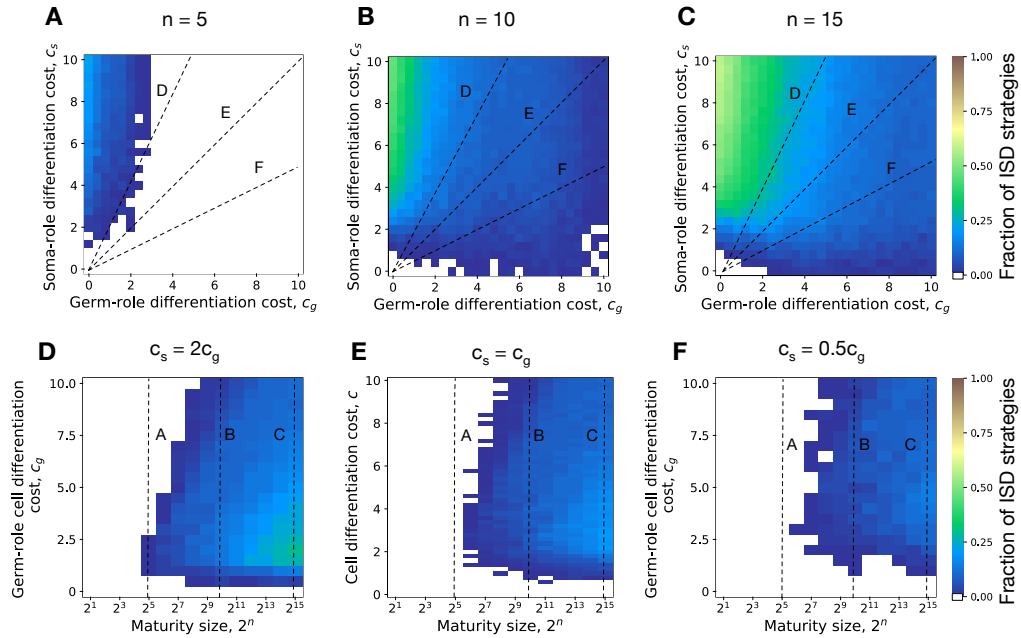


Figure 3.9: **Evolution of irreversible somatic differentiation at unequal cell differentiation costs.** **A-C.** The fraction of 200 random composition effect profiles promoting ISD at various cell differentiation costs ( $c_s$ ,  $c_g$ ) at fixed maturity size  $n = 5$  (panel A), 10 (B), and 15 (C). Larger maturity sizes promote the evolution of ISD across all cell differentiation costs. **D-F.** The fraction of composition effect profiles promoting ISD at unequal cell differentiation costs  $c_s/c_g = 2$  (panel D),  $c_s/c_g = 1$  (E), and  $c_s/c_g = 0.5$  (F). Even with unequal differentiation costs, the minimal maturity size allowing the evolution of ISD stays roughly the same —  $2^5 - 2^6$  cells. Dashed lines indicate overlap between panels.

# Evolution of reproductive strategies in incipient multicellularity

---

A manuscript based on this chapter is in preparation: Yuanxiao Gao, Yuriy Pichugin, Chaitanya Gokhale and Arne Traulsen, “Evolution of reproductive strategies in incipient multicellularity”.

### **Abstract**

Multicellular organisms can potentially enjoy a notable diversity in terms of reproductive strategies to reproduce offspring with varying sizes and composition, compared to their unicellular ancestors. In reality, only a few reproductive strategies are prevalent. So far, little is known about the evolution of reproductive strategies in multicellularity. Here, we develop a stage-structured population model to probe evolutionary growth advantages of reproductive strategies in incipient multicellularity. In the model, populations take different reproductive strategies, whose performance is evaluated by the growth rates of populations they are in. We seek the optimal reproductive strategy, who leads to the largest growth rate to a population. Considering the effects of organism sizes and cellular interactions, we found that distinct reproductive strategies could perform uniquely or equally optimal under different conditions. Among them, the binary-splitting reproductive strategy is unique optimal. Our results show that organism size and cellular interaction in the form of a threshold effect play crucial roles in shaping the reproductive strategies in nascent multicellularity. Our model sheds light on understanding the mechanism driving the evolution of reproductive strategies in complex multicellularity. Meanwhile, beyond multicellularity, we suggest a crucial factor in the evolution of repro-

ductive strategies among unicellular species - organism size.

### 4.1 Introduction

The evolution of multicellularity has been viewed as a major evolutionary transition, which has happened repeatedly across prokaryotes to eukaryotes [Bonner, 1998, Grosberg and Strathmann, 2007, Rokas, 2008, Claessen et al., 2014, Sebe-Pedros et al., 2017, Brunet and King, 2017]. Along with the increased organism size, phenotypically heterogeneous organisms emerged through cell differentiation [McCarthy and Enquist, 2005, Arendt, 2008, Brunet and King, 2017]. The reproductive mode may transform with the change of size and organism composition. In principle, multicellular organisms could reproduce multiple offspring with distinct cell numbers and organism composition in contrast to their unicellular ancestors [Michod and Roze, 1999, Ratcliff et al., 2012, ?, Pichugin et al., 2019, Gao et al., 2019]. The number of reproductive modes can rapidly increase with organism size. For example, for an organism containing three cells, two reproductive strategies are possible: reproducing three single-celled offspring or a single-celled offspring and a two-celled offspring. However, only a few of reproductive strategies are prevalent across the tree of life. Some prominent examples abound such as binary fission reproducing two single-celled organisms, multiple fission reproducing many single-celled organisms simultaneously [Suresh et al., 1994, Angert, 2005, Flores and Herrero, 2010], the fragmentation reproducing some many-celled propagules [Ratcliff et al., 2012] and a special bottleneck reproductive strategy, a multicellular organism reproducing a single-celled offspring repeatedly [Grosberg and Strathmann, 1998, Wolpert and Szathmary, 2002, Brunet and King, 2017]. We tackle this disparity between conjecture and reality in terms of the number of reproductive strategies in primitive multicellularity.

The origin and evolution of reproductive strategies are little understood, and only a few of them have been considered in previous work. The fragmentation mode of re-

producing many-celled propagules has been investigated to understand cell death in yeast [Libby et al., 2014] or to understand the advantages of multicellular life experiencing a unicellular stage [Grosberg and Strathmann, 1998, Michod and Roze, 1999]. The mechanism of the life cycle transition from unicellular to multicellular has been examined, however, the underlying reproductive strategies are still unknown [Staps et al., 2019]. Recently, reproductive strategies have been explored for phenotypically homogeneous organisms, where an organism is allowed to take mixed reproductive strategies [Pichugin et al., 2019]. The mixed reproductive strategy allows an organism to have different fragmentation modes in a population. For example, a three-celled organism can reproduce three single-celled offspring or a one-celled offspring and a two-celled offspring. The pure reproductive strategy only has one fragmentation mode. Pure reproductive strategies have been studied in phenotypically heterogeneous organisms, but the size effect has not been considered [Gao et al., 2019]. Meanwhile, cellular interactions have been only considered in linear frequency-dependence. Therefore it is still unclear how organism size and cellular interactions beyond linear forms, can shape reproductive strategies.

The organism size confers multifarious advantages to organisms [Kaiser, 2001, Carroll, 2001], such as to avoid predators [Fisher et al., 2016, Kapsetaki and West, 2019], and to incentivise the division of labour [Carroll, 2001, Matt and Umen, 2016]. Meanwhile, disadvantages in organism growth have also been suggested due to different reasons, such as less space [Libby et al., 2014] or less light [Kapsetaki and West, 2019]. Furthermore, the organism size can affect reproductive strategies as early as in nascent multicellularity [Michod, 2007, Solari et al., 2013, Ratcliff et al., 2012, Libby et al., 2014]. Field observations are ambiguous about the effects of organism size [Yamamoto and Shiah, 2010, Nielsen, 2006, Li et al., 2014, Wilson et al., 2006, Li and Gao, 2004, Wilson et al., 2010]. The effects of size have also been considered in investigating reproductive strategies in phenotypically heterogeneous organisms [Pichugin et al., 2019]. But only the monotonous increasingly functional forms of size on growth rate has been considered. Here, we consider a wide scope of size effects that can increase, decrease or no change of an organism's

## Chapter 4. Evolution of reproductive strategies in incipient multicellularity

---

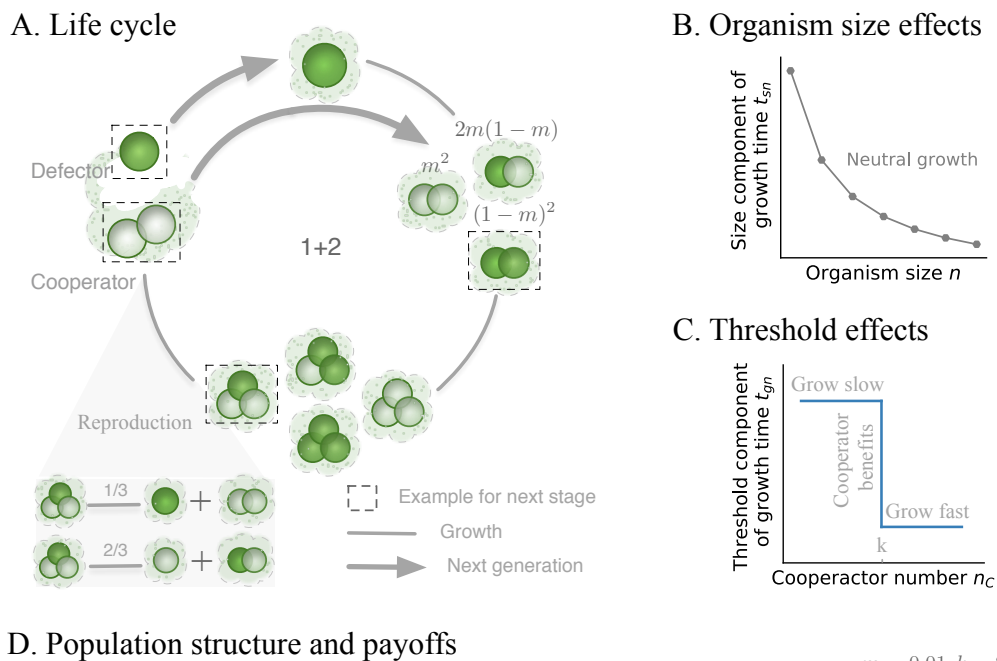
growth rate. And the functional forms of size effects are broader.

The previous studies have shown that cellular interactions can transform reproductive modes [Kaiser, 2001, Solari et al., 2013, Ratcliff et al., 2012]. For example, a new phenotype with a higher death rate leads to the reproduction of reproducing propagule among yeast *Saccharomyces cerevisiae*. Phenotypically heterogeneous organisms could provide diverse cellular interaction forms, considering different organism composition. The effects of linear frequency-dependent organism composition have been investigated [Gao et al., 2019]. Here we study the cellular interaction that depends on the minimum threshold of a certain phenotype of an organism. This cellular interaction form has frequently been observed in nature. For example, to differentiate heterocyst for cyanobacteria in response to nitrogen depletion, a spatial pattern is found, in which one heterocyst surrounded by thresholds of vegetative cells between 10 to 20 [Kumar et al., 2010, Flores and Herrero, 2010]. In the genus *Volvox*, along with the germ-soma differentiation [Matt and Umen, 2016], a pattern of cell distribution has been observed, in which the thresholds of germ cells are between 1 to 20 and soma cells are between 500 to 42,000 [Shelton et al., 2012]. In our model, we study the threshold effects of the organism composition.

Thus in heterogeneous multicellular organisms both size and composition could affect growth rates. We develop a theoretical model to address the evolution of reproductive strategies considering the effects of size and threshold. The size effects could increase or decrease organism growth. The threshold effects depend on the number of cells with a certain phenotype. We assume that organisms grow fast if they meet a given threshold. Organisms in a population share the same reproductive strategy. Thus, reproductive strategies compete with each other via their population growth rates. The optimal reproductive strategy is the one with the largest population growth rate. We found that reproductive strategies can co-exist or can dominate others under different conditions. The unique optimal reproductive strategy is always the one reproducing two offspring.

## 4.2 Model

We consider multiple infinite large populations. Each population possesses a unique reproductive strategy for all of its organisms. Reproductive strategies are the partition of integers. For example, for an integer  $N = 3$ , there are two reproductive strategies:  $1 + 1 + 1$  and  $1 + 2$ . In the population with reproductive strategy  $1 + 2$ , mature organisms with three cells can only reproduce a single-celled newborn organism and a two-celled



**D. Population structure and payoffs**

$m = 0.01; k = 2$

Reproductive strategy	1+2				
Newborn state	$\{1, 0\}$ 	$\{0, 1\}$ 	$\{2, 0\}$ 	$\{1, 1\}$ 	$\{0, 2\}$ 
Payoff	0	—	0	0	—
	—	$-c$	—	$-c$	$b - c$
Average payoff	0	$-c$	0	$-\frac{c}{2}$	$b - c$
Mature state prospect					
Long-term prospect		Intermediate beneficial			Beneficial

Figure 4.1: **Schematic of life cycle and reproduction.** **A.** Examples of life cycles of two organisms ( $\{1,0\}$  and  $\{0,2\}$ ) in the population with reproductive strategy  $1+2$ . In the shaded area, we show the potential random offspring organisms and their probabilities, see Appendix 4.5.1 for the calculation. **B.** Organism size effects on the growth time of organisms. The grey dots are an example of the neutral condition, where populations have the same growth rate. **C.** Threshold effects on the growth time of organisms. In an organism, when the cooperator number  $n_C$  meets the contribution threshold  $k$ , the threshold component of growth time  $t_{gn}$  will decrease according to the cooperator benefits via a volunteer dilemma game, see the explanation in the main text. **D.** An example of a population's newborn organisms and their payoffs under the threshold effect. Newborn organisms of the population with reproductive strategy  $1+2$  are shown. Each cell's payoff in an organism and organisms' average payoffs are listed under the contribution threshold  $k=2$ . The mature state prospect describes the most likely cell composition of each newborn organism. The long-term prospect of newborn organisms is classified into "beneficial" and "intermediate beneficial", see the detail in main text.

newborn organism. Furthermore, a reproductive strategy determines the organism sizes at newborn stages and mature stages. For a reproductive strategy  $n_1 + n_2 + \dots + n_M$ , its population has newborn organisms with different sizes of  $n_i$  ( $i \in [1, M]$ ) and the maturity size of newborn organisms  $N$ , where  $N = \sum_{i=1}^M n_i$ . Inspired by the viability investment of germ cells and somatic cells in the genus *Volvox*, we consider organisms consisting of two cell phenotypes: cooperator and defector [Kirk, 2001, 2005, Matt and Umen, 2016]. Therefore, a population composes of different organisms considering their sizes and organism composition. For example, the population with reproductive strategy  $1+2$  has five types of newborn organisms:  $\{1,0\}$ ,  $\{0,1\}$ ,  $\{2,0\}$ ,  $\{1,1\}$ , and  $\{0,2\}$ , where  $n_D$  and  $n_C$  in  $\{n_D, n_C\}$  represent the number of defectors and cooperators respectively, see Fig 4.1D. Each newborn organisms grow by one cell increment at a time. During each increment, a cell with higher fitness is selected to divide and two daughter cells are produced. Each daughter cell can switch to another phenotype independently with a cell-type switching probab-



ility, which is constant in our model  $m = 0.01$ . After reaching their maturity size  $N$ , organisms reproduce via random fragmentation in terms of organism composition, and the probabilities of forming different newborn organisms are calculated in Appendix 4.5.1. These life cycles, from newborn stages to reproductive stages, are repeated by organisms in a population, see Fig 4.1A.

We assume that organisms in populations grow independently without density dependence, thus populations follow exponential growth [Tuljapurkar and Caswell, 1997]. The population growth rate  $\lambda$  can be calculated by the leading eigenvalue of a matrix, based on the offspring numbers and the growth time of organisms [De Roos, 2008, Gao et al., 2019], see Appendix 4.5.2. Since there are no death events in a population, so the offspring number of each organism is constant, which depends on its reproductive strategy. For example, under strategy 1 + 2, all organisms reproduce two offspring after reproduction. Thus, the population growth rate is determined by the growth time of its organisms from being newborn to mature. We additionally assume that reproduction is instantaneous and no time is needed for this process. We consider that population growth time is determined by the size and composition of its organisms as

$$T = \sum^n t_n = \sum^n (t_{sn} \times t_{gn}) \quad (4.1)$$

where the  $t_n$  is the cell increment time for an organism growing from size  $n$  to  $(n + 1)$ . The  $t_{sn}$  and the  $t_{gn}$  are the size component and the threshold component contributing to  $t_n$ . Next, we introduce how we design the  $t_{sn}$  and the  $t_{gn}$ , respectively.

The size component  $t_{sn}$  depends on the cell number of an organism. Under neutral condition  $t_{sn}^0 = \gamma \ln \frac{n+1}{n}$ , all populations have the same growth rate 1, where  $\gamma$  is a positive constant [Gao et al., 2019]. Without lose of generality, we chose  $\gamma = 1$ . To seek the size effects beyond the neutral conditions, we screen a large number of values of  $t_{sn}$  around the neutral condition ( $t_{sn}^0$ ), see Fig 4.1B and Fig 4.2A. We refer to  $\chi_n = \chi_n$  as normalised cell increment components, where  $n \in [1, N]$ . When  $\chi_n = 1$ , we recover the neutral condition.

The threshold component  $t_{gn}$  depends on the number of cooperator cells in an organ-

## Chapter 4. Evolution of reproductive strategies in incipient multicellularity

---

ism. An organism grows faster if the number of cooperators meets a given threshold  $k$ , Fig 4.1C. There are many methods to construct the compositional threshold effect. Here we choose the volunteer dilemma game. Consider an organism consisting of  $n$  cells with  $n_D$  defectors and  $n_C$  cooperators. When cooperator number  $n_C$  meets the contribution threshold  $k$ , each cell can get a  $b$  payoff. Each cooperator always bears a cost  $c$  and defectors pay no costs, see Eq4.2 and Fig 4.1D.

$$P_D(n_C) = \begin{cases} b & n_C \geq k \\ 0 & n_C < k \end{cases} \quad (4.2)$$

$$P_C(n_C) = P_D(n_C) - c.$$

Meanwhile, the cell payoffs affect the division probability among these two phenotypes, i.e. which phenotypic cell is more likely to divide, see Eq.4.3.

$$D_D = \frac{n_D e^{wP_D}}{n_D e^{wP_D} + n_C e^{wP_C}} \quad (4.3)$$

$$D_C = \frac{n_C e^{wP_C}}{n_D e^{wP_D} + n_C e^{wP_C}},$$

where  $D_D$  and  $D_C$  are the division probability for defectors and cooperators, respectively.  $w$  is the intensity of selection. The composition component  $t_{gn}$  is defined by the payoff of  $P_D$  and  $P_C$ , see Eq4.4.

$$t_{gn} = \left( \frac{n_D e^{wP_D} + n_C e^{wP_C}}{n_D + n_C} \right)^{-1}. \quad (4.4)$$

To check the compositional threshold effects, we screen the threshold values of  $k$ .

### 4.3 Results

#### 4.3.1 The effects of organism sizes on reproductive strategies

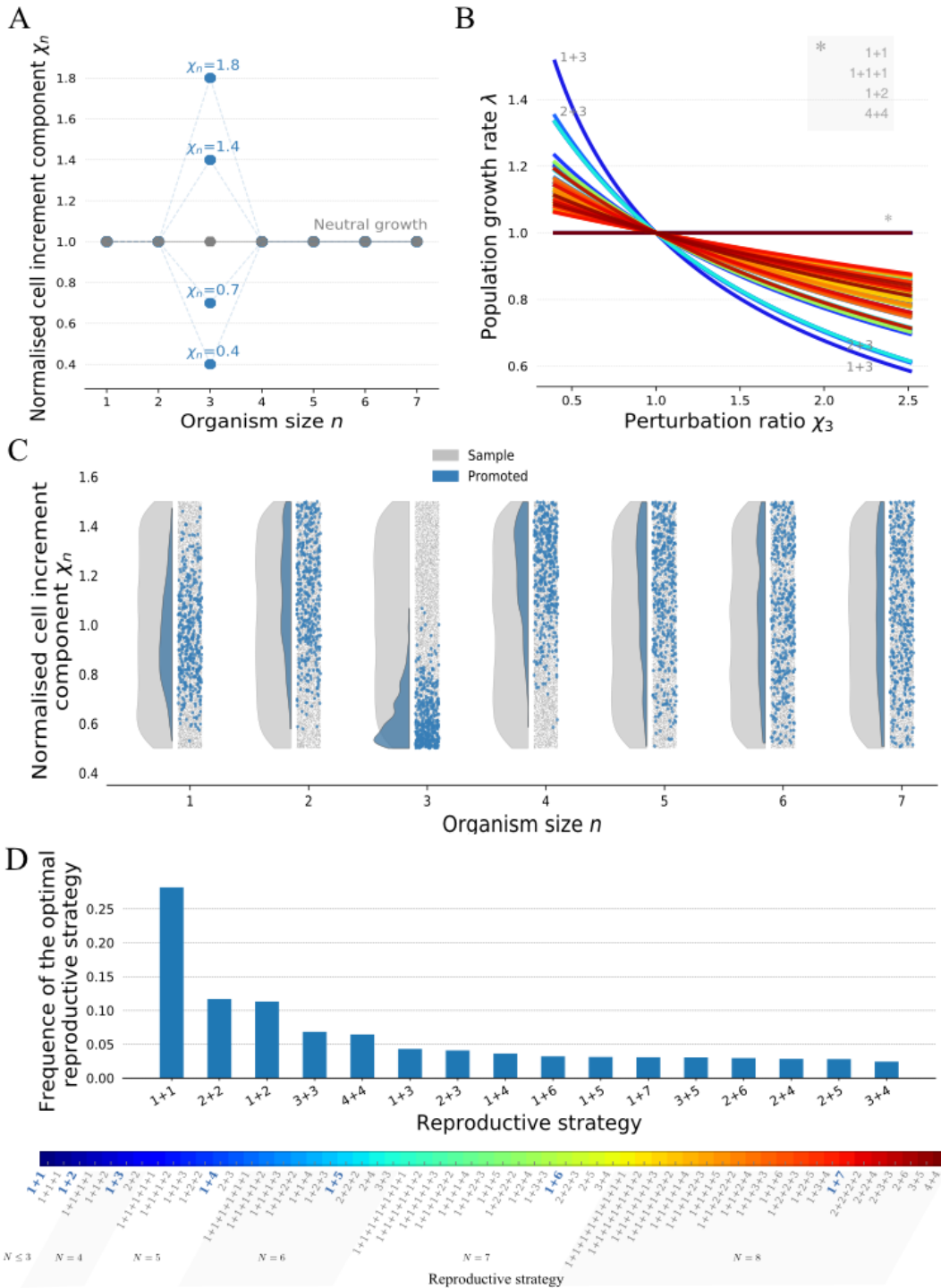


Figure 4.2: **The performance of reproductive strategies under organism size effects.** **A.** A diagram of perturbations at size  $n = 3$ . Grey dots are the conditions for neutral population growth i.e.  $\chi_n = 1$ . Blue dots are the perturbed dots at size 3 with different degrees, where  $0.4 \leq \chi_3 \leq 2.5$ . **B.** Population growth rates of reproductive strategies under perturbations at size  $n = 3$ . The asterisk \* shows the unaffected reproductive strategies. **C.** The distribution of  $\chi_n$  that promotes the reproductive strategy 1 + 3 (in blue) among samples (in grey).  $\chi_n$  are drawn from the uniform distribution and  $\chi_n \in [0.5, 1.5]$ . A sequence of  $\chi_n$  ( $n \in [1, 7]$ ) is randomly chose at a time for finding the optimal reproductive strategy. 10000 sequences are investigated in total. **D.** The frequency of observed optimal reproductive strategies under size effects. Parameters for all panels, each population growth rate has been calculated by running its organisms' life cycles for 5000 times. The growth rates of 58 populations are calculated under maturity size  $N \leq 8$ . In the lower legend panel, the reproductive strategies highlighted in blue are the optimal ones under a size perturbation.

We investigate size effects by perturbing a cell increment component  $\chi_n$  from neutral conditions  $\chi_n = 1$ , see Fig 4.2A. Under a perturbation at size  $n$ , we found that the performance is affected (population growth rate deviates from 1) for the reproductive strategies, whose populations contain organisms going through the size  $n$ . Since population growth rate is inversely related to growth time, thus the perturbation of  $\chi_n$  is either advantageous ( $\chi_n < 1$ ) or adverse ( $\chi_n > 1$ ) for population growth. A reproductive strategy is referred as being promoted (suppressed), when its population growth rate is greater (smaller) than the neutral growth rate 1. Under  $\chi_n < 1$ , the reproductive strategy is promoted if its population's organisms go through the size under perturbations, see Fig 4.2B. The promoted extent ( $|\lambda - 1|$ ) of a reproductive strategy is positively related to the perturbation degree, i.e the deviation value  $\chi_n$  from 1. The performance of reproductive strategies is unaffected when their populations' organisms do not go through the size under perturbations. For  $\chi_n > 1$ , reproductive strategies which are promoted for  $\chi_n < 1$  are suppressed. Likewise, the suppressed extent of the reproductive strategies is positively related to the perturbation degree. Among these affected populations, we found that the reproductive

strategy  $1 + n$  is the most affected one under perturbations at size  $n$ . Because its population contains  $n$ -celled newborn organisms, which mature at size  $n + 1$ , thus its growth time depends on  $\chi_n$ . Therefore, under the condition of  $\chi_n < 1$  and  $\chi_k = 1$  ( $k \neq n, k \in [1, N]$ ), the reproductive strategy  $1 + n$  is unique optimal. At the same time, the reproductive strategy  $1 + n$  is the most suppressed one for  $\chi_n > 1$ , see Fig 4.2B. Analogous to reproductive strategy  $1 + n$ , the reproductive strategy  $2 + n$  is the second most affected strategy, as a result of its population containing organisms going through size stage from  $n$  to  $n + 2$ . Similarly, the structure of a population decides its reproductive strategy's performance for other affected strategies.

Under the general size effect by perturbing  $\chi_n$  from 1 for different  $n$ , we found that the normalised cell increment components determines the optimal reproductive strategies. We observed that the populations of optimal reproductive strategies contain the organisms, who mostly undergo the size  $n$  possessing the small  $\chi_n$ , and meanwhile mostly do not contain the size  $n$  possessing the large  $\chi_n$ , see Fig 4.2C D. The small normalised cell increment components are the relatively smaller normalised cell increment components, see Fig 4.2D and the analytical proof for reproductive strategies with  $N \leq 3$  in Appendix 4.5.3. Furthermore, we found that only the binary-splitting reproductive strategy (reproducing two offspring with varying sizes) can be the unique optimal one, see Fig 4.2E and the analytical proof in Appendix 4.5.4. In a population, the fastest-growing newborn organisms largely decide its population growth rate. However, the fastest-growing newborn organisms in a population with a multiple-splitting reproductive strategy can always be found in another population with a binary-splitting reproductive strategy. For example, the reproductive strategy  $1 + 1 + 2$  cannot outcompete reproductive strategies of  $1 + 1$ ,  $1 + 2$ , or  $2 + 2$  at the same time under any conditions of  $t_{sn}$ , where  $n \in 1, 2, 3$ . Additionally,  $1 + 1$  is the most frequently observed strategy in binary-splitting reproductive strategies, see Fig 4.2E. Because of  $1 + 1$  is the only reproductive strategy that depends on a single cell increment component, i.e.  $t_{s1}$ . Therefore, under the randomly chose  $t_{sn}$  ( $n \in [1, N]$ ),  $1 + 1$  has a higher frequency to turn into the optimal one. Generally, reproductive strategies have lower chances to be optimal when they make or

organisms undergoing many cell increment stages. But this numerical conclusion needs to be further confirmed by taking large sampling size, especially for the binary-splitting reproductive strategies with frequencies lower than 0.01.

### 4.3.2 The effects of thresholds on reproductive strategies

Under the threshold effect, newborn organisms of a population with cooperator number  $n_C \geq k$  have large payoffs and thus have shorter growth time, see Eq 4.2 and Eq 4.4. In a population, newborn organisms contribute relatively different to the growth rate. We take the newborn organisms in the population with strategy 1 + 2 as an example, see Fig 4.1D. Under contribution threshold  $k = 2$ ,  $\{0, 2\}$  grows fastest as its cooperator number starts with 2.  $\{0, 1\}$  is the second fastest growing newborn organisms as it may gain benefits by producing 2 cooperators during growth. The rest newborn organisms of  $\{1, 0\}$ ,  $\{1, 1\}$  and  $\{2, 0\}$  grow relative slow because their cooperator numbers are less likely to meet contribution threshold  $k$ . For convenience, we refer to newborn organisms in a population as “beneficial” ones if  $n_C \geq k$  and “intermediate beneficial” ones if  $n_C < k$  and  $n_D = 0$ . The growth rate of a population depends primarily on its beneficial newborn organisms and secondly on its intermediate beneficial newborn organisms. The homogeneous newborn organisms are more populous than the heterogeneous ones, because for the cell-type switching probability  $m = 0.01$ , populations mostly contain homogeneous newborn organisms in the long run, see Appendix Fig 4.5.

Under the threshold effects, the unique optimal reproductive strategies are the binary-splitting ones with the maximum maturity size, see Fig 4.3A. Furthermore, the optimal reproductive strategies can be classified into three categories: the multiple optimal, the symmetric binary-splitting  $\frac{N}{2} + \frac{N}{2}$  (or  $\frac{N+1}{2} + \frac{N-1}{2}$ ) and the asymmetric binary-splitting with a  $k$ -celled newborn organism  $(N - k) + k$ . Under  $k = 1$ , all reproductive strategies are multiple optimal, see Fig 4.3A B and C. Since all populations contain beneficial newborn organisms, the population growth rates of reproductive strategies are close to each other. As  $k$  increases, the symmetric binary-splitting reproductive strategies  $\frac{N}{2} + \frac{N}{2}$  (or

### 4.3. Results

$\frac{N+1}{2} + \frac{N-1}{2}$ ) are optimal under  $1 < k \leq \frac{1}{2}N$ , see Fig 4.3A B. As the beneficial newborn organisms benefit a population most, therefore, the optimal reproductive strategy's population should consist of the beneficial newborn organisms, whose newborn sizes are larger or equal to the contribution threshold  $k$ . But, among them, we found that  $\frac{N}{2} + \frac{N}{2}$  (or  $\frac{N+1}{2} + \frac{N-1}{2}$  are the optimal ones, rather than others, such as  $k + k + k$  or  $(k + 1) + k$ . This is due to the intrinsic population structure of reproductive strategies and the effects of

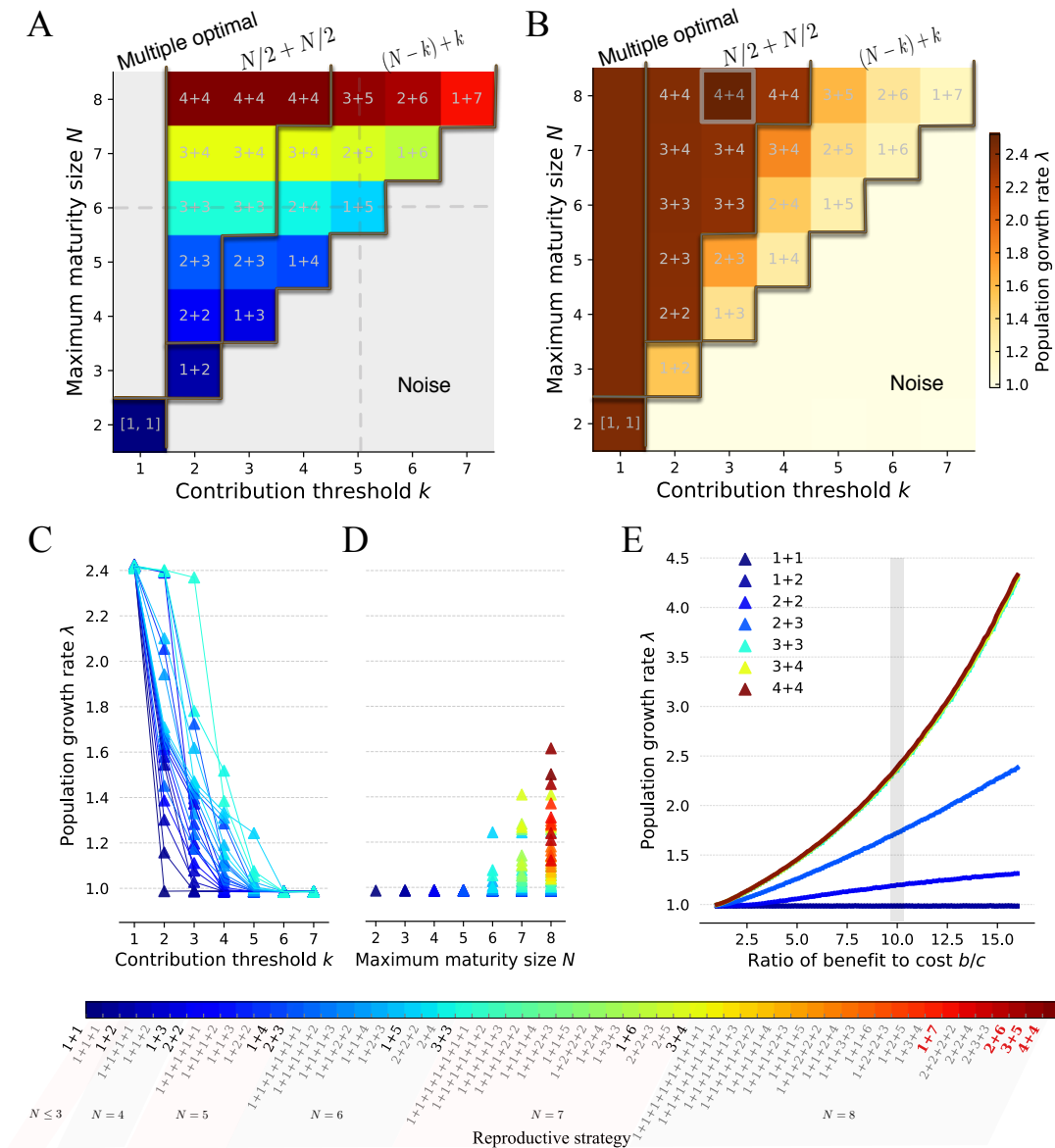


Figure 4.3: **Optimal population profile across different thresholds.** **A.** The optimal reproductive strategies across the contribution threshold  $k$  and the maturity size  $N$ , where  $k < 8$  and  $N \leq 8$ . The grey dashed lines indicate parameter space in panel C and D. The dark brown dashed lines (panel A and B) are the boundaries between the multiple optimal strategies, the symmetric binary reproductive strategies, the asymmetric binary reproductive strategies and noise sections. **B.** The growth rates of the optimal reproductive strategies in panel A. The shaded area with maximum maturity size 8 and  $k = 3$  is the parameter space we investigated all populations featured by symmetric binary reproductive strategies in panel E. **C.** Population growth rates of reproductive strategies with  $N \leq 6$  are shown across different contribution threshold  $k$ . **D.** Population growth rates under contribution threshold  $k = 5$  are shown across different maturity size  $N \leq 8$ . **E.** The growth rates of populations with symmetric binary reproductive strategy are shown across to varying ratio of benefit to cost. In the lower legend panel, the optimal populations appeared in panel A are highlighted in black. The unique optimal populations under the threshold effect are highlighted in bold and red. Parameters  $w = 0.1$ ,  $b = 10$ ,  $c = 1$  and  $m = 0.01$ .

the relatively small cell-type switching probability  $m = 0.01$ . We take  $4 + 4$  and  $3 + 3$  at  $k = 3$  as an example to explain it. Under  $k = 3$ , the population of  $4 + 4$  contains the beneficial newborn organisms  $\{0, 4\}$ . The population of  $3 + 3$  contains the beneficial newborn organisms  $\{0, 3\}$ . Thus the two populations should have the same growth rate asymptotically without considering  $m$ . But cell-type switch leads to a decline in the proportion of beneficial newborn organisms in a population. This decline gives rise to an advantage for populations with larger maturity size. For example, if there is a cell-type switching event for  $\{0, 4\}$  and  $\{0, 3\}$ , then after their reproduction  $\{1, 3\}$  and  $\{1, 2\}$  can be reproduced, respectively.  $\{1, 3\}$  is still a beneficial newborn organism, but not for  $\{1, 2\}$ . Population growth rates are close to each other among binary-splitting reproductive strategies under  $m = 0.01$ . But the differences amplify with the increasing ratios of benefit to cost, see Fig 4.3E. Finally, when  $\frac{1}{2}N < k < N$ , the reproductive strategies  $(N - k) + k$  become optimal, see Fig 4.2 A. When  $\frac{1}{2}N < k$ , each population can have at most one beneficial



newborn organism. Next, we explain why  $(N - k) + k$  is optimal. For intermediate beneficial newborns, a larger newborn size gives rise to a better performance in growth. As smaller newborn organisms need more time to reach  $k$  cells than large ones. We take  $1 + 1 + 2$  for  $k = 3$  as an example.  $\{0, 1\}$  undergoes two cell increment stages with negative average payoffs, while  $\{0, 2\}$  only undergoes one stage. As a result, the best strategy is the one, whose population contains a beneficial newborn organism with size  $k$  and an intermediate beneficial newborn organism with size  $N - k$ .

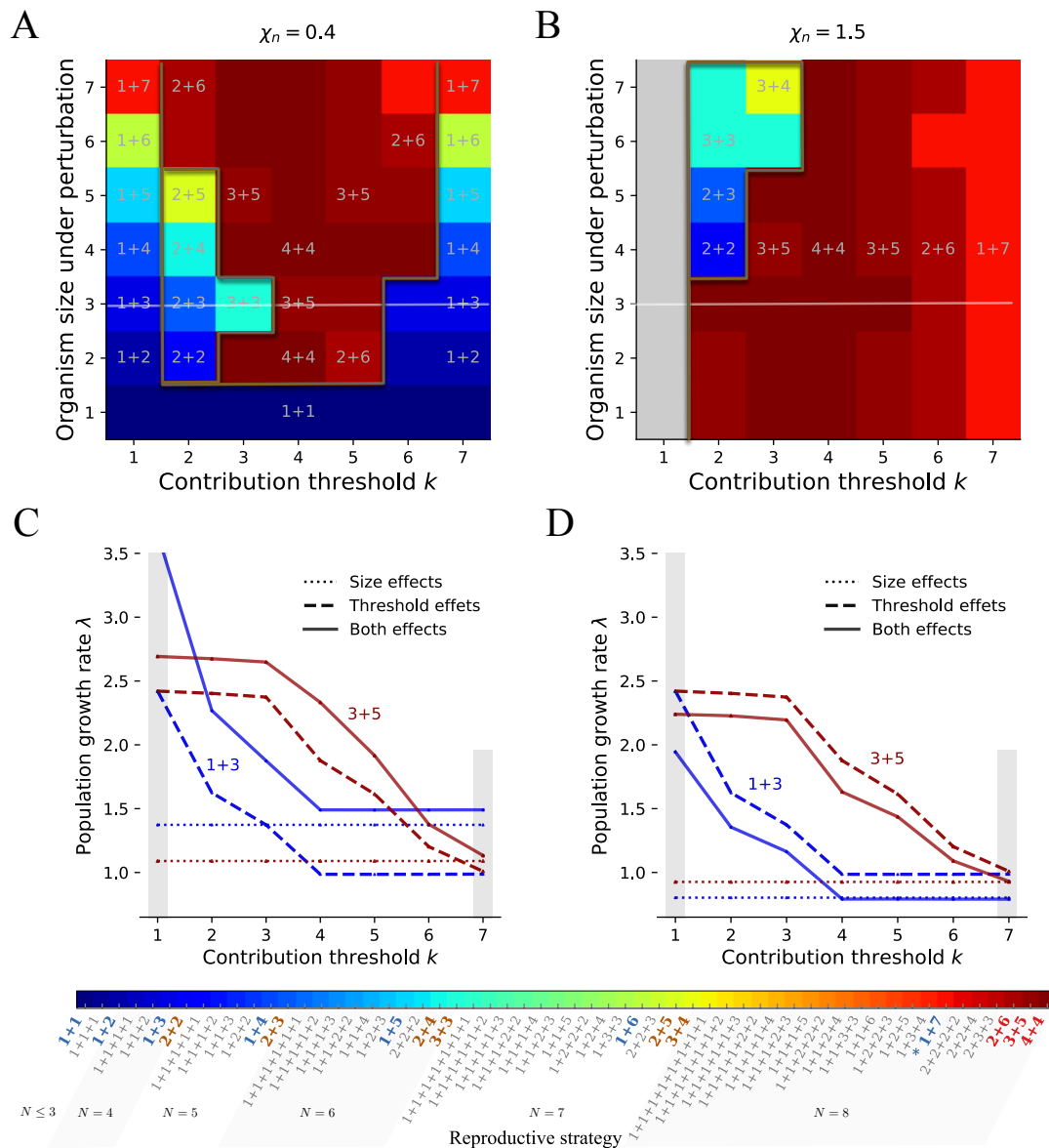
Additionally, our results show that reproductive strategies are suppressed progressively with the growing contribution threshold  $k$ , but it is promoted with the maturity size  $N$  and the ratio of benefit to cost  $\frac{b}{c}$ , see Fig 4.3B C D and E. Population growth rates decrease with increasing  $k$  because the numbers of beneficial and intermediate beneficial newborn organisms decrease. Especially, when  $k \geq N$ , all reproductive strategies fall into noisy and their populations grow in the same low growth rate, see Fig 4.3A B. Contrarily, the increasing maturity size  $N$  increases population growth rates, as more newborn organisms become beneficial ones in a population. It is also manifest that population growth rates increase with the ratio of benefit to cost.

### 4.3.3 The effects of organism sizes and thresholds on reproductive strategies

We next find the optimal reproductive strategies under the size and threshold effects. For simplicity, we only investigate the size effects of a single perturbation at a time. We found all binary-splitting reproductive strategies can be unique optimal, see Fig 4.4A and B. Under the size and threshold effects, new optimal binary-splitting reproductive strategies emerge, including  $2 + 2$ ,  $2 + 3$ ,  $2 + 4$ ,  $2 + 5$ ,  $3 + 3$  and  $3 + 4$ . Furthermore, under the beneficial size perturbation  $\chi_n = 0.4$ , we found  $1 + n$  ( $n \in [1, 7]$ ) is optimal both at small and large contribution threshold  $k$ , see Fig 4.4A and B. Because under small  $k$ , reproductive strategies' populations could possess similar higher growth rates. Under large  $k$ , pop-

## Chapter 4. Evolution of reproductive strategies in incipient multicellularity

ulations share a similar low growth rate, Fig 4.2B and Fig 4.3C. Therefore, under these two scenarios, the performance of reproductive strategies depends remarkably on size effects, especially when  $k = 1$ , see Fig 4.4A C. Consequently, the size perturbation plays a more important role here, leading the reproductive strategy  $1 + n$  becoming optimal, where  $n \in [1, 7]$ . The advantages arise for the newly emerged binary-splitting reproductive strategies in the intermediate values of  $k$ , suggesting that it is an outcome of the trade-



**Figure 4.4: The binary-splitting reproductive strategies are optimal under the effects of size with a single perturbation and threshold.** **A.** Optimal reproductive strategies under the effects of single advantages size perturbations and thresholds. **B.** Optimal reproductive strategies under the effects of single adverse size perturbations and thresholds. In panel A and B, the perturbation only occurs at a single size at a time. The dark brown lines indicate the boundaries of optimal reproductive strategies that appeared under a single perturbation, threshold effects and both. Note that  $1 + 7$  can be unique optimal strategy under either a single perturbation or threshold effects. The grey area indicates the multiple-optimal reproductive strategies, whose populations do not contain newborn organisms undergoing the perturbed size. The white lines indicate the perturbation being investigated at size  $n = 3$  in panel C and D. **C D** The population growth rates of reproductive strategies of  $1 + 3$  and  $3 + 5$  under the effects of a size perturbation at  $n = 3$ , threshold and both, respectively. In A and C,  $\frac{r_n}{r_0} = 0.4$ . In B and D,  $\frac{r_n}{r_0} = 1.5$ . Parameters of all panels  $w = 0.1$ ,  $b = 10$ ,  $c = 1$ ,  $m = 0.01$ . In the legend panel, the strategies in blue are the optimal ones under the size effect of a single perturbation. The strategies in red are the optimal ones under the threshold effects. The strategies in brown are the newly emerged optimal ones under both the size effect of a single size perturbation and the threshold effects.

off between the two factors. Under the adverse size perturbation  $\chi_n = 1.5$ , we found the reproductive strategy  $1 + n$  cannot be the optimal one at small  $k$ , see Fig 4.4B. Because the adverse size perturbation leads to the poor performance of reproductive strategies being influenced, see Fig 4.2B and Fig 4.4D. The optimal reproductive strategies arise when they enjoy the benefit of threshold effects and meanwhile avoid the disadvantages from the adverse size effects. For example,  $3 + 3$  outcompetes  $4 + 4$  at the size perturbed at  $n = 7$  and  $k = 2$ . As the adverse size perturbation decreases the population growth rate of  $4 + 4$ , but have no influences on that of  $3 + 3$ . Thus the performance of a reproductive strategy is an outcome of the trade-off between the effects of size and threshold. Our results indicate that all binary-splitting reproductive strategies can readily evolve under the joint effects of the size and the threshold.

## 4.4 Discussion

More theoretically potential reproductive strategies are conceivable for multicellularity, yet little attention has been paid to their evolution. To understand that, we developed a theoretical model considering the effects of size and cell interaction of a threshold effect on the evolution of reproductive strategies. Only under the size effects, reproductive strategies are determined by the normalised cell increment components. The small normalised cell increment components can promote the performance of reproductive strategies. The large ones have the opposite effect. We found only the binary-splitting reproductive strategies can be the unique optimal ones. Specifically, only the binary-splitting reproductive strategy  $1 + n$  is the optimal one under a single size perturbation, where  $n$  is the size under perturbations. Only under the threshold effect, the optimal reproductive strategies are determined by the contribution threshold and the cell-type switching probability. We found only the reproductive strategies with maximum maturity size  $x + (N - x)$  are unique optimal, where  $x \in [1, N)$ . Under the joint effects of the size with a single perturbation and the threshold, we found that all binary-splitting reproductive strategies can be the unique optimal ones. Our conclusions suggest only the binary-splitting reproductive strategies can be unique optimal under either condition. The single size perturbation suggests that the binary-splitting reproductive strategies can readily evolve in multicellularity under influences of joint factors.

Our conclusion that the unique optimal reproductive strategies are binary-splitting ones under size effects coincides with the results in our previous work [Gao et al., 2019]. Meanwhile, this conclusion has been obtained when considering the costs of reproduction [Pichugin and Traulsen, 2020]. Moreover, our results show that the special bottleneck reproductive strategy  $1 + N$  can be unique optimal under either condition, which may indicate a new advantage of it other than the previously stated advantages of decreasing the mutation load and regulating the cell conflict [Grosberg and Strathmann, 1998, Michod and Roze, 1999]. Our results also show that multiple reproductive strategies are optimal

simultaneously under some conditions. This conclusion coincides the phenomenon that one species can possess several reproductive strategies simultaneously in nature, such as cyanobacteria, which has the reproductive strategies of binary fission, budding and multiple fission [Angert, 2005, Flores and Herrero, 2010]. The frequently observed reproductive strategy  $1 + 1$  among binary-splitting reproductive strategies indicates it is the best strategy under unpredicted size effects.

In our model, the size form is quite flexible, which could have positive or negative effects on organism growth even at the same size stage. This model set is corresponding to the field studies concerning the size effect [Yamamoto and Shiah, 2010, Nielsen, 2006, Li et al., 2014, Wilson et al., 2006, Li and Gao, 2004, Wilson et al., 2010], where the size and the organism growth have been shown to have positive, negative or no relationships. This size set covers a wide range of size functional forms, including the special form that previous work has been investigated [Pichugin et al., 2019]. We delineated the threshold effect in a multiplayer volunteer game because the game theory has been frequently used to depict the biological interactions ranging from social foraging to cancer development [Maynard Smith and Price, 1973, Tomlinson, 1997, Dugatkin and Reeve, 2000, Nowak and Sigmund, 2004, Nowak, 2006b, McNamara and Leimar, 2020]. However, the volunteer's dilemma in our model is quite different from the previous one, where people are interested in finding the equilibria either for pure strategies or mixed strategies [Archetti, 2009]. In our work, the individual cell only plays a pure strategy, which is manifest via its cell phenotype. The strategy can change after cell divisions by producing another type with a probability.

We chose the cell-type switching probability  $m = 0.01$  because, in reality, the emergence of a new cell type usually caused by environmental pressures in multicellularity, which suggests it happens less likely naturally [Gallon, 1992, Claessen et al., 2014]. This relative small cell-type switching probability leads to a relatively homogeneous population, who mostly contains homogeneous newborn organisms, see Fig 4.5. Homogeneous newborn organisms consisting of cooperators dominate a population if the maturity size

## Chapter 4. Evolution of reproductive strategies in incipient multicellularity

---

of organisms meets a given contribution threshold. Otherwise, homogeneous newborn organisms consisting of defectors dominate the population. If contribution threshold is between the maximum size of newborn organisms and the maturity size of organisms for a population, then the intermediate beneficial newborn organisms dominate the population. Although heterogeneous beneficial newborn organisms are the fastest growing ones, they are not the populous ones in a population. The conclusions that binary-splitting reproductive strategies are the unique optimal strategies is similar to the previous conclusion under relative small  $m$ . There the binary-splitting reproductive strategies appeared when the cell interactions are both beneficial among cells [Gao et al., 2019]. But here we relax the constrain of the small intensity of selection.

A limitation we assumed in our model is that the organisms are formed in a clonal way, by considering their advantages of purging deleterious mutations and reducing conflicts among cells [Grosberg and Strathmann, 1998, 2007]. Nevertheless, the multicellularity could also form in an aggregative way, usually responding to the adverse environment [Claessen et al., 2014, Brunet and King, 2017]. In this scenario, the stickiness of cells and organisms could significantly impact reproductive strategies [Amado et al., 2018, Brunet and King, 2017, Staps et al., 2019]. Thus, a new mechanism needs to implement. Our model provides a framework for it in terms of reproductive strategies. Future work needs to address how and to what extent this stickiness factor could impact reproductive strategies in aggregative multicellularity.

Our model is flexible to extend to include more multicellular traits, as we deliberately simplified them to focus on reproductive strategies. We first simplified the number of cell types into two and then only considered the asexual life cycle of multicellularity. Multicellular organisms could comprise more cell types or sexual life cycles [Leu et al., 2020, Nishii and Miller, 2010]. These traits can be incorporated into our model as well. Our work in simple multicellularity may yield insights into the evolving mechanism of reproductive strategies in complex multicellularity. Since complex multicellularity formed through simple multicellular structures at beginning [Knoll, 2011, Nagy et al.,

2018]. Furthermore, although we investigated the reproductive strategies in multicellularity containing many cells, our model is general to depict the diverse reproductive strategies in unicellular organisms, which increase the sizes of their single cells rather than the cell number to reproduce offspring [Angert, 2005]. Our model implies that analogous to the size in multicellularity, the mass of single cells plays an important role in shaping the evolution of reproductive strategies.

## 4.5 Appendix

### 4.5.1 The probability distribution of offspring

We show the calculation of the probabilities of producing different types of offspring for an mature organism  $\{i, j\}$ , where  $i + j = N$ . The probability to produce the offspring type  $\{i', j'\}$  ( $i' + j' < N$ ) is calculated by

$$P_{\{i',j'\}} = \frac{C_i^{i'} C_j^{j'}}{C_N^{i'+j'}}. \quad (4.5)$$

We take the mature organism  $\{1,2\}$  in a population with reproductive strategy  $1+2$  as an example. There are five newborn organisms:  $\{1,0\}$ ,  $\{0,1\}$ ,  $\{2,0\}$ ,  $\{1,1\}$  and  $\{0,2\}$ . The probabilities to reproduce different offspring from the mature organism  $\{1,2\}$  are shown in Fig 4.5.

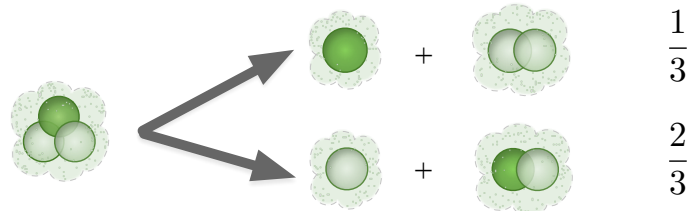


Figure 4.5: **The probabilities of producing different offspring for the mature organism  $\{1,2\}$  in the population reproductive strategy  $1+2$ .** The organism  $\{1,2\}$  has the probability of  $\frac{1}{3}$  to produce a newborn organism containing a defector and a newborn organism containing two cooperators. It has the probability of  $\frac{2}{3}$  to produce a newborn organism containing a cooperator and a newborn organism containing one cooperator and one defector.



### 4.5.2 Population growth rate

We illustrate the calculation of the growth rate in a population. For the reproductive strategy  $n_1 + n_2 + \dots + n_M$  with maturity size  $N$ , its population consists of the newborn organisms with sizes  $n_i$ , where  $0 \leq n_i < N$  and  $\sum_{i=1}^M n_i = N$ . As two cell types are considered, thus for each size  $n_i$  contains  $n_i + 1$  organisms with different numbers of cooperators. Thus, there are totally  $\sum_{i=1}^M (n_i + 1)$  different types of newborn organisms in a population with maturity size  $N$ . For example, the population with reproductive strategy  $1 + 2$  totally have 5 newborn organisms, see Fig 4.1D. The population growth rate depends on the growth of all newborn organisms with different cell composition. We assume that at the initial state, a population contains one organism of each newborn. We track the growth time and offspring number of each newborn organism. We use  $T_{ij}$  and  $N_{ij}$  to show that the growth time and the number of  $i$  type newborn organism reproduces  $j$  type offspring. The growth time  $T_{ij}$  depends on both the current size and the organism composition of defectors and cooperators by Eq 4.1. The offspring number  $N_{ij}$  depends on the cell-type switching probability and the cell division probabilities of each phenotype. Then the population growth rate can be calculated by the largest eigenvalue of the matrix

$$A = \sum a_{ij}, \quad (4.6)$$

where  $a_{ij} = N_{ji} e^{-\lambda T_{ji}}$  [De Roos, 2008, Gao et al., 2019]. We obtain the conclusions of the size effect analytically when maturity size is not greater than 3, see Appendix 4.5.3. We obtain the conclusion that only the binary-splitting reproductive strategies are the unique optimal analytically under size effects, see Appendix 4.5.4. Other cases are investigated by simulation. In our simulation, we simulate the growth trajectory of each organism in a population. In each trajectory, we record its growth time  $T_{ji}$  and its offspring  $N_{ji}$  of an organism with newborn type  $i$  reproducing the newborn organism type  $j$ . And then we run 5000 times to get the element  $a_{i,j} = \frac{\sum_z^{5000} N_{ji}^z e^{-\lambda T_{ji}^z}}{5000}$ , where  $T_{ji}^z$  and  $N_{ji}^z$  are the values in  $z$ -th single run.

### 4.5.3 Analytical prove of the general size effects on populations with maturity size $N \leq 3$

There are three reproductive strategies when maturity size  $N \leq 3$ :  $1 + 1$ ,  $1 + 1 + 1$  and  $1 + 2$ . We prove that under general cell division sequence  $[t_{s1}, t_{s2}]$ , the optimal population is determined by the beneficial size stages. The beneficial size stages have the relative small normalised values of  $\chi_n$ , see the explanation in main text. Thus, the reproductive strategy  $1 + 1$  is optimal when  $\chi_1 < \chi_2$  (size  $n = 1$  is the beneficial size stage) and the reproductive strategy  $1 + 2$  is the optimal when  $\chi_1 > \chi_2$  (size  $n = 2$  is the beneficial size stage). The population growth rates  $\lambda_{[1,1]}$ ,  $\lambda_{[1,1,1]}$ , and  $\lambda_{[1,2]}$  of  $1 + 1$ ,  $1 + 1 + 1$  and  $1 + 2$  respectively can be calculated through the matrix  $A$  in Appendix 4.5.2,

$$\lambda_{[1,1]} = \frac{\ln 2}{t_{s1}} \quad (4.7)$$

$$\lambda_{[1,1,1]} = \frac{\ln 3}{t_{s1} + t_{s2}} \quad (4.8)$$

$$e^{-\lambda_{[1,2]}(t_{s1}+t_{s2})} + e^{-\lambda_{[1,2]}t_{s2}} - 1 = 0. \quad (4.9)$$

$\lambda_{[1,2]}$  cannot be calculated directly, but it satisfies Eq 4.9. The population growth rate is positive, as there is no cell death in our model setting. As  $t_{sn}^0 = \ln \frac{n+1}{n}$ , thus  $[t_{s1}, t_{s2}] = [\chi_1 t_{s1}^0, \chi_2 t_{s2}^0] = [\chi_1 \ln 2, \chi_2 \ln \frac{3}{2}]$ , where  $\chi_1, \chi_2 > 0$ .

We first prove that the reproductive strategy  $1 + 1$  is the optimal one under the condition of  $\chi_1 < \chi_2$ , i.e.  $\chi_1 < \chi_2$ . Since  $\frac{\chi_1}{\chi_2} < 1$ , thus

$$\begin{aligned} \frac{\lambda_{[1,1]}}{\lambda_{[1,1,1]}} &= \frac{\frac{\ln 2}{\chi_1 \ln 2}}{\frac{\ln 3}{\chi_1 \ln 2 + \chi_2 \ln \frac{3}{2}}} \\ &= \frac{1}{\ln 3} \frac{\chi_1 \ln 2 + \chi_2 \ln \frac{3}{2}}{\chi_1} \\ &= \frac{1}{\ln 3} \left( \ln 2 + \frac{\chi_1}{\chi_2} \ln \frac{3}{2} \right) \end{aligned}$$

$$\begin{aligned}
 &> \frac{1}{\ln 3} \left( \ln 2 + \ln \frac{3}{2} \right) \\
 &= 1.
 \end{aligned}$$

Thus  $\lambda_{[1,1]} > \lambda_{[1,1,1]}$  for  $\chi_1 < \chi_2$ , indicating the reproductive strategy 1 + 1 performing better than the reproductive strategy 1 + 1 + 1. Next we prove that  $\lambda_{[1,1]} > \lambda_{[1,2]}$  for  $\chi_1 < \chi_2$ . If  $\lambda_{[1,2]} > \lambda_{[1,1]} = \frac{1}{\chi_1}$ , then

$$\begin{aligned}
 e^{-\lambda_{[1,2]}(t_{s1}+t_{s2})} + e^{-\lambda_{[1,2]}t_{s2}} - 1 &= e^{-\lambda_{[1,2]}(\chi_1 \ln 2 + \chi_2 \ln \frac{3}{2})} + e^{-\lambda_{[1,2]}\chi_2 \ln \frac{3}{2}} - 1 \\
 &< e^{-\ln 2 - \lambda_{[1,2]}\chi_2 \ln \frac{3}{2}} + e^{-\lambda_{[1,2]}\chi_2 \ln \frac{3}{2}} - 1 \\
 &= \frac{3}{2} e^{-\lambda_{[1,2]}\chi_2 \ln \frac{3}{2}} - 1 \\
 &= \frac{3}{2} \left( \frac{2}{3} \right)^{\lambda_{[1,2]}\chi_2} - 1.
 \end{aligned}$$

Because the left side is 0 due to Eq 4.9, we get

$$\begin{aligned}
 \frac{3}{2} \left( \frac{2}{3} \right)^{\lambda_{[1,2]}\chi_2} - 1 &> 0 \\
 \left( \frac{2}{3} \right)^{\lambda_{[1,2]}\chi_2} &> \frac{2}{3} \\
 \lambda_{[1,2]}\chi_2 &< 1 \\
 \lambda_{[1,2]} &< \frac{1}{\chi_2} < \frac{1}{\chi_1},
 \end{aligned}$$

which contradicts the assumption of  $\lambda_{[1,2]} > \lambda_{[1,1]} = \frac{1}{\chi_1}$ . Thus  $\lambda_{[1,1]} > \lambda_{[1,2]}$  for  $\chi_1 < \chi_2$ , which shows the reproductive strategy 1 + 1 is optimal under  $\chi_1 < \chi_2$ .

Then, we prove that the reproductive strategy 1 + 2 is the optimal one under the condition of  $\chi_1 > \chi_2$ , i.e.  $\chi_1 > \chi_2$ . Since  $\frac{\chi_2}{\chi_1} < 1$ , thus

$$\begin{aligned}
 \frac{\lambda_{[1,1,1]}}{\lambda_{[1,1]}} &= \frac{\frac{\ln 3}{\chi_1 \ln 2 + \chi_2 \ln \frac{3}{2}}}{\frac{\ln 2}{\chi_1 \ln 2}} \\
 &= \frac{\chi_1 \ln 3}{\chi_1 \ln 2 + \chi_2 \ln \frac{3}{2}}
 \end{aligned}$$

$$= \frac{\ln 3}{\ln 2 + \frac{\chi_2}{\chi_1} \ln \frac{3}{2}}$$

$$> 1.$$

Thus  $\lambda_{[1,1,1]} > \lambda_{[1,1]}$  for  $\chi_1 > \chi_2$ , indicating the reproductive strategy 1 + 1 + 1 performing better than the reproductive strategy 1 + 1. Next we prove  $\lambda_{[1,2]} > \lambda_{[1,1,1]}$  for  $\chi_1 > \chi_2$ . If  $\lambda_{[1,2]} < \lambda_{[1,1,1]} = \frac{\ln 3}{\chi_1 \ln 2 + \chi_2 \ln \frac{3}{2}}$ , then

$$e^{-\lambda_{[1,2]}(t_{s1} + t_{s2})} + e^{-\lambda_{[1,2]}t_{s2}} - 1 = e^{-\lambda_{[1,2]}(\chi_1 \ln 2 + \chi_2 \ln \frac{3}{2})} + e^{-\lambda_{[1,2]}\chi_2 \ln \frac{3}{2}} - 1$$

$$> e^{-\ln 3} + e^{-\lambda_{[1,2]}\chi_2 \ln \frac{3}{2}} - 1$$

$$= \left(\frac{2}{3}\right)^{\lambda_{[1,2]}\chi_2} - \frac{2}{3}.$$

Because the life side is 0 due to Eq 4.9, thus we get

$$\left(\frac{2}{3}\right)^{\lambda_{[1,2]}\chi_2} - \frac{2}{3} < 0$$

$$\lambda_{[1,2]}\chi_2 > 1$$

$$\lambda_{[1,2]} > \frac{1}{\chi_2}.$$

Since  $\frac{\chi_1}{\chi_2} > 1$ , thus

$$\frac{\frac{1}{\chi_2}}{\frac{\ln 3}{\chi_1 \ln 2 + \chi_2 \ln \frac{3}{2}}} = \frac{\chi_1 \ln 2 + \chi_2 \ln \frac{3}{2}}{\chi_2 \ln 3}$$

$$= \frac{1}{\ln 3} \left( \frac{\chi_1}{\chi_2} \ln 2 + \ln \frac{3}{2} \right)$$

$$> 1.$$

This result  $\frac{1}{\chi_2} > \frac{\chi_1 \ln 3}{\chi_1 \ln 2 + \chi_2 \ln \frac{3}{2}}$  contradicts the assumption  $\lambda_{[1,2]} < \frac{\ln 3}{\chi_1 \ln 2 + \chi_2 \ln \frac{3}{2}}$  and the conclusion  $\lambda_{[1,2]} > \frac{1}{\chi_2}$ . Thus the reproductive strategy 1 + 2 is the optimal one under the condition of  $\chi_1 > \chi_2$ .

Furthermore, our results of the size perturbation is a special case of this general

conclusion by set  $\chi_1 = 1$  or  $\chi_2 = 1$ . Under this condition, the reproductive strategy 1 + 1 is optimal when  $n = 1$  is under perturbation and the reproductive strategy 1 + 2 is optimal when  $n = 2$  is under perturbation. The results of size effects further show that the optimal reproductive strategy is the binary-splitting ones (1 + 1 or 1 + 2) instead of the multiple-splitting one (1 + 1 + 1). The multiple-splitting reproductive strategy is the optimal one when  $\chi_1 = \chi_2 = x^*$ , where the populations of reproductive strategies of 1 + 1, 1 + 1 + 1 and 1 + 2 possess the same growth rate  $\frac{1}{x^*}$ .

#### 4.5.4 Only the binary-splitting reproductive strategies can be the optimal one under size effects

For a reproductive strategy  $n_1 + n_2 + \dots + n_M$  with maturity size  $N = \sum_{i=1}^M n_i$ . Since  $n_i$  might equal to  $n_j$ , the types of newborn organisms could small than  $M$ . We suppose there are  $M'$  types of newborn organisms, where  $M' \leq M$ . As under the size effects, the cell types do not impact the newborn organisms. Thus, there are only size differences between newborn organisms. Therefore  $N_{ji} = N_i$  for  $j \in [1, M']$ . Equation Eq 4.6 can be written in Eq 4.10.

$$\begin{vmatrix} N_1 e^{-\lambda T_1} - 1 & N_1 e^{-\lambda T_2} & \dots & N_1 e^{-\lambda T_{M'}} \\ N_2 e^{-\lambda T_1} & N_2 e^{-\lambda T_2} - 1 & \dots & N_2 e^{-\lambda T_{M'}} \\ \vdots & \vdots & \ddots & \vdots \\ N_{M'} e^{-\lambda T_1} & N_{M'} e^{-\lambda T_2} & \dots & N_{M'} e^{-\lambda T_{M'}} - 1 \end{vmatrix} = 0. \quad (4.10)$$

Next, we simplify the determinant in the left size of Eq 4.10. In  $i$ th row, we add the values of the first row multiplied by  $\frac{N_i}{N_1}$ , where  $i \in [2, M']$ . We get

$$\begin{vmatrix} N_1 e^{-\lambda T_1} - 1 & N_1 e^{-\lambda T_2} & \dots & N_1 e^{-\lambda T_{M'}} \\ \frac{N_2}{N_1} & -1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \frac{N_{M'}}{N_1} & 0 & \dots & -1 \end{vmatrix} = 0. \quad (4.11)$$

## Chapter 4. Evolution of reproductive strategies in incipient multicellularity

---

Then we let the first column add  $i$ -th column multiplied by  $\frac{N_i}{N_1}$ , where  $i \in [2, M']$ . We get

$$\begin{vmatrix} \sum_{i=1}^{M'} N_i e^{-\lambda T_i} - 1 & N_1 e^{-\lambda T_2} & \dots & N_1 e^{-\lambda T_{M'}} \\ 0 & -1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & -1 \end{vmatrix} = 0. \quad (4.12)$$

Then we get

$$\sum_{i=1}^{M'} N_i e^{-\lambda T_i} - 1 = 0, \quad (4.13)$$

where  $i \in [1, M']$ . For the reproductive strategy  $n_1 + n_2 + \dots + n_M$  with  $N = \sum_{i=1}^M n_i$ , Eq 4.13 can be written in the following equation

$$\sum_{i=1}^M e^{-\lambda T_{n_i}} - 1 = 0, \quad (4.14)$$

Where  $T_{n_i}$  is the growth time for an organism from newborn size  $n_i$  to its maturity size  $N$ . Since  $N_i$  is the repetitive times of a newborn organism. For example,  $N_1 = 3$  for the reproductive strategy  $3 + 3$ , as there is only one type of newborn organisms.

To prove that only binary-spitting reproductive strategies are the optimal and unique ones, we use a similar method as that in [Pichugin and Traulsen, 2020]. We show three reproductive strategy  $S_1 = n_1 + n_2 + \dots + n_M$ ,  $S_2 = n_1 + n_2 + \dots + n_M$  and  $S_3 = n_1 + n_2$ , where  $N = \sum_{i=1}^M n_i$ . We use  $\lambda_1$ ,  $\lambda_2$ , and  $\lambda_3$  to denote the growth rates of  $S_1$ ,  $S_2$  and  $S_3$ , respectively. The growth rates can be calculated in the following equations

$$f_1(\lambda) = e^{-\lambda T_{(n_1, N)}} + e^{-\lambda T_{(n_2, N)}} + \sum_{i=3}^N e^{-\lambda T_{(n_i, N)}} - 1 = 0 \quad (4.15)$$

$$f_2(\lambda) = e^{-\lambda T_{(n_1 + n_2, N)}} + \sum_{i=3}^N e^{-\lambda T_{(n_i, N)}} - 1 = 0 \quad (4.16)$$

$$f_3(\lambda) = e^{-\lambda T_{(n_1, n_1 + n_2)}} + e^{-\lambda T_{(n_2, n_1 + n_2)}} - 1 = 0. \quad (4.17)$$

Since the growth time  $T$  is positive, thus the above equations are monotonically decreasing functions. We multiply Eq 4.17 by  $e^{-\lambda T_{(n_1+n_2, N)}}$ . Since  $T_{(x,y)} + T_{(y,z)} = T_{(x,z)}$ , we get

$$f'_3(\lambda) = e^{-\lambda T_{(n_1, N)}} + e^{-\lambda T_{(n_2, N)}} - e^{-\lambda T_{(n_1+n_2, N)}} = 0. \quad (4.18)$$

Thus,  $f_1(\lambda) = f_2(\lambda) + f'_3(\lambda) = 0$ . Hence, we have either  $\lambda_1 = \lambda_2 = \lambda_3$  or  $f_2(\lambda_1)f'_3(\lambda_1) < 0$  at  $\lambda_1$ . If  $f_2(\lambda_1) < 0$  and  $f'_3(\lambda_1) > 0$ , we get  $\lambda_2 < \lambda_1 < \lambda_3$ . If  $f_2(\lambda_1) > 0$  and  $f'_3(\lambda_1) < 0$ , we get  $\lambda_3 < \lambda_1 < \lambda_2$ . Thus, the unique optimal reproductive strategies are always the binary-splitting ones.

#### 4.5.5 Newborn organisms distribution at the stationary growth stage of populations

To understand the newborn organism future prospect on population growth at the population stationary stage. We made a full screen of newborn organism distribution for all populations across varying cell-type switching probability  $m$  and the contribution threshold  $k$ . The newborn organism distribution is calculated by the right eigenvector of the matrix  $A = \sum a_{ij}$  at the point with the largest eigenvalue [De Roos, 2008]. We found that the beneficial newborn organisms always dominate their population when cell-type switching probability is small, see Fig 4.5. The multiple beneficial newborn organisms jointly dominate their population, see the population 1 + 2 in Fig 4.5B. If there are no beneficial newborn organisms in a population, the intermediate beneficial newborn organisms will dominate, like the intermediate beneficial newborn organism  $\{0, 1\}$  in the population 1 + 1 + 1 under  $k = 2$ , see Appendix Fig 4.5 B. Otherwise, under the condition lacking both of them (when  $k \geq N$ ), then a population will be dominant by its neutral newborn organisms rather than adverse newborn organisms because of their payoffs, see 1 + 1 under  $k = 2$  in Fig 4.5 B. Due to the future prospect of the newborn organisms and the small type switching rate, in the subsequent investigation of the population growth rate, we mainly focus on the behaviour of the beneficial newborn organisms and then the intermediate beneficial newborn organisms.

## Chapter 4. Evolution of reproductive strategies in incipient multicellularity

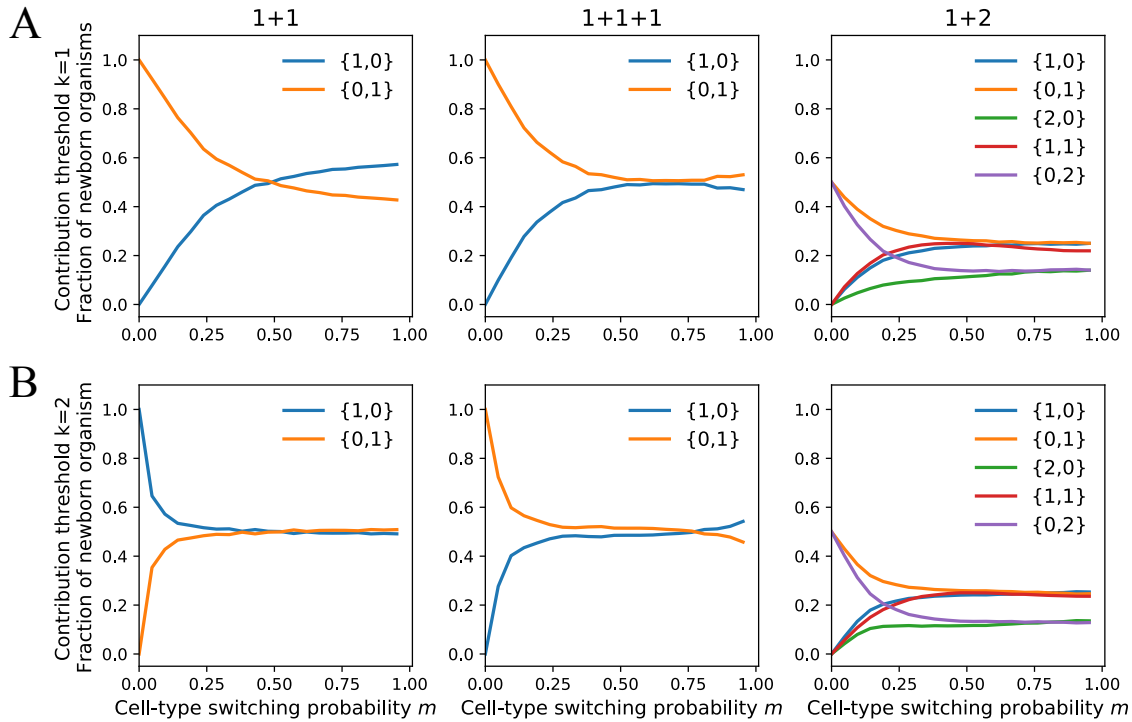


Figure 4.6: **Examples of the newborn organism distribution of populations across varying cell-type switching probability  $m$  and contribution threshold  $k$ .** **A.** Newborn organisms of populations with reproductive strategy of  $1+1$ ,  $1+1+1$  and  $1+2$  are investigated under  $m \in [0, 1]$  and  $k = 1$ . **B.** Newborn organisms of populations with reproductive strategy of  $1+1$ ,  $1+1+1$  and  $1+2$  are investigated under  $m \in [0, 1]$  and  $k = 2$ . Parameters  $w = 0.1$ ,  $m = 0.01$ ,  $b = 10$  and  $c = 1$ .



## Chapter 5

# Summary and outlook

---

### 5.1 Summary

In this thesis, I found that cellular interactions in hierarchical organisms can shape the characteristics of organisms in terms of formation, division of labour and reproductive mode. These three major questions were addressed in structured populations by different models and methods. Throughout this thesis, I assumed that selection acts on population growth. Population growth rates were calculated analytically (or numerically) by characteristic equations. In the first model (in Chapter 2) and the third model (in Chapter 4), the concepts of payoff matrices in evolutionary game theory were adopted to describe individual interactions. In Chapter 2, the model uses a general  $2 \times 2$  game form, while in Chapter 3, model uses a specific volunteer game. Furthermore, weak intensity of selection was assumed in the first model to obtain the analytical results. This constraint was relaxed in the third model where population growth rates were numerically calculated. The Monte Carlo method for sampling was used in the second model (in Chapter 3) and the third model (in Chapter 4) to capture the potential functional forms in high dimensional spaces.

In Chapter 2, I performed an extensive investigation of the competition of life cycles driven by cellular interactions in organisms. The results showed that multicellular life cycles can emerge under a wide range of parameters, see Fig 2.2 and Fig 2.4. Furthermore, among the huge variety of life cycles, only eight multicellular life cycles were found to be evolutionarily optimal. The vast majority of games promoted either of two very specific classes of life cycles: multiple fission (reproducing only single-celled offspring, such as  $1+\dots+1$ ) or binary-splitting (reproducing exactly two strictly multicellular daughter groups of the identical or almost identical size, such as  $2+2$  and  $2+3$ ).

## Chapter 5. Summary and outlook

---

I found that social dilemma games may not promote the evolution of life cycles with a single-cell bottleneck. Naive intuition would suggest that life cycles with unicellular offspring are likely to be favoured by all social dilemmas, as a single cell bottleneck is an effective way to police defectors. However, in our model, I found that social dilemmas may lead to the evolution of any of eight life cycles (Fig 2.4). This is due to the structure of the population and the cell-type switching probability, which allows a cell type to continuously produce another cell type. The cell-type switching probability increases the phenotypical heterogeneity in an organism, especially when the probability is around 0.5. The strategy of a cell is determined by the cell-type switching probability because each cell player's strategy was assigned via its phenotype.

In chapter 3, I investigated the conditions favouring an extreme altruism form of one cell type completely losing its reproductive ability. Two cell types were considered: somatic cell (representing cooperators) and germ cell (representing defectors). I sought the conditions under which somatic cells completely sacrifice their reproductive abilities, i.e. somatic cells only produce somatic cells. Our results showed that irreversible somatic differentiation emerges under the conditions of both somatic cells' benefits and cell differentiation costs in larger multicellular organisms. Somatic cells' benefits mean the somatic cells contribute to organism growth and the contributions of somatic cells is additive. Cell differentiation costs mean the decreasing in organism growth result from cell differentiation. Counterintuitively, only under the condition of somatic benefits, irreversible somatic differentiation cannot emerge. The underlying reason is the trade-off between the viability and fertility of an organism. When there is a benefit from somatic cells, an organism can exploit somatic cells gaining benefits (grow fast) by producing somatic cells. But at the late stage of organism development, the optimal strategy for an organism is to produce as many germ cells as possible. Due to this trade-off, the best strategy for an organism is to let germ cells produce somatic cells and let somatic cells produce germ cells. Thus, only under somatic cells' benefits, the optimal strategies are the reversible somatic differentiation rather than the irreversible ones.

In our results, the contribution threshold  $x_0$  is one of the parameters that determine somatic cells' benefits functions.  $x_0$  implies the proportion at which somatic contribution starting to contribute to organism growth. The  $x_0$  played an important role in the emergence of irreversible somatic differentiation. That is, most irreversible somatic differentiation emerges under the functions of somatic cells' benefits with a certain small contribution threshold  $x_0$ . In reality, species in genus *Volvox* exemplify the necessity for the contribution threshold  $x_0$ . The smaller *Gonium* consists of identical cells, but without division of labour, whereas the slightly larger *Pleodorina* shows a partial division of labour and the larger *Volvox carteri* exhibits a complete germ-soma division of labour in reproduction [Matt and Umen, 2016], revealing an underlying threshold for the emergence of division of labour in terms of somatic cells' proportions. The somatic cells' benefits functions with four features enhance the chance to capture the potential diverse somatic contribution functions in reality, which suggests the phenomena of the rarity of irreversible somatic differentiation [Rodrigues et al., 2012] [Cooper and West, 2018].

In chapter 4, I investigated organism size effects and threshold effects on reproductive strategies of multicellularity based on the first model (in Chapter 2). Our results show that reproductive strategies can co-exist under different conditions. This conclusion coincides with the phenomenon that one species can possess several reproductive strategies simultaneously in nature, such as cyanobacteria, which has the reproductive strategies of binary fission, budding and multiple fission [Angert, 2005, Flores and Herero, 2010]. Moreover, our results show that the bottleneck reproductive strategy that repeatedly reproduces one offspring can be optimal and unique under either condition. This conclusion indicates the new advantages for the bottleneck reproductive strategies other than the previously stated advantages of decreasing the mutation load and regulating cell conflicts [Grosberg and Strathmann, 1998, Michod and Roze, 1999].

Our results also show that only the binary-splitting reproductive strategies can be optimal and unique, see Fig 4.2, Fig 4.3, and Fig 4.4. Among them, the bottleneck reproductive strategy (producing a single-celled offspring repeatedly) is commonly adop-

ted for complex species, such as plants and human. Other binary-splitting reproductive strategies are not frequently observed in nature. Besides, under the size effects, the reproductive strategy  $[1, 1]$  is the most frequently observed strategy among binary-splitting ones. This indicates reproductive strategy  $[1, 1]$  is the best strategy under the environment with unpredicted size effects.

Finally, I chose a fixed cell-type switching probability  $m$ , where  $m = 0.01$  in our model. This assumption increases the degree of heterogeneity in organisms compared with homogeneous organisms. And it also increases the cellular interactions between cells with different types to a certain extent. Therefore, a population is consisting of organisms with different cell composition. As the organisms with only cooperators can grow faster if their cooperator numbers meet a given threshold, they are more abundant in a population. Otherwise, organisms with only defectors are more abundant in a population. The results of optimal reproductive strategies are robust to the cell-type switching probability, especially for large organisms.

## 5.2 Open questions

In the models of this thesis, I considered organisms growing in rich environments, which led to populations growing exponentially. While in reality, we know populations will be limited by resources and they will at most reach carrying capacities. Vanessa Ress has investigated the evolution of different reproductive strategies, where populations compete for resources. In an environment with insufficient nutrition, logistic growth is more natural to describe population growth. Therefore, a question arises: Can we get the same conclusions by considering finite populations? Meanwhile, what kind of methods should we adopt? New approaches that can evaluate the performance of reproductive strategies and conditions for ISD (irreversible somatic differentiation) are needed. In the following two paragraphs, I introduce two possible methods for investigating the reproductive strategies of organisms. Similar methods can be used to investigate ISD,

as long as one replaces reproductive strategies (in Chapter 2 and 4) by developmental strategies (in Chapter 3).

Let us consider a finite population consisting of organisms, which may have different reproductive strategies. One idea for finite population models is introducing internal competition between organisms with different reproductive strategies. Lotka-Volterra equations could be an ideal mathematical method to describe the competition. We assume the population only contains few newborns of each reproductive strategy initially (see newborn types in Chapter 4). The birth rate of each organism should be constrained by the population carrying capacity. Meanwhile, the birth rate of each organism is proportional to its fitness. The fitness of an organism can be defined by its exponential population growth rate which is calculated in the density-independent models (in Chapter 2 and 4). The death rate of each organism is constant for each organism. Finally, we can investigate the proportions of different reproductive strategies among organisms at a strategy stable state. The reproductive strategy with the highest proportion is the optimal one.

Another idea is using a Moran process to describe a finite population. We can estimate the performance of two reproductive strategies by their fixation probabilities [Traulsen and Nowak, 2006]. Fixation probability of a strategy is the probability that a single organism with this strategy invades a population, where organisms share a resident strategy. In a pairwise comparison, one reproductive strategy is better when it has a larger fixation probability. Finally, we can compare all reproductive strategies through the pairwise comparison. But one should note that in our model one reproductive strategy are related to several different newborn organisms, thus one should handle the death process carefully to keep the population size constant.

In the models of this thesis, I considered cells interacting in a well-mixed organism, where there are no spatial structures. Cells can directly interact with each other in an organism without taking their physical distances into account. In nature, organisms are shaped in various forms, such as filament, planes or spheres. Different shapes constrain

## Chapter 5. Summary and outlook

---

the direct interactions for cells physically located far away from each other. Instead of interactions among all cells, only the neighbour cells can directly interact with each other. Spatial structures could also play a crucial role in organisms [Nowak, 2006b, Libby et al., 2014, Yanni et al., 2020], and it remains an open question to know how it can impact cellular interactions. To address this concern, a model including the interaction ranges among cells in an organism is required. I will next provide some thoughts for implementing such a model.

Yanni et al. [2020] proposed graph structures to illustrate cellular interactions in an organism. In a graph, nodes represent the cells of an organism and links connect two interacting cells. Similarly, we can use the same method to investigate the effects of different spatial structures on reproductive strategies by including spatial structures in organisms. We can set a fixed population structure for all organisms and keep all other assumptions unchanged in our models. Then, we can investigate the best population structure for a reproductive strategy of interest (for Chapter 2 and Chapter 4) or for ISD (for Chapter 3). To find the optimal organism structure for a reproductive strategy or ISD, the structure can change the number of links, the pattern of links and the cellular interaction strength (link strength).

For the reproduction of multicellular organisms, I assumed that cells in a mature organism randomly fragment to form offspring. Since only cooperators benefit an organism, it is reasonable to assume that the fragmentation relies on a specific mechanism. For example, cooperators may produce some materials into their environment to stick to other cells, thus cooperators are more likely to stay together than defectors [Rainey and Rainey, 2003b]. To model organisms with such reproduction preferences, a new description is needed. I will propose an approach for a scenario including a fragmentation preference in the next paragraph.

We consider a situation where cooperators can release a sticky material that keeps cells staying together after organism fragmentation. An organism containing more cooperators is thus likely to reproduce larger organisms. It seems reasonable to set a public

goods game for cooperators and defectors, where cooperators pay a cost but defectors don't. Under these assumptions, our classification of reproductive strategies becomes invalid, because the organisms in a population may not follow a fixed reproductive strategy any more. Nevertheless, we can borrow the idea from previous work to adopt mixed reproductive strategies for organisms [Pichugin et al., 2017]. Initial organisms have all potential fragmentation modes, and we track the dynamics of the population based on parameters of cell stickiness and different cellular interactions forms. Finally, we can obtain conclusions by observing the fractions of existing reproductive strategies at their stable state. A similar method can be used to investigate ISD.

In the models of this thesis, I investigated clonal organisms by considering their advantages of purging deleterious mutations and reducing conflicts among cells [Grosberg and Strathmann, 1998, Michod and Roze, 1999]. Nevertheless, in nature, multicellular organisms could also form by aggregation, usually responding to adverse environments [Claessen et al., 2014, Brunet and King, 2017]. The role of aggregation promotes cooperation has been investigated theoretically [Tarnita et al., 2013, Garcia et al., 2015]. In the models with cell aggregation, the stickiness of cells in organisms could significantly impact reproductive strategies [Amado et al., 2018, Brunet and King, 2017, Staps et al., 2019]. Thus, a new mechanism needs to be introduced. Our model provides a framework for classifying reproductive strategies. Future work needs to address how and to what extent the stickiness of cells could impact reproductive strategies in aggregative multicellularity. In the next paragraph, I will discuss a model considering aggregative organisms.

In a population, when organisms can grow through aggregation, the generation time of organisms will be heavily impacted. An organism can reach its mature size quickly by aggregating other large organisms. In organisms with two cell types, it is natural to consider different stickiness abilities. The stickiness of an organism can be defined as the average one across cells. Or it can be defined as other forms, such as a function of cell composition. The reproductive strategies of organisms and the other settings are the same as in our models (see Chapter 2 and Chapter 4). Finally, we can investigate the ef-

## Chapter 5. Summary and outlook

---

fects of stickiness on reproductive strategies by observing their fractions at strategy stable state. I expect the results will be the same for ISD under the design of aggregative organisms because only defectors (germ cells) can be aggregated during growth. Therefore, all developmental strategies are affected at the same time, thus the relative performance of ISD strategies is likely to be unchanged.

Since I focused on the reproductive strategies of organisms in this thesis, I deliberately simplified other traits, such as multiple cell types and sexual life cycles [Nishii and Miller, 2010]. Future models can incorporate more cell types or sexual life cycles. Our results are still attainable if there is no gene recombination happening during gamete fusion, but conclusions are unknown if gene recombination is involved. Then we ask: Would the conclusions in our models still hold when considering organisms with sexual life cycles? Furthermore, we could ask under which environmental conditions sexual life cycles can outperform asexual life cycles. In the next paragraph, I discuss the potential outcomes by considering organisms with multiple cell types and sexual life cycles.

The scenario is more complicated for organisms with more than two cell types, as cell types are corresponding to cellular interaction strategies. There are total  $n$  strategies in a population with  $n$  cell types, where  $n > 2$ . Multiplayer games with  $n$  strategies could be an approach to construct such a population [Gokhale and Traulsen, 2014]. Cellular interactions among  $n$  cell types can be captured by a payoff matrix. The size of the payoff matrix depends on  $n$  and the cell number of an organism. The parameters in the payoff matrix will lead to a high dimensional model, which is highly likely to yield diverse outcomes in terms of the optimal reproductive strategy. For organisms with sexual life cycles, not all reproductive strategies in our model are valid. As sexual life cycles start from the fusion of two gametes, the offspring of organisms can only be single cells. Under this condition, we can only investigate the reproductive strategies of binary fission and multiple fission (see Chapter 2 and Chapter 4). One also needs to design a mechanism to determine the sex of offspring of an organism. The mechanism could include some ecological factors based on the experimental observation [Nishii and Miller, 2010].



For example, under nitrogen starvation, *Volvox carteri* reproduces equal-sized gametes with different types. By doing so, one can investigate the effects of sexual life cycles on reproductive strategies. A similar method can be used in investigating ISD.

Overall, this work has yielded insights into the mathematical mechanism of the effects of cell interactions on the competitiveness of organisms with different reproductive strategies. In this work, I have studied cell interactions based on game-theoretic methods and populations based on stage-structured methods. In future work, I plan to extend my models with density-dependence, spatial structures and complex life cycles.

# Bibliography

---

A. Amado, C. Batista, and P. R. Campos.

A mechanistic model for the evolution of multicellularity.

Physica A: Statistical Mechanics and its Applications, 492:1543–1554, 2018.

(Cited on pages [67](#), [81](#), [110](#), and [127](#).)

E. R. Angert.

Alternatives to binary fission in bacteria.

Nature Reviews Microbiology, 3(3):214–224, 2005.

(Cited on pages [92](#), [109](#), [111](#), and [123](#).)

M. Archetti.

The volunteer’s dilemma and the optimal size of a social group.

Journal of Theoretical Biology, 261(3):475–480, 2009.

(Cited on page [109](#).)

D. Arendt.

The evolution of cell types in animals: emerging principles from molecular studies.

Nature Reviews Genetics, 9(11):868, 2008.

(Cited on page [92](#).)

R. Axelrod and W. D. Hamilton.

The evolution of cooperation.

Science, 211:1390–1396, 1981.

(Cited on page [11](#).)

D. Bakthavatsalam and R. Gomer.

The secreted proteome profile of developing dictyostelium discoideum cells.

Proteomics, 10(13):2556 – 2559, 2010.

(Cited on page [80](#).)

J. Bonner.

The Cellular Slime Molds.

Princeton University Press, Princeton, NJ, 1959.

(Cited on page [20](#).)

J. Bonner.

The origins of multicellularity.

Integrative Biology, 1:27–36, 1998.

(Cited on pages [1](#) and [92](#).)

H. Brandt, C. Hauert, and K. Sigmund.

Punishing and abstaining for public goods.

Proceedings of the National Academy of Sciences USA, 103(2):495–497, 2006.

(Cited on page [38](#).)

M. Broom and J. Rychtář.

Game-Theoretical Models in Biology.

Chapman and Hall/CRC, 2013.

(Cited on page [35](#).)

T. Brunet and N. King.

The origin of animal multicellularity and cell differentiation.

Developmental cell, 43(2):124–140, 2017.

(Cited on pages [1](#), [92](#), [110](#), and [127](#).)

S. B. Carroll.

Chance and necessity: the evolution of morphological complexity and diversity.

Nature, 409(6823):1102, 2001.

(Cited on pages [1](#), [81](#), and [93](#).)

H. Caswell.

Matrix population models.

Sinauer Associates, Sunderland MA, 2nd edition, 2001.

## Bibliography

---

(Cited on page 5.)

- D. Claessen, D. E. Rozen, O. P. Kuipers, L. Sogaard-Andersen, and G. P. van Wezel.  
Bacterial solutions to multicellularity: a tale of biofilms, filaments and fruiting bodies.  
Nat Rev Micro, 12(2):115–124, 2014.  
(Cited on pages 1, 21, 92, 109, 110, and 127.)

- G. Cooper and S. A. West.  
Division of labour and the evolution of extreme specialization.  
Nature ecology & evolution, 2018.  
(Cited on pages 21, 67, 80, and 123.)

- M. Czarna, G. Mathy, A. MacCord, R. Dobson, W. Jarmuszkiewicz, C. Sluse-Goffart, P. Leprince, E. De Pauw, and F. Sluse.  
Dynamics of the dictyostelium discoideum mitochondrial proteome during vegetative growth, starvation and early stages of development.  
Proteomics, 10(1):6 – 22, 2010.  
(Cited on page 80.)

- S. De Monte and P. B. Rainey.  
Nascent multicellular life and the emergence of individuality.  
Journal of biosciences, 39(2):237 – 248, 2014.  
(Cited on page 20.)

- A. M. De Roos.  
Demographic analysis of continuous-time life-history models.  
Ecology Letters, 11(1):1–15, 2008.  
(Cited on pages 9, 26, 97, 113, and 119.)

- A. Diekmann.  
Volunteer's dilemma.  
Journal of conflict resolution, 29(4):605–610, 1985.  
(Cited on page 12.)

- T. DuBuc, C. Schnitzler, E. Chrysostomou, E. McMahon, Febrimarsa, J. Gahan, T. Buggie, S. Gornik, S. Hanley, S. Barriera, P. Gonzalez, A. Baxevanis, and U. Frank.  
Transcription factor ap2 controls cnidarian germ cell induction.  
Science, 367(6479):757–762, 2020.  
(Cited on page [67](#).)
- L. A. Dugatkin and H. K. Reeve.  
Game theory and animal behavior.  
Oxford University Press on Demand, 2000.  
(Cited on page [109](#).)
- M. R. Easterling, S. P. Ellner, and P. M. Dixon.  
Size-specific sensitivity: applying a new structured population model.  
Ecology, 81(3):694–708, 2000.  
(Cited on pages [7](#) and [8](#).)
- S. P. Ellner and M. Rees.  
Integral projection models for species with complex demography.  
The American Naturalist, 167(3):410–428, 2006.  
(Cited on page [8](#).)
- E. Erten and H. Kokko.  
From zygote to a multicellular soma: body size affects optimal growth strategies under cancer risk.  
Evolutionary Applications, 13(7):1593 – 1604, 2020.  
(Cited on pages [67](#) and [81](#).)
- S. Ficici and J. Pollack.  
Effects of finite populations on evolutionary stable strategies.  
In D. Whitley, D. Goldberg, E. Cantu-Paz, L. Spector, I. Parmee, and H.-G. Beyer, editors, Proceedings GECCO, pages 927–934, Morgan-Kaufmann, San Francisco, 2000.  
(Cited on page [13](#).)

## Bibliography

---

R. Fisher, T. Bell, and S. A. West.

Multicellular group formation in response to predators in the alga *Chlorella vulgaris*.

Journal of evolutionary biology, 29(3):551–559, 2016.

(Cited on pages [81](#) and [93](#).)

E. Flores and A. Herrero.

Compartmentalized function through cell differentiation in filamentous cyanobacteria.

Nature Reviews Microbiology, 8(1):39, 2010.

(Cited on pages [21](#), [92](#), [94](#), [109](#), and [123](#).)

J. H. Fowler.

Altruistic punishment and the origin of cooperation.

Proceedings of the National Academy of Sciences USA, 102(19):7047–7049, 2005.

(Cited on page [38](#).)

J. Gallon.

Tansley review no. 44. reconciling the incompatible: N<sub>2</sub> fixation and O<sub>2</sub>.

New Phytologist, pages 571–609, 1992.

(Cited on pages [70](#) and [109](#).)

Y. Gao, A. Traulsen, and Y. Pichugin.

Interacting cells driving the evolution of multicellular life cycles.

PLoS Computational Biology, 15(5):e1006987, 2019.

(Cited on pages [18](#), [70](#), [83](#), [92](#), [93](#), [94](#), [97](#), [108](#), [110](#), and [113](#).)

J. García and A. Traulsen.

Leaving the loners alone: Evolution of cooperation in the presence of antisocial punishment.

Journal of Theoretical Biology, 307:168–173, 2012.

(Cited on page [38](#).)

T. Garcia, L. G. Brunnet, and S. De Monte.

Differential adhesion between moving particles as a mechanism for the evolution of social groups.

PLoS Comput Biol, 10(2):e1003482, 02 2014.

(Cited on page [38](#).)

T. Garcia, G. Doulcier, and S. De Monte.

The evolution of adhesiveness as a social adaptation.

eLife, 4:e08595, 2015.

doi: 10.7554/eLife.08595.

(Cited on pages [38](#) and [127](#).)

A. Gardner.

The Price equation.

Current Biology, 18:R198–R202, 2008.

(Cited on page [2](#).)

S. Gavrillets.

Rapid transition towards the division of labor via evolution of developmental plasticity.

PLoS Computational Biology, 6(6):e1000805, 2010.

(Cited on pages [21](#) and [67](#).)

P. Godfrey-Smith.

Darwinian Populations and Natural Selection.

Oxford University Press, 2009.

(Cited on page [19](#).)

C. S. Gokhale and A. Traulsen.

Evolutionary multiplayer games.

Dynamic Games and Applications, 4:468–488, 2014.

(Cited on pages [12](#) and [128](#).)

H. J. Goldsby, A. Dornhaus, B. Kerr, and C. Ofria.

## Bibliography

---

Task-switching costs promote the evolution of division of labor and shifts in individuality.

Proceedings of the National Academy of Sciences, 109(34):13686–13691, 2012.

(Cited on page 67.)

H. J. Goldsby, D. B. Knoester, B. Kerr, and C. Ofria.

The effect of conflicting pressures on the evolution of division of labor.

PloS one, 9(8):e102713, 2014.

(Cited on page 67.)

R. K. Grosberg and R. R. Strathmann.

One cell, two cell, red cell, blue cell: the persistence of a unicellular stage in multicellular life histories.

Trends in ecology & evolution, 13(3):112 – 116, 1998.

(Cited on pages 70, 92, 93, 108, 110, 123, and 127.)

R. K. Grosberg and R. R. Strathmann.

The evolution of multicellularity: A minor major transition?

Annual Review of Ecology, Evolution, and Systematics, 38:621–654, 2007.

(Cited on pages 1, 19, 66, 70, 92, and 110.)

A. Hallmann.

Evolution of reproductive development in the volvocine algae.

sexual plant reproduction, 24(2):97 – 112, 2011.

(Cited on page 81.)

W. Hamilton.

Extraordinary sex ratios.

Science, 156:477–488, 1967.

(Cited on page 12.)

K. Hammerschmidt, C. J. Rose, B. Kerr, and P. B. Rainey.

Life cycles, fitness decoupling and the evolution of multicellularity.



Nature, 515(7525):75–79, 2014.

(Cited on page 20.)

G. Hardin.

The tragedy of the commons.

Science, 162:1243–1248, 1968.

(Cited on page 20.)

C. Hauert, S. De Monte, J. Hofbauer, and K. Sigmund.

Volunteering as red queen mechanism for cooperation in public goods games.

Science, 296:1129–1132, 2002.

(Cited on page 38.)

C. Hauert, A. Traulsen, H. Brandt, M. A. Nowak, and K. Sigmund.

Via freedom to coercion: the emergence of costly punishment.

Science, 316:1905–1907, 2007.

(Cited on page 38.)

C. Hilbe, M. A. Nowak, and K. Sigmund.

The evolution of extortion in iterated prisoner’s dilemma games.

Proceedings of the National Academy of Sciences USA, 110:6913–6918, 2013.

(Cited on page 34.)

J. Hofbauer and K. Sigmund.

Evolutionary Games and Population Dynamics.

Cambridge University Press, Cambridge, UK, 1998.

(Cited on pages 13 and 35.)

I. Ispolatov, M. Ackermann, and M. Doebeli.

Division of labour and the evolution of multicellularity.

Proceedings of the Royal Society of London B: Biological Sciences, 279(1734):1768–1776,  
2012.

(Cited on pages 21 and 67.)

## Bibliography

---

D. Kaiser.

Building a multicellular organism.

Annual Review of Genetics, 35(1):103–123, 2001.

(Cited on pages [93](#) and [94](#).)

S. E. Kapsetaki and S. A. West.

The costs and benefits of multicellular group formation in algae.

Evolution, 2019.

(Cited on pages [81](#) and [93](#).)

K. Kaveh, C. Veller, and M. A. Nowak.

Games of multicellularity.

Journal of Theoretical Biology, 403:143 – 158, 2016.

(Cited on pages [21](#) and [38](#).)

K. Kikuchi.

Cellular differentiation in pleodorina californica.

Cytologia, 43(1):153 – 160, 1978.

(Cited on pages [79](#) and [81](#).)

D. L. Kirk.

Germ–soma differentiation in volvox.

Developmental biology, 238(2):213–223, 2001.

(Cited on page [96](#).)

D. L. Kirk.

A twelve step program for evolving multicellularity and a division of labor.

BioEssays, 27(3):299–310, 2005.

(Cited on pages [66](#), [82](#), and [96](#).)

A. H. Knoll.

The early evolution of eukaryotes: a geological perspective.

Science, 256(5057):622–627, 1992.

(Cited on page [20](#).)

A. H. Knoll.

The multiple origins of complex multicellularity.

Annual Review of Earth and Planetary Sciences, 39:217–239, 2011.

(Cited on page [110](#).)

V. Koufopanou.

The evolution of soma in the volvocales.

The American Naturalist, 143(5):907 – 931, 1994.

(Cited on page [81](#).)

D. P. Kroese, T. Brereton, T. Taimre, and Z. I. Botev.

Why the monte carlo method is so important today.

Wiley Interdisciplinary Reviews: Computational Statistics, 6(6):386–392, 2014.

(Cited on page [14](#).)

K. Kumar, R. A. Mella-Herrera, and J. W. Golden.

Cyanobacterial heterocysts.

Cold Spring Harbor perspectives in biology, 2(4):a000315, 2010.

(Cited on page [94](#).)

R. Lanfear.

Do plants have a segregated germline?

PLoS biology, 16(5):e2005439, 2018.

(Cited on page [67](#).)

J.-Y. Leu, S.-L. Chang, J.-C. Chao, L. C. Woods, and M. J. McDonald.

Sex alters molecular evolution in diploid experimental populations of *s. cerevisiae*.

Nature ecology and evolution, 4(3):453–460, 2020.

(Cited on page [110](#).)

M. Li, W. Zhu, X. Dai, M. Xiao, G. Appiah-Sefah, and P. N. Nkrumah.

## Bibliography

---

Size-dependent growth of microcystis colonies in a shallow, hypertrophic lake: use of the rna-to-total organic carbon ratio.

Aquatic ecology, 48(2):207–217, 2014.

(Cited on pages [93](#) and [109](#).)

Y. Li and K. Gao.

Photosynthetic physiology and growth as a function of colony size in the cyanobacterium *Nostoc sphaeroides*.

European Journal of Phycology, 39(1):9–15, 2004.

(Cited on pages [93](#) and [109](#).)

E. Libby, W. C. Ratcliff, M. Travisano, and B. Kerr.

Geometry shapes evolution of early multicellularity.

PLoS Computational Biology, 10(9):e1003803, 2014.

(Cited on pages [20](#), [93](#), and [126](#).)

H. Marchant.

Colony formation and inversion in the green alga *Eudorina elegans*.

Protoplasma, 93(2-3):325 – 339, 1977.

(Cited on page [81](#).)

G. Matt and J. Umen.

*Volvox*: A simple algal model for embryogenesis, morphogenesis and cellular differentiation.

Developmental biology, 419(1):99–113, 2016.

(Cited on pages [71](#), [81](#), [93](#), [94](#), [96](#), and [123](#).)

J. Maynard Smith.

Evolution and the Theory of Games.

Cambridge University Press, Cambridge, 1982.

(Cited on pages [12](#) and [35](#).)

J. Maynard Smith and G. R. Price.

The logic of animal conflict.

Nature, 246:15–18, 1973.

(Cited on pages 10 and 109.)

J. Maynard Smith and E. Szathmáry.

The major transitions in evolution.

W. H. Freeman, Oxford, 1995.

(Cited on page 1.)

M. C. McCarthy and B. J. Enquist.

Organismal size, metabolism and the evolution of complexity in metazoans.

Evolutionary Ecology Research, 7(5):681–696, 2005.

(Cited on page 92.)

J. M. McNamara and O. Leimar.

Game theory in biology: concepts and frontiers.

Oxford University Press, USA, 2020.

(Cited on page 109.)

R. E. Michod.

Cooperation and conflict in the evolution of individuality. i. multilevel selection of the organism.

The American Naturalist, 149:607–645, 1997.

(Cited on page 21.)

R. E. Michod.

Evolution of individuality during the transition from unicellular to multicellular life.

Proceedings of the National Academy of Sciences, 104(suppl 1):8613–8618, 2007.

(Cited on pages 67, 80, and 93.)

R. E. Michod and D. Roze.

Cooperation and conflict in the evolution of individuality iii.

## Bibliography

---

In C. Nehaniv, editor, Mathematical and Computational Biology: Computational Morphogenesis, Hierarchical Complexity, and Digital Evolution, volume 26, pages 47 – 92. American Mathematical Society, 1999.

(Cited on pages [2](#), [92](#), [93](#), [108](#), [123](#), and [127](#).)

R. E. Michod and D. Roze.

Some aspects of reproductive mode and origin of multicellularity.

Selection, 1(1-3):97 – 110, 2001.

(Cited on page [21](#).)

K. Mikhailov, A. Konstantinova, M. Nikitin, P. Troshin, L. Rusin, V. Lyubetsky, Y. Panchin, A. Mylnikov, L. Moroz, S. Kumar, and V. Aleoshin.

The origin of metazoa: a transition from temporal to spatial cell differentiation.

Bioessays, 31(7):758 – 768, 2009.

(Cited on page [66](#).)

J. Morand-Ferron and J. L. Quinn.

Larger groups of passerines are more efficient problem solvers in the wild.

Proceedings of the National Academy of Sciences, 108(38):15898–15903, 2011.

(Cited on page [81](#).)

L. G. Nagy, G. M. Kovács, and K. Krizsán.

Complex multicellularity in fungi: evolutionary convergence, single origin, or both?

Biological Reviews, 93(4):1778–1794, 2018.

(Cited on page [110](#).)

S. L. Nielsen.

Size-dependent growth rates in eukaryotic and prokaryotic algae exemplified by green algae and cyanobacteria: comparisons between unicells and colonial growth forms.

Journal of plankton research, 28(5):489–498, 2006.

(Cited on pages [93](#) and [109](#).)

I. Nishii and S. M. Miller.

Volvox: simple steps to developmental complexity?  
Current opinion in plant biology, 13(6):646–653, 2010.  
(Cited on pages [110](#) and [128](#).)

M. A. Nowak.

Five rules for the evolution of cooperation.  
Science, 314:1560–1563, 2006a.  
(Cited on pages [2](#), [3](#), [21](#), and [34](#).)

M. A. Nowak.

Evolutionary dynamics: Exploring the equations of life.  
Harvard University Press, 2006b.  
(Cited on pages [35](#), [109](#), and [126](#).)

M. A. Nowak and K. Sigmund.

Evolutionary dynamics of biological games.  
Science, 303:793–799, 2004.  
(Cited on pages [11](#), [13](#), [35](#), and [109](#).)

M. A. Nowak, A. Sasaki, C. Taylor, and D. Fudenberg.

Emergence of cooperation and evolutionary stability in finite populations.  
Nature, 428:646–650, 2004.  
(Cited on pages [3](#) and [13](#).)

S. Okasha.

Evolution and the levels of selection.  
Oxford Univ. Press, 2006.  
(Cited on pages [2](#) and [19](#).)

S. Y. Ow, T. Cardona, A. Taton, A. Magnuson, P. Lindblad, K. Stensjö, and P. C. Wright.

Quantitative shotgun proteomics of enriched heterocysts from nostoc sp. pcc 7120 using 8-plex isobaric peptide tags.  
Journal of Proteome Research, 7(4):1615 – 1628, 2008.

## Bibliography

---

(Cited on page [80](#).)

J. M. Pacheco, F. C. Santos, M. O. Souza, and B. Skyrms.

Evolutionary dynamics of collective action in n-person stag hunt dilemmas.

Proceedings of the Royal Society B, 276:315–321, 2009.

(Cited on page [34](#).)

Y. Pichugin and A. Traulsen.

Evolution of multicellular life cycles under costly fragmentation.

PLOS Computational Biology, 16(11):e1008406, 2020.

(Cited on pages [38](#), [108](#), and [118](#).)

Y. Pichugin, J. Peña, P. Rainey, and A. Traulsen.

Fragmentation modes and the evolution of life cycles.

PLoS Computational Biology, 13(11):e1005860, 2017.

(Cited on pages [3](#), [20](#), [21](#), [24](#), [25](#), [27](#), [38](#), [52](#), [70](#), and [127](#).)

Y. Pichugin, H. Park, and A. Traulsen.

Evolution of simple multicellular life cycles in dynamic environments.

Journal of the Royal Society Interface, 16:154, 2019.

(Cited on pages [3](#), [92](#), [93](#), and [109](#).)

G. R. Price.

Extension of covariance selection mathematics.

Annals of human genetics, 35(4):485–490, 1972.

(Cited on page [2](#).)

P. B. Rainey and B. Kerr.

Cheats as first propagules: a new hypothesis for the evolution of individuality during the transition from single cells to multicellularity.

BioEssays, 32(10):872 – 880, 2010.

(Cited on pages [20](#) and [37](#).)



P. B. Rainey and K. Rainey.

Evolution of cooperation and conflict in experimental bacterial populations.

Nature, 425(6953):72–74, 2003a.

(Cited on page [20](#).)

P. B. Rainey and K. Rainey.

Evolution of cooperation and conflict in experimental bacterial populations.

Nature, 425(6953):72, 2003b.

(Cited on page [126](#).)

J. Ramaley.

Buffon’s noodle problem.

The American Mathematical Monthly, 76(8):916–918, 1969.

(Cited on page [14](#).)

A. Rashidi, D. E. Shelton, and R. E. Michod.

A darwinian approach to the origin of life cycles with group properties.

Theoretical Population Biology, 102:76 – 84, 2015.

(Cited on page [20](#).)

W. C. Ratcliff, R. F. Denison, M. Borrello, and M. Travisano.

Experimental evolution of multicellularity.

Proceedings of the National Academy of Sciences USA, 109(5):1595–1600, Jan 2012.

(Cited on pages [20](#), [39](#), [81](#), [92](#), [93](#), and [94](#).)

W. C. Ratcliff, M. D. Herron, K. Howell, J. T. Pentz, F. Rosenzweig, and M. Travisano.

Experimental evolution of an altering uni- and multicellular life cycle in *chlamydomonas reinhardtii*.

Nature Communications, 4(2742), November 2013a.

(Cited on page [20](#).)

W. C. Ratcliff, J. T. Pentz, and M. Travisano.

## Bibliography

---

Tempo and mode of multicellular adaptation in experimentally evolved *Saccharomyces cerevisiae*.

Evolution, 67(6):1573–1581, 2013b.

(Cited on page 20.)

J. F. M. Rodrigues, D. J. Rankin, V. Rossetti, A. Wagner, and H. C. Bagheri.

Correction: Differences in Cell Division Rates Drive the Evolution of Terminal Differentiation in Microbes.

PLoS computational biology, 8(5), 2012.

(Cited on pages 21, 67, and 123.)

A. Rokas.

The origins of multicellularity and the early history of the genetic toolkit for animal development.

Annual review of genetics, 42:235–251, 2008.

(Cited on page 92.)

V. Rossetti, B. E. Schirrmeister, M. V. Bernasconi, and H. C. Bagheri.

The evolutionary path to terminal differentiation and division of labor in cyanobacteria.

Journal of Theoretical Biology, 262(1):23–34, 2010.

(Cited on pages 67 and 80.)

V. Rossetti, M. Filippini, M. Svercel, A. D. Barbour, and H. C. Bagheri.

Emergent multicellular life cycles in filamentous bacteria owing to density-dependent population dynamics.

Journal of The Royal Society Interface, 8(65):1772–1784, 2011.

doi: 10.1098/rsif.2011.0102.

(Cited on page 19.)

D. Roze, R. E. Michod, and A. E. M. J. Wade.

Mutation, multilevel selection, and the evolution of propagule size during the origin of multicellularity.

The American Naturalist, 158(6):638 – 654, 2001.

(Cited on pages [3](#), [21](#), and [38](#).)

C. Rueffler, J. Hermisson, and G. P. Wagner.

Evolution of functional specialization and division of labour.

Proceedings of the National Academy of Sciences USA, 109:E326–E335, 2012.

(Cited on page [67](#).)

A. Sambamurty, editor.

A textbook of algae.

IK International Pvt. Limited, 2005.

(Cited on pages [79](#) and [80](#).)

G. Sandh, M. Ramstrom, and K. Stensjo.

Analysis of the early heterocyst cys-proteome in the multicellular cyanobacterium *Nostoc punctiforme* reveals novel insights into the division of labor within diazotrophic filaments.

BMC Genomics, 15(1):1064, 2014.

(Cited on page [80](#).)

A. Sebe-Pedros, B. M. Degnan, and I. Ruiz-Trillo.

The origin of metazoa: a unicellular perspective.

Nature Reviews Genetics, 18(8):498, 2017.

(Cited on pages [1](#) and [92](#).)

D. E. Shelton, A. G. Desnitskiy, and R. E. Michod.

Distributions of reproductive and somatic cell numbers in diverse volvox (chlorophyta) species.

Evolutionary ecology research, 14:707, 2012.

(Cited on pages [70](#) and [94](#).)

## Bibliography

---

C. A. Solari, J. O. Kessler, and R. E. Goldstein.

A general allometric and life-history model for cellular differentiation in the transition to multicellularity.

The American Naturalist, 181(3):369–380, 2013.

(Cited on pages [67](#), [93](#), and [94](#).)

M. Staps, J. van Gestel, and C. Tarnita.

Emergence of diverse life cycles and life histories at the origin of multicellularity.

Nature ecology & evolution, 3(8):1197 – 1205, 2019.

(Cited on pages [93](#), [110](#), and [127](#).)

K. Suresh, J. Howe, G. Ng, L. Ho, N. Ramachandran, A. Loh, E. Yap, and M. Singh.

A multiple fission-like mode of asexual reproduction in *blastocystis hominis*.

Parasitology research, 80(6):523–527, 1994.

(Cited on page [92](#).)

E. Szathmáry and J. M. Smith.

The major evolutionary transitions.

Nature, 374(6519):227–232, 1995.

(Cited on page [1](#).)

C. E. Tarnita, C. H. Taubes, and M. A. Nowak.

Evolutionary construction by staying together and coming together.

Journal of Theoretical Biology, 320(0):10–22, 2013.

(Cited on pages [20](#) and [127](#).)

P. D. Taylor and L. B. Jonker.

Evolutionarily stable strategies and game dynamics.

Mathematical Biosciences, 40:145–156, 1978.

(Cited on page [13](#).)

B. Thomas and H.-J. Pohley.

ESS-theory for finite populations.

Biosystems, 13(3):211–221, Jan. 1981.

(Cited on page [13](#).)

A. Tomitani, A. H. Knoll, C. Cavanaugh, and T. Ohno.

The evolutionary diversification of cyanobacteria: molecular–phylogenetic and paleontological perspectives.

Proceedings of the National Academy of Sciences, 103(14):5442–5447, 2006.

(Cited on page [20](#).)

I. Tomlinson.

Game-theory models of interactions between tumour cells.

European Journal of Cancer, 33(9):1495–1500, 1997.

(Cited on page [109](#).)

A. Traulsen and M. A. Nowak.

Evolution of cooperation by multilevel selection.

Proceedings of the National Academy of Sciences, 103(29):10952–10955, 2006.

(Cited on pages [2](#), [3](#), [4](#), and [125](#).)

A. Traulsen, C. Hauert, H. De Silva, M. A. Nowak, and K. Sigmund.

Exploration dynamics in evolutionary games.

Proceedings of the National Academy of Sciences USA, 106:709–712, 2009.

(Cited on page [38](#).)

S. Tuljapurkar and H. Caswell, editors.

Structured-Population Models in Marine, Terrestrial, and Freshwater Systems.

Chapman & Hall, 1997.

(Cited on pages [5](#), [7](#), [8](#), [9](#), and [97](#).)

D. Tverskoi, V. Makarenkov, and F. Alekserov.

Modeling functional specialization of a cell colony under different fecundity and viability rates and resource constraint.

PLOS ONE, 13(8):e0201446, 2018.

## Bibliography

---

(Cited on page 67.)

J. van Gestel and C. E. Tarnita.

On the origin of biological construction, with a focus on multicellularity.

Proceedings of the National Academy of Sciences USA, 114(42):11018–11026, 2017.

(Cited on page 20.)

M. van Veelen.

On the use of the Price equation.

Journal of Theoretical Biology, 237:412–426, 2005.

(Cited on page 2.)

M. van Veelen.

Group selection, kin selection, altruism and cooperation: When inclusive fitness is right and when it can be wrong.

Journal of Theoretical Biology, 259:589–600, 2009.

(Cited on page 21.)

J. von Neumann and O. Morgenstern.

Theory of Games and Economic Behavior.

Princeton University Press, Princeton, 1944.

(Cited on page 10.)

J. S. Waters, C. T. Holbrook, J. H. Fewell, and J. F. Harrison.

Allometric scaling of metabolism, growth, and activity in whole colonies of the seed-harvester ant *Pogonomyrmex californicus*.

The American Naturalist, 176(4):501–510, 2010.

(Cited on page 81.)

J. Weesie.

Asymmetry and timing in the volunteer's dilemma.

Journal of Conflict Resolution, 37:569–590, 1993.

(Cited on page 12.)

J. W. Weibull.

Evolutionary Game Theory.

MIT Press, Cambridge, 1995.

(Cited on pages [34](#) and [35](#).)

M. Willensdorfer.

On the evolution of differentiated multicellularity.

Evolution, 63(2):306–323, 2009.

(Cited on pages [67](#) and [80](#).)

A. E. Wilson, W. A. Wilson, and M. E. Hay.

Intraspecific variation in growth and morphology of the bloom-forming cyanobacterium *Microcystis aeruginosa*.

Applied and Environmental Microbiology, 72(11):7386–7389, 2006.

(Cited on pages [93](#) and [109](#).)

A. E. Wilson, R. B. Kaul, and O. Sarnelle.

Growth rate consequences of coloniality in a harmful phytoplankter.

PLoS One, 5(1):e8679, 2010.

(Cited on pages [93](#) and [109](#).)

L. Wolpert and E. Szathmáry.

Multicellularity: evolution and the egg.

Nature, 420(6917):745–745, 2002.

(Cited on page [92](#).)

M. Wynne and H. Bold.

Introduction to the Algae: Structure and Reproduction.

Prentice-Hall, Incorporated, 1985.

(Cited on page [81](#).)

Y. Yamamoto and F.-K. Shiah.

## Bibliography

---

Variation in the growth of *microcystis aeruginosa* depending on colony size and position in colonies.

Annales de Limnologie-International Journal of Limnology, 46(1):47–52, 2010.

(Cited on pages [93](#) and [109](#).)

D. Yanni, S. Jacobeen, P. Márquez-Zacarías, J. S. Weitz, W. C. Ratcliff, and P. J. Yunker.

Topological constraints in early multicellularity favor reproductive division of labor.

Elife, 9:e54348, 2020.

(Cited on page [126](#).)

E. C. Zeeman.

Population dynamics from game theory.

Lecture Notes in Mathematics, 819:471–497, 1980.

(Cited on page [13](#).)



# List of Publications

---

In this thesis,

- published: Yuanxiao Gao, Arne Traulsen, Yuriy Pichugin. Interacting cells driving the evolution of multicellular life cycles. PLoS Computational Biology 2019.
- accepted: Yuanxiao Gao, Hye Jin Park, Arne Traulsen and Yuriy Pichugin. Evolution of irreversible somatic differentiation. eLife 2021.
- submitted: Yuanxiao Gao, Yuriy Pichugin, Chaitanya Gokhale and Arne Traulsen, Evolution of reproductive strategies in incipient multicellularity.

Further publication not part of this thesis:

- Zhang W, Gao Y, Long M, Shen B. Origination and evolution of orphan genes and de novo genes in the genome of *Caenorhabditis elegans*. Sci China Life Sci. 2019 Apr;62(4):579-593. doi: 10.1007/s11427-019-9482-0.

# Acknowledgements

---

From the day I started my PhD study until now writing this dissertation, I have received tremendous support and assistance.

Foremost, I wish to express my sincere gratitude to my supervisor Arne Traulsen, who encouraged me to start my PhD studying in this field. He was extremely generous and patient to me at an early stage. During my study, I was always motivated by him when facing setbacks. Without his persistent help, I could not achieve today's progress and finish this thesis. And at the same time, he also created a very harmonious atmosphere in our working environment, where I am always supported by other people both in work and life.

Besides my supervisor, I would like to sincerely thank Yuriy Pichugin, who gave me immensely help on all my work. He provided me with the inspiration for the topic of multicellularity and numerous constructive suggestions during my work. My sincere thanks also go to Hyejin Park and Chaitanya Gokhale, who helped me in designing models and writing papers.

I also thank Florence Bansept and Nikhil Sharma, who helped me proofread this thesis. Michael Raatz helped me translate the abstract section. Furthermore, I thank my colleagues in the Theory department Group: Roman Ulises Zapien Campos, Saumil Shah, Jenna Gallie, Maria Bargas i Ribera, Dave Rogers, Primrose Boynton, Laura Hinder-sin, Mario Santer, Prateek Verma, Stefano Giaimo, Michael Sieber, Hildegard Uecker, and Hanna Schenk for modelling design ideas, stimulating codes, introducing related biological stories, exchanging ideas and polishing presentations. In particular, I thank Da Zhou, who deepened my understandings of the general framework of game dynamics.

Meanwhile, I would like to thank my family: my parents Jianxin Gao and Qiaoping Chen, my husband Wenyu Zhang and my daughter Jiayin Zhang. Especially, I am grateful to Wenyu Zhang for his encouragement and support during the past three years.

Last but not the least, I appreciate the additional financial support to me as a female scientist with a kid from the Max Planck Society, the Christiane Nüsslein-Volhard Foundation and the European Society for Evolutionary Biology.

# Declaration

---

I hereby declare that I have written this dissertation completely on my own. The work was guided by my PhD supervisor Arne Traulsen. Meanwhile, Yuriy Pichugin gave me great help to finish this work. Michael Raatz helped me translate the abstract section. The work has been done after registration for the degree of Dr. rer. nat. at the University of Lübeck. Furthermore, I confirm that no other sources have been used other than those specified in the dissertation itself. This dissertation, in the same or similar form, has not been submitted to any other doctoral degree committee yet.