# Ecological and evolutionary processes at multiple scales of biological organization

*Studies on the experimental evolution of dispersal and on theoretical cancer dynamics*

विद्या वाचस्पति की
उपाधि की अपेक्षाओं की आंशिक पूर्ति में प्रस्तुत शोध
प्रबंध

A thesis submitted in partial fulfilment of the requirements of the degree of

Doctorate of Philosophy

द्वारा / By


विभीषण बा / **Vibishan B.**

पंजीकरण सं. / Registration No.: **20163448**

शोध प्रबंध पर्यवेक्षक / Thesis Supervisor: **Prof Sutirth Dey**

IISER PUNE

भारतीय विज्ञान शिक्षा एवं अनुसंधान संस्थान
पुणे

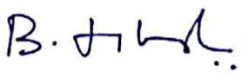
INDIAN INSTITUTE OF SCIENCE EDUCATION AND RESEARCH PUNE


2024

*To the earnestly dear and now departed,*
*for the wind beneath my wings.*

# Declaration

I declare that this written submission represents my ideas in my own words and where others' ideas have been included, I have adequately cited and referenced the original sources. I declare that I have acknowledged collaborative work and discussions wherever such work has been included. I also declare that I have adhered to all principles of academic honesty and integrity, and have not misrepresented or fabricated or falsified any idea/data/fact/source in my submission. I understand that violation of the above will be cause for disciplinary action by the Institute and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed.



Vibishan B.

Roll number: 20163448

Date: February 29, 2024

# Certificate

Certified that the work incorporated in the thesis entitled "Ecological and evolutionary processes at multiple scales of biological organization" submitted by Vibishan B. was carried out by the candidate, under my supervision. The work presented here or any part of it has not been included in any other thesis submitted previously for the award of any degree or diploma from any other University or institution.



Prof. Sutirth Dey

(Supervisor)

Date: 29 February, 2024

# Acknowledgements

This was always going to be a rocky road, but I have been lucky enough to have several people along the way for signposts and springboards. Foremost among those is my supervisor, Prof Sutirth Dey. He took a chance on me at a time when I could have well been a liability, and the fact that I managed to produce the work that I have owes a very great to his faith in my ability to do science. I can only hope that I continue to meet his high standards for scientific rigor and diligence. By extension, I also owe a great deal to the institution that is the Population Biology Lab (PBL). In many ways, even though Sutirth took me on, PBL has been on the business end of the deal and every single person here has been instrumental in my growth, as a person and as a scientist. For their patience and their unflinching support, I will forever be grateful.

Akshay and Selva have been my comrades-in-arms through some of the most turbulent and demanding times of my life. Akshay's unyielding commitment to science is exceeded only by his unbridled zeal to help. I'm grateful to him for being the voice of reason and good sense looking over my shoulder, especially in those moments I did not realize I needed one-he has saved me many a terrible mistake. From terrible food choices to cringe-worthy TV shows, Selva is an unparalleled source of joy, comfort, and acceptance. My only regret lies in not gaining her friendship much earlier. Likewise, Aarcha, Akhila, Kasturi, Sayantan, and Ruchitha all made my time in this lab come alive with fun and companionship. Thanks to Saumil, Imran, Puneeth, Harsha, Saismit, Satavisha, John T, and Aditya for keeping the spark of theory active with me-I hope the theoreticians' circle will soon rise again. PBL seniors Shraddha, Sudipta, Yashraj and Abhishek continue to guide us when we most need guidance. Newer colleagues Anish and Chinmay have been delightful labmates.

Prof Milind Watve was the first person to cultivate my interests in evolutionary thinking, and his singular capacity for insight is a subject of constant bewilderment and wonder for me. I'm grateful to him for getting me started on thinking about cancer, and all my Watve lab colleagues who have remained friends through the long years, for giving me a very liberal sandbox in which to explore my interests in physiology, ecology, and evolution. Manawa,

Shubhankar, and Akanksha are the best co-conspirators I could have ever asked for.

The Department of Biology at IISER Pune has been a pillar of support and I am constantly surprised by my luck in the number of cherished friends I have made here. Nagaraj, Aurnab, Nishad, and Soumen, among many others, have been incredibly generous with their time and commiseration, and I hope the years ahead look upon our friendship kindly. Girish, Deepak, and Ramana have never failed to produce an encouraging word or a nudge in the right direction, and I'm grateful for their support. The discussion on neurotransmitters in the dispersal syndrome study is a result of Prof Subhedar's infectious enthusiasm and magnanimity with his time. The Biology office staff, the Academic office staff, and the housekeeping staff at IISER Pune cannot be thanked sufficiently for all that they do to facilitate our work on a daily basis. Financial help found its way to me at very opportune moments from two sources-the Trimurti Fabricator-Twenty Twenty Interior Design endowment, and the Scivic Engg India-Innoplexus Consultancy endowment-and I am grateful to both.

In true Yoda fashion, Anand has never failed to put me down when I deserved to be, while somehow also managing to be a source of strength and succour-may his jowls remain aquiver with sage wisdom and sass. Some friendships are truly interdisciplinary, and Sudha, Mridula, Poornima, Aswin, and Alok have broadened my horizons for what friendship could look like. For the horrors and warmth of these last two years in their society, I am grateful. I have been luckier than most in having gained care and affection in not one, but two departments in IISER. The HSS clan-Swapna, Anuja, Kalyani, and everyone else-have fondly adopted me as a surrogate grad student in the humanities, and my life has been made better by their company and generosity. My friends from before IISER-Prathik, Rosh, Ju, Vaish-have been my touchstones for personal growth.

In a very real way, music and the arts have been the anchors that have kept me grounded, and an entire legion of people must be credited for being the carriers of this stability-Harsha, Farhan, Anuvind, Malavika, Abhi, Kitchi, Soumya, Neelay, Rutwik, the talented teams of *Naqaab* as well as *the Sun was White, the Moon Disobedient*, and the diverse membership of the erstwhile IISER Pune Choir. My gurus, Dr Sushruti and Smt Vedavalli, have given me invaluable life lessons that have become a part of my constitution. It is hard to describe the larger-than-life influence John Mathew has had in my life; it suffices to say that I look forward to years more of singing, talking, and laughing with him. Some of my happiest moments have been spent screaming my lungs out with these people and they will remain very close to my heart.

No one knows more about me than Ram and Alaka, and in everything I do, I carry a little bit of them with me. I doubt if I will ever receive a greater honour than being made an

honorary Moitro, and I have Alaka, Mini and Cocoa to thank for that. My family have borne the abrasiveness of my stress when it boils over, and have seen me through this program almost by the sheer force of their will and love. Nandan is our little ray of sunshine. I will never not be thankful.

## Collaborators

Formally, my work in this thesis has included the professional involvement of several collaborators, without whom the studies presented here would not have been possible. This includes my PBL colleagues who have been instrumental in data collection, analysis, and writing, as well as students and mentees at the undergraduate and Master's levels, whose work is a part of the results in this thesis. I have therefore added a note at the beginning of each chapter, documenting the contributions of my collaborators in the corresponding study *vis-à-vis* my own role thereof.

**“ஓசை பெற்று உயர் பாற்கடல் உற்று. ஒரு**
**பூசை முற்றவும் நக்குபு புக்கென.**
**ஆசை பற்றி அறையலுற்றேன்...**

**“நொய்தின் நொய்ய சொல் நூற்கலுற்றேன்—**
**என்னை!”**

கம்ப ராமாயணம், பால காண்டம், அவையடக்கம் 1-2

***“To the sonorous Ocean of milk, a cat harkening,***
***thinks to drink its every last drop.***
***In such avarice do I embark upon this narration***
***…***

***“The humblest of humble words do I weave—***
***alas!”***

*Kamba Rāmāyaṅam, Bāla kāṅdam, Poet’s Humility 1-2*

# Contents

**Declaration** . . . . . **v**

**Certificate** . . . . . **vii**

**Acknowledgements** . . . . . **ix**

**Synopsis** . . . . . **xix**

**I Main material** **1**

**Chapter 1**
**Introduction** . . . . . **3**

**Chapter 2**
**Selection for dispersal under larval malnutrition results in a non-monotonic kernel** . . . . . **11**
2.1 Introduction . . . . . 12
2.2 Materials and methods . . . . . 15
2.3 Results . . . . . 20
2.4 Discussion . . . . . 27

**Chapter 3**
**Malnourished dispersers show sexually-dimorphic trait correlations under dispersal selection** . . . . . **33**
3.1 Introduction . . . . . 34
3.2 Materials and methods . . . . . 36
3.3 Results . . . . . 40
3.4 Discussion . . . . . 44

**Chapter 4**
**A resource-based mechanistic framework for castration-resistant prostate cancer (CRPC)** . . . . . **51**
4.1 Introduction . . . . . 52

4.2 Methods . . . 54
4.3 Results . . . 57
4.4 Discussion . . . 63

**Chapter 5**
**Context-dependent selection as the keystone in the somatic evolution of cancer . . . 69**
5.1 Introduction . . . 70
5.2 The "bad luck" model . . . 72
5.3 Models with selection on mutants . . . 73
5.4 Sensitivity of predictions . . . 77
5.5 Differential testable predictions of the models . . . 79
5.6 Discussion . . . 80

**Chapter 6**
**A mechanistic interpretation of patterns in cancer risk across biological scales 83**
6.1 Introduction . . . 84
6.2 Methods . . . 86
6.3 Results . . . 92
6.4 Discussion . . . 100

**Chapter 7**
**Conclusions and future directions . . . 105**

**II Supplementary Material 113**

**Chapter S1**
**Modified formulation used for larval malnutrition in fruit flies . . . 115**

**Chapter S2**
**Statistical analyses for Chapter 2 . . . 117**
S2.1 Generation 42 . . . 118
S2.2 Generation 66-67 . . . 125
S2.3 Generation 93 . . . 133

**Chapter S3**
**Statistical analyses for Chapter 3 . . . 135**
S3.1 Physiological traits . . . 136
S3.2 Behavioral traits . . . 141

**Chapter S4**
**Parameterization of the resource consumption framework for CRPC . . . . 147**
S4.1 Derivation of cell growth parameters . . . . . 148
S4.2 Normalization of resource dynamics . . . . . 148

**Chapter S5**
**Supplementary analyses for Chapter 5 . . . . . 151**
S5.1 Alternative distributions of $g$ . . . . . 152
S5.2 Sensitivity of model predictions to $k$ . . . . . 152
S5.3 The nature of the relationship between cancer incidence and cell number . . 152

**Chapter S6**
**Supplementary Figures . . . . . 155**
S6.1 Supplementary Figures for Chapter 2 . . . . . 156
S6.2 Supplementary Figures for Chapter 4 . . . . . 160
S6.3 Supplementary Figures for Chapter 5 . . . . . 169

**List of publications/manuscripts under preparation . . . . . 203**

# List of Figures

2.1 Increase in dispersal propensity due to dispersal selection . . . . . 21
2.2 Progressive increase in dispersal ability . . . . . 22
2.3 Model-free estimates of the dispersal kernel shape . . . . . 23
2.4 Gamma function fit to empirical dispersal kernels . . . . . 24
2.5 Spatial extent across generations . . . . . 26
2.6 Measures of long-distance dispersal . . . . . 27

3.1 Physiological trait responses due to dispersal selection under larval malnutrition . . . . . 42
3.2 Locomotor phenotype under dispersal selection . . . . . 43
3.3 Impairment in mating traits due to dispersal selection . . . . . 44
3.4 Aggressive behavior is not associated with dispersal evolution . . . . . 45

4.1 Schematic of the prostate cancer model . . . . . 54
4.2 Steady state behavior under identical resource use efficiencies . . . . . 59
4.3 Steady state behavior under mismatched resource use efficiencies . . . . . 60
4.4 Producer-consumer balance with varying oxygen supply . . . . . 62

4.5 Abundance regimes of oxygen vs testosterone . . . . . 63

5.1 Cancer incidence predictions for the "bad luck" model . . . . . 74
5.2 Cancer incidence predictions for the CIS model . . . . . 76
5.3 Randomizing $p$ in the CIS model . . . . . 77
5.4 Cancer incidence predictions for the CDS model . . . . . 78

6.1 Sequence of events in the model and model components . . . . . 89
6.2 Cell compartment composition as a function of ontogenetic strategy . . . 94
6.3 Conditions for tapering deficit dynamics . . . . . 95
6.4 Maximum clone length corresponds to conditions of tapering deficit dynamics 97
6.5 developmental viability is related to mutation accumulation . . . . . 100
6.6 Mutant clone growth across model parameters . . . . . 104

S6.1.1 Percentiles of the mean-subtracted distances for both populations . . . . . 156
S6.1.2 Multiple LDD measures across generations . . . . . 157
S6.1.3 Dispersal kernels after mean subtraction . . . . . 158
S6.1.4 Fitted equation parameters for the Gamma distribution . . . . . 159
S6.2.1 Fixing the testosterone consumption rate for $T^+$ . . . . . 160
S6.2.2 Cell growth time series when neither resource is supplemented externally . 161
S6.2.3 Cell growth time series when only oxygen is supplemented externally . . . 162
S6.2.4 Cell growth time series when only testosterone is supplemented externally 163
S6.2.5 Cell growth time series when both resources are supplemented externally . 164
S6.2.6 Overall population viability under a range of resource supply conditions . 165
S6.2.7 Time series corresponding to Figure 4.4A for T supply rate 0.0001 $\text{min}^{-1}$ . 166
S6.2.8 Time series corresponding to Figure 4.4A for T supply rate 0.001 $\text{min}^{-1}$ . 167
S6.2.9 Time series corresponding to Figure 4.4B . . . . . 168
S6.3.1 Alternate distributions of $g$: Gumbel . . . . . 169
S6.3.2 Alternate distributions of $g$: uniform . . . . . 170
S6.3.3 Randomizing $k$ in the context-independent selection case . . . . . 171
S6.3.4 Randomizing $k$ in the context-dependent selection case . . . . . 172
S6.3.5 Effect of $k$ in the context-independent selection case . . . . . 173
S6.3.6 Effect of $k$ in the context-dependent selection case . . . . . 174
S6.3.7 Distribution of slopes . . . . . 175
S6.3.8 Distribution of Pearson's $r$ for reduced datasets . . . . . 175

# List of Tables

3.1 Dispersal syndrome under standard vs. poor larval nutrition . . . . . 50

4.1 Parameters of the CRPC resource consumption model . . . . . 67

6.1 Numerical example of different paces of differentiation . . . . . 89
6.2 List of free parameters explored across simulations . . . . . 91

S1.1 Modified banana-jaggery medium recipes . . . . . 116

S5.1 Residuals of cancer risk against *lscd* . . . . . 154

# Synopsis

**Thesis title**: Ecological and evolutionary processes at multiple scales of biological organization
**Name**: Vibishan B.
**Roll number**: 20163448
**Name of supervisor**: Prof Sutirth Dey
**Department**: Biology
**Date of registration**: 01/08/2016
Indian Institute of Science Education and Research (IISER), Pune

Ecological and evolutionary processes play roles of fundamental importance across levels of organization in biology. In this thesis, I document five studies that, between them, investigate questions of ecological and evolutionary importance in the context of two broad biological phenomena: dispersal and cancer. Part I consists of 7 chapters, organized as follows.

Chapter 1 is a general introduction, in which I expand upon the importance of ecology and evolution in biology, and place in context the motivations for the questions addressed in subsequent chapters.

Chapters 2 and 3 deal with my work in the experimental evolution of dispersal in lab populations of fruit flies. Dispersal is recognized as an important aspect of an organism's life, and has been studied extensively in a wide variety of experimental and theoretical contexts. The affective scope of dispersal can be understood in terms of two prevalent concepts in the literature: the dispersal kernel and the dispersal syndrome (Clobert et al., 2012).

In Chapter 2, I focus on the former. Dispersal kernels are essentially probability distributions of the spatial locations of an individual or propagule in a given population. But the apparent simplicity of the definition of the kernel belies its important role in determining several aspects of population ecology including invasion dynamics, spatial structuring, metapopulation stability, and phenotypic polymorphism (Kot et al., 1996, Travis and

Dytham, 1999, Higgins and Cain, 2002, Lindström et al., 2011). Notwithstanding this importance, it remains unclear unclear if and how the evolution of kernel features is affected by environmental factors. In Chapter 2, I address this question using laboratory populations of *Drosophila melanogaster* selected for greater dispersal under a protein-deficient larval diet. At three different time points in their evolutionary history, we measure a range of dispersal-related properties in these dispersal-selected flies and the corresponding unselected controls. This includes measurements of the full dispersal kernel in these populations at two different generation points. Our results reveal that dispersal-selected populations initiate dispersal more often and travel further on average than unselected controls. Interestingly, despite showing a significant change in the overall kernel shape, selected populations do not show a higher frequency of long-distance dispersers (LDDs). However, it is worth pointing out this result only holds when the dispersal kernels of the two populations are corrected for their respective average dispersal distances. This raises practical questions with regard to the study of long-distance dispersal as it indicates that short-distance dispersal, as reflected in the population average, must be treated independently from long-distance dispersal. As individuals that travel significantly greater distances than the population average, LDDs are known to be key drivers of population expansion and their limitation could have potential consequences for long-term population dynamics (Andow, 2022). On the whole, it is possible that the separation of the evolution of short- and long-distance dispersal could be a consequence of larval malnutrition, but this would require further experimental confirmation. We also find that dispersal evolution under larval malnutrition leads to a non-monotonic dispersal kernel, with the highest frequencies at intermediate distances, which is fairly uncommon in the realm of insect dispersal. The evolved shape of the dispersal kernel and the limited LDD phenotype therefore point to significant qualitative shifts in the evolution of dispersal due to a change in larval nutrition. Our results altogether provide a starting point for further investigation of context dependence in dispersal evolution.

This chapter is reproduced from a manuscript that is currently under review and can be found on the preprint database, bioR$\chi$iv as below:
**B. Vibishan**, A. Malwade, S. R. Gupta, A. Koner, V. Khodake, and S. Dey. 2023. "Selection for dispersal under larval malnutrition results in a non-monotonic kernel in *Drosophila melanogaster*." bioRxiv. doi: 10.1101/2023.10.26.564216

Chapter 3 focuses on the second key concept in the study of dispersal i.e., the dispersal syndrome. While the evolution of dispersal is often cast in terms of the collection of traits that is linked to a dispersive strategy (Bowler and Benton, 2005), it is unclear if and how these syndromes of co-evolving traits respond to sustained environmental shifts. To address this question, I return to the fly populations selected for greater dispersal under larval malnutrition and examine if they have also evolved unique trait correlations alongside higher dispersal capacity compared to the unselected controls. Our results in this chapter

demonstrate that selecting for dispersal under larval malnutrition has led to multiple correlations with dispersal in both physiological and behavioral traits. In particular, we find that these trait correlations are organized in sex-specific ways such that the overall dispersal syndromes are quite different between males and females. Comparing these results from the current study with earlier results of correlated trait evolution with dispersal under standard nutrition, we are able to demonstrate clearly that several features of the dispersal syndrome have evolved uniquely under larval malnutrition, as dispersers selected under standard nutrition did not show similar trait correlations at a similar stage in their evolutionary history. The documented shift in the dispersal syndrome due to larval malnutrition suggests that the poor nutrition background has led to the evolution of a disperser phenotype with a distinctly poor outlook for successful invasion and settlement. These results lead us to discuss two key points of further research: (1) evolved trait correlations with dispersal offer a valuable substrate for further investigation of the underlying mechanisms of dispersal (Campos-Candela et al., 2019), and (2) the qualitative differences in the dispersal syndrome attributable to the larval nutrition environment clearly indicates that context-dependent lability of the dispersal syndrome must be studied more closely, particularly when this lability affects the consequences of dispersal in a post-movement settlement context.

The results from this chapter are currently under preparation for publication:
**B. Vibishan**, A. Malwade, A. Koner, V. Khodake, and S. Dey. 2023. "Malnourished dispersers show sexually-dimorphic trait correlations under dispersal selection." *(in prep)*

Chapters 4-6 are a series of studies investigating various aspects of ecological and evolutionary dynamcis in cancer, across multiple scales of biological organization. All three studies were carried out as collaborative efforts-Chapters 4 and 6 involved efforts from Master's students in the lab, while Chapter 5 stems from work done by me with a previous advisor.

Cancer therapy often leads to the selective elimination of drug-sensitive cells from the tumor. This can favor the growth of cells resistant to the therapeutic agent, ultimately causing a tumor relapse (Gallaher et al., 2018, Hansen and Read, 2020). Castration-resistant prostate cancer (CRPC) is a well-characterized instance of this phenomenon (Zhang et al., 2017, Cunningham et al., 2018). In CRPC, after systemic androgen deprivation therapy (ADT), a subset of drug-resistant cancer cells autonomously produce testosterone, thus enabling tumor regrowth. A previous theoretical study has shown that such a tumor relapse can be delayed by inhibiting the growth of drug-resistant cells using biotic competition from drug-sensitive cells (Cunningham et al., 2018). In this context, the centrality of resource dynamics to intra-tumor competition in the CRPC system indicates clear scope for the construction of theoretical models that can explicitly incorporate the underlying mechanisms of tumor ecology (Kotler and Brown, 2020). Along these lines, in Chapter 4, I collaborate

with a Master's student to build a modified logistic framework to model cell-cell interactions in terms of the production and consumption of resources. Our results show that the steady state composition of CRPC can be understood as a composite function of the availability and utilization efficiency of two resources-oxygen and testosterone. In particular, we show that the effect of changing resource availability or use efficiency is conditioned by their general abundance regimes. Testosterone typically functions in trace amounts and thus affects steady state behavior of the CRPC system differently from oxygen, which is usually available at higher levels. Our data thus indicate that explicit consideration of resource dynamics can produce novel and useful mechanistic understanding of CRPC. Furthermore, such a modeling approach also incorporates variables into the system's description that can be directly measured in a clinical context. This is therefore a promising avenue of research in cancer ecology that could lead to therapeutic approaches that are more clearly rooted in the biology of CRPC.

This chapter is reproduced from a manuscript that is currently under review and can be found on the preprint database, bioR$\chi$iv as below:
**B. Vibishan**, B. V. Harshavardhan, and S. Dey. 2023. "A Resource-Based Mechanistic Framework for Castration-Resistant Prostate Cancer (CRPC)." bioRxiv. doi: 10.1101/2023.09.19.558379

Chapter 5 remains in the context of cancer in humans, but examines evolutionary phenomena that affect population-level patterns in cancer incidence. Somatic evolution of cancer involves a series of mutations, and attendant changes, in one or more clones of cells (Nunney, 1999, Hanahan and Weinberg, 2000, Gerlinger et al., 2014). An entire class of models, starting from around the middle of the last century, have attempted to relate these population-level incidence patterns to cancer progression mechanisms at the cellular level (Armitage and Doll, 1954). Although these efforts have crystallized into a multi-stage model of carcinogeneis, questions still remain in terms of whether current frameworks explain trends in cancer incidence sufficiently. The diversity of cancer mechanisms posited by these frameworks can be encapsulated by three broad conceptual models, as follows. A "bad luck" type model assumes chance accumulation of mutations, without considering the growth of mutant cell clones in the tissue micro-environment (Calabrese and Shibata, 2010). The clonal expansion model assumes, on the other hand, that any mutation leading to partial loss of control of cell proliferation will give a selective advantage to the mutant (Nowell, 1976). This selective advantage is then thought to drive the progression of cancer. However, a number of experiments show that an intermediate pre-cancer mutant has only a conditional selective advantage (Cao, 2010, Rundqvist et al., 2013), and given that tissue micro-environmental conditions differ across individuals, this selective advantage to a mutant could be widely distributed over the population. These are broad starting points that reflect prevalent ideas about the nature of cancer mechanisms, but it is currently unclear how these differ in their

predictive power. In a study designed and carried out with Prof Milind Watve, I evaluate these three models, namely "bad luck", context-independent, and context-dependent selection, in a comparative framework, on their ability to predict patterns in total incidence, age-specific incidence, the stem cell number-incidence relationship and other known phenomena associated with human cancers. Our results show that among the factors considered in the model, context dependence of somatic evolution is necessary and sufficient to explain observed epidemiological patterns, and that cancer evolution is largely selection-limited, rather than mutation-limited. Based on this understanding, we identify a wide range of physiological, genetic and behavioral factors influencing the tissue micro-environment that could potentially be the source of this context dependence of pre-cancer somatic mutation fitness. The identification and targeting of these micro-environmental factors that influence the dynamics of selection in the soma offer new possibilities for cancer prevention.

This chapter is largely reproduced from the peer-reviewed article below:
**B. Vibishan**, and M. G. Watve. 2020. "Context-Dependent Selection as the Keystone in the Somatic Evolution of Cancer." Scientific Reports 10 (1): 4223. doi: 10.1038/s41598-020-61046-7

In Chapter 6, I begin with the simple question that could be asked across all mammalian species, if not all animals: do more cells lead to more cancer? The allometry of cancer risk with size appears to be scale-specific, as the inter-specific, intra-specific, and inter-organ patterns in cancer risk scaling are distinct from each other (Abegglen et al., 2015, Nunney, 2018). For instance, while cancer risk barely changes with increasing body sizes across species, larger individuals of the same species are known to be at a higher risk of cancer in their lifetimes. This is in addition to the variety of known cancer risk trends with age and lifespan, some of which have been discussed in the previous chapter. Notwithstanding this heterogeneity however, mammals also broadly share developmental features across taxa, and the twin processes of cell division and differentiation are central to the dynamics of tissue architecture and turnover in all mammals (Pepper et al., 2007, Gerlinger et al., 2014). It is possible then that a better understanding of the common processes across these biological scales could lead to a clearer understanding of how and why cancer risk varies with cell number in a scale-specific manner. Chapter 6 investigates this possibility, for which I collaborate with a Master's student to build an agent-based model of the development of a cell compartment from a single cell. The model allows us to investigate the scaling of cellular organization with the size of the overall cell compartment through division and differentiation dynamics. Our data suggest that developmental constraints on rates of cell differentiation affect both the size of the actively-dividing cell pool and the average division frequency to produce a cell compartment of a given size. We also show that the same developmental constraints determine how and when the accumulation of neutral mutations increases with the cell compartment size. The accumulation of non-neutral mutations, on the other hand,

offers an interesting case of antagonism between the actions of mutation and differentiation in a developing cell compartment. On the whole, our model framework is able to produce a wide range of trends in cancer risk with cell number across different parts of its parameter space, thus reflecting the observed heterogeneity in cancer risk-size allometries across biological scales. Our results provide a theoretical framework that allows us to elucidate the conditions under which having more cells does lead to higher cancer risk, and importantly, the conditions under which it does not.

Chapter 7 is the concluding chapter, in which I reiterate the main findings from each of these studies briefly, before touching upon more speculative implications in each case based on recent findings emerging in the literature. These lead to potential directions that future work could take. In the case of the CRPC model, ongoing work in the lab has addressed some of these future directions and I briefly discuss the most promising of these preliminary results.

Part II of the thesis is an extensive supplement with detailed information regarding the modified fly food recipe and the statistical analyses used in the fly studies (Chapters S1, S2, and S3), parameterization and other assumptions underlying the CRPC model (Chapter S4), and details of sensitivity analysis for the cancer incidence study as well as supporting analysis of other epidemiological data (Chapter S5). Supplementary figures for all the studies are given together in the last chapter in Part II (Chapter S6).

It is worth noting here that each of the chapters in Part I stems from the results of studies that were designed and intended to stand independently of the others. Over the course of their execution, I have worked extensively with diverse people for analysis and writing, and it is perhaps inevitable that the writing styles may occasionally be inconsistent between one chapter and the next. I have tried to maintain uniformity where feasible, but have refrained from substantially rewriting published work.

# References

Abegglen, L. M., A. F. Caulin, A. Chan, K. Lee, R. Robinson, M. S. Campbell, W. K. Kiso, D. L. Schmitt, P. J. Waddell, S. Bhaskara, S. T. Jensen, C. C. Maley, and J. D. Schiffman. 2015. Potential Mechanisms for Cancer Resistance in Elephants and Comparative Cellular Response to DNA Damage in Humans. JAMA 314:1850. doi: 10.1001/jama.2015.13134.

Andow, D. A. 2022. Dispersal kernel of green rice leafhopper estimated from truncated data. Population Ecology 65:111–120. doi: 10.1002/1438-390X.12141.

Bowler, D. E., and T. G. Benton. 2005. Causes and consequences of animal dispersal strategies: relating individual behaviour to spatial dynamics. Biological Reviews 80:205–225. doi: 10.1017/S1464793104006645.

Calabrese, P., and D. Shibata. 2010. A simple algebraic cancer equation: Calculating how cancers may arise with normal mutation rates. BMC Cancer 10:3. doi: 10.1186/1471-2407-10-3.

Campos-Candela, A., M. Palmer, S. Balle, A. Álvarez, and J. Alós. 2019. A mechanistic theory of personality-dependent movement behaviour based on dynamic energy budgets. Ecology Letters 22:213–232. doi: 10.1111/ele.13187.

Cao, L. 2010. Environmental and Genetic Activation of a Brain-Adipocyte BDNF/Lep- tin Axis Causes Cancer Remission and Inhibition. Cell 142. doi: 10.1016/j.cell.2010.05.029.

Clobert, J., M. Baguette, T. G. Benton, and J. M. Bullock. 2012. Dispersal Ecology and Evolution. Oxford University Press. ISBN: 978-0-19-960889-8.

Cunningham, J. J., J. S. Brown, R. A. Gatenby, and K. Staňková. 2018. Optimal control to develop therapeutic strategies for metastatic castrate resistant prostate cancer. Journal of Theoretical Biology 459:67–78. doi: 10.1016/j.jtbi.2018.09.022.

Gallaher, J. A., P. M. Enriquez-Navas, K. A. Luddy, R. A. Gatenby, and A. R. Anderson. 2018. Spatial Heterogeneity and Evolutionary Dynamics Modulate Time to Recurrence in Continuous and Adaptive Cancer Therapies. Cancer Research 78:2127–2139. doi: 10.1158/0008-5472.CAN-17-2649.

Gerlinger, M., N. McGranahan, S. M. Dewhurst, R. A. Burrell, I. Tomlinson, and C. Swanton. 2014. Cancer: Evolution Within a Lifetime. Annual Review of Genetics 48:215–236. doi: 10.1146/annurev-genet-120213-092314.

Hanahan, D., and R. Weinberg. 2000. The hallmarks of cancer. Cell 100:57–70.

Hansen, E., and A. F. Read. 2020. Cancer therapy: Attempt cure or manage drug resistance? Evolutionary Applications 13:1660–1672. doi: 10.1111/eva.12994.

Higgins, S. I., and M. L. Cain. 2002. Spatially realistic plant metapopulation models and the colonization–competition trade-off. Journal of Ecology 90:616–626. doi: 10.1046/j.1365-2745.2002.00694.x.

Kot, M., M. A. Lewis, and P. Van Den Driessche. 1996. Dispersal Data and the Spread of Invading Organisms. Source: Ecology 77:2027–2042.

Kotler, B. P., and J. S. Brown. 2020. Cancer Community Ecology. Cancer Control 27:1073274820951776. doi: 10.1177/1073274820951776.

Lindström, T., N. Håkansson, and U. Wennergren. 2011. The shape of the spatial kernel and its implications for biological invasions in patchy environments. Proceedings of the Royal Society B: Biological Sciences 278:1564–1571. doi: 10.1098/rspb.2010.1902.

Nowell, P. 1976. The clonal evolution of tumor cell populations. Science 194. doi: 10.1126/science.959840.

Nunney, L. 1999. Lineage selection and the evolution of multistage carcinogenesis. Proceedings of the Royal Society of London. Series B: Biological Sciences 266:493–498. doi: 10.1098/rspb.1999.0664.

———. 2018. Size matters: Height, cell number and a person's risk of cancer. Proceedings of the Royal Society B: Biological Sciences 285:20181743. doi: 10.1098/rspb.2018.1743.

Pepper, J. W., K. Sprouffske, and C. C. Maley. 2007. Animal Cell Differentiation Patterns Suppress Somatic Evolution. PLOS Computational Biology 3:e250. doi: 10.1371/journal.pcbi.0030250.

Rundqvist, H., M. Augsten, A. Strömberg, E. Rullman, S. Mijwel, P. Kharaziha, T. Panaretakis, T. Gustafsson, and A. Östman. 2013. Effect of Acute Exercise on Prostate Cancer Cell Growth. PLoS ONE 8:e67579. doi: 10.1371/journal.pone.0067579.

Travis, J. M. J., and C. Dytham. 1999. Habitat persistence, habitat availability and the evolution of dispersal. Proceedings of the Royal Society of London. Series B: Biological Sciences 266:723–728. doi: 10.1098/rspb.1999.0696.

Zhang, J., J. J. Cunningham, J. S. Brown, and R. A. Gatenby. 2017. Integrating evolutionary dynamics into treatment of metastatic castrate-resistant prostate cancer. Nature Communications 8:1816. doi: 10.1038/s41467-017-01968-5.

# Part I

# Main material

# 1 | Introduction

In 1963, Niko Tinbergen would write a paper that was to become paradigmatic for the philosophy of biology (Tinbergen, 1963). His central thesis in this paper was the delineation of the understanding of behavior into four fundamental levels of inquiry-mechanism, function, evolution and phylogeny-and while his framing was intended for the investigation of animal behavior, it has since grown into a general conceptual framework of causation in biology (Bateson and Laland, 2013, Nesse, 2013). Within this framework, the twin disciplines of ecology and evolutionary biology are implicated in multiple levels of causation in biological phenomenon. As the study of organisms and their interactions with each other as well as their environments, ecology is naturally a part of the mechanisms underlying the manifestation of a trait insofar as it is affected by such interactions, while also informing the immediate functional consequences of the trait and the scope of its evolution (Levin et al., 2009). The study of evolutionary biology in turn involves the investigation of the mechanisms by which heritable variation in phenotypes changes the composition of populations of organisms over the course of generations, and while it is generally thought to occur over much longer timescales than ecological phenomena, the two are often mutually interconnected (Losos et al., 2014). Together, ecology and evolutionary biology form the substrate for much of the current understanding of the broadest patterns in living organisms, from local phenomena occurring over short timescales to global patterns developing over geological time periods (Losos et al., 2013). Given the vast diversity of spatial and temporal scales involved, the importance of ecology and evolution in constructing a complete understanding of life on the Earth cannot be overstated, and new avenues of research are constantly being uncovered in biology that illuminate fresh scope for the application of ecological and evolutionary perspectives. This thesis is the result of investigations into two such biological phenomena-dispersal and cancer-both of which, in their own ways, have been the grounds for interesting expansions of ecological-evolutionary thinking.

## Dispersal

All organisms must undergo spatial displacement. Whether as mobile individuals or mobile propagules generated by sessile individuals, all populations spread in space (Clobert et al., 2012). While individual organisms engage in myriad kinds of movement in their lifetimes in response to a wide range of internal and external cues (Nathan et al., 2008), populations of organisms generally tend to spread and expand in space. Specifically, when this expansion has consequences for gene flow across landscape, a biological phenomenon called dispersal is said to occur that is increasingly recognized as being of fundamental importance to the ecology, life history and evolution of the organism (Ronce, 2007, Bonte and Dahirel, 2017). Despite the fact that the study of dispersing organisms began in the middle of previous century at the latest (Dobzhansky and Wright, 1943, Wallace, 1966), much about this complex process remains unexplained. The complexity in dispersal stems from the fact that it inte-

grates several aspects of organismal biology across physiology and behavior in its functioning (Renault, 2020), and that it has manifold consequences for phenotypic evolution (Travis and Dytham, 1999), population dynamics (Kot et al., 1996) and long-term stability under changing environmental conditions (Cayuela et al., 2020). However, it is also this complexity that underlies the centrality of dispersal in the life history of an organism, as it plays a significant role in energetic allocation and the survival strategy of the organism. While this situates the study of dispersal as an indispensable aspect of ecology and evolutionary biology, the rapid and drastic shifts in environmental conditions brought on by global anthropogenic climate change have made it more urgent to develop a comprehensive understanding of dispersal, as it represents one of the primary means of tracking favorable habitats (Massot et al., 2008).

Due to the diverse kinds of movement organisms engage in, the empirical characterization of dispersal is often challenging (Lustenhouwer et al., 2023), and the dispersal kernel offers a straightforward way to quantify several key features of dispersal within a single measurement framework. The kernel is typically defined as the frequency distribution of dispersing individuals or propagules against distance from a fixed point source (Clobert et al., 2012) and its importance in understanding the spatial spread, composition and evolution of populations is well-recognized (Kot et al., 1996). However, an intriguing gap in current understanding concerns the extent to which the shape of the dispersal kernel can shift over time within a given population, even though such shifts would be highly consequential for the population. Chapter 1 situates this gap in the context of current literature more broadly, and describes a series of experimental evolution studies carried out in the lab in which large, outbred populations of *Drosophila melanogaster* are subjected to selection for greater dispersal capacity. These studies seek to address the question of whether and how dispersal kernels could evolve in response to environmental change, and the results indicate that larval nutrition could play a decisive role in the form and scope of dispersal evolution in the fruit fly. Our data show that when selected for dispersal under poor larval nutrition, the dispersal kernel evolved towards a unimodal shape with an identifiable peak in frequency at intermediate distances and a clear reduction in the tendency for long-distance dispersal. These trends are placed in the context of previous results of dispersal evolution under standard larval nutrition from the same experimental system (Tung et al., 2018*b*), and the implications for population ecology and evolution are discussed.

The results in the first chapter demonstrate a clear increase in the dispersal capacity of the selected flies relative to the control flies, even though poor larval nutrition seems to have placed significant constraints on long-distance dispersal. Due to its complexity and energetic costs, evolution of dispersal can have significant impacts for the correlated evolution of many other physiological and behavioral traits. While correlated evolution of traits with dispersal has been widely documented (Renault, 2020), it is a matter of some debate in the field as

to whether such collections of traits co-evolving with dispersal constitute a "syndrome" that could be predictive of dispersal status (Buoro and Carlson, 2014). While strong associations have been identified in plant dispersal particularly with seed morphology (Levin et al., 2003), "syndromes" of comparable robustness have eluded identification in animal systems due to much higher variability (Clobert et al., 2009). Chapter 2 addresses the question of the dispersal syndrome in our experimental system, and builds on the robust response to selection to consider the broader consequences of dispersal selection for phenotypic evolution. The results therein illustrate several trait correlations which are unique to dispersal selection under poor larval nutrition, but also show that this set of correlated traits are not entirely consistent with previously measured trait correlations of dispersal under standard larval nutrition (Tung et al., 2018*a*). This allows us to demonstrate the lability of so-called dispersal "syndromes" within a controlled mesocosm environment and further indicates the limited ability of correlated trait evolution to be predictive of dispersal status. On the other hand, a closer examination of the observed trait correlations in the current study with those under standard nutrition suggests intriguing possibilities regarding the mechanisms of dispersal evolution in our experimental system, some of which are discussed in light of available empirical evidence.

## Cancer

All organisms must die. In the wild, mortality is predominantly extrinsic, be it from predators, infections or lack of resources (David and Zimmerman, 2010). Uniquely in multicellular organisms, however, cancer represents an internal source of mortality by which some cells within the organism can escape the regulatory machinery to grow disproportionately at the expense of normal cells (Weinberg, 1998). In the human context, the excess consumption of bodily resources by rogue cancer cells underlies most of the pathology of the disease (Kareva, 2011, Markert and Vazquez, 2015), and it is their resemblance to many kinds of normal cells in the organism that makes them so difficult to target and eliminate (Hanahan and Weinberg, 2000, 2011). Much of the conventional understanding regarding the origin and spread of cancer in humans has centered around the acquisition of genetic mutations and their consequences for altered cell signaling (Weinberg, 1998), and it is only recently that the application of eco-evo perspectives has gained prominence (Aktipis and Nesse, 2013, Maley et al., 2017). Research over the past two decades has revealed cancers to be microcosms of a range of ecological and evolutionary processes at several scales of biological organization, from cellular and somatic ecology and evolution (Kareva, 2011, Madan et al., 2022) to macro-evolutionary trends across species (Albuquerque et al., 2018). The three studies on cancer-related phenomena in this thesis are investigations of such processes in three separate contexts-resource competition between different kinds of cancer cells within a growing tumor, consequences of different cancer progression mechanisms for population-level

cancer incidence in humans, and the ontogenetic balance between cell division, differentiation and mutation in determining the scale-specific allometries of cancer risk in mammals.

The first study concerns the ecology of within-tumor interactions between different types of cells within castration-resistant prostate cancer (CRPC). Testosterone signaling plays an important role in the progression of prostate cancer, with several cell types emerging over the course of the disease with varied dependencies on the hormone (Watson et al., 2015). The cellular phenotype space of CRPC has been delineated functionally into three cell types whose interactions provide a reasonable description of the progression of the disease: a non-producing consumer of testosterone, a producer-consumer of testosterone and a non-producing non-consumer of testosterone. Earlier theoretical work using the classical logistic framework from population ecology has shown that the competitive and amensal interactions between these three cell types reflects the clinical progression of CRPC in response to treatments that target testosterone-dependent cell types within the tumor (Cunningham et al., 2018). However, the parameterization of phenomenological models does not always correspond to clinically-measurable quantities and even in those cases where the models have been fit to longitudinal data from human sources, the mechanistic details of the underlying interactions remain largely unclear. In this context, theoretical ecology offers alternative frameworks to model multi-species interactions based on explicit formulations of resource dynamics to describe interactions in mechanistic terms. In Chapter 3, we develop one such model framework for CRPC consisting of the three cell types mentioned above along with two resources-oxygen and testosterone-with interactions between the cell types actualized through resource abundance rather than interaction coefficients. Our data allows us to construct an understanding of the steady state behavior of CRPC in terms of the supply, consumption and utilization efficiencies of the two resources. This also leads to the identification of specific resource supply conditions that can enable the survival of testosterone-dependent, drug-sensitive cell types, as well as the elucidation of how the effects of resource supply are modulated by resource utilization efficiency and initial population size. The implications of such a mechanistic framework are discussed in the context of the clinical monitoring of CRPC as well as the scope for resource-based interventions in novel treatment strategies rooted in tumor ecology like adaptive therapy.

Even as the CRPC study offers a window into the functional importance of ecological processes in cancer, the accumulation of mutations over the course of cancer progression has been understood explicitly as an evolutionary process in a somatic context (Wu et al., 2016). The study of mutation accumulation first began as a way to explain age-related trends in population-level cancer incidence among humans (Armitage and Doll, 1954, Knudson, 1971), and the current paradigm of cancer progression as driven by sequential mutation accumulation is largely an extension of these early *ad hoc* models. With the advent of modern

molecular biology and high-throughput data platforms (Gerstung et al., 2020), the study of cancer epidemiological data as a predictive framework for mechanisms of cancer progression has become somewhat marginalized, albeit with some notable exceptions (Frank, 2007, Meza et al., 2008, Harding et al., 2012, Tomasetti et al., 2017). Chapter 4 presents one such study in which three alternative hypotheses are considered that differ in their assumptions regarding the dynamics of mutation accumulation across individuals in a simulated population. Each of these hypotheses also entails different predictions of how cancer incidence is expected to vary with age at the population level, and we use a combination of mathematical and agent-based modeling to compare these predictions qualitatively to actual epidemiological incidence data. Our results show that the additive fitness effects of subsequent mutations are an important consideration for predicting the age of inflection at which population-level incidence begins to rise. On the other hand, the late-life decline in population-level cancer incidence could only be explained when the pace of mutation accumulation is assumed to be different across different individuals in the population. Surprisingly, this rather straightforward result has not been shown in clear terms so far, which implies that population-level heterogeneity and somatic evolution together dictate the emergence of age-related trends in cancer incidence. The chapter concludes with an overview of the current empirical evidence regarding physiological and other factors that could contribute to such heterogeneity and are amenable to intervention.

The importance of somatic evolution has been clearly emphasized in the context of the growth and spread of human cancers, largely in terms of mutational processes (Aktipis and Nesse, 2013). However, while mutations represent one way in which somatic cell fitness can change, developmental biology has long since established that differentiation entails not only cell physiological changes to enable various bodily functions, but also a gradient of the rates of cell division (Tetteh et al., 2015). Interestingly, all mammals broadly share developmental features across taxa, and the twin processes of cell division and differentiation are central to the dynamics of tissue architecture and turnover across all of them (Hindersin et al., 2016, Derényi et al., 2020, Erten and Kokko, 2020). On the other hand, the allometry of cancer risk with cell number appears to be scale-specific, as the inter-specific, intra-specific, and inter-organ patterns in cancer risk scaling are distinct from each other (Albanes et al., 1988, Noble et al., 2015, Hochberg and Noble, 2017, Gorbunova et al., 2014, Peto, 2015). Chapter 5 thus takes up the question of what determines the circumstances under which having more cells increases the risk of cancer. Using an individual-based model of ontogenetic growth, we identify a complex interplay between cell division, differentiation and mutation that together dictate whether and to what extent the risk of cancer scales with the number of actively-dividing cells. Our results pinpoint parameter combinations that produce all possible allometries, such that cancer risk either increases, decreases or is unaffected by an increase in cell number. The limited empirical evidence is then reviewed

in the context of model results, to speculate on how cancer risk and cell number could be related mechanistically at each level of biological organization.

The final chapter reviews the main findings from each of these studies and identifies promising and logical lines of investigation that are apposite in each case based on the current state of each corresponding field. On the whole, the place occupied by ecology and evolutionary biology in study of living things continues to expand, both in the context of better understanding and mitigating the effects of anthropogenic climate change and in developing new applications of their principles in the realms of health, physiology and medicine. The results and ideas advanced in this thesis are an attempt at furthering this expansion, and pushing these boundaries in interesting and fruitful directions.

# 2 | Selection for dispersal under larval malnutrition results in a non-monotonic kernel

## Key contributions

This study has been designed by Sutirth Dey and me, as part of my PhD work. I have designed and run the experiments for generations 42 and 66-67, and Swaraj Ranjan Gupta has designed and run the experiments for generation 93. Akshay Malwade, Anish Koner and Vinayak Khodake have participated in all stages of data collection, and the former two have provided critical inputs on a later draft of a manuscript resulting from the study. Aarcha Thadi, Subhashree Subhadarshini, Rakesh M, Pranjali Singh, Pratik Joshi, and Ruchitha B G have been instrumental with their support with stock maintenance and data collection. The MD-MC selection experiment has been initiated by Nikita Sabnis, although she has not participated in the design or execution of the current study. I have analyzed the data and have written the manuscript with critical inputs from Sutirth Dey. The following chapter has been reproduced from this manuscript, with minor formatting changes.

The manuscript is currently under review at a refereed publication and is available on bioR$\chi$iv at https://doi.org/10.1101/2023.10.26.564216.

## 2.1 Introduction

Dispersal is one of the primary processes that enable organisms to track favorable conditions in space and time, making it an important consideration in the study of how organisms respond to rapid global climate change (Kokko and López-Sepulcre, 2006, Massot et al., 2008, Clobert et al., 2012). Although it is a complex, multi-component phenomenon, dispersal is usually divided into three distinct stages: (1) emigration/departure, (2) transfer or inter-patch movement, and (3) immigration and settlement (Bowler and Benton, 2005). Each stage involves specific environmental pressures, and the ability of the dispersing propagule to respond effectively to these varied pressures determines its eventual viability (Bonte et al., 2012).

In the departure phase, for instance, the tendency to leave the natal patch is often associated with selection to reduce kin competition (Hamilton and May, 1977), or avoid inbreeding (Bengtsson, 1978, Charlesworth and Charlesworth, 1987) and the effects of demographic stochasticity (McPeek and Holt, 1992, Travis and Dytham, 1998). Similarly, the settlement phase requires accurate assessments of patch quality and resource abundance (Bowler and Benton, 2005), as well as the ability to withstand competition from the natives of the new patch (Léna et al., 1998). Accordingly, organisms are known to respond to broader landscape features like the distribution of suitable habitats (Matter and Roland, 2002, Schneider et al., 2003, Entling et al., 2011), as well as the presence of conspecifics (Serrano and Tella, 2003, Clobert et al., 2009) in making decisions regarding leaving the natal patch or settling at a given patch. Furthermore, movement between patches is particularly vulnerable to a range of biotic and abiotic environmental influences that can be the causes of both mortality (DiLeo et al., 2022) as well as adaptive change (Travis and Dytham, 2002, Poethke et al., 2011, Baines et al., 2020). Factors like temperature and resource availability can act as filters that pose extrinsic mortality risks and thus limit the spatial extent over which a population can viably spread (Denno et al., 1996), while conspecific density has been shown to affect dispersal distances within (Bitume et al., 2013) and across generations (Bitume et al., 2014). On the other hand, constraints during the movement phase could impose further requirements on dispersing individuals, including the development of physiological and morphological features, and behaviors (Zera and Denno, 1997, Bell et al., 2005). This would effectively determine which subset of a given population is able to carry out dispersal, and could affect the spatial segregation of dispersers and non-dispersers over a given landscape. Across its stages therefore, the process of dispersal is acted upon by a range of factors that establish the context in which dispersal occurs, all of which eventually determines the spatial distribution of the population.

An essential feature of dispersal that also offers a straightforward assessment of the spatial distribution of a population is the dispersal kernel (or the dispersal distance kernel), which is defined as the distribution of the frequency of propagules/individuals with distance from the source or a given natal patch (Clobert et al., 2012). The shape and range of the kernel are important features that dictate the speed of population spread under a range of patch distributions (Kot et al., 1996, Clark et al., 2001, Lindström et al., 2011, Ochocki and Miller, 2017), metapopulation structure and dynamics (Travis and Dytham, 1998, Higgins and Cain, 2002, Baines et al., 2020), population stability (Murrell et al., 2002), and phenotypic polymorphism (Travis and Dytham, 1999, Bonte et al., 2010). On the one hand, theoretical investigation using a variety of modeling frameworks has led to significant phenomenological and mechanistic understanding of the ecological and evolutionary consequences of the kernel features like shape and scale (Travis et al., 2010, Lindström et al., 2011). Moreover, while empirical measurements of dispersal are limited by logistic and practical considerations, field data over the years have documented a variety of dispersal kernels across different study systems and ecological contexts (Clobert et al., 2012). To some extent, theoretical studies have also explored the evolution of dispersal-related traits, including the dispersal kernel shape, under a range of conditions (Poethke et al., 2011, Fronhofer et al., 2015). On the other hand, empirical studies investigating the potential for evolutionary and/or adaptive change in the dispersal kernel itself for a given population are relatively scarce. Heritable variation has been reliably identified in multiple aspects of the dispersal phenotype (Saastamoinen, 2008) and in select cases, the evolution of dispersal phenotypes has been demonstrated explicitly in the context of specific environmental changes (Stevens et al., 2012). Given that the dispersal kernel is a cumulative reflection of a large number of individual dispersal events, it is therefore reasonable to expect that adaptive change in the underlying dispersal traits could also lead to measurable shifts in the dispersal kernel. The dynamics of such shifts in the kernel and its potential context dependence remain under-explored.

Previous work has shown that selection for increased dispersal in large populations of the fruit fly, *Drosophila melanogaster* can lead to much more leptokurtic (fat-tailed) kernels than the corresponding control populations (Tung et al., 2018*b*). More importantly, the selected populations in that study had a higher frequency of long-distance dispersers (LDDs), i.e., individuals that travel distances much further than the average of a population and are known to play a disproportionate role in the spatial spread and viability of a population (Petrovskii and Morozov, 2009). LDDs have profound impacts on several aspects of population dynamics and evolution. Both theoretical and empirical studies indicate that dispersal kernels with fatter tails, and thus higher frequencies of LDDs, lead to an accelerating invasion front in an expanding population (Kot et al., 1996, Andow, 2022). The shape of the kernel tail in particular has been shown to be relevant to dispersal-related responses to environmental disturbances (Bianchi et al., 2009). LDDs could therefore play a key role in how organ-

isms respond to progressive habitat fragmentation and other sustained environmental shifts (Murrell et al., 2002). Higher absolute distances covered also relates to the spatial extent of a population (Kot et al., 1996, Murrell et al., 2002, Caton et al., 2022), which corresponds to the physical distances over which a population is seen to exist.

In this context, the distance over which an organism moves should depend on its physiological conditions and various behavioral patterns (Meylan et al., 2002, Clobert et al., 2009, Wijerathna and Evenden, 2019), which in turn will depend on various biotic and abiotic conditions in the natal patch as well as the movement landscape. For example, natal patch quality affects behavioral variability in the initiation of dispersal (Vinatier et al., 2011), as well as decisions related to exploration and settlement (Bonte et al., 2012, Jones et al., 2019). At the landscape level, spatial heterogeneity in resource availability could interact with the energetic costs of searching movements to make the evolution of optimal search strategies highly context-dependent (Zollner and Lima, 1999). Mathematical modeling has also shown that landscape structure, in terms of patch size, quality, and permanence can affect the nature of selection on dispersal (North et al., 2011). At the individual level, the disperser's nutritional status and the mobilization of existing energy reserves are both key determinants of the initiation and success of dispersal events (Saastamoinen and Rantala, 2013, Wijerathna and Evenden, 2019). All these factors can potentially lead to qualitatively different kinds of dispersal kernels under varying nutritional contexts. Does this mean that changing the food available to a population selected for dispersal can alter the shape of the evolved dispersal kernel?

In fruit flies specifically, it is well-known that larval nutritional status plays a crucial role in the amounts of energetic reserves available to the adult fly at eclosion (Bakker, 1959, Robertson, 1960), which would subsequently determine the fly's capacity to actively travel long distances in the absence of food or water (Bonte et al., 2012). Larval nutrition is relatively easier to control over evolutionary timescales as well as single-generation experiments in laboratory populations of fruit flies (Prasad and Joshi, 2003, Tennessen et al., 2014, Tatar et al., 2014). Thus, *Drosophila melanogaster* offers a convenient system to investigate the context dependence of kernel evolution.

In this study, we select multiple populations of fruit flies for higher ambulatory dispersal under conditions of larval malnutrition (20% reduction of dietary protein, Supplementary Material S1 for details). The primary aim here is to look for signs of altered patterns of kernel evolution due to dispersal selection under larval malnutrition, as opposed to dispersal selection under standard larval nutrition. This involves measurements of the dispersal kernel in both selected and unselected populations across multiple generations of experimental evolution, allowing us to characterize several aspects of the dispersal phenotype, and follow

their response to dispersal selection as well as larval malnutrition. The data arising out of these experiments reveal that larval nutrition has significant effects on the extent and nature of dispersal evolution. In particular, while selection for dispersal under larval malnutrition leads to the evolution of a higher capacity for dispersal and greater overall spatial extent, the data shown here suggest that the evolution of long-distance dispersal under larval malnutrition could be a more nuanced outcome. Moreover, control populations not selected for dispersal, but under larval malnutrition, also showed an increase in some components of dispersal. The results of this study thus implicate natal nutrition as a major influence on kernel evolution and spatial spread of a population, and further substantiate the current understanding of the context dependence of dispersal evolution.

## 2.2 Materials and methods

### 2.2.1 Fly maintenance and selection protocol

The maintenance procedure of the selected and control flies were similar to previous studies (see Tung et al., 2018*b*, and the supplementary material therein). Briefly, in the previous study, two populations had been derived from each of four ancestral outbred populations ($DB_i$). The so-called $VB_{1-4}$ populations were selected for higher ambulatory dispersal and the so-called $VBC_{1-4}$ acted as the corresponding controls that did not undergo selection for dispersal. Populations that shared a subscript (for example, $VB_1$ and $VBC_1$) were related by ancestry (i.e. derived from $DB_1$), and were treated as a block in statistical analyses.

In the current study, from each $VBC_i$ ($i \in [1-4]$), we derived two more populations-$MD_i$ (Malnourished Dispersers) and MC (Malnourished Control). As detailed below, MDs were selected for higher ambulatory dispersal using a setup identical to that in the VB selection experiment, while MCs were maintained as corresponding unselected controls. Importantly, while VBs were selected for higher dispersal under standard nutrition, MDs underwent selection for higher dispersal under a larval malnourishment regime, which consisted of a modified banana-jaggery medium with one-third the amount of dry yeast powder as the standard recipe (Supplementary Material S1 and Supplementary Table S1.1). This corresponds to a $\sim 20\%$ reduction in the amount of protein, and initial observations showed that this leads to a slight decrease in egg-to-adult survivorship and an increase in the time to eclosion (personal observation). This is consistent with the known effects of increased larval crowding that have been demonstrated in the context of larval malnutrition, both by overall dilution of all dietary components as well as protein deficiency (Kolss et al., 2009, Lee et al., 2008).

Eggs from both MD and MC were reared until eclosion in this protein-deficient diet, at a density of about 350 eggs per 60mL of food in plastic milk bottles. The adult flies were subjected to dispersal selection on the 12th day after egg collection. The dispersal selection apparatus consisted of a set of two transparent plastic containers, one of which served as the source and the other as the sink. They were connected by transparent plastic piping (internal diameter, $\sim$ 1cm) that served as the dispersal path. The path extended to some distance inside the sink container to reduce back flow of flies from the sink. Neither the source nor the dispersal path contained food or water sources, while a strip of moist cotton was placed in the sink to prevent desiccation after dispersal. Starting from 2m in the first generation, the path length used for dispersal selection was increased intermittently across generations. This served as an approximation of progressively increasing habitat fragmentation, leading to a path length of 54m by the $93^{rd}$ generation when the last of the assays reported here were performed.

On day 12 from egg collection, adult MD flies of both sexes were introduced into the source container and allowed to migrate for a period of 6h, or until about 50% of the flies had reached the sink, whichever occurred first. In parallel, MC flies were introduced into plastic containers identical to the source containers of the dispersal setup that were not connected to a dispersal path. Halfway through dispersal selection, MC flies were given water through a moist piece of cotton, to mimic the desiccation conditions faced by the MD flies that only get water after successfully reaching the source. Following selection, MD flies that reached the sink and all the MC flies were transferred to Plexiglass population cages (25cm $\times$ 20cm $\times$ 15cm) with standard banana-jaggery medium. Yeast supplementation was provided on day 13 and a fresh food surface was provided overnight for oviposition on the evening of day 14. Eggs for the next generation were collected the following day, leading to a 15-day life cycle.

$MD_i$-$MC_i$ with the same subscript $i$ were related by ancestry to the corresponding $VBC_i$, were assayed together and considered a single block in statistical analyses. All populations were maintained at a breeding population size of roughly 2400 flies, in discrete generation cycles under uniform environmental conditions comprising a temperature of 25°C and 24h light conditions. All assays were performed after one generation of common garden rearing under the larval malnourishment regime, but without dispersal selection, to minimize the effects of non-genetic parental effects due to dispersal selection. In order to avoid larval crowding, eggs for pre-assay common garden rearing were collected at a density of about 350 per bottle containing about 60mL of food, and for the assays, about 30 eggs per vial containing about 6mL of food.

### 2.2.2 Kernel experiments and calculations

For the purposes of this study, the dispersal kernel was measured as the number of flies traveling up to a certain distance along a one-dimensional path within a given time period. This corresponds to the accepted definition of a dispersal distance kernel across a variety of study systems (Clobert et al., 2012). Kernels were measured thrice over the period of selection, at generations 42, 66-67 (blocks 1 and 2 in 66, and 3 and 4 in 67) and 93. For each kernel, a source-path-destination setup similar to that used for dispersal selection was assembled, consisting of detachable segments. A 20m path was used in generation 42 with a combination of 0.5m and 1m segments, and a 60m path was used in generation 66-67 comprising only of 1m segments. In keeping with the selection protocol, on day 12 from egg collection, a mixed-sex cohort of about 1000 adult flies of the respective population were introduced into the source and allowed to disperse for 6h. In generations 42 and 66-67, no food was provided in the source container and no water was provided in the destination container. At the end of 6h, the dispersal setup was disassembled and each individual segment was sealed before being placed in a hot air oven at $\sim$ 40°C for about 30 minutes to heat-kill the flies, after which the numbers of male and female flies in each segment were recorded. All the flies from a given segment were considered to have traveled the same distance, calculated as the distance from the source up to the midpoint of the corresponding segment. The $93^{rd}$ generation kernel was performed with a 2m path and no water in the destination container, but included about 25mL of standard banana jaggery food in the source container. The goal of this shorter kernel was to test whether or not dispersal evolution under larval malnutrition resulted in phenotype-dependent dispersal. Since the proportion of flies leaving the source container in presence of food is a sufficient measure for this purpose, a shorter path length was considered appropriate. Three replicate kernels were run for each population in generations 42 and 93, and two replicates were run in generation 66-67, and all of them were conducted two blocks at a time due to logistic constraints.

The above data were used to derive the following characteristic quantities that together provide a simple description of the dispersal kernel: (1) dispersal propensity, calculated as the proportion of total flies that left the source container by the end of the 6h period; (2) dispersal ability, which represents the average distance traveled by the flies of a given population and was calculated as $\Sigma_i \frac{n_i\ d_i}{n_i}$, where $n_i$ is the number of flies in the $i$th segment and $d_i$ is the distance from the source to the midpoint of the $i$th segment. Note that ability was only calculated for flies that left the source container in order to prevent biases arising from different frequencies of flies staying in the source between the two populations; (3) Higher moments including the standard deviation, skew and kurtosis, to describe the overall shape and symmetry of the kernel; and (4) The $90^{th}$ percentile of the distribution of distances traveled, which gives an indication of the spatial extent of the kernel. The shorter kernel

from generation 93 produced only sex-separated estimates of dispersal propensity for each population. Calculations were performed for each block separately, averaging over within-block replicates.

Nonlinear regression was used to further characterize any divergence in the kernel properties between selected and control populations by fitting a curve to the distribution of distances traveled. While the negative-exponential function has been a popular choice for insect dispersal kernels (Taylor, 1978, Hill et al., 1996, Clobert et al., 2012), visual inspection of our data suggested that the MD kernel was non-monotonic while the MC kernel seemed to decay in a more prototypical monotonic decline. The Gamma distribution can mathematically produce both non-monotonic peaks and monotonic declines, and has been applied to a variety of dispersal-related processes (Van Houtan et al., 2007, Guttal et al., 2011). The following form was used to fit the distribution of frequency of flies with distance:

$$f(d) = \frac{c}{\Gamma(k)\ \theta^k}\ d^{k-1}\ e^{-\frac{d}{\theta}}, \tag{2.1}$$

where $f(d)$ is the frequency of flies (ranged [0, 1]) at a distance $d$ from the source. This form included three free parameters whose values must be determined by optimization: $\theta$, the horizontal scale parameter, $k$, the shape parameter, and $c$, an additional multiplicative factor introduced by us to scale the distribution along the vertical axis (Van Houtan et al., 2007, Clobert et al., 2012). As with dispersal ability, the source compartment was ignored for curve fitting. Nonlinear least-squares regression was carried out to fit Equation 2.1 to kernel data from generations 42 and 66-67 using the `nls_multstart()` function from the `nls.multstart` package (Padfield and Matheson, 2020, version 1.2.0) in `R` (R Core Team, 2021, version 4.1.2). Since convergence of a nonlinear regression is highly sensitive to the starting values of the free parameters, `nls_multstart()` was used to run the regression with 500 initial values picked randomly from a given range and the best model was chosen based on AIC (Akaike information criteria) values. Visual inspection of plots of the residuals did not reveal any significant biases in any of the fits, except for the tail of the MC kernel where zero inflation might have contributed to some overestimation towards the end of the kernel (not shown). The Gamma function was fit to frequencies averaged across blocks for the illustrative plots in Figure 2.4, while the fitted parameter values of the Gamma function as well as other statistical information, separated by block and sex, are given in Supplementary Material S2.

As mentioned earlier, long-distance dispersal is primarily studied in the context of individuals that move significantly beyond the average dispersal distance out of a given population. Previous work has indirectly characterized such individuals of a given population based on the distance beyond which the most vagrant 1% of individuals in the population are found

(Tung et al., 2018*b*), which is essentially an estimate of the spatial extent of the population. This was obtained for both the MD and MC populations from their respective frequency distributions by calculating the 99th percentile of all distances traveled. While this is in keeping with our prior work, differences in long-distance dispersal between MD and MC were additionally characterized based on the extremeness of dispersal events in each kernel relative to the corresponding average distances traveled. Hampel filters are commonly used to identify data points that could be considered outliers based on the median absolute deviation and a fixed cutoff value, and are particularly preferred for being agnostic to the exact shape of the underlying distribution (Wilcox, 2010). The frequency of extreme dispersal events could therefore be reflected in the number of outliers identified by the Hampel filter. For the first approach to characterize long-distance dispersal therefore, outlier identification was carried out with a custom `R` script for a Hampel filter with a cutoff value of 2.5 (Wilcox, 2010) and the number of outliers detected was examined in the MD and MC kernels. In the second approach to long-distance dispersal, the aim was to identify whether or not an observed change in the extreme part of the kernel is equivalent to an underlying change in the average of the kernel, or is incommensurate with the latter. To do this, the average of distances above the $95^{th}$ distance percentile was calculated for both MD and MC-this represents the absolute magnitude of the top 5% of all dispersers. This extreme value was then scaled by the corresponding overall average distance, to normalize the extreme value against the mean. If selection for dispersal in MD has led to a higher proportion of long-distance dispersers (LDDs) in addition to greater dispersal in general, this scaled extreme value of the kernel would be higher for MD than MC, in excess of any underlying difference in the average distance moved.

### 2.2.3 Statistical analyses

Unless specified otherwise, all analyses were performed over block-segregated, sex-separated mean values of the response variable, averaging only over the measurement replicates within each block. Data from different generations were analyzed separately such that generation was never included as a factor in any part of the analysis. As detailed below, all models were fit to averages over measurement replicates within a block, and data were never averaged across blocks or sex in any of the analysis, except for the demonstrative plot of the Gamma-fitted dispersal kernels in Figure 2.4. All analyses were carried out on the `R` platform (R Core Team, 2021, version 4.1.2).

For dispersal propensity, mean numbers of flies that left the source and those that stayed in the source were used as "Success" and "Fail" values in a mixed-effects GLM with binomial error distributions and logit link function, with selection and sex as fixed effects, and block as a random intercept. Mean values of dispersal ability were fit to a Gaussian mixed-effects

GLM with selection and sex as fixed effects, and block as a random intercept. Both GLMs were fit using the `lmerTest` package (Kuznetsova et al., 2017, version 3.1-3), and functions from the `emmeans` (Lenth, 2023, version 1.8.4-1) and `psych` (Revelle, 2022, version 2.2.9) packages were used to calculate marginal differences between group means and effect sizes respectively for pairwise comparisons. For the latter, values of Cohen's $d$ were interpreted as large, medium and small for $|d| \geq 0.8$, $0.8 > |d| \geq 0.5$, and $|d| < 0.5$ respectively (Cohen, 2013). Higher moments of the kernel including SD, skew and kurtosis were compared across the four blocks between MD and MC using Wilcoxon's rank-sum tests using the `wilcox.test()` function from the `stats` package (R Core Team, 2021, version 4.1.2).

The procedure for nonlinear regression of the Gamma distribution function has been elaborated earlier. Goodness of fit of the Gamma distribution was also tested using $\chi^2$ and Kolmogorov-Smirnov tests using the `chisq.test()` and `ks.test()`functions from the `stats` package. Fitted values of the Gamma distribution parameters, $\theta$, $k$ and $c$ were used to compare the properties of the kernel. For an unscaled Gamma distribution, the mean and SD can be expressed as $k\theta$ and $\theta\sqrt{k}$ respectively. From Equation 2.1, the vertical scaling factor $c$ can be used to derive a corrected value of $\theta$ that corresponds to the canonical scale parameter such that, $\theta' = \frac{\theta}{\sqrt[k]{c}}$. The mean and SD for the scaled Gamma distribution in Equation 2.1 can therefore be calculated as $k\theta'$ and $\theta'\sqrt{k}$ respectively. $k$ and $\theta'$ can be meaningfully compared between MD and MC, for each of which, Gaussian mixed-effects GLMs were used with selection and sex as fixed effects, and block as a random intercept. Sexual dimorphism in overall kernel shape was investigated using population-wise block-segregated Kolmogorov-Smirnov tests to test the null hypothesis that the male and female kernels are derived from the same distribution.

The unscaled and scaled means of distances above the $95^{th}$ percentile were compared across the four blocks between MD and MC using Wilcoxon's rank-sum tests using the `wilcox.test()` function from the `stats` package.

## 2.3 Results

### 2.3.1 Dispersal components respond to selection under larval malnutrition

The flies selected for dispersal (i.e. MDs) had greater propensity compared to the corresponding controls (i.e. MCs) (Figure 2.1), with the selection × sex interaction being significant at both generation 42 ($\chi^2_1 = 104.6$, $p < 0.001$) and 66-67 ($\chi^2_1 = 46.15$, $p < 0.001$). Pairwise comparisons in generation 42 showed that while MD propensity was higher than that of MC, the difference among females was much greater than that among males (MD vs.

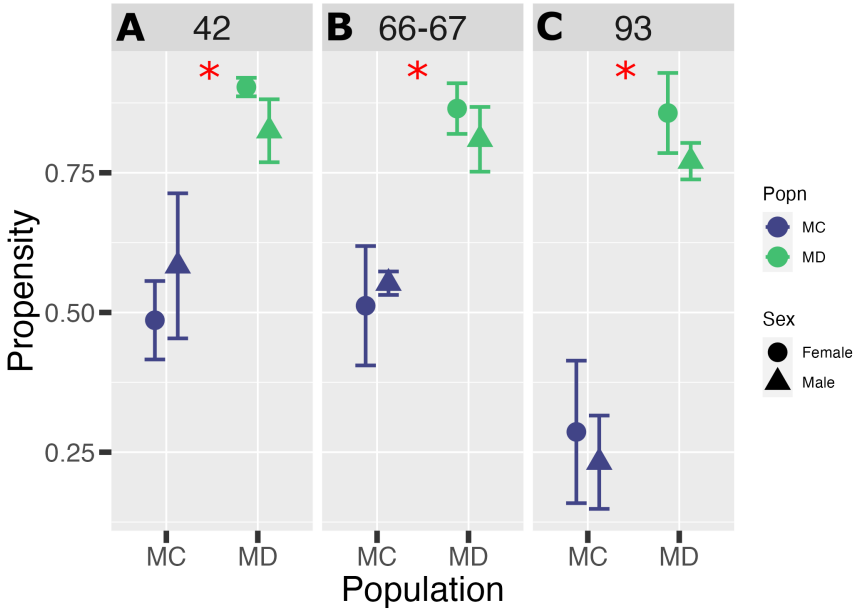



***Figure 2.1:* Increase in dispersal propensity due to dispersal selection**. Assays for generations (A) 42 and (B) 66-67 were performed without food in the source while generation (C) 93 had food in the source. All points are ensemble means ± SD over the four blocks. * denotes a large effect difference (see Supplementary Material S2 for details)

MC, Cohen's $d = 9.4$ for females vs. 2.8 for males; Supplementary Material S2). Data from the $93^{rd}$ generation further show that MC propensity is condition-dependent, and drops to its lowest value in the presence of food even as MD propensity remains largely unchanged (Figure 2.1C). This indicates that while dispersal propensity in MC flies is affected by the presence of food, propensity in MD flies is constitutively high regardless of the presence or absence of food.

Figure 2.2 shows that on an average, the MDs traveled a longer distance compared to the MCs ($\chi^2_1 = 178.67$, $p < 0.001$ in generation 66-67). Comparing trends across generations, it is clear that there is a nearly three-fold increase in ability between the two generations not only in MDs, but also MCs, despite the latter not being selected for dispersal (*cf* Figures 2.2A and B; see Discussion for a possible reason for the latter observation). A notable consequence of this increase is that the dispersal ability of MC at generations 66-67 has in fact exceeded that of MD from generation 42.

### 2.3.2 Kernel shape progressively diverges over ∼ 20 generations of selection

As Figure 2.2 shows, between generations 42 to 66-67, average dispersal ability had increased considerably, in both MC and MD. This suggests that the shape of the overall dispersal kernel could have likewise shifted, both between MD and MC, and over the course

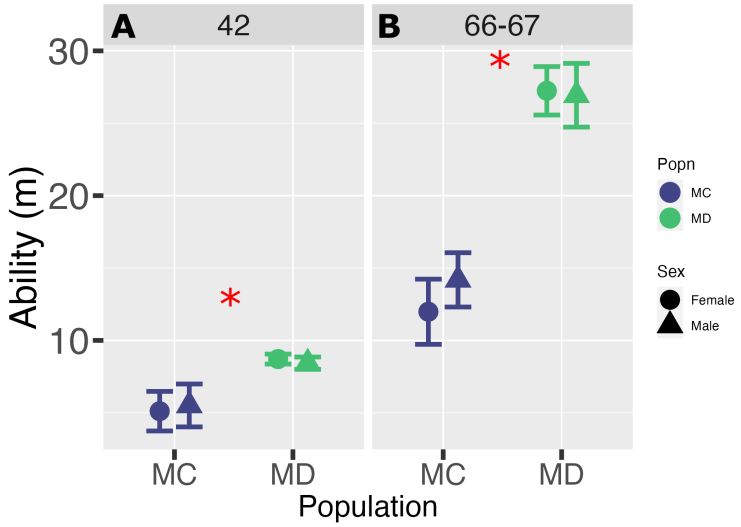



***Figure 2.2:*** **Progressive increase in dispersal ability**. Data only from generations (A) 42 and (B) 66-67 as dispersal ability was not measured in generation 93. All points are ensemble means ± SD over the four blocks. * denotes a large effect difference (see Supplementary Material S2 for details).

of 20 generations of selection.

Figure 2.3 presents a comparison of multiple shape characteristics of the dispersal kernel in terms of the higher moments of the distribution between MD and MC, in generations 42 (Figure 2.3A) and 66-67 (Figure 2.3B). Figure 2.4 shows the actual change in overall kernel shape over the same period. Between MD and MC, in both generations, the MD kernel is associated with lower skew and kurtosis than MC (Figures 2.3B and C, both columns), while a trend (not statistically significant) is seen towards higher standard deviation in MD than MC (Figure 2.3A, Supplementary Material S2). In both generations, MC specifically has a considerable right skew whereas MD kernels are more symmetric. Kurtosis is calculated here as the absolute value, so that a kurtosis value of 3 corresponds to a normal distribution. The data suggest that the MD kernel, although more apparently bell-shaped, is rather platykurtic as its tails decay faster than those of a normal distribution. Although the moment values between the two generations cannot be compared within a single statistical framework, visual examination of the two columns in Figure 2.3 indicates that both the standard deviation and the $90^{th}$ distance percentile of both MD and MC kernels have increased. This is consistent with the trends in average dispersal ability in Figure 2.2, and suggests that over the course of 20 generations of selection, both populations have developed the capacity to spread over much greater physical distances, i.e. increased their spatial extent (*sensu* Kot et al., 1996, Murrell et al., 2002).

The differences seen in the higher moments between both MD and MC, as well as generations 42 and 66-67, are reflected in the shapes of the MC and MD kernels. The MD

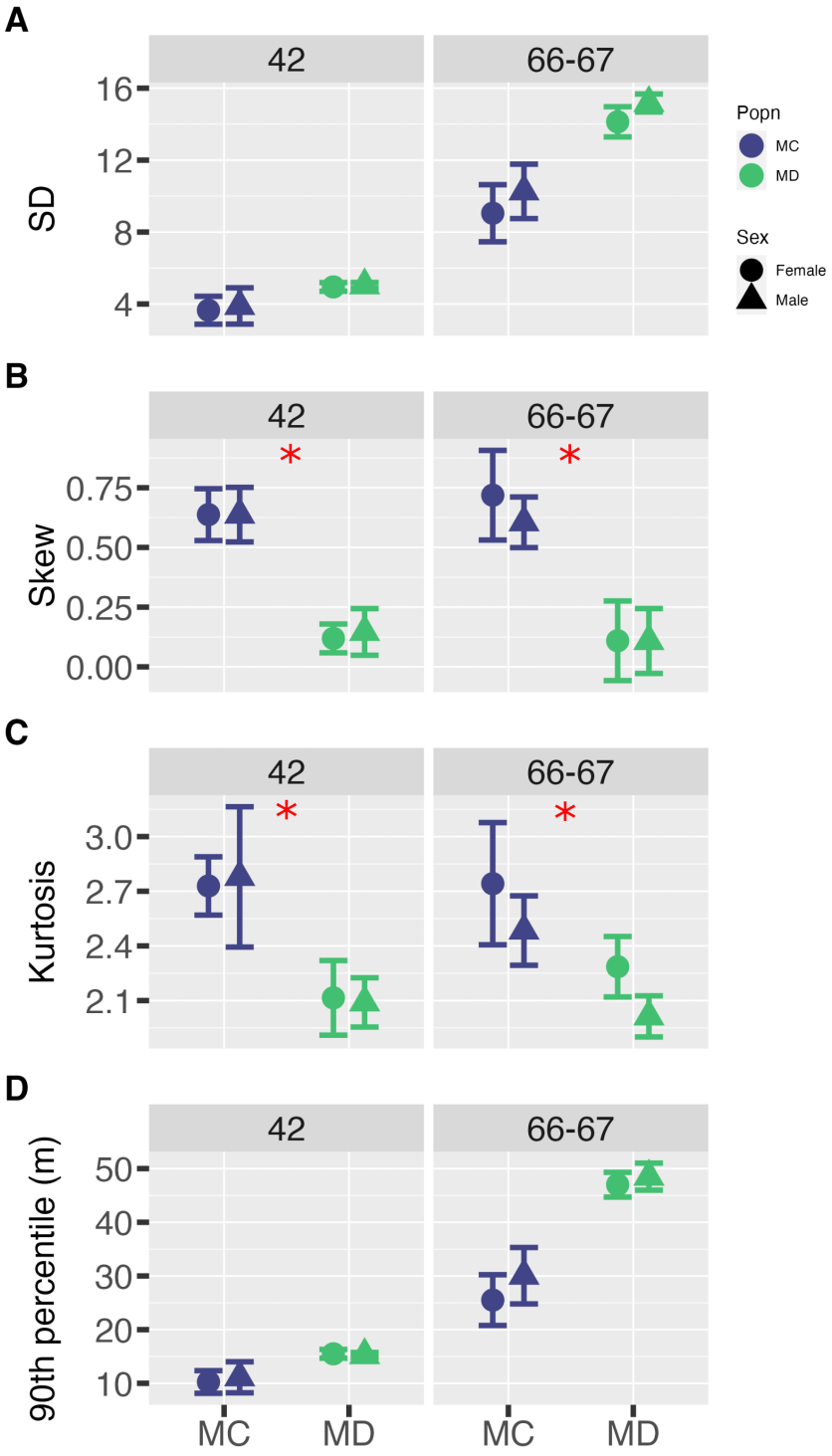



***Figure 2.3:*** **Model-free estimates of the dispersal kernel shape**. Higher moments of the dispersal kernel, including (A) the standard deviation (SD), (B) skew, and (C) absolute kurtosis, along with (D) the $90^{th}$ percentile of the absolute distance moved by each population. All points are ensemble means ± SD over the four blocks. * $p < 0.05$, Wilcoxon's rank-sum test (see Supplementary Material S2 for full information).

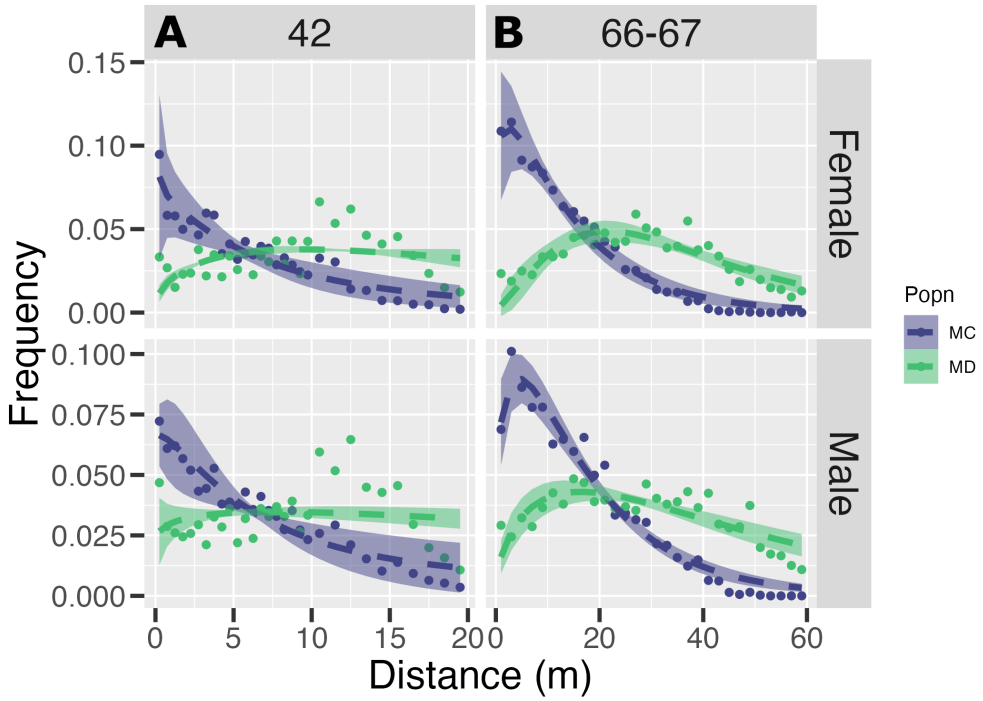



***Figure 2.4:*** **Gamma function as given in Equation 2.1 fit to the distribution of frequencies with distance**. The points are mean observed frequencies across the four blocks, the dashed lines are corresponding predicted values based on the fit and the shaded region is the standard deviation of the predicted values, all calculated as ensembles over the four blocks. Goodness of fit was assessed using Kolmogorov-Smirnov and $\chi^2$ tests to compare the observed frequency with model predictions (see Supplementary Material S2 for goodness of fit, and a list of the fit parameter and AIC values, separated by block and sex).

frequency distribution stretches much further to the right than the MC distribution, and that from generation 42 to 66-67, the physical length of both kernels have increased (Figure 2.4, note the different x-axis limits between 2.4A and B). The fitted values of the Gamma distribution parameters are consistent with this change in shape, as both $k$ and $\theta'$ are on average smaller for MC than MD (Supplementary Figure S6.1.3, Supplementary Material S2). A comparison between generations 42 and 66-67 (*cf* Figures 2.4A and B) shows that the overall divergence between MD and MC has progressively increased over evolutionary time such that the difference in kernel shape between MD and MC is much clearer at generations 66-67 than at generation 42. The fact that the MD population kernels are clearly non-monotonic is noteworthy as empirical insect kernels in literature have been predominantly monotonic declines described by negative exponential or Gaussian tails (Koch et al., 2012).

### 2.3.3 Incipient sexual dimorphism in a subset of dispersal components

Among the various quantities derived from the dispersal kernel, sexual dimorphism in the difference between MD and MC is apparent in propensity while a trend may be emerging in ability (Figures 2.1 and 2.2). The selection × sex interaction for propensity is significant in both generation 42 and 66-67 (Supplementary Material S2). It is clear from the corresponding odds ratios that in MD, male propensity is lower than that of females, while the trend is reversed in MC. A similar trend is also indicated for ability, although neither the effects from the GLMM nor the differences between pairwise means is statistically significant. None of the higher moments calculated from the frequency distribution showed evidence of sexual dimorphism in either generation (full information can be found in Supplementary Material S2).

With respect to the Gamma functions fitted to the dispersal kernels, Supplementary Figure S6.1.3 suggests that differences in $\theta'$ and $k$ between MD and MC could be sex-specific. On the other hand, block-wise Kolmogorov-Smirnov tests revealed no significant differences in the overall shape of the kernel between males and females in either population (full information can be found in Supplementary Material S2).

### 2.3.4 Evolution of long-distance dispersal

Earlier work showed that populations selected for dispersal under standard nutrition had also showed a higher spatial extent, measured as the distance beyond which the top 1% traveling flies could be found (Tung et al., 2018*b*). In order to address this possibility in the current populations, the $99^{th}$ distance percentiles were compared between MD and MC. This corresponds to the estimate of spatial extent used in the previous study, and the trends indicate that as with dispersal evolution under standard nutrition, the MD flies have evolved a much greater spatial extent and could thus be more capable of long-distance dispersal compared to MC flies (Figure 2.5).

On the one hand, the $99^{th}$ distance percentile serves as a simple indicator of the overall spatial extent of the dispersal kernel. On the other hand, when comparing the MD and MC dispersal kernels after subtracting their respective means, the righthand tails of both distributions appear to overlap with each other (Supplementary Figure S6.1.3). This highlights an important aspect of long-distance dispersal, as the "extremeness" of the distance traveled by an individual in a given population can be seen in the context of the average distance moved for that population. Long-distance dispersal can therefore also be conceptualized in terms of the extreme dispersal events occurring in a population that far exceed

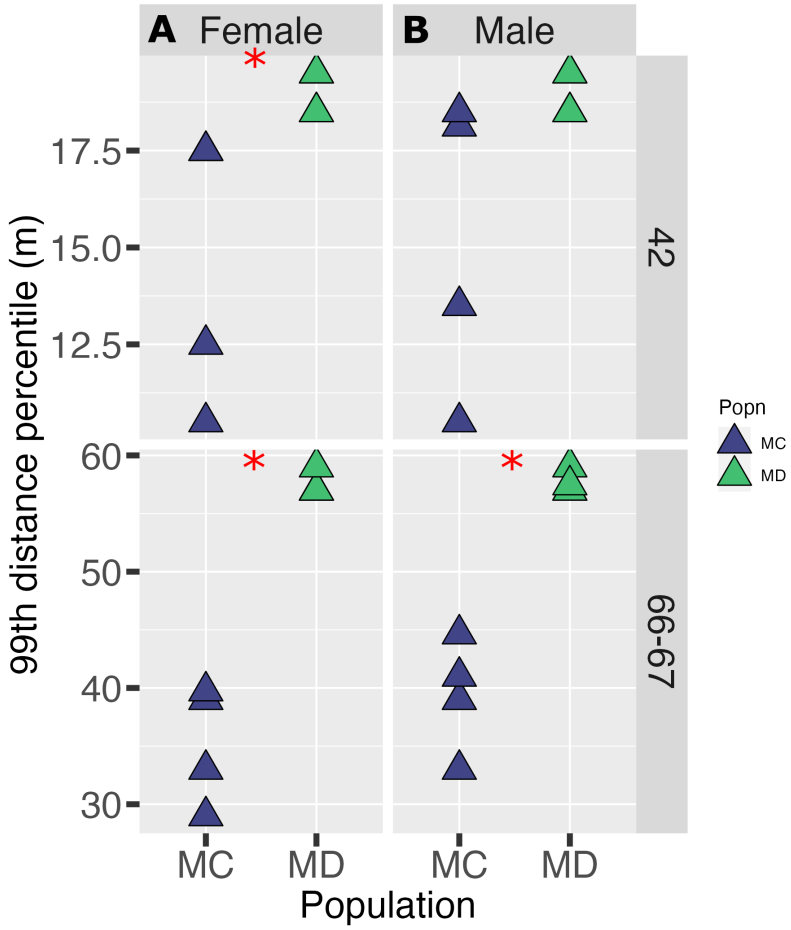



***Figure 2.5:* Spatial extent across generations**. $99^{th}$ percentile of distances traveled by flies of either population, in generations 42 (top row) and 66-67 (bottom row), (A) for females, and (B) for males. * $p < 0.05$, Wilcoxon's rank-sum test.

the observed average dispersal distance for that population. In order to quantify this, multiple methodologies were used, including a Hampel filter for outlier detection. With a cutoff value of 2.5, the Hampel filter failed to identify any outliers at all in the MD kernel while a varying but nonzero number of outliers were identified in MC (Figure 2.6A). This suggests that the overall frequency of long-distance dispersal events could be lower in MD than in MC. the unscaled mean of the top 5% distances traveled is significantly higher for MD than MC (Figure 2.6B, see Supplementary Material S2 for details), which is consistent with the existing difference in the average ability (Figure 2.2B). However, when scaled by the overall mean, the top 5% of distances traveled by MC flies are significantly higher than those of MD (Figure 2.6C, Supplementary Material S2). This is a further indication of the "long-tail" of the MC kernel, as MC flies travel greater distances relative to their mean than MD flies do relative to theirs. The data in Figure 2.6 altogether indicate that the MD kernel is composed of fewer long-distance dispersal events, and while the distances traveled by MD flies are on average much higher than MC, the latter are more extreme dispersers. Finally, Supplementary Figure S6.1.1 shows an alternate way of making the same comparison, based on the

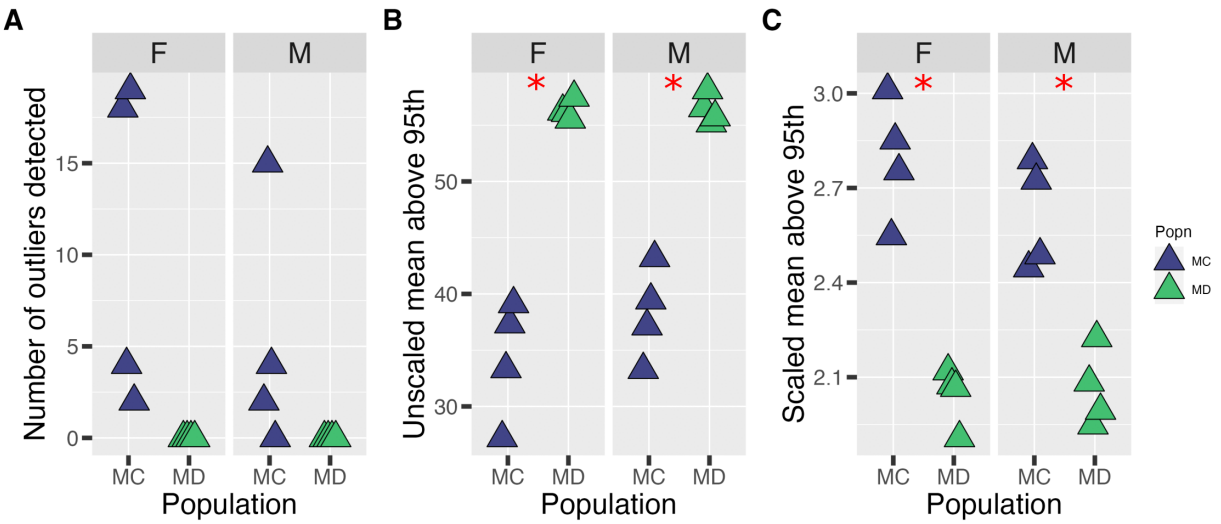



***Figure 2.6:* Measures of long-distance dispersal**. (A) Number of outliers inferred using the Hampel filter with a cutoff value of 2.5; (B) Mean of absolute distances traveled above the $95^{th}$ percentile; and (C) mean of distances traveled above the $95^{th}$ percentile scaled by the mean distance over the entire kernel only for generation 66-67; * $p < 0.05$ by Wilcoxon's rank-sum test. Trends in generation 42 are qualitatively the same and shown in Supplementary Figure S6.1.2, with detailed statistical information in Supplementary Material S2.

frequency distribution of all distances after having the corresponding mean subtracted from the entire distribution. Comparing the percentiles of mean-subtracted distances between MD and MC, no clear difference emerges at long distances from the source. In fact, towards the right hand extreme, the percentiles of MD and MC converge, indicating that the shapes of the kernel tails are very similar after accounting for the difference in distance moved.

To summarize, a simple metric for long-distance dispersal based on the spatial extent of the population shows that MD flies selected for dispersal also evolve greater long-distance dispersal. Using a more nuanced analysis, it is also clear that by accounting for a change in the mean of the underlying dispersal distribution, the increase in long-distance dispersal in MD can be attributed largely to the increase in average dispersal. Taken together, dispersal events in MD do not seem to be as extreme as expected relative to its average distance, whereas dispersal events in MC, despite being shorter in absolute physical distances, are considerably more extreme relative to its average distance.

## 2.4 Discussion

This study intended to examine how the dispersal kernel would evolve if organisms were subjected to a diet with relatively less protein content (compared to a standard laboratory diet). Several insights emerge from the data presented so far.

### 2.4.1 Evolution of short- and long-distance dispersal under larval malnutrition

We find that selection for dispersal can lead to increases in both dispersal propensity and ability, as shown consistently by the MD flies (Figures 2.1 and 2.2). Furthermore, the increase in dispersal propensity is also seen to be phenotype-dependent, as it is not affected by changing the availability of food at the source. Both of these outcomes are consistent with the results of dispersal selection under standard nutrition (Tung et al., 2018*b*), and it is therefore difficult to pinpoint a clear effect of the nutritional environment of dispersal selection based on these results alone. Estimates of the spatial extents of both the MD and MC kernels also indicate that MD flies may be more capable long-distance dispersers than MC flies (Figure 2.5). However, controlling for the average distance moved in each population, we find that the MD dispersal events are less extreme than the dispersal events in MC (Figure 2.6, Supplementary Material S2). These trends may be viewed in the context of a potential physiological maximum that acts as a hard upper limit for ambulatory dispersal in flies. In this light, it is possible that MC flies, which travel shorter absolute distances on average than MDs, are far from this hypothetical upper limit, and thus the MC populations could have the potential for highly vagrant individuals to travel extremely long distances. On the other hand, MD flies could be collectively closer to this hypothetical upper limit, thus restricting the potential in the MD populations for vagrant individuals to cover unexpectedly large distances beyond the population average. While the current data are consistent with such a scenario, this hypothesis must be investigated more closely. More broadly, our results also highlight the importance of conceptual clarity in inferences regarding the evolution of long-distance dispersal. While a metric based on the absolute spatial extent confirms that dispersal-selected flies have also evolved long-distance dispersal, estimates that account for changes in the underlying average dispersal distances allow us to draw more nuanced conclusions regarding the limits of long-distance dispersal evolution in our system. Closer attention must therefore be paid to the interpretative scope and limitations of different conceptualizations of long-distance dispersal.

Our results show modest indications that under larval malnutrition, dispersal evolution could have proceeded in different ways over short and long spatial scales, thus requiring a distinction between short- and long-distance dispersal from each other. Individual-based modeling (IBM) has previously suggested that landscape features such as habitat availability and fragmentation can lead to independent evolution of short- and long-distance dispersal (Bonte et al., 2010). Specifically, it was found that both increased habitat availability and evenly-distributed habitats favor the evolution of long-distance dispersal, whereas increasing habitat fragmentation and lower habitat availability generally discourages long-distance dispersal. In this context, it is relevant to note that both the current and the previous study

on dispersal selection (Tung et al., 2018*b*) involve progressively increasing path lengths in a reflection of habitat fragmentation. Taken together, these two studies indicate that while long-distance dispersal can evolve under conditions of habitat fragmentation under standard nutrition (Tung et al., 2018*b*), the extent of long-distance dispersal is clearly constrained by severity of nutrient limitation in the larval and adult stages preceding dispersal. It is thus clear that theoretical expectations of dispersal evolution under changing landscape regimes, albeit a valuable starting point, need closer consideration of local ecological factors that can attenuate these effects. An indication of an attenuation of this kind can be seen in the documented effects of poor larval nutrition on LDD in our experimental system. As mentioned earlier, we speculate that dispersal evolution under nutrient limitation has brought MD flies closer to their intrinsic physiological capacity for dispersal, thus impeding the evolution of long-distance dispersal (Figures 2.6B and C).

The current study thus documents a potential nutrition-driven separation of components of the kernel, which in turn is consistent with the broader dispersal literature. For example, meta-analyses of dispersal data from mammalian taxa (Whitmee and Orme, 2013) have identified that the predictive power of a wide range of ecological, life history and phylogenetic variables differs considerably between median and maximum dispersal distances, indicating that different parts of the dispersal kernel may be responding to different environmental and physiological factors. Disparate changes in dispersal components has been also documented empirically in butterflies, in which oviposition strategies and other routine movements correlated more strongly with short-distance dispersal (Hovestadt et al., 2011, Stevens et al., 2012). The insights from our study could therefore inform dispersal ecology beyond the laboratory settings of the fly system. It is worth noting that the analyses underlying these conclusions (Figure 2.6) have not yet been replicated with the data from the previous dispersal selection experiment. This is an essential step in further elucidating the role of larval malnutrition in the evolution of short- and long-distance dispersal.

### 2.4.2 Larval malnutrition drives higher dispersal in the absence of dispersal selection

Across taxa and ecological contexts, natal patch conditions has been shown to affect the overall tendency of organisms to initiate and carry out dispersal, as the tendency to leave is known to be associated both with declining quality of the natal patch as well as higher frequency of competitors within the natal patch (Muller-Landau et al., 2003, Fragoso et al., 2003, Bowler and Benton, 2005, Latty and Reid, 2010, DiLeo et al., 2022). Poethke et al. (2011) in particular find that density dependence in decisions related to emigration and/settlement could lead to long dispersal distances using a simulation of a step-wise dispersal process. Trends in the dispersal characteristics of MC between generations 42 and

66-67 offer further indications of the ecological role of natal nutrition. Our data show that under larval malnutrition, MC flies have developed a marked increase in both dispersal propensity (Figure 2.1) and ability (Figure 2.2) despite not being selected for dispersal. It is also worth pointing out that propensity and ability could be independent components of dispersal and as such, affect different aspects of the dispersal process (Fronhofer et al., 2014, Bonte et al., 2012). Interestingly, a recent study has demonstrated that fruit flies under larval malnutrition through protein reduction show increased activity of certain dopamine signaling pathways leading to higher locomotor activity in a dopamine receptor-dependent manner (Zúñiga-Hernández et al., 2023). Therefore, it is possible that the larval malnutrition experienced by the MC flies had led to selection for increased levels of dopamine, which in turn could have increased their locomotor activity across generations. Although it is not possible for us to speculate further on this aspect based on the current data, it is a promising line for future investigation.

### 2.4.3 Insect dispersal kernels can be non-monotonic

Greater average dispersal (Figure 2.2) and limited long-distance dispersal (Figure 2.5 and 2.6) among the MDs together could cause the corresponding kernel to evolve towards a bell-shaped distribution, such that the frequencies of organisms are highest at intermediate distances from the source. On the other hand, the MC kernel retains its overall declining trend despite increasing in spatial extent (Figure 2.4). Prior theoretical work has shown that when the costs of dispersal increase with distance traveled, dispersal distributions could evolve to show peak probabilities of dispersal at intermediate distances from the source (Rousset and Gandon, 2002), i.e., non-monotonocity. Individual-based models have also been used to demonstrate the evolution uni- and bi-modal dispersal kernels in a population of sessile organisms under varying conditions of competition, dispersal costs, and maternal investment (Fronhofer et al., 2015). Empirical studies in plant metapopulations have also found that the persistence time of metapopulations varies non-linearly with the distance between sub-populations that dispersing seeds must traverse, such that intermediate distances are associated with the greatest persistence (Molofsky and Ferdy, 2005). While these observations are broadly consistent with the non-monotonicity of the MD dispersal kernel, their implications for the underlying mechanisms that lead to such kernel shapes in our system remain unclear. Our selection apparatus requires dispersal to occur under desiccation (Mishra et al., 2022), which could form the basis of increasing costs with dispersal distance, although it is difficult to say whether or not this increase is likely to be linear. Such increasing costs could be exacerbated by larval malnutrition, suggesting further possibilities for the evolution of non-monotonic dispersal kernels.

Non-monotonic kernels entail an uneven distribution of individuals across space, which could in turn lead to further spatio-temporal variability in the biotic and abiotic conditions experienced by the dispersers. While the evolution of dispersal is known to be impacted by environmental variability (Denno et al., 1996), spatial structure (Travis et al., 2010) and presence of other organisms (Muller-Landau et al., 2003) individually, the possibility of multiple such factors being tied together by the shape of the dispersal kernel is relatively unexplored, particularly at short to intermediate distances from the natal patch. It could thus be of interest to examine whether and how non-monotonic dispersal kernels can enforce spatial structures in ecological interactions, and what its consequences are for population ecology in general.

# 3 | Malnourished dispersers show sexually-dimorphic trait correlations under dispersal selection

## Key contributions

This study has been designed by Sutirth Dey and me as part of my PhD work. I have run all the experiments, and Akshay Malwade, Anish Koner and Vinayak Khodake have participated in all stages of data collection. Aarcha Thadi, Subhashree Subhadarshini, Rakesh M, Pranjali Singh, Pratik Joshi, and Ruchitha B G have been instrumental with their support with stock maintenance and data collection. The MD-MC selection experiment has been initiated by Nikita Sabnis, although she has not participated in the design or execution of the current study. I have analyzed all the data and have written the following chapter with critical inputs from Sutirth Dey.

## 3.1 Introduction

Dispersal is a complex process that integrates several aspects of an organism's physiology and behavior into how it responds to changing environmental conditions (Ronce, 2007). Dispersal has a well-recognized role in determining the spatiotemporal distribution of populations over ecological timescales (Nathan et al., 2008, Kot et al., 1996), and has long been considered a key factor that affects the stability and dynamics of natural populations (Clark et al., 2001). More recently, the role of dispersal at the level of the individual organism has also come to the fore with the growing understanding of dispersal as a key component of organismal life history (Whitmee and Orme, 2013, Renault, 2020). The latter role is generally thought to stem from the fact that dispersal is both integrative across multiple axes of an organism's biology and is energy-intensive (Bonte et al., 2012). Thus, the evolution of dispersal is likely to affect, and be affected by, its role within the organism's life history strategy (Bonte and Dahirel, 2017). Furthermore, changes in dispersal-related features can potentially alter other aspects of the organism's physiology and behavior, and vice versa (Renault, 2020).

In the context of rapid global climate change, predicting the responses of natural populations to changing environments has emerged as an urgent necessity (Kokko and López-Sepulcre, 2006). The centrality of dispersal in a large number of ecological and evolutionary processes has generated a great deal of interest in developing ways of predicting the direction and consequences of dispersal in a wide variety of systems (Clobert et al., 2012). One particular framework that has been useful in this regard conceptualizes dispersal as a three-stage process (Bowler and Benton, 2005, Bonte et al., 2012). The first stage in the dispersal process is emigration, i.e. when the focal individual leaves the natal patch. This is followed by transfer, i.e. when the individual moves across the landscape, looking for suitable habitat patches. The final stage is that of immigration or settlement, when the individual identifies a destination patch, reaches there and tries to establish itself there. This framework is helpful in placing individual traits of organisms in the context of their role in dispersal. For instance, morphological traits, ranging from sensory structures that enable decision making at various dispersal phases (Kautz et al., 2014, Merckx and Van Dyck, 2007, Edelaar et al., 2008) to physical features related to facilitating actual movement (Zera and Denno, 1997, Nardi et al., 2008), are known to play important roles across all three phases of dispersal. The energetic demands of the transfer phase require a range of physiological processes to ensure sufficient resource availability during inter-patch movement (Zera and Zhao, 2004). There are likewise a wide range of behavioral traits that are linked to specific stages of dispersal-exploratory movement is an important aspect of emigration (Cote et al., 2010, Korsten et al., 2013), locomotor activity and sampling behavior dictate how organisms collect information about the environment (Kempenaers and Valcu, 2017, Arnold et al., 2017, Hanski et al., 2006), and

aggression is one of a spectrum of social behaviors, ranging from cooperative to antagonistic, that are involved in the initiation and success of dispersal (Duckworth and Badyaev, 2007, Guerra and Pollack, 2010). With such a range of traits implicated in the dispersal process, it is straightforward to see that the evolution of a dispersal phenotype has the potential to change any of these traits, while the values of certain traits could in turn limit the evolution of dispersal. The concept of a dispersal syndrome seeks to capture this complex interplay of traits by positing that the evolution of dispersal is invariably accompanied by covariation in a suite of traits that are under selection *en bloc* (Clobert et al., 2009, 2012).

Dispersal syndromes are an active area of investigation in ecology and evolution across several taxa (Clobert et al., 2012), but it remains unclear if the predictive function that is expected out of their study can be satisfactorily met. Much of the uncertainty stems from the fact that the component parts of any potential dispersal syndrome are highly labile, as organismal traits can respond to environmental changes to varying degrees (Bowler and Benton, 2005, Vinatier et al., 2011, Jones et al., 2019). Given this scope for plasticity, it is likely that the composition of a dispersal syndrome would also be highly context-dependent. Moreover, classical life history theory also posits that changes that affect the overall energy budget of the organism could have a significant impact on the relative energetic allocation to various traits and thus create the potential for tradeoffs (Losos et al., 2014). While this is generally accepted in principle (Bonte et al., 2012), empirical investigation of the potential for lability in dispersal syndromes due to environmental changes is still sparse.

Experimental evolution offers the possibility of studying both the evolution of dispersal and its correlated responses in other traits within a single, well-controlled study system (Lustenhouwer et al., 2023). Previous work has demonstrated that lab populations of *Drosophila melanogaster* selected for higher dispersal respond readily and robustly to selection (Tung et al., 2018*b*). In addition to evolving a higher capacity for dispersal, selected populations also show correlated changes in a suite of behavioral and metabolomic traits that all point to the emergence of a dispersal syndrome (Tung et al., 2018*a*). While the previous study was carried out under conditions of standard larval nutrition (standard banana jaggery medium, recipe in the Supplementary Information), more recent work has replicated this selection experiment in an independent set of fly populations under conditions of poor larval nutrition. Initial results from this second selection experiment have shown that flies selected for dispersal under poor larval nutrition also evolved a much higher capacity for dispersal than corresponding unselected control flies (see previous chapter). This background now allows us to address the question whether or not the trait correlations underlying the dispersal syndrome are conserved despite the shift in the nutritional regime, given that the direct response to dispersal selection has remained robust. This is the subject of the current investigation, in which fly populations selected for dispersal under larval malnutrition are examined in terms

of a range of physiological and behavioral traits relative to unselected controls. The data presented here indicate not only that selection for dispersal leads to a distinctly different phenotype from the control population in terms of the traits studied here, but also that the overall dispersal syndrome obtained under larval malnutrition is qualitatively different from that under standard larval nutrition.

## 3.2 Materials and methods

### 3.2.1 Fly maintenance and selection protocol

Detailed descriptions of the maintenance procedure of the ancestral populations have been published elsewhere (Tung et al., 2018*a*, and Chapter 2). Briefly, from each of the four ancestral $DB_i$ populations, two populations were derived to give rise to $VB_i$ (Vagabond) and $VBC_i$ (VB-Control). VBs were selected for higher dispersal and VBCs acted as the corresponding controls that did not undergo selection for dispersal. All populations were maintained at a breeding population size of roughly 2400 flies, in discrete generation cycles, under uniform environmental conditions comprising a temperature of 25°C and 24h light conditions. In the current study, from each $VBC_i$, we derived two more populations: $MD_i$ (Malnourished Dispersers) and $MC_i$ (Malnourished Control). Thus, $MD_i$-$MC_i$ with the same subscript $i$ were related by ancestry to the corresponding $VBC_i$. MDs were selected for higher dispersal using a setup identical to that in the VB selection experiment, while MCs were maintained as corresponding controls not selected for dispersal. The MD-MC populations were raised on a modified banana-jaggery medium with one-third the amount of dry yeast powder as the standard recipe on which VB and VBCs were raised (see Supplementary Material S1 for the recipe). Eggs from both MD and MC were reared in this protein-deficient diet until eclosion, and the adults underwent selection for dispersal on the $12^{th}$ day from egg collection.

As with VB-VBC, the selection apparatus consisted of a set of two transparent plastic containers, one of which served as the source and the other as the sink. The source and the sink were connected by a transparent plastic pipe that served as the dispersal path. Neither the source nor the dispersal path contained food or water sources, while a strip of moist cotton was placed in the sink to prevent desiccation after dispersal. On day 12 from egg collection, adult MD flies of both sexes were introduced into the source container and allowed to migrate for a period of 6h, or until about 50% of the flies had reached the sink, whichever occurred first. During this period, the MC flies were kept in a source container (of identical dimension as the MDs), without any opening for dispersal. Thus, all the MC flies were allowed to lay eggs for the next generation, which implied that they did not undergo selection for dispersal. Post selection, the adults were transferred to population

cages with standard banana-jaggery medium. The adults were provided with live yeast-paste on day 14 and eggs for the next generation were collected on day 15 from the previous egg collection, leading to a 15-day life cycle. The path length used for dispersal selection was increased every few generations to simulate the process of progressively increasing habitat fragmentation. At the start of the experimental evolution, the path length was 2m, whereas by the $83^{rd}$ generation (when the last of the assays reported here were performed) the length had increased to 50m. Since $MD_i$-$MC_i$ with the same subscript were related by ancestry, they were assayed together and considered a single block in all statistical analyses. All assays were performed after one generation of common garden rearing in the larval malnutrition diet but without dispersal selection to minimize the effects of non-genetic parental effects due to dispersal selection. In order to avoid larval crowding, eggs for common garden rearing were collected at a density of about 350 per bottle containing about 60mL of food, and for the assays, about 30 eggs per vial containing about 6mL of food.

Assays to measure various physiological and behavioral traits were performed to elucidate if selection for dispersal under larval malnutrition had led to correlated responses between different aspects of life history. In all cases, assays were performed two blocks at a time due to logistic constraints.

### 3.2.2 Physiological assays

Dry body weight was measured as a proxy for body size in generation 66-77. 12 days post egg collection, flies were sorted by sex under mild $CO_2$ anesthesia, and then placed in 1.5mL Eppendorf tubes, with 20 individuals per tube and 10 such replicates for each sex and population. The flies were killed by flash freezing in liquid nitrogen and dried in the tubes at 60°C for 72h. Following this, the weight of each Eppendorf tube was measured to the closest 0.1mg on a digital weighing balance, first with and then without the flies. The average body weight per fly was calculated as: $\frac{\text{Tube weight with flies - Weight of empty tube}}{\text{\# of flies in the tube}}$.

Fecundity was measured as the number of eggs laid per male-female pair in each population in generation 82-83, assayed two blocks at a time. On day 14 post egg collection, 50 male-female pairs were collected under mild $CO_2$ anesthesia and each pair was placed in a 50mL Falcon tube with holes for ventilation. The lid contained a small food cup that served as a flat surface for oviposition. These set-ups were left undisturbed for 14 hours in a well-lit environment maintained at 25°C. After 14h, the flies were discarded and the number of eggs laid on the food cup was counted under a stereomicroscope.

Development time was measured in generation 66-67 as the time taken for eclosion from the time of egg collection. Each population was given an egg-laying period of about 14h, after which exactly 30 eggs were collected and placed in vials containing about 6mL of the

malnutrition food. 10 replicate vials were set up for each population and the vials were monitored for eclosion starting from around day 8 after egg collection, once pupae had started to darken. Freshly eclosed flies were subsequently collected every 2h under mild $CO_2$ anesthesia, and the numbers of males and females emerging at each time point were recorded.

Desiccation resistance was measured in generation 66-67 as the time it took for a fly to die due to lack of food and water. For each population, same-sex groups of 10 flies each were collected on day 12 from egg collection under mild $CO_2$ anesthesia and placed in fresh glass vials without any food or moisture. 10 such replicate vials each of males and females were set up for each population, and monitored every 2h. At each time point, the number of flies alive was recorded.

### 3.2.3 Behavioral assays

Locomotor activity and resting behavior of adult male flies was assayed in generation 66-67 using the Drosophila Activity Monitoring (DAM2) data collection system (Trikinetics Inc., Waltham, MA) following standard protocols (Chiu et al., 2010, and Text S1.2.1 therein). On day 12 from egg collection, individual males were orally aspirated singly into small glass tubes (length 6.5cm, diameter 5mm), with 32 replicates per population. These tubes were placed in the monitoring apparatus such that two perpendicular IR beams pass through the center of each glass tube. Activity was recorded for 6h after a 15 minutes period for acclimation. The activity for a given fly was estimated as the average number of times the fly crossed the IR beam per hour (Chadov et al., 2015). Continuous inactivity for five minutes or more was considered as sleep/rest and the amount of time spent in rest was calculated as a proportion of the total recording duration (Hendricks et al., 2000, Chiu et al., 2010).

Aggressive behavior was studied in generation 66-67 by means of male-male fighting bouts between MD and MC within each block (e.g. $MD_1$ male against $MC_1$ male and so on). Following published protocols (Yurkovic et al., 2006), approximately 30 eggs were collected from each population and placed in vials containing about 6mL of the malnutrition food. The first few flies to eclose were discarded to collect flies from the eclosion peak and the vials were subsequently checked once every 6h (Yurkovic et al., 2006). As fruit flies take about 8h from eclosion to become sexually active, a 6h monitoring period would ensure that the flies collected would be virgins (Prasad and Joshi, 2003). Flies eclosing in each 6h interval were collected under mild $CO_2$ anesthesia and the males were isolated singly in food vials with about 2mL of food until the day of the assay. This isolation was necessary to prevent social contact with other males as this could lead to the formation of hierarchies through aggressive interactions and affect their performance during the assay (Chen et al., 2002, Kravitz and Fernandez, 2015). About 24h before the assay, the males from both populations

were exposed to daylight fluorescent pigments (DayGlo) by loading a small amount of the pigment powder on the cotton plug of the vial. This coated the fly in the corresponding color, enabling easy identification. On day 12 from egg collection, arenas were assembled in 12-well clear plastic plates (Corning) with a piece of standard banana jaggery food in the center of the well. A decapitated female from an unrelated baseline population was stuck to the center of the food piece with yeast paste and one male of each population was introduced into the arena by aspiration. The food piece and the decapitated female acted as proximate cues to induce the males to fight (Yurkovic et al., 2006). The arenas were recorded using video cameras for a period of 1 hour under well-lit conditions. Since we needed to use two color pigments for distinct identification of MD vs. MC flies within each arena, both populations were assayed in both color schemes to control for any possible biases in behavior due to the specific coloring. For each block therefore, 18 replicate arenas were assayed in which the MD male was loaded with a green pigment and the MC male with pink, and 18 more replicate arenas were assayed in which the MD male was pink and the MC male was green. This lead to 36 replicates of MD-MC fights in each block, 18 in each color scheme. Aggression behavior was scored in terms of victory in fighting bouts, after disregarding the first 5 minutes due to acclimation. Following previous studies (Yurkovic et al., 2006), a given fly was considered the winner of a fight if it successfully chased the other fly out of the food piece three times in a row.

Mating traits were measured in generation 66-67. Following previous studies (Syed et al., 2020, Maggu et al., 2021), virgin males and females were collected for each population from the eclosion peak as described above and isolated in same-sex vials of 8 flies each with about 2mL food. On the day of the assay, one male and one female were introduced into a vial containing about 2mL of 1.3% agar and their interactions were recorded on video for a period of 1h. For each population, 24 such replicate vials were recorded. The solidified agar allowed for the mitigation of desiccation stress. From the recordings, the times when mating began and ended were noted for each vial in which mating occurred within the observed period. Three metrics were calculated from this data to characterize the mating properties of each population. Mating latency, i.e., the time taken for initiation of mating, was calculated as the difference between the mating start time and the video start time. Copulation duration was calculated as the difference between the mating start and end times, and mating propensity was calculated as the proportion of male-female pairs that mated during the period of observation.

### 3.2.4 Statistical analyses

Except for the aggression data, all other statistical analyses were carried out using general or generalized linear mixed-effect models with the `lmerTest` package (Kuznetsova et al.,

2017) in `R` (R Core Team, 2021, version 4.1.2). Dry body weight was analyzed in a mixed-effects general linear model (GLM) with Gaussian error distributions, considering selection and sex as fixed effects, and block as a random intercept. Average activity counts per hour, mating latency and copulation duration were likewise analyzed in separate Gaussian mixed-effects GLMs with selection as a fixed factor and block as a random intercept. Rest fraction and mating propensity were both calculated as proportions, and were therefore analyzed using mixed-effects generalized linear models (GLMs) with binomial errors and a logit link function, with selection as a fixed effect and block as a random intercept. For rest fraction, the response variable consisted of the total duration spent in rest defined as success and the total active duration defined as failure. For mating propensity, the response variable consisted of the number of pairs that mated defined as success and the number that failed to mate defined as failure. For the time to desiccation (10 adults per vial) and time to eclosion (30 eggs per vial) assays, we analyzed the vial means using mixed-effects Gaussian GLMs. Time to eclosion was calculated from the time of assay setup on the day of egg collection up to whenever each fly eclosed. For these GLMs, selection and sex were considered fixed effects, with block as a random intercept. Significance of individual effects or interactions from each model was inferred based on $\chi^2$ statistics from analysis of deviance at the 5% significance level. In all cases, further comparisons of group means within the model were carried out using pairwise differences from the `emmeans` package (Lenth, 2023, version 1.8.4-1) and Cohen's $d$ was calculated using appropriate functions from the `psych` package to estimate effect sizes (Revelle, 2022, version 2.2.9). All group mean comparisons and effect sizes were calculated as the MD-MC difference, and could be positive (MD > MC), negative (MD < MC) or zero. The absolute values of effect size were therefore interpreted as large, medium and small for $|d| \geq 0.8$, $0.8 > |d| > 0.5$, and $|d| < 0.5$ respectively (Cohen, 2013).

The response variable from the aggression assay was the number of wins for a given population. This was pooled across the two color schemes, and compared between MD and MC using a two-sided Wilcoxon's rank-sum test using the `wilcox.test()` function from the `stats` package (R Core Team, 2021, version 4.1.2).

## 3.3 Results

The aim of this study is to investigate the consequences of dispersal selection when applied on a background of a protein-deficient larval diet. Earlier studies have elucidated the effects of dispersal selection under standard nutrition (Tung et al., 2018*b*,*a*). However, malnutrition could alter the allocation of resources across various life processes and thus potentially change the overall patterns of trait correlations and trade-offs. This is examined here by measuring a suite of physiological and behavioral traits in fruit fly populations selected for increased ambulatory dispersal, and comparing their trait values with those of unselected control

populations.

### 3.3.1 Physiological trait responses to dispersal selection are sexually dimorphic

Figure 3.1 shows a summary of all responses in physiological traits because of dispersal selection under larval malnutrition. While sexual dimorphism in adult fly body weight is well-known (Prasad and Joshi, 2003), we also found a significant selection × interaction in the mixed-effects GLM ($\chi^2_1 = 30.11$, $p < 0.0001$; Figure 3.1A). The average body weight of the females of the dispersal-selected populations (MDs) was significantly less than those of the controls i.e., MCs (Cohen's $d = -1$, $t_{153}$ for mean difference = -8.838, $p < 0.0001$), although there was no significant difference among the males ($d = -0.19$, $t_{153}$ for mean difference = -1.078, $p = 0.2827$). Adult body weight in flies could affect fecundity through the total available biomass (Honěk, 1993, Prasad and Joshi, 2003, Tu and Tatar, 2003). Fecundity is therefore measured here as the number of eggs laid in a 14h period on the $15^{th}$ day after egg collection. Figure 3.1B suggests that fecundity is lower in MD than MC, and the corresponding mixed-effects model showed a significant effect of selection ($\chi^2_1 = 4.182$, $p = 0.0408$). While this is in line with the difference in female body weight between MD and MC, the effect size corresponding to the difference in fecundity is rather small ($d = -0.06$). This could therefore be an emerging trend that may progressively diverge with further selection.

Patterns in development time are less straightforward (Figure 3.1D). The mixed-effects GLM reveals no significant effect of selection or sex, and the selection × sex interaction is not statistically significant ($\chi^2_1 = 1.529$, $p = 0.2163$; see Supplementary Material S3 for details). However, pairwise comparison of group means within each sex reveals a significant difference in time to eclosion between MD and MC only in males, albeit with a small effect size ($d = -0.41$, $t_{153}$ for mean difference = $-2.556$, $p = 0.0116$). Females show an even smaller effect size and insignificant mean difference in their development times ($d = -0.15$, $t_{153}$ for mean difference = $-0.807$, $p = 0.4208$). Visual inspection of Figure 3.1C suggests that there could be a considerable difference in the mean time to eclosion between MD and MC, particularly in the males, but one point each in MD males and females appears to be an outlier, which corresponds to the mean time for block 4. Removing this point from the analysis leads to a significant selection × sex interaction ($\chi^2_1 = 7.549$, $p = 0.006$) and large effect sizes for the MC vs MD difference in time to eclosion ($d = -1.3$, $t_{134}$ for mean difference = $-4.422$, $p < 0.0001$ for females; $d = -1.9$, $t_{134}$ for mean difference = $-8.8494$, $p < 0.0001$ for males). Therefore, male MDs could be developing faster than MC males, while a similar trend may be emerging in the females.

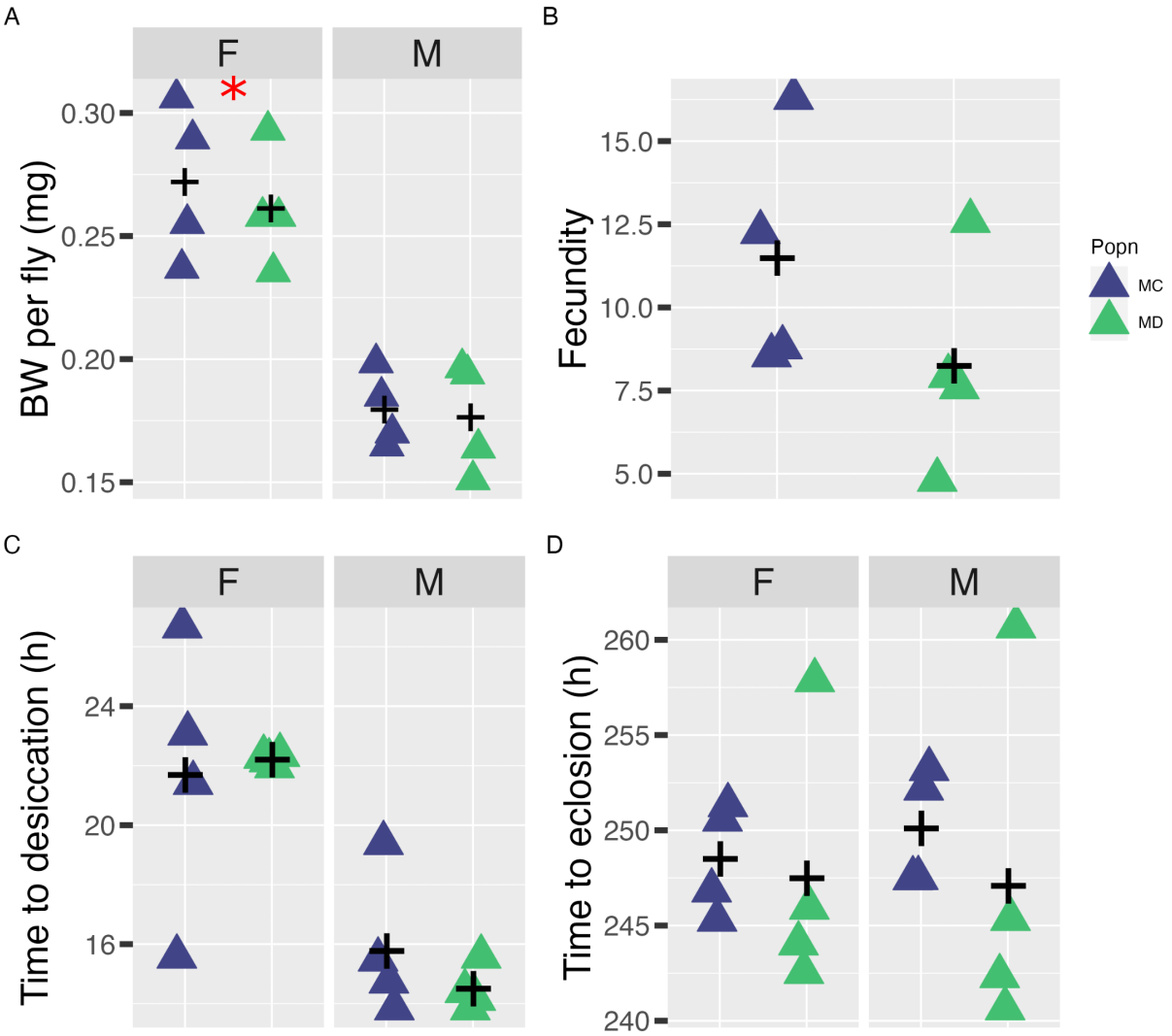



***Figure 3.1:*** **Physiological trait responses due to dispersal selection under larval malnutrition**. (A) dry body weight measured in generation 66-67, except for block 2 which was measured in generation 77; (B) fecundity measured as clutch size per mating pair in generation 82-83; (C) desiccation resistance measured in generation 66-67 as the time taken to death under desiccation; and (D) developmental time measured as the time period between egg collection to adult eclosion in generation 66-67. All points are average values within each block and the black plus sign is the average of the block means. * denotes a large effect difference (see main text and Supplementary Material S3 for further details).

Finally, time to death by desiccation shows a distinctly different MD vs MC trend among the sexes (Figure 3.1D). Both the main effect of sex ($\chi^2_1 = 109.99$, $p < 0.0001$) and the selection × sex interaction are significant ($\chi^2_1 = 5.41$, $p = 0.02$). The main effect of sex is not surprising as *Drosophila* females are heavier and could have greater biomass to survive longer under desiccation (Clark and Doane, 1983, Graves et al., 1992). However, the MD females show a trend towards longer time to desiccation, although the difference is not statistically significant, with only a modest effect size ($d = 0.16$, $t_{152}$ for mean difference = 0.926, $p = 0.3558$). Among males, MDs died significantly faster under desiccation than MC

($d = -0.61$, $t_{152}$ for mean difference $= -2.359$, $p = 0.0196$).

### 3.3.2 Behavioral correlates of dispersal evolution under larval malnutrition

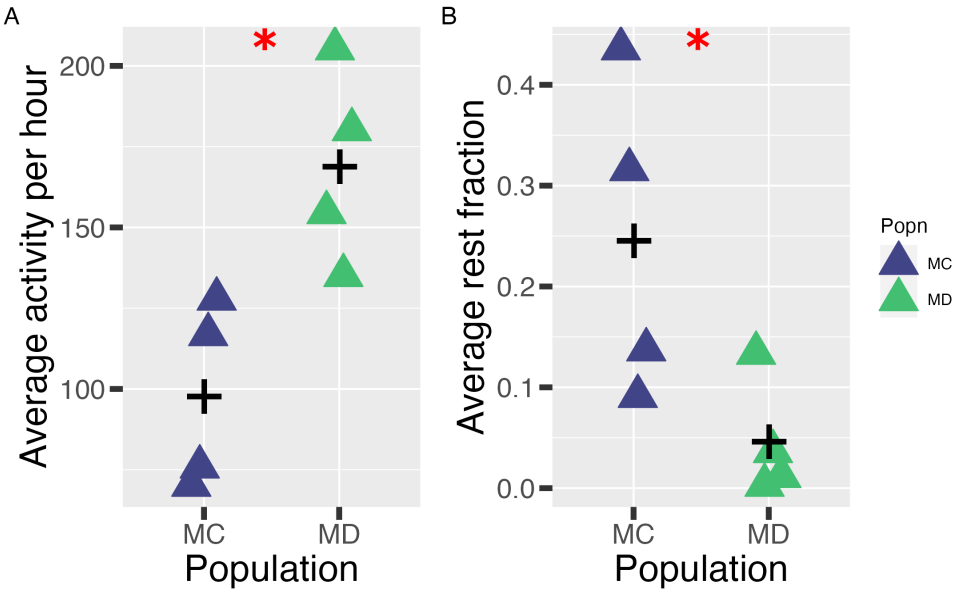



***Figure 3.2:*** **Locomotor phenotype under dispersal selection**. Changes in (A) activity counts per hour and (B) fraction of rest duration due to dispersal selection under larval malnutrition; all measurements were made after 66-67 generations of selection. All points are average values within each block and the black plus sign is the average of the block means. * denotes a large effect difference (see main text and Supplementary Material S3 for further details).

Average activity levels are clearly elevated in MD relative to MC (Figure 3.2A) with a significant selection effect ($\chi^2_1 = 109.88$, $p < 0.0001$) and a large effect size ($d = 1.27$, $t_{251}$ for mean MC vs MD difference $= 10.482$, $p < 0.0001$). Likewise, Figure 3.2B shows that the rest fraction, which the proportion of the monitoring period spent in a rest state, is considerably reduced in MD relative to MC ($\chi^2_1 = 6352.72$, $p < 0.0001$, $d = -0.84$). Together, this points to a more active, restless locomotor phenotype in MD relative to MC due to dispersal selection, paralleling what has been seen earlier due to dispersal selection under standard nutrition (Tung et al., 2018*a*).

Average mating latency is slightly elevated in the MDs relative to that of MCs (Figure 3.3A), although the difference is not statistically significant from the model ($\chi^2_1 = 0.0901$, $p = 0.764$) or from pairwise comparisons of the group means, with the latter producing a very small effect size ($d = -0.05$, $t_{91.6}$ for mean difference $= 0.286$, $p = 0.016$). On the

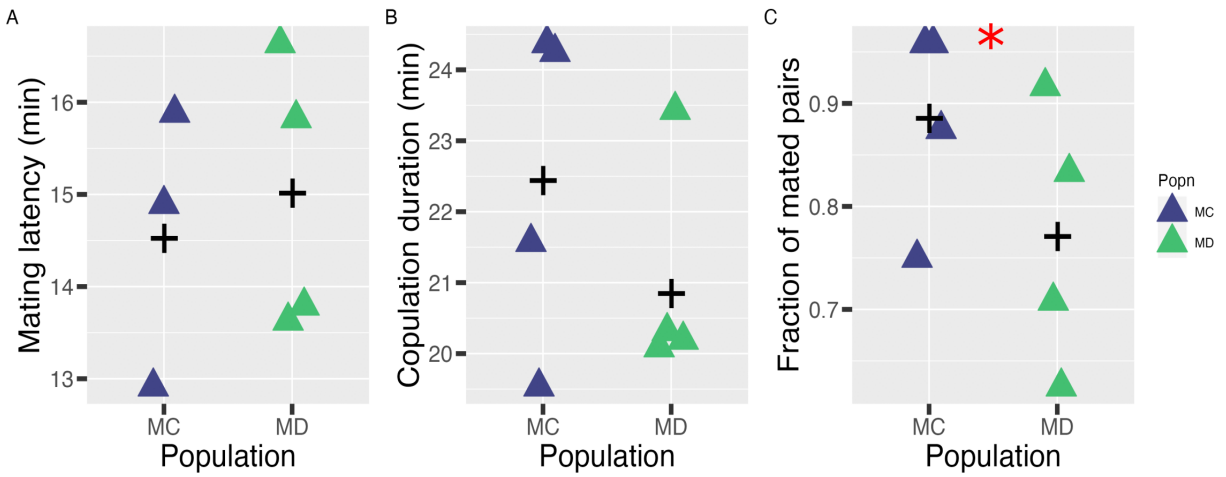



***Figure 3.3:*** **Impairment in mating traits due to dispersal selection**. Changes in (A) mating latency, (B) copulation duration and (C) mating propensity between MD and MC; all measurements were made after 66-67 generations of selection, and all points are average values within each block and the black plus sign is the average of the block means. * denotes a large effect difference (see main text and Supplementary Material S3 for further details).

other hand, MDs had a significantly shorter copulation duration ($\chi^2_1 = 5.794$, $p = 0.016$) even though the corresponding effect size is small ($d = -0.39$, $t_{155}$ for mean difference $= -2.404$, $p = 0.0174$; Figure 3.3B). Mating propensity, defined as the fraction of pairs that mated during the experiment, shows a significant effect of selection ($\chi^2_1 = 4.279$, $p = 0.0385$) and a comparison of the group means shows that the odds of mating propensity for MD are lower than MC (MD/MC odds ratio $= 0.435$, $z = -2.069$, $p = 0.0389$, $d = -1.15$). Finally, the average number of fights won by MC vs. MD is similar, and a Wilcoxon's rank-sum test confirms that the two populations do not differ in their winning tendencies (W $= 12.5$, $p = 0.2425$; Figure 3.4).

## 3.4 Discussion

The above data demonstrate that several physiological and behavioral traits have co-evolved under selection for dispersal. Closer examination of such a suite of traits evolving in association with each other could help further elucidate the nature of dispersal evolution in our selection regime. Moreover, while the trends reported here for MD and MC have been obtained under conditions of larval malnutrition, previous work has already provided a broad understanding of the ecology and evolution of dispersal under conditions of standard larval nutrition (VB vs. VBC, see Tung et al., 2018*a*). It could therefore be useful to compare the trends in the current study to those reported earlier to examine if larval nutrition plays a broader role in the nature and direction of dispersal evolution.

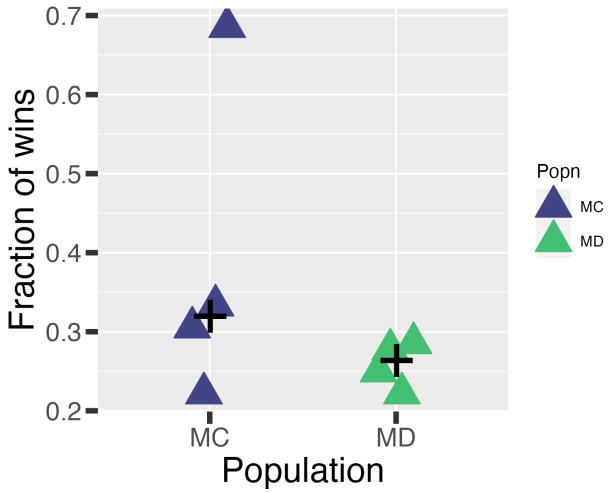



***Figure 3.4:*** **Aggressive behavior is not associated with dispersal evolution**. Number of chase-aways between MD and MC; all measurements were made after 66-67 generations of selection, and all points are number of wins within each block and the black plus sign is the median number of wins for each population across the four blocks.

### 3.4.1 Emerging sexual dimorphism in physiology due to dispersal selection

We found the dry body weight of the dispersal-selected (MD) females to be lower than those of the control (MC) females (Figure 3.1A). As dry body weight is typically used as a proxy for body size and, to some extent, body condition (Knapp, 2016), the reduction in body weight could imply that the MD females are showing signs of a cost of dispersal selection. An association between body size and dispersal has been documented across taxa, with or without distinct dispersal polymorphisms (Zera and Denno, 1997, Braendle et al., 2006, Jenkins et al., 2007, Asplen, 2018). In general, body size is expected to be positively associated with dispersal, either through mechanical improvements in the efficiency of locomotion (Nardi et al., 2008, Lancaster and Downes, 2017, Berwaerts et al., 2006, Ohlberger et al., 2006) or the availability of greater energetic reserves (Reim et al., 2019). On the other hand, a negative association of body size with active dispersal has been found in some cases in which a larger or heavier body increases the energetic cost of locomotion (Dahirel et al., 2021). The net association between body size and dispersal in active dispersers can thus be seen as the resultant of the extent to which a larger body size leads to potentially greater energetic reserves as opposed to increased energetic demands. Lower body weight among female dispersers in our study could be indicative of the latter case, where large body size under a poor nutrition regime might be a substantial energetic demand for dispersers. In turn, this could represent a direct cost of dispersal evolution.

The reduction in body weight in the MDs is consistent with their lower fecundity (Figure 3.1B), as the number of eggs laid is generally expected to be positively correlated with fe-

male body weight in *Drosophila* (Prasad and Joshi, 2003). Extending the reasoning for the body weight difference then, lower fecundity in MD could indicate a further cost of dispersal evolution. However, lower fecundity in the dispersers in our study could also be a different form of the flight-oogenesis syndrome, which signifies the reduction and/or postponement of reproductive investment *vis-à-vis* dispersal (Roff, 1990). Although closely associated with wing dimorphism in migrating insects (Mole and Zera, 1993, Zera and Denno, 1997, Lorenz, 2007), temporal separation between reproduction and dispersal has been recorded in monomorphic migrant butterflies (Stefanescu et al., 2021). Allocation differences have also been demonstrated between dispersal and fecundity between range margins and cores, particularly favoring dispersive strategies at margins at the expense of reproductive investment (Hughes et al., 2003). Hormone-signaling based modification of metabolism serves as a common mechanistic link between dispersal and reproduction for wing-dimorphic crickets (Zera and Zhao, 2004) and paper wasps (Southon et al., 2020), while processes are yet to be identified that could elucidate a dispersal-reproduction tradeoff in monomorphic insects. Given that female dispersers in this study show a trend consistent with such a trade-off (Figures 3.1A and B), this offers a suitable experimental system in which to address the corresponding physiological mechanisms.

Although MD males do not differ from MC males in terms of their dry body weight (Figure 3.1A), they do show a trend towards lower desiccation resistance (Figure 3.1D). This suggests that, unlike with the females, the cost of dispersal evolution could be manifesting in terms of desiccation resistance rather than body weight in the males. While body weight is generally positively associated with stress tolerance (Clark and Doane, 1983, Arnett and Gotelli, 2003, Knapp, 2016), desiccation resistance in particular is closely related to the total body levels of glycogen rather than the actual biomass (Graves et al., 1992). It is therefore possible for MD and MC males to differ in terms of their glycogen content without a significant difference in net biomass, which would explain the difference in desiccation resistance between the two populations. A possible explanation for the evolution of a male-specific desiccation resistance could be the differences between the sexes in how dispersal selection is actualized through our selection protocol. Since oviposition for the next generation happens exclusively after dispersal, female reproductive success in our experiments is contingent on successfully reaching the destination patch. On the other hand, male reproductive success can be realized through mating at any point after reaching sexual maturity, regardless of whether those mating opportunities arise before, during or after the dispersal event. It is thus possible that selection on males to reach the destination patch successfully is weaker than that on females, which might reduce the need for improved desiccation resistance in the former. It is interesting to note here that experiments in the previous chapter have shown that among the MDs, males have a lower tendency to initiate dispersal than females. Taken together, these trends indicate that the sexes have responded differently to dispersal selection.

Finally, our data indicate that there were no effects of selection for dispersal on the development times of either sex (Figure 3.1C). However, this lack of effect was driven entirely by one population (MD4). When we remove the data for that one population, dispersers of either sex tend to develop faster than non-dispersers. Within the context of a so-called pace-of-life syndrome, higher dispersal is expected to be associated with a faster growth rate (Réale et al., 2010), which is consistent with the developmental time responses in both sexes in our study. In general, adult body weight is known to be positively associated with the time taken for eclosion in Drosophila (Prasad and Joshi, 2003, Tu and Tatar, 2003). Therefore, all else being equal, our MD flies are expected to be lighter than the MC flies, which was borne out in the females, but not in the males (Figure 3.1A).

On the other hand, while females in our study seem to show a cost of dispersal evolution through body weight, males seem to manifest the cost of dispersal in terms of desiccation resistance. This indicates that despite some common developmental associations, dispersal selection has also led sex-specific physiological responses. In this context, it is important to note that whether or not a faster growth rate leads to lower adult body weight is known to be dependent on the development phase in which growth acceleration has taken place (Vijendravarma et al., 2012). It is therefore possible that disperser males and females could have evolved growth acceleration in different developmental phases. This in turn could involve different sets of physiological mechanisms operating in a sex-specific manner, leading to sex-specific costs of dispersal. If substantiated, this would add to a growing pool of knowledge regarding sexual dimorphism in dispersal evolution (Mishra et al., 2018, Li and Kokko, 2019).

### 3.4.2 Behavioral correlates of dispersal suggest the involvement of dopamine signaling in dispersal evolution

Given the way in which dispersal selection has been carried out in this study, locomotor activity represents one of the most direct indicators of the response to selection. As expected, the dispersal selected populations had greater locomotor activity over a 6h period than unselected controls (Figure 3.2A). Likewise, the fact that selected flies tend to rest for a smaller part of the total duration (Figure 3.2B) points to a more active and restless behavioral phenotype, which could be considered indicative of a general tendency arising due to dispersal selection. The current data also suggest a general impairment in mating characteristics associated with dispersal evolution. This is made clear by the reduction in average duration of first copulation (Figure 3.3B) and overall propensity to mate (Figure 3.3C), as well as increase in the time to first mating (Figure 3.3A) for the MD flies.

Behaviors like locomotor activity, wakefulness, and mating characteristics are known to be controlled by monoamine signaling pathways in the brain, which includes dopamine, octopamine and serotonin (Crickmore and Vosshall, 2013, Nall and Sehgal, 2014). In fruit flies, dopamine activity in particular has been shown to be a key mediator of increased locomotor activity and reduced sleep (Kume et al., 2005, Sitaraman et al., 2015). Furthermore, recent evidence has shown that poor larval nutrition can significantly increase adult activity by upregulating dopaminergic neuronal pathways (Zúñiga-Hernández et al., 2023). This is consistent with personal observations, which indicate that average levels of locomotor activity are generally elevated in both MCs and MDs relative to other fly populations raised on standard nutrition (unpublished data). Therefore, it is tempting to speculate that the increased activity in the MD flies might be driven by increased levels of dopamine. However, a scrutiny of the trends reveals that the situation might be more complicated.

In the MD flies, higher locomotor activity is seen to be associated with no changes in mating latency (Figure 3.3A), but a significant reduction in copulation duration (Figure 3.3B) and mating propensity (Figure 3.3C). However, increased dopamine levels in adult flies are expected to enhance mating receptivity in females (Neckameyer, 1998) and courtship behavior in males (Lim et al., 2018, Love et al., 2023) in fruit flies, both of which (singly or in combination) should lead to a reduction in the mating latency and an increase in the mating propensity. Thus, it appears as if the trends in the mating-related traits do not match with those expected from an increase in dopamine levels. However, further complications await.

Forcible induction of higher dopamine signaling within the first few hours of adult eclosion has been shown to impact neuronal development negatively, leading to impaired mating behavior later in life (Kayser et al., 2014). This suggests that while increased dopamine signaling generally enhances locomotor activity, its consequences for mating performance depend on when in the fly life cycle the dopamine levels are elevated. Selection for dispersal in our system happens within 2-3 days from eclosion, and the locomotor activity of the flies is high at this time (Figure 3.2A). This suggests that dopamine activity could also be elevated during the same time, which in turn could potentially explain impaired mating in the MDs. However, it remains unclear if the phase of higher locomotor activity identified in our flies corresponds sufficiently to the critical time period of early adult life in which the importance of dopamine signaling has been demonstrated (Kayser et al., 2014). Further investigation would be thus required, both to confirm elevated dopamine activity in our disperser flies early in life and to establish a causal role of this increased signaling in impaired mating. It is worth mentioning that the neuronal processes surrounding dopamine signaling in the fruit fly are complex, involving multiple classes of both excitatory and inhibitory receptors that are implicated in a wide variety of functions (Zhang et al., 2016). The above speculative mechanistic connection between locomotor activity and mating could therefore be

an interesting starting point towards understanding the underlying mechanisms of dispersal evolution. This is also in line with current thinking on the elucidation of so-called behavioral syndromes, which hinges on the ability to identify physiological mechanisms that connect otherwise disparate behaviors (Campos-Candela et al., 2019).

### 3.4.3 Dispersal "syndromes" under standard vs. poor larval nutrition

Finally, we examine the findings of this study in the context of an earlier work on the evolution of dispersal under standard nutrition in *Drosophila melanogaster* (Tung et al., 2018*a*, Mishra et al., 2018, 2022). The two studies used flies from the same genetic background, the same dispersal set up, similar environmental conditions, and mostly assayed similar traits, thus making them eminently comparable. Given the difference in the nutritional background, this comparison could shed light on the context dependence of dispersal evolution and its correlates.

Table 3.1 shows a summary of both behavioral and physiological traits whose correlated evolution with dispersal were compared between standard and poor larval nutrition. It is clear that with the exception of desiccation resistance, locomotor activity and rest duration, all other traits have responded differently to dispersal selection between the two nutritional regimes (note that development time is not included in this table as comparable measurements were not available from the standard nutrition populations). Taken together, there are clear indications that protein deficiency through the larval stages represents a sufficiently severe stress for several trade-offs to develop in terms of the physiological traits. The shift in the difference in male-male aggression between MD and MC, apart from being an additional tradeoff of dispersal, could be of interest regarding the mechanism of dispersal evolution under larval malnutrition. The flies selected for dispersal under standard nutrition had elevated levels of octopamine, which was thought to be the reason for their increased aggression (Tung et al., 2018*a*). The lack of a similar increase in aggression in the MD flies (Figure 3.4) suggests that octopamine signaling may not be as involved in dispersal evolution under malnutrition.

| **Trait** | **Poor nutrition (66-83)** | **Standard nutrition (49-70)** |
|---|---|---|
| Dry body weight | Lower in females | No difference |
| Early-life fecundity | Lower | No difference |
| Desiccation resistance | Lower in males | Lower |
| Locomotor activity | Higher ** | Higher ** |
| Rest duration | Lower ** | Lower ** |
| Male-male aggression | No difference | Higher |
| Mating characteristics | Latency ↑, Duration ↓, Propensity ↓ | No difference |

***Table 3.1:*** **Comparing trait correlations of dispersal evolution under varying conditions of larval nutrition**. The second and third columns describe the change in each trait in the disperser against the non-dispersers; lower therefore implies that the trait value was lower in the dispersers than the non-dispersers, and so on. The numbers in parentheses indicate the generation numbers over which data were collected in each case. Poor nutrition studies were carried out in the MD-MC populations (second column), while standard nutrition studies were carried out in the VB-VBC populations (third column), as published elsewhere (Tung et al., 2018*b*). ** Only measured in males.

# 4 | A resource-based mechanistic framework for castration-resistant prostate cancer (CRPC)

## Key contributions

This study has been designed by Sutirth Dey and me, as part of my PhD work. A subset of the study also formed the Master's thesis project of Harshavardhan B V, who has written the original codes for the simulations, plotting and analysis, and has contributed part of the data in the final study. I have run additional simulations, and have written a manuscript based on the results of the study with critical input from Sutirth Dey and Harshavardhan. The following chapter has been largely reproduced from this manuscript. While some of the data and analysis in the current chapter appear in Harshavardhan's Master's thesis (B V, 2021), I have placed them in a different scientific narrative, and have included newer data and analyses produced by me as part of the study in its current form. Likewise, while the current chapter uses the same system of differential equations as in Harshavardhan's Master's thesis, the model description itself has been re-written for the current chapter and is not a reproduction of Harshavardhan's Master's thesis.

The manuscript is currently under review at a refereed publication and can be accessed on bioR$\chi$iv at https://doi.org/10.1101/2023.09.19.558379.

## 4.1 Introduction

Prostate cancer is among the leading causes of mortality in men (Cancer Society, 2023), and both its clinical progression and therapy have been the subject of intense theoretical investigation (Basanta et al., 2012, Zhang et al., 2017, Cunningham et al., 2018, West et al., 2020, Zhang et al., 2022). Research over the past few years has revealed an important role for intra-tumor competition in prostate cancer progression, particularly in drug-resistant relapse in advanced castration-resistant prostate cancer (CRPC) (Zhang et al., 2017, Gallaher et al., 2018, Brady-Nicholls et al., 2020). While intra-tumor competition is often driven by a diversity of cancer cell types occupying the tumor microenvironment (Gatenby et al., 2009, Kareva et al., 2015, Carreira et al., 2014, Gedye and Navani, 2022, Madan et al., 2022), such cellular heterogeneity also gains therapeutic relevance when a subset of cancer cells in the tumor are resistant to a given drug treatment. Investigations of such tumor populations comprised of a diversity of drug sensitivity phenotypes have demonstrated that drug-resistant relapse under conventional treatment approaches is primarily due to selective elimination of drug-sensitive cells from the tumor by therapy, which in turn allows drug-resistant cells to proliferate more freely (Gatenby et al., 2009, Marusyk et al., 2014, Sahoo et al., 2021, Farrokhian et al., 2022). Current research efforts are therefore aimed at the design of alternative therapeutic regimens that can prevent such competitive release of drug-resistant cells across a variety of cancer types (Hansen and Read, 2020, West et al., 2020).

Prostate cancer is well-characterized in terms of its clinical progression (Gordetsky and Epstein, 2016, Montironi et al., 2018) and the emergence of advanced castration-resistant prostate cancer (CRPC) is considered to be driven primarily by cancer cell types that are resistant to chemical suppression of systemic androgen growth factors through androgen deprivation therapy (ADT) (Culig et al., 1999, Grasso et al., 2012, Fontana et al., 2022). In this context, progress towards alternative strategies to treat CRPC that can better manage drug resistance has been driven to a great degree by a model by Cunningham et al. (Cunningham et al., 2018), which used a logistic framework to cast CRPC as a system of three cell types in density-dependent competition with each other. Uniquely for CRPC, one of these cell types can drive tumor expansion by autonomous production of the androgen, testosterone, which two of the three tumor cell types require for their growth. This adds a commensal dimension to the existing competitive interactions in the system between producers and consumers of testosterone. Cell-autonomous production of testosterone is also thought to drive the clinical progression of prostate cancer towards castration resistance by enabling tumor growth without systemic testosterone supply, which, as mentioned earlier, is directly relevant to the development of CRPC following androgen-deprivation therapy (ADT).

Cunningham et al. (Cunningham et al., 2018) found that the interaction strengths between the cancer cell types, as described by the interaction coefficients, could be used to classify the dynamics and steady states of the model as "Best Responders", "Responders" or "Non-responders", to a drug that inhibits testosterone production. Only the testosterone-dependent cell types are sensitive to the inhibitory drug. Therefore, in absence of inter-conversion between these cell types through mutation, a higher frequency of the testosterone-independent cell type will lead to an increasingly drug-resistant tumor population. The relative frequencies of testosterone-dependent cell types therefore determines the "responsiveness" category to which the tumor population is assigned. As with any logistic framework, the steady state stability and cell-type composition in such a model are determined by the values of the interaction coefficients. Thus, the stratification of model behavior based on drug responsiveness is essentially the identification of qualitatively different points in the parameter space for interaction coefficients.

The Cunningham et al. model represents the use of a classic ecological model of interspecific competition to functionalize tumor heterogeneity within a framework for therapeutic intervention in CRPC. While models of this kind provide valuable phenomenological insights into the progression of CRPC, it is difficult to relate their findings to physiological processes operating within a tumor community. Parameters like the interaction coefficients are usually not directly informative about the underlying mechanisms of cell-cell interactions, and the lack of direct mechanistic links to tumor processes could limit the translational applicability of such models.

Classical theoretical ecology, on the other hand, offers other conceptual frameworks that more explicitly account for the mechanisms of inter-specific interactions. Resource-consumer models represent one such framework within theoretical ecology that cast interspecific interactions entirely in terms of production, supply and consumption of, and competition for, physically explicit resources (Tilman, 1980, Grover, 1997, Muscarella and O'Dwyer, 2020). Their mechanistic roots arguably makes these models harder to parameterize than a logistic framework, which might explain their limited use in cancer systems so far (Kareva et al., 2015). Nevertheless, a resource dynamics based approach offers considerable potential to relate tumor ecology to measurable quantities within the tumor microenvironment.

In this study, we add a mechanistic element to the phenomenological description of the Cunningham et al. model by including the temporal dynamics of two resources-oxygen and testosterone. Oxygen, or the lack thereof (i.e. hypoxia), has been implicated in at least three separate cancer hallmarks including deregulated cellular energetics, angiogenesis, and metastasis (Hanahan and Weinberg, 2011), which makes it a biologically-pertinent inclusion in this model. As recognized by earlier work, testosterone is indispensable to any mechanistic

description of the prostate tissue, cancerous or otherwise, as androgens are an integral part of prostate homeostasis (Mohler et al., 2004, Page et al., 2006, Calistro Alvarado, 2010). Furthermore, the demonstrated relevance of within-tumor testosterone production for the progression of advanced prostate cancer (Titus et al., 2005, Montgomery et al., 2008, Stanbrough et al., 2006, Watson et al., 2015) makes it a justifiable addition to the model.

The inclusion of resource dynamics in our model entails additional processes like cellular resource production and consumption that must be parameterized, and wherever possible, we have determined these parameter values based on available empirical measurements of the corresponding processes. Within this theoretical framework, our results enable us to construct an understanding of the steady state behavior of the CRPC system in terms of the differential resource use properties of three distinct types of prostate cancer cells. Our data broadly demonstrate that resource consumption processes could potentially elucidate underlying mechanistic links within a system that is otherwise largely phenomenological. In particular, we find that the efficiency of resource use and abundance regimes of resources both play important roles in determining the steady-state composition of the tumor. We also identify resource use efficiency as an interesting aspect of tumor-intrinsic properties that could potentially be used to predict the sensitivity and responsiveness of a given tumor to testosterone-targeting therapy.

## 4.2 Methods

### 4.2.1 Model framework

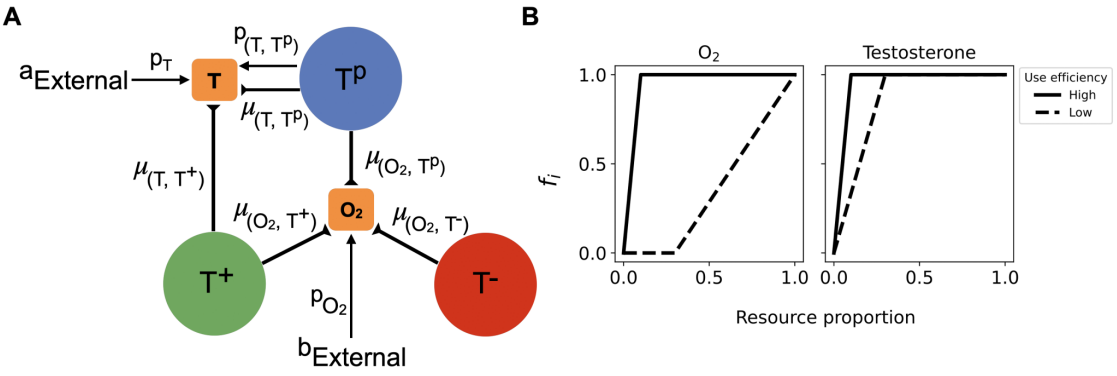



***Figure 4.1:*** **Schematic showing the components of the current model**. (A) Three cell types of CRPC, with oxygen as the common resource and testosterone (Test) as the -produced resource. (B) Functional forms used to describe the two resource use efficiencies. HE-high use efficiency, LE-low use efficiency. See main text for descriptions of the cell types and resource use efficiencies.

Figure 4.1 shows a schematic with the main components of the current model. Following the Cunningham et al. model, we use three separate logistic equations (Equation 4.1) with density-dependent competition to describe the dynamics of each of the three types of cancer cells: $T^+$, which requires testosterone for growth but cannot synthesize it; $T^p$, which also requires testosterone for growth but is capable of autonomous testosterone production; and $T^-$, which neither requires nor produces testosterone. We use two ODEs to describe the dynamics of two resources-oxygen (Equation 4.2) and testosterone (Equation 4.3)-that contribute directly to cell growth (Figure 4.1A).

In order to connect resource levels to cell abundances, we model the cell type-specific carrying capacities as dynamic functions of real-time resource levels according to their respective consumption modalities detailed further below, such that competitive interactions in the model are realized through the levels of each resource. The complete set of equations are as follows:

$$\frac{dy_i}{dt} = r_i\ y_i\ \left(1 - \frac{\sum_i y_i}{1 + K_{i,max}\ f_i(O_2)\ f_i(T)}\right)\ - \delta_i\ y_i,\ \text{for } i \in \{T^+,\ T^p,\ T^-\} \tag{4.1}$$

$$\frac{dO_2}{dt} = p_{O_2} - \sum_i \mu_{O_2,i}\ y_i - \lambda_{O_2} O_2,\ \text{for } i \in \{T^+,\ T^p,\ T^-\} \tag{4.2}$$

$$\frac{dT}{dt} = p_T\ y_{T^p} - \sum_i \mu_{T,i}\ y_i - \lambda_T T,\ \text{for } i \in \{T^+,\ T^p\} \tag{4.3}$$

where $r_i$, $y_i$ and $\delta_i$ are the growth rate, abundance and death rate for cell type $i$, and $p_j$, $\mu_{j,\ i}$ and $\lambda_{j,\ i}$ are the production, consumption and decay rates for cell type $i$ and resource $j$. We model oxygen as a common resource that is supplied externally in a cell-independent manner reflecting baseline blood supply. It is consumed by all three cell types and leaks out of the tumor tissue at some constant rate. Since castration-resistant prostate cancer arises after some period of androgen deprivation therapy (Mohler et al., 2004), we initially assume that there is no external supply of testosterone, such that any testosterone in the system is derived solely from production by $T^p$. As explained below, this is an assumption that we eventually relax to consider CRPC with external testosterone supply. Since $T^-$ is testosterone-independent, testosterone consumption is only due to $T^p$ and $T^+$, in addition to some constant leakage out of the tumor as with oxygen. Note that more realistic resource supply conditions may also include fluctuating rates of supply, but the lack of empirical data at sufficient spatio-temporal resolution precluded us from simulating such resource supply regimes.

$K_{i,max}$ is the maximum carrying capacity for each cell type and this is further tuned by the availability of resources through the response function, $f_i(res)$. As a heuristic based on other models (Ghaffarizadeh et al., 2018), we assume $f_i$ to be a piecewise linear function valued in [0, 1], such that it remains at zero for resource levels below a certain lower threshold value. Once the lower threshold is crossed, $f_i$ increases linearly up to one at some upper threshold value, and stays at one thereafter. Figure 4.1B shows the general shape of $f_i$ for both resources and this is discussed further in the following section, but we note here that this form allows the effective carrying capacity, as represented by the whole of the denominator in Equation 4.1, to respond dynamically to the amount of resources available in the environment. When $f_i(res) = 0$, the denominator is 1, and for abundant resource availability such that $f_i(res) = 1$, the denominator approaches $K_{i,max}$.

### 4.2.2 Parameterization and exploration

Table 4.1 gives a complete list of all the parameter values and their ranges used in the model. As noted above, we have tried to derive as many of these from empirical sources as possible, such that our model is rooted as far as possible in the biology of the system.

Growth and death rates for the cell types are in units of $\text{min}^{-1}$ and derived from the doubling times of corresponding cell lines (see Table 4.1 and Supplementary Text S4). As the growth and death rates are different between the cell types, we derive $K_{i,max}$ for each cell type grown individually without resource limitations (i.e. $f_i(T) = f_i(O_2) = 1$) to achieve the same effective carrying capacity of 10000 (Equation S4.2). We rescale all resource levels such that a resource level of 1 corresponds to the empirically measured steady state value of that resource in prostate cancer tissue. This enables us to sidestep a detailed treatment of the stoichiometry of two very different resources. Setting Equation 4.2 to zero, we rescale empirically-derived cell-type specific consumption rates and decay rate for oxygen, and use these to calculate the steady state supply rate for oxygen assuming a population of only $T^-$ cells. Similarly, setting Equation 4.3 to zero, we use rescaled values of the production rate and natural decay rates for testosterone to calculate testosterone consumption rate for $T^p$. This rate is then used to determine the testosterone consumption rate for $T^+$. We restrict all equations to a biologically plausible space by requiring that $y_i = \frac{dy_i}{dt} = 0,\ if\ y_i < 1$ and only allowing non-zero population and resource abundances. Details of the parameterization and normalization are given in the Supplementary Text S4.

The efficiency of resource conversion into cellular growth is an important aspect of any resource-based conception of growth and competition. Resource efficiency is equally relevant to cancer, simply due to the diversity of niches and corresponding metabolic phenotypes within any solid tumor (Zheng, 2012, Birsoy et al., 2014, Loponte et al., 2019). Since environmental heterogeneity is well-documented in solid tumors, the availability and efficiency

of use of oxygen could vary substantially between different regions of a tumor (Zheng, 2012). Similarly, variation has been identified across prostate cancer cell lines from several sources in their response to the availability and amount of testosterone in the system (Gregory et al., 2001, Chuu et al., 2011). Resource use efficiency could therefore underlie distinct metabolic strategies within a heterogeneous tumor population. We consider two distinct kinds of such metabolic strategies in our model for each resource, based on the use efficiency of each resource. We choose two distinct forms of the function $f_i$, corresponding to how much resource is required to realize the maximum carrying capacity in Equation 4.2. The $f_i$ for high use efficiency (Figure 4.1B, blue line) reaches 1 at a lower resource level than that for low efficiency (Figure 4.1B, green line), thus requiring a lesser amount of resource to realize maximum carrying capacity.

In addition to resource use, we also explore the effect of initial population size on the steady state composition of the system. This stems from the expectation that as the sole producer of testosterone, absolute abundance of $T^p$ could have a significant impact on the balance of resource-based competition in the system.

All ODEs were simulated numerically until a plateau was observed, using the LSODA algorithm provided by the `scipy.integrate.ode` function in Python. Initial data suggested that a plateau was reached within $\sim$ 1000 days of model time for most parameter combinations. We therefore ran all simulations for 1000 days initially and subsequently extended the simulation time for those cases where a steady state had not been reached. Supplementary Figures S6.2.2-S6.2.5 show the time series for all the steady states examined in the main text along with the final run time in each case. All simulations reached steady state within 2500 days, except for only one combination of low oxygen use efficiency and high testosterone use efficiency, and initial population size of 500, which was simulated for 5000 days. Custom Python scripts have been used for all the simulations, data analysis and plotting.

## 4.3 Results

Our model aims to investigate intra-tumor cell dynamics in CRPC in terms of the availability of and competition for resources. Specifically, we study the effects of changing resource availability under varying conditions of resource use efficiency and population size. Considering two distinct levels of supply for both oxygen and testosterone, we define four resource supply states under which we examine steady state tumor composition. The first state is when **neither** resource is supplemented externally, testosterone production is entirely from $T^p$, and oxygen supply is at a low level, as in a hypoxic tumor. In the second state, **only oxygen** is supplemented and we use a higher value of the oxygen supply rate, while testosterone production remains restricted to $T^p$. In the third state, there is **only testosterone**

supplementation, implemented by the inclusion of a constant supply term independent of $T^p$ in Equation 4.3, while oxygen supply rate is low. In the fourth state, **both** resources are supplemented, oxygen supply rate is high, and external supply of testosterone is included.

We evaluate model outcomes based on whether or not $T^p$ and $T^+$ are able to persist at steady state. In the context of therapy, we consider the competitive exclusion of the testosterone-dependent cell types by $T^-$ an unfavorable outcome, the reasons for which will be discussed later. We classify model behavior into four distinct tumor responses based on the particular resource supply state in which steady-state persistence of $T^p$ and $T^+$ occurs. The first is a T-type response, which signifies that external supplementation of testosterone is sufficient for $T^p - T^+$ persistence at steady state. The second is an O-type response, which signifies that external supplementation of $O_2$ alone is sufficient for $T^p - T^+$ persistence. The third is a TO-type response signifying that external supplementation of both resources is required for $T^p - T^+$ persistence. The fourth response is called an N-type one, and indicates that $T^p - T^+$ persistence is not possible under any of the resource supply states tested here. Unless otherwise stated, all simulations are initialized with equal numbers of all three cell types and resource use efficiencies are identical across all three cell types. In the following sections, we illustrate how both resource use efficiency and supply states can play key roles in determining the overall tumor response as defined above.

### 4.3.1 Steady state composition under high resource use efficiency: a mix of tumor responses

Figure 4.2 shows the steady-state composition of the tumor across the four resource supply states when both resources are used at the same efficiency, either both high (Figure 4.2A) or both low (Figure 4.2B). At high use efficiency (Figure 4.2A), for any initial population size, $T^p - T^+$ are competitively excluded by $T^-$ when neither resource is supplemented externally. This is intuitive as $T^p$ and $T^+$ are dependent on both testosterone and oxygen for growth, and lack of external supplementation of both resources puts them at a competitive disadvantage relative to $T^-$ which are limited only by oxygen. At population size 500, external supply of testosterone is sufficient to ensure $T^p - T^+$ persistence, indicating a T-type response. However, for higher initial population sizes, external supply of either or both resources can ensure $T^p - T^+$ persistence, suggesting a T- and O-type tumor response. Both these responses can be understood by identifying the limiting resource in each case. In the absence of external testosterone supply, $T^p$ cells are the only source of testosterone in the system. A small initial population with few $T^p$ cells and no external testosterone supply is therefore limited by testosterone, leading to a T-type tumor response. Larger initial population sizes also entail higher $T^p$ frequency, thus relieving testosterone limitation. Under these circumstances, external supply of either one of the resources is sufficient to support

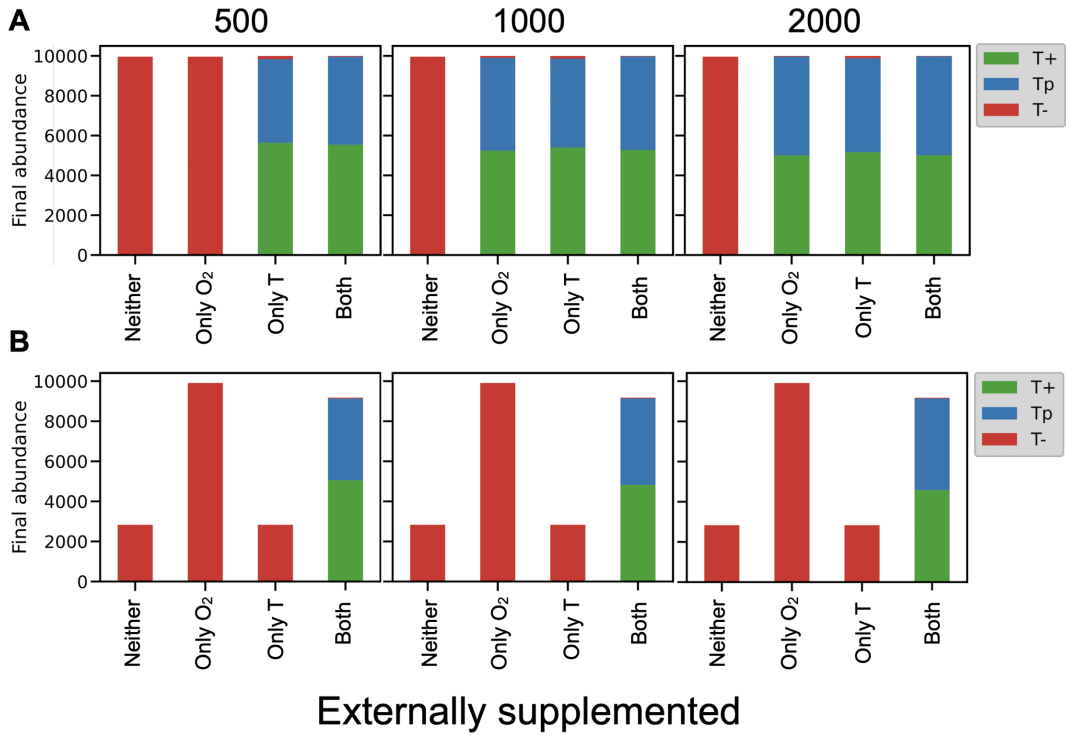



***Figure 4.2:*** Steady state behavior under identical use efficiencies of both resources. Abundances of all three cell types, under (A) high use efficiency of both resources, and (B) low use efficiency of both resources; initial population sizes are given on top of each subpanel and equal numbers of each cell type were seeded in the beginning. Without external supplementation, $p_{O_2} = 0.06$ min$^{-1}$, $O_2$ at $t = 0$ was set to 0, and testosterone was produced by $T^p$ at a rate $p_T = 5 \times 10^{-7}$ min$^{-1}$ per cell. Under external supplementation, $p_{O_2} = 0.11$ min$^{-1}$, $O_2$ at $t = 0$ was set to 0.5 and an additional supply of 0.001 min$^{-1}$ was added to Equation 4.3. $T$ at $t = 0$ was always 0. High use efficiency (i.e. panel A) usually leads to either a T- or T/O-type tumor response depending on the initial population size. Low use efficiency (i.e. panel B) produces an obligate TO-type tumor response regardless of initial population size. Further elaboration on tumor response types are in the main text.

$T^p - T^+$ growth and steady-state persistence. Taken together, our data indicate that high resource use efficiency generally favors $T^p - T^+$ persistence at steady state under a range of resource supply conditions, although this tendency is affected by the initial population size.

Under low use efficiency of both resources, Figure 4.2B shows that supply of both resources is necessary for $T^p - T^+$ persistence at steady state, indicating a TO-type tumor response. Contrary to high efficiency (Figure 4.2A), the TO-type response under low use efficiency is unaffected by the initial population size. We also observe that under low efficiency, low supply of oxygen leads to a smaller steady-state population size while the lack of external testosterone supply does not have such an effect on the steady-state population size. This disparity is explicable based on the functional form of low use efficiency of oxygen, which allows only a fraction of the maximum carrying capacity to be realized under the oxygen-limiting conditions imposed by limited external supply. It is also worth noting that since

testosterone use efficiency is also low, $T^p - T^+$ growth is limited strongly by both oxygen and testosterone at all times, and therefore requires external supplementation of both resources for their persistence. This leads to a TO-type response.

### 4.3.2 Mismatched resource use efficiencies reveal the lability of tumor responses

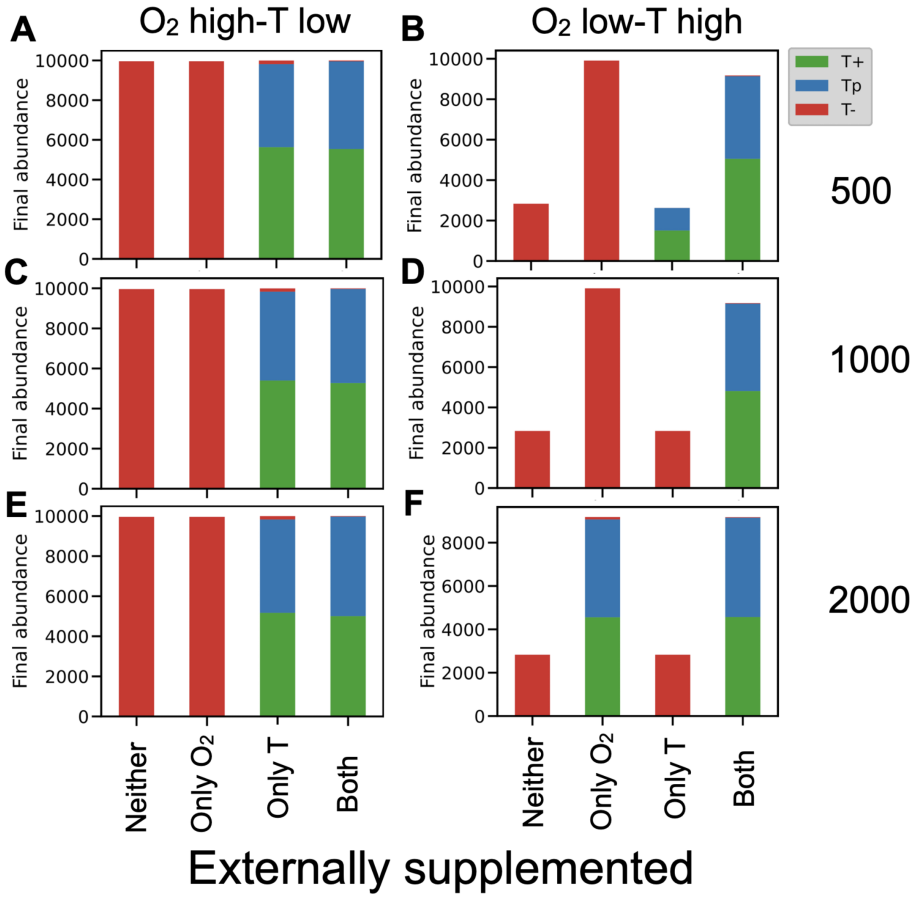



***Figure 4.3:*** **Steady state behavior under mismatched resource use efficiencies.** Abundances of all three cell for (A, C, E) low use efficiency of testosterone only and (B, D, F) low use efficiency of oxygen only; initial population sizes vary across rows as given on the right end of each row, and equal numbers of each cell type were seeded in the beginning. Resource supplies were parameterized as mentioned earlier. T stands for testosterone. When testosterone use efficiency is low (A, C, E), tumor responses are obligately T-type regardless of initial population size. On the other hand, low use efficiency of oxygen (B, D, F) leads to shifting tumor responses, from T-type for small initial population size (B), to TO-type response (D), to an O-type response (F). See main text for further explanation.

So far, we have only dealt with identical use efficiency of both resources- either both high or both low. We now consider responses to resource supplies of tumors with one resource used at high efficiency and the other at low efficiency. As Figure 4.3 (left) shows, low

testosterone use efficiency is the simpler of the two in terms of tumor responses, as it leads to an obligate T-type response. This implies that $T^p - T^+$ persistence requires external testosterone supply regardless of the initial population size, and as Figures 4.3A, C and E make clear, testosterone is both necessary and sufficient for $T^p - T^+$ persistence. As observed earlier, lower testosterone abundance does not affect the steady state population size.

Low use efficiency of oxygen coupled with high testosterone use efficiency presents a more complicated picture. As shown in Figure 4.3 (panels B, D and F), steady state population size responds to low oxygen abundance as expected, leading to a smaller final population. However, when oxygen and/or testosterone are supplied, tumor responses in terms of steady state composition show a greater degree of variability. Following the trend across Figure 4.3B, D and F from top to bottom, at small initial population size (Figure 4.3B), the lower frequency of $T^p$ cells leads to testosterone limitation and the tumor shows a T-type response. Given that $T^p - T^+$ are already limited by testosterone, external supply of oxygen under small initial population size favors $T^-$. Intermediate initial population size increases the $T^p$ frequency, but in the absence of external testosterone supply, testosterone production by $T^p$ alone is insufficient for $T^p - T^+$ persistence. Since oxygen is already limiting, this leads to a TO-type response overall (Figure 4.3D). Further increase in the initial population size increases $T^p$ frequency, and this increase ameliorates the initial testosterone limitation. However, higher total cell number under low oxygen use efficiency makes oxygen the limiting resource in turn, and this leads to an O-type response (Figure 4.3F). Low use efficiency of oxygen therefore reveals the potential for population size to determine which resource is actually limiting and how tumor populations respond to resource supply states.

### 4.3.3 Oxygen supply rate could affect producer-consumer cell abundance balance

Figure 4.4 shows part of a finer exploration of the resource-supply parameter space, in which we vary the supply rates of both resources over a range of values. Under low use efficiency of only oxygen (Figure 4.4A), we find that increasing the supply rate of oxygen increases the proportion of $T^+$ at steady state relative to $T^p$ (*cf* Figure 4.4A and B). Low efficiency use of testosterone, on the other hand, limits the viable resource supply region, such that population growth occurs for only the highest testosterone supply rate when its use efficiency is low (Figure 4.4B; note that the x-axis here has only the highest value of testosterone supply rate). It is also worth noting that in the supply cases shown here, the steady-state population is almost entirely comprised of $T^p$ and $T^+$, indicating that under sufficient resource abundance, $T^-$ is competitively excluded. As further elucidated in Supplementary Figure S6.2.6, below some lower threshold supply rates of oxygen and testosterone, $T^p - T^+$ are both driven to extinction at steady state. While the data are not shown here,

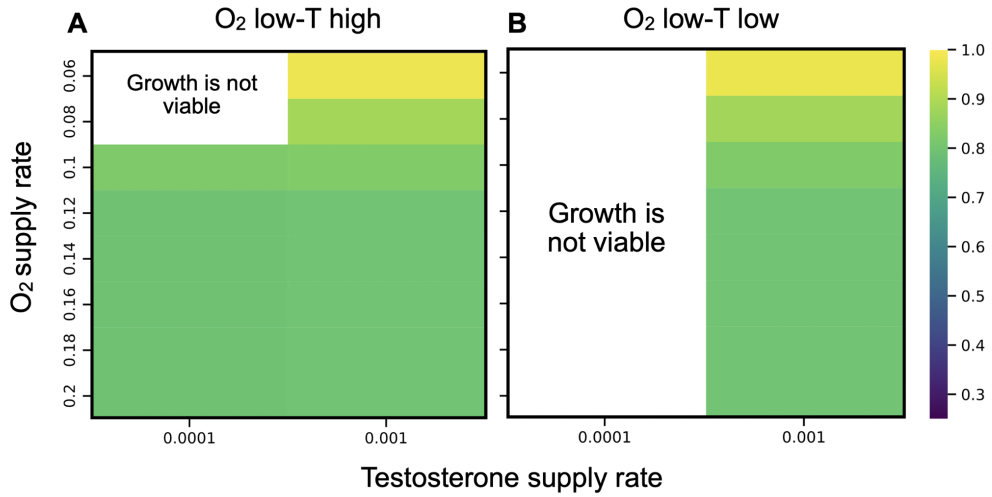



***Figure 4.4:*** **Producer-consumer balance with varying oxygen supply**. Steady state ratio of $T^p$ against $T^+$ across a range of supply rates of oxygen and testosterone for low use efficiency of oxygen and either (A) high use efficiency or (B) low use efficiency of testosterone. Note that lower supply rates of either resource, below 0.0001 $\text{min}^{-1}$ for testosterone and below 0.06 $\text{min}^{-1}$ for oxygen, could not support any cell growth at all and are therefore not shown here. Empty spaces marked as "Growth is not viable" are resource supply states for which $T^p$ and $T^+$ both go extinct. Oxygen supply rates on the y-axis increase from top to bottom. In all cases, simulations were initialized with equal numbers of all three cell types, with an initial population size of 500 and run for 2500 days.

we also find that below some other threshold supply of oxygen lower still, all three cell types are driven to extinction. Testosterone supply and its use efficiency therefore seem to result in binary outcomes in terms of steady-state $T^p - T^+$ persistence. On the other hand, under conditions when testosterone availability is sufficient for - persistence, higher oxygen availability favors a greater proportion of $T^+$ at steady state.

### 4.3.4 Oxygen and testosterone function at different abundance regimes

The previous sections show that across resource efficiencies, oxygen and testosterone have qualitatively different effects on steady state population sizes and tumor compositions. To investigate these differences in further detail, we show part of the parameter exploration for the choice of upper and lower thresholds for $f_i(O_2)$ and $f_i(T)$ in Figure 4.5A-B, for a population of $T^p$ cells alone. The heat maps clearly demonstrate that when testosterone is only produced by $T^p$ and oxygen is supplied externally, $T^p$ growth is viable across the whole parameter space of upper and lower thresholds for oxygen (Figure 4.5A). However, $T^p$ goes extinct for testosterone lower thresholds of 0.2 or greater (Figure 4.5B). Comparing this with a representative simulated time series of resource levels in Figure 4.5C, we see that in the initial few days, testosterone levels remain slightly below 0.2 whereas oxygen levels increase

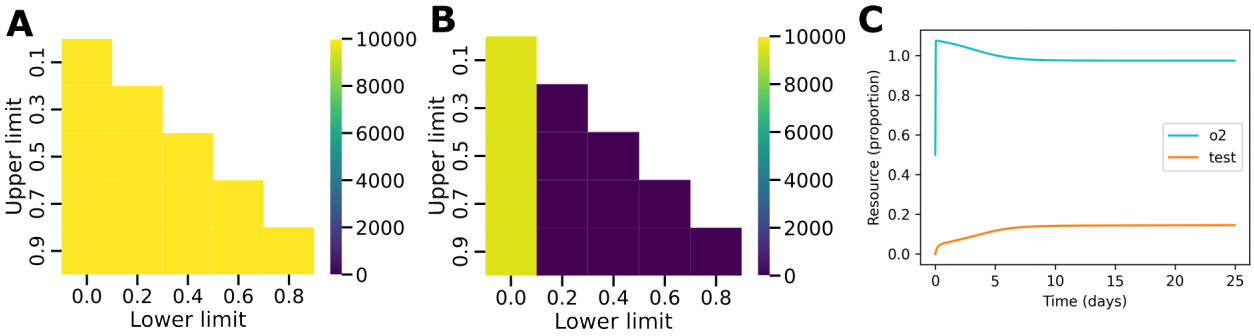



***Figure 4.5:* Abundance regimes of oxygen vs. testosterone**. Steady state abundances of a population of $T^p$ cells grown without resource limitation and in the absence of other cell types, under a range of thresholds for $f_i$ of (A) oxygen and (B) testosterone, and (C) representative time series of resource dynamics in the first 25 days. For panels (A) and (B), upper and lower threshold values correspond to the upper and lower inflection points of $f_i$ as shown in Figure 4.1B. The missing upper diagonal values of (A) and (B) are not valid threshold values of $f_i$ as the lower limit must be less than the upper limit by definition. For (C), equal numbers of all three cell types were initialized, initial total population size was 2000, lower and upper thresholds were 0 and 1 respectively for $f_i(O_2)$, 0 and 0.1 for $f_{T^p}(T)$, and 0 and 0.05 for $f_{T^+}(T)$, and $p_{O_2} = 0.11$ min$^{-1}$.

rapidly to stabilize close to one. This distinction in turn stems from the fact that oxygen and testosterone are produced and consumed at vastly different rates in the system, leading to different abundances at steady state as well as different dynamics towards the steady state. Figure 4.5 altogether demonstrates that the difference in steady state resource abundance corresponds to the parameter space of $f_i(O_2)$ and $f_i(T)$ where cell growth is viable. The extent to which the upper and lower thresholds can be varied for $f_i(T)$ is therefore much more restricted than for $f_i(O_2)$, primarily due to differences in the underlying resource dynamics.

## 4.4 Discussion

The current study goes beyond previous investigations (Cunningham et al., 2018) by allowing us to describe system behavior in terms of the availability and consumption of resources that are pertinent to the biology of CRPC. In the absence of external supplementation, oxygen in our model is usually present at much higher levels than testosterone (Figure 4.5), which is consistent with measurements from the tumor microenvironment (Dillard et al., 2008, Stewart et al., 2010). As $T^p$ and $T^+$ are simultaneously limited by two resources, they are driven to extinction in many parts of the resource efficiency-supply space (Figures 4.2 and 4.3). While low resource use efficiency is detrimental to $T^p - T^+$ survival overall (Figure 4.2B), overcoming this through external supply of the limiting resource is context-dependent. Figure 4.3A shows that when testosterone use efficiency is low, external supply of testosterone is always sufficient to ensure $T^p - T^+$ survival. When oxygen use efficiency is low, Figure 4.3B shows that the initial population size determines which resource is limiting and

can therefore support $T^p - T^+$ survival if supplied externally. Figure 4.4 shows a specific part of the resource supply-efficiency parameter in which oxygen supply seemingly favors a higher abundance of $T^+$ at steady state. Since $T^+$ has a slightly higher doubling rate than $T^p$, it is possible that its growth could be favored by the presence of higher resource abundance. However, an examination of the corresponding time series (Supplementary Figures S6.2.7-S6.2.9) suggests that the simulated time window of 2500 days could have been insufficient for the growth curves of $T^p$ and $T^+$ to reach a clear plateau in some cases. However, as the time series also indicate, the tumor composition changes so slowly after 2500 days that any effect of supply rates on cell type abundances would be of limited clinical importance.

On the whole, our data suggest that resource use efficiencies could serve as potential tumor-intrinsic properties that could be predictive of tumor composition at steady state. In the case of CRPC, the steady-state composition serves as an indicator of whether the tumor is likely to respond to resource-targeting therapy. As $T^p$ and $T^+$ are both testosterone-dependent cell types, their persistence at steady state leads to a tumor that would respond to testosterone inhibition. On the other hand, a higher frequency of $T^-$ at steady state leads to a tumor that is less responsive to testosterone-targeting therapy as $T^-$ cells are not dependent on testosterone for growth. Given that the development of personalized cancer therapy is a subject of active research (Martin et al., 2015, Hansen et al., 2017, Marigorta et al., 2017, Gedye and Navani, 2022), our data indicate that the elucidation of resource use strategies in a tumor could form the basis of designing more effective, personalized therapeutic strategies for CRPC cases depending on the extent of resource dependencies within the tumor. We also speculate that the abundance regime in which a resource operates could be an important factor in determining its therapeutic relevance. Testosterone in our model operates at a much lower abundance than oxygen. On one hand, this identifies testosterone as a clear target for therapeutic approaches involving resource supply as it is invariably limiting. However, we also find that model behavior is highly sensitive to changes in testosterone availability, with very sharp thresholds for steady-state $T^p - T^+$ persistence (Supplementary Figure S6.2.6). To what extent these thresholds are an artifact of our parameterization of resource use efficiencies is unclear, and this is a subject of ongoing investigation.

Although we use clinically relevant parameter regimes for our data, our model results have not been fit to longitudinal data from patient samples, unlike some other models in this context (Zhang et al., 2022). Consequently, rather than providing precise quantitative predictions about the progression and dynamics of CRPC, our results point towards the qualitative patterns and behaviors of such a system. These results therefore could be of value for growing developments in cancer management strategies. As the focus of cancer treatment shifts from complete cure to effective control, identifying and delimiting new ways to achieve such effective control is of vital importance. Cytotoxic chemotherapy weakens

intra-tumor competition by depleting the tumor of sensitive cells, but our results suggest that in cases where drug sensitivity is accompanied by unique resource use strategies, sensitive cells could be supported by external supply of the corresponding resource that they use exclusively. This could offer an effective counterweight to chemotherapy by strengthening intra-tumor competition, thus potentially extending the time for which the same therapy can be administered fruitfully. Broadly, these inferences stemming from our model are supported by recent work on a classical resource-consumer model of estrogen and glucose consumption in breast cancer, which used an $R^*$-based conceptual framework to discuss the effects of resource dynamics on intra-tumor competition (Kareva and Brown, 2021). The authors identify resource supply and use efficiencies as key mediators overall tumor dynamics, which is consistent with our results in CRPC. Specifically for prostate cancer, it has been observed that tumors of an LNCaP-derived cell line in athymic mice undergo significant regression and necrosis when the mouse is given a testosterone bolus through an implanted pellet (Umekita et al., 1996). Therapeutic use of testosterone also offers the possibility of improvements in patient-reported Quality of Life indicators that are typically adversely affected by prolonged androgen deprivation (Niraula et al., 2016). The design and optimization of dosing regimens that can use both cytotoxic drugs and externally-supplied resources could therefore be a promising direction for the treatment of CRPC (Nishiyama and Hoshii, 2016, Teply et al., 2018).

It is undoubtedly a huge challenge to reliably pinpoint which resources are limiting and to what extent, given the considerable complexity of the tumor microenvironment. However, even simple identification of tumors in which a resource-based intervention could be effective is a reasonable step forward. For instance, tumors corresponding to high resource efficiencies in our model are likely to be excellent candidates for a resource-based treatment strategy as they are likely to respond to changing resource availabilities without complete loss of the drug-sensitive population. The mapping of our model outcomes to therapeutic feasibility is currently limited both because the model is not fitted to clinical data and because our conceptualization of resource efficiencies is entirely heuristic. That said, our data indicate that a theoretical framework that integrates available information on the biology of resource use in a tumor with its growth dynamics could considerably expand our current toolkit for cancer control. This view has seen increasing advocacy recently (Kareva et al., 2015, Kareva and Brown, 2021) and has particular relevance for endocrine cancers, most of which frequently have a hormone or growth factor signaling axis that is crucial for cancer progression, as exemplified by the importance of EGFR signaling in breast cancer (Osborne et al., 1985, Kareva and Brown, 2021), testosterone signaling in prostate cancer (Mohler et al., 2004), or insulin signaling in pancreatic cancer (Chan et al., 2014, Archetti et al., 2015). More such signaling axes will undoubtedly be identified in the years to come and in each case, tumor composition and size could be seen as a result of the interplay between such signaling and

other energetic resources, which offers much scope for further modeling effort.

Realizing the translational value of resource use models in cancer treatment depends on the ability to identify and monitor relevant cellular phenotypes and resource availability. One potentially interesting avenue are the so-called "radiomics" approaches that can integrate quantitative imaging data with additional information about spatial structure, genomics and gene expression to help with diagnosis and monitoring (Gillies et al., 2016). Such analyses have been applied to classify prostate cancer and benign masses accurately, and further separate different kinds of tumors with the same overall Gleason score ($3+4$ vs $4+3$, for example (Wibmer et al., 2015, Fehr et al., 2015)). The possibility of obtaining such information before and after a therapeutic intervention represents unprecedented scope in the future for precise, patient-driven treatment that can respond to drug-induced ecological and evolutionary changes in a tumor. Theoretical models must therefore be geared to make use of such granular data from the tumor microenvironment to enable an ecologically-informed treatment approach.

| Parameter | Description | | | Value(s) | Source(s) |
|---|---|---|---|---|---|
| $r_i$ | Population growth rate of cell type $i$ | | $T^+$ | $2.84 \times 10^{-3}$ $\text{min}^{-1}$ | Equation S4.1 |
| | | | $T^p$ | $2.79 \times 10^{-3}$ $\text{min}^{-1}$ | |
| | | | $T^-$ | $6.23 \times 10^{-4}$ $\text{min}^{-1}$ | |
| $\delta_i$ | Population death rate of cell type $i$ | | $T^+$ | $2.5 \times 10^{-3}$ $\text{min}^{-1}$ | (Jain et al., 2011) |
| | | | $T^p$ | $2.5 \times 10^{-3}$ $\text{min}^{-1}$ | |
| | | | $T^-$ | $1.6 \times 10^{-4}$ $\text{min}^{-1}$ | |
| $K_{i,max}$ | Cell-type specific carrying capacity | | $T^+$ | $8.35 \times 10^4$ | Equation S4.2 |
| | | | $T^p$ | $9.62 \times 10^4$ | |
| | | | $T^-$ | $1.34 \times 10^4$ | |
| $p_{res}$ | Production rate of resource, either as bulk or by cells | $O_2$ | | 0.11 $\text{min}^{-1}$ | Equations S4.3 and S4.4 |
| | | T | | $5 \times 10^{-7}$ $\text{min}^{-1}\text{cell}^{-1}$ | |
| $\mu_{res,i}$ | Uptake of resource $res$ by cell type $i$ | $O_2$ | $T^+$ | $1.63 \times 10^{-6}$ $\text{min}^{-1}\text{cell}^{-1}$ | (Hail et al., 2010), Equation S4.3 |
| | | | $T^p$ | $1.63 \times 10^{-6}$ $\text{min}^{-1}\text{cell}^{-1}$ | |
| | | | $T^-$ | $1.04 \times 10^{-6}$ $\text{min}^{-1}\text{cell}^{-1}$ | |
| | | T | $T^+$ | $2.34 \times 10^{-8}$ $\text{min}^{-1}\text{cell}^{-1}$ | |
| | | | $T^p$ | $6.00 \times 10^{-8}$ $\text{min}^{-1}\text{cell}^{-1}$ | |
| | | | $T^-$ | 0 $\text{min}^{-1}\text{cell}^{-1}$ | |
| $\lambda_{res}$ | Decay rate of resource | | $O_2$ | 0.100 $\text{min}^{-1}$ | (Ghaffarizadeh et al., 2018), (see Jain et al., 2011, and supplements therein) |
| | | | $test$ | 0.004 $\text{min}^{-1}$ | |
| $\tau_{d,i}$ | Doubling time of cell type $i$ | | $T^+$ | 34 hr | (Bairoch, 2018) |
| | | | $T^p$ | 40 hr | |
| | | | $T^-$ | 25 hr | |
| $y_i^*$ | Equilibrium population size without competition | | | 10000 | Assumed |
| $res^*$ | Equilibrium tissue levels of resource | $O_2$ | | 2.5 mmHg | (Stewart et al., 2010, Titus et al., 2005) |
| | | T | | 3.74 pmol/g tissue | |

***Table 4.1:*** **List of all parameters used in the current model**. T stands for testosterone. Units of all resource-related parameters do not include dimensions of absolute abundance or concentration as these are normalized against $res^*$, and absolute values can be obtained by multiplying the parameter by $res^*$. Cell-type specific values of $K_{i,max}$ correspond to the value for which an effective carrying capacity of 10000 is realized for that cell type growing without other competitors and abundant resource availability. Details of normalization of resource parameters and derivation of growth parameters are given in the Supplementary Text S4.

# 5 | Context-dependent selection as the keystone in the somatic evolution of cancer

## Key contributions

This study has been designed by Milind Watve and me, as part of my PhD work. I have written the original codes for the simulations and have contributed to the analytical models. Milind Watve has built the analytical models. We have both co-written a manuscript based on the study results and Milind Watve has revised it addressing reviewers' concerns. The following chapter has been largely reproduced from this publication with minor formatting and grammatical changes.

The manuscript has been published at the peer-reviewed journal, *Scientific Reports*, and can be accessed at https://doi.org/10.1038/s41598-020-61046-7.

## 5.1 Introduction

Mathematical models of somatic evolution in cancer have been in development for the past several decades, with a strong focus on mutational processes (Armitage and Doll, 1954, McFarland et al., 2013, Blokzijl, 2016, Mina, 2017, Tomasetti and Vogelstein, 2015, Tomasetti et al., 2017). All cancers are necessarily a combination of different types of genomic changes including point mutations, aneuploidy and other chromosomal aberrations. The cancer phenotype has a large number of distinguishing characters, encapsulated by the notion of the "hallmarks" of cancer (Hanahan and Weinberg, 2000, Schäfer and Werner, 2008, Hanahan and Weinberg, 2011). The wide range of characteristics that these hallmarks include makes it astonishing that so many alterations in cell properties come together in cancers purely out of chance.

Cancer is modeled as a multistage process but the dynamics of the multistage process is still largely unknown. The series of genetic or epigenetic changes have been considered to co-occur by chance alone by the "bad-luck" paradigm. The clonal expansion paradigm assumes that every oncogenic mutation on the way to a cancerous phenotype causes the mutant clone to expand, and as the mutant population increases, the probability of a second oncogenic mutation increases proportionately (Nowell, 1976). Implicit in this theory is the assumption that every oncogenic mutation has a selective advantage over the normal cell. Since most changes involved in carcinogenesis relate to evading growth regulatory mechanisms, it is considered logical that any mutation that allows for such evasion will have a selective advantage. However, evidence has been accumulating over the past few years that the fitness advantage of a mutant is largely dependent on the tissue micro-environment (Hanahan and Coussens, 2012, Pietras and Östman, 2010). Studies in mice (Cao, 2010) and humans (Rundqvist et al., 2013) have demonstrated the effect of contingent factors, such as behavioral profiles and lifestyle parameters, on cancer progression. With changes in the lifestyle parameters, the age adjusted incidence for many types of cancers has increased significantly in recent decades (Supplementary Material from Wu et al., 2016). These findings provide clear indications that the selective forces which determine mutant clone fitness can vary considerably in response to many factors that vary across individuals, leading to context-dependent clonal expansion of potentially oncogenic mutants. The role of age induced changes in the microenvironmental selective forces in somatic evolution has been modeled as well as demonstrated (Rozhok et al., 2014, 2016, Rozhok and DeGregori, 2016, 2019, Henry et al., 2010, Henry, 2015, Vas, 2012, Vas et al., 2012). But the role of individual differences in microenvironment at any given age has not been adequately incorporated by cancer models so far.

In this paper, we compare three modeling approaches, two of which explicitly include selection at different levels: (1) random mutagenesis, or the "bad luck" hypothesis, (2)

context-independent selection (CIS) on mutant clones within individuals, and (3) context-dependent selection (CDS), where the selective forces vary depending upon the tissue micro-environmental context differentially in every individual. Genetic mutations are a fundamental aspect of multi-stage models of carcinogenesis. In our models however, we use the term mutation more broadly, to denote any change that is heritable within a given cell lineage, genetic or epigenetic.

Since well-curated data are available for human cancer incidence patterns through the SEER databases (Siegel et al., 2016), we develop models of these three processes, and compare their predictions with the epidemiological picture of cancer in the human population. A good working model of somatic evolution of cancer needs to explain the following epidemiological features:

1. Total and age-specific incidence: The pooled incidence of all cancer types lies well below 30%, while age-specific and cumulative incidence patterns show more variations between cancer types (Siegel et al., 2016). Interestingly, recent analyses have shown for several cancers that the age-specific incidence rates decline late in life in humans as well as in rodents (Pompei et al., 2001, Harding et al., 2012, Belikov, 2017) in contrast with a previous belief of a monotonic power law increase in incidence with age (Armitage and Doll, 1954, Meza et al., 2008). Some theoretical explanations for this decline have been suggested (Frank, 2007, Zhang, 2018), but biologically interpretable and translationally useful insights are yet to be obtained. The late-life decline causes the cumulative incidence to saturate with age at a small percentage of the population size. No matter the lifespan, the proportion of cancer in the population can never reach 100%, but saturate at a finite limit.

2. Incidence vs. cell number: The relationship between cancer risk and stem cell number (as the lifetime number of stem cell divisions) has been kept in the spotlight by recent work by Tomasetti et al. (Tomasetti and Vogelstein, 2015, Tomasetti et al., 2017) and its criticism by others (Wu et al., 2016, Zhang, 2018). Tomasetti et al.'s studies indicate that the risk of cancer increases linearly with the lifetime number of stem cell divisions on a log-log scale. The authors argue that since cancer risk is strongly correlated to the number of division events, the predominant cause of sporadic cancers must be non-preventable mutagenesis. Since good data are available on this relationship, if the three models make differential predictions about the relationship, it can be used to evaluate the models comparatively. Empirical data on this relationship of mutation rate versus incidence are less common, as mutation rates are difficult to measure reliably. Therefore it is difficult to use this relationship as a testable prediction although some insights are possible (Frank, 2007, Hao et al., 2016).

3. Non-mutagenic carcinogens: There are several agents, including hormones and growth factors that increase cancer risk without affecting the basal mutation rate (Tennant, 1993). The activity of these agents, and their signature in epidemiological patterns, are both important in building a complete framework of explaining cancer etiology.

4. Peto's paradox and similar observations: This relates to the incidence-cell number relationship, as cancer risk is seen not to scale with body size or cell number across species (Nagy et al., 2007), but may correlate with the latter within a species (Noble et al., 2015). A wide range of explanations have been offered for this observation (Tollis et al., 2017), and in a modeling context, these explanations can come from the model itself (intrinsic), or involve extrinsic factors, such as evolved cancer defenses, that are not part of the basic model framework.

## 5.2 The "bad luck" model

This hypothesis assumes that the required set of driver mutations accumulate in a cell by chance alone. This may happen over a period of time, or in a single large-scale event, like chromothripsis (Stephens, 2011). Regardless of whether mutations accumulate sequentially or otherwise, the "bad luck" model does not assume selection of any kind on mutations over the course of somatic evolution. A typical bad luck model is given by Calabrese and Shibata (Calabrese and Shibata, 2010), and Zhang et al. (Zhang, 2018). We use the following model for life time probability of cancer parameterized by age:

$$p_{can} = 1 - (1 - (1 - (1 - p)^{logS+d.a})^k)^S \tag{5.1}$$

where $S$ is the number of stem cells, $d$ the stem cell division rate, $k$ the number of driver events required for cancer onset, and $p$ is the mutation rate per division. We use $S$ for a steady state stem cell number and $n$ (see below) for total number of stem cell divisions up to a given age.

### 5.2.1 Predictions from the "bad luck" model

From Equation 5.1, the cancer incidence shows a threshold relationship with the number of stem cells as well as the mutation rate. The cumulative incidence also increases with age in a threshold relationship saturating only at 100% (Figure 5.1). The sharpness of the threshold as measured by the slope of the curve at the point of inflection (i.e. at cumulative incidence of 50%) is inversely proportional to the threshold age. The age-slope inverse relationship is not affected by stem cell number and mutation rates but at larger $k$ the threshold becomes sharper. The threshold relationship has been obtained by earlier models as well (McFarland et al., 2013) but its implication in terms of testable predictions of the models have not been

discussed or used. Because of the threshold relationship, if cancers were not lethal, then we would have obtained a bimodal relationship of incidence with stem cell numbers. For tissues with stem cell numbers above the threshold, the incidence would have been near 100% and for tissues with low numbers close to none. Since cancers are fatal, we would expect that the first tissue to exceed the threshold would kill almost 100% of the population before cancers could develop in any other tissues. Since we do not see such patterns in the population, we are forced to restrict ourselves to the part of the parameter space in which the threshold for any tissues lies beyond the maximum lifespan. Under these conditions, the incidence increases monotonically with age in a power law relationship with no signs of late age decline. The expected relationship between stem cell number and cancer incidence across different tissues is a straight line on a log-log plot with a slope very close to unity. Lifespan and mutation rates affect the position of the line but the slope remains unaltered. This can be taken as a testable prediction of the bad luck model.

## 5.3 Models with selection on mutants

For models that involve selection, we choose to explore the effects of CIS and CDS through a simulation-based stochastic framework.

We use a linear process to model the sequential accumulation of mutations in a population of stem cells. We begin by considering the development of a generalized tissue compartment in each organism starting from one stem cell, with mutation rate per cell generation per locus, $p$, growing logistically to a carrying capacity. Conceptually the carrying capacity is the same as standing stem cell number in adult tissue $S$. We use the discrete logistic equation:

$$m_{i,t} = m_{i,t-1} + m_{i,t-1} * g_i * (\frac{S - \sum_{i=0}^{k} m_{i,t-1}}{S}) - m_{i,t-1} * d \tag{5.2}$$

Here, $m_{i,t}$ is the size of the $i^{th}$ mutant population at time, $t$, with $i = 0$ being the non-mutant cell population, $g_i$ the corresponding logistic growth rate, $d$ the common death rate, and $k$ is the threshold number of oncogenic mutations required for cancer onset. As the organism develops into an adult, net growth in the stem cell compartment saturates, but reaches a dynamic equilibrium between cell death and renewal. The stem cell population can be reduced, either by death of stem cells or differentiation, as reflected by $d$ in Equation 5.2. The replacement of the lost cells by either mutants or non-mutants is a function of their growth rates. We simulate new mutation events stochastically; the probability of at least one cell mutating is given by $1 - (1 - p)^{m_{i,t}}$, and if this probability exceeds a random number between 0 and 1, a new $(i+1)^{th}$ mutant population is initiated. Each new oncogenic mutation could give a growth advantage over older cell populations, leading to successive cycles of clonal expansion in which the newer population gradually replaces older cells through

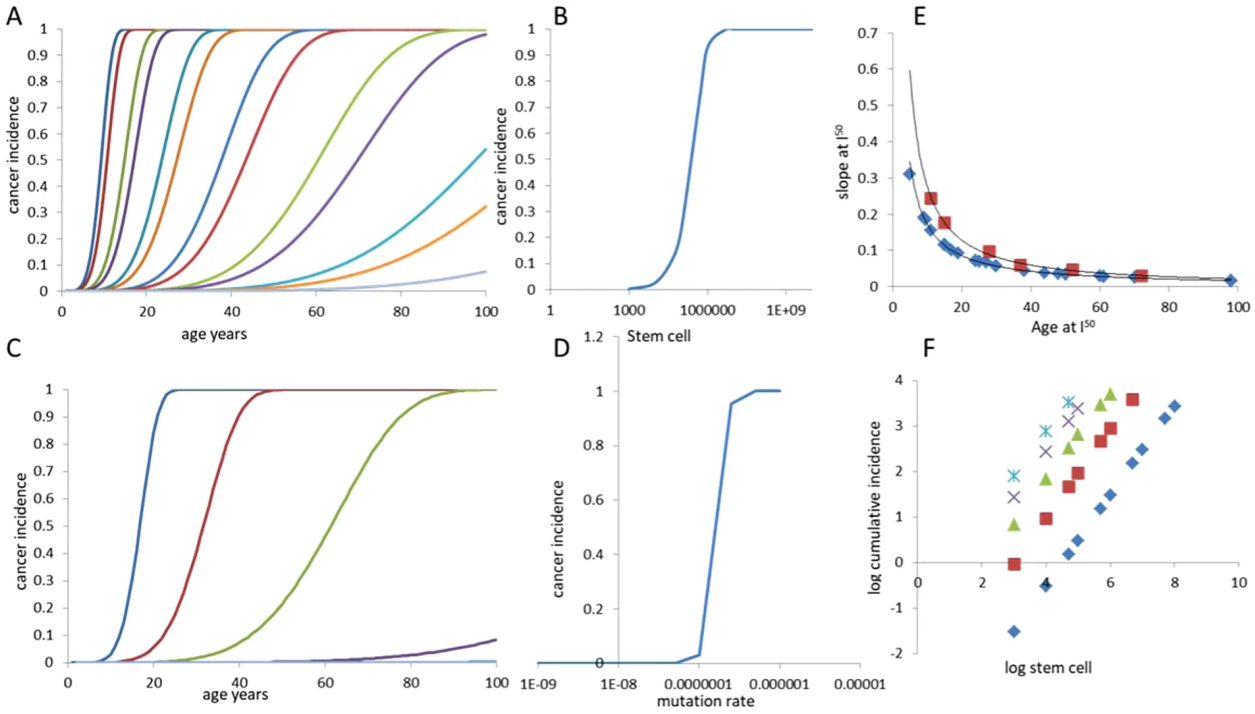



***Figure 5.1:*** **The nature of relationships between age, stem cell number, mutation rate and cancer incidence by the bad luck model**. The cumulative incidence of cancer increases with age in a threshold relationship. (A) The stem cell number affects the threshold age, but in all cases the trend saturates only at 100% incidence. (B) At any given lifespan, the stem cell number holds a threshold relationship with cancer incidence. For tissues with stem cell numbers below the threshold, the cancer probability is close to zero and above the threshold it rises to almost unity. (C) The mutation rate also influences the threshold age for a given cancer type and given a fixed lifespan, the mutation rates holds a threshold relationship with incidence. (D) The age threshold and the sharpness of the threshold measured by the slope at 50% cumulative incidence ($I_{50}$) have an inverse relationship (E) shown at $k = 5$ and 8 respectively. If we take only those cancers for which the threshold lies beyond the maximum lifespan, the relationship of stem cell number with lifetime cancer incidence on a log-log plot is linear (F) with the slope = 1, with lifetime incidence increasing for both higher $n$ and higher $p$. Lifespan and other parameters change the position of the line, but the slope remains unaltered. Shown in the graph are analytical results. Simulations at population size 10000 and above follow the analytical predictions very closely. Parameters: For A and B, $p = 10^{-6}$, $n$ ranges from $10^3$ to $10^{10}$ (curves right to left). For C and D, $n = 10^9$ and $p$ ranges from $10^{-6}$ to $10^{-9}$ (curves left to right). In all results displayed here, $k = 5$ and $d = 1000$.

competitive exclusion. We simulate this linear evolution process until $k$ mutations have been accumulated, which is the assumed threshold for cancer onset. Death of the individual occurs either at cancer onset when the $k^{th}$ mutation occurs, or at the end of the assumed lifespan of 100 years, whichever happens first. This simulation is repeated independently for a population of 10000 individuals, and the population-level cancer incidence is recorded, along with the age of onset.

### 5.3.1 Choice of parameter range

In order to standardize the discrete logistic simulation, we assume the time unit to be one day per logistic growth step. Most human organs complete development and maturation with in the first 10–20 years of the lifespan, and the final carrying capacity achieved is the adult stem cell number, ranging between $10^6$ and $10^{11}$ across different tissues (Supplementary Material from Tomasetti and Vogelstein, 2015). Given the final population size and the time taken to reach it, a simple calculation based on the logistic equation shows the required growth rate for a non-mutant stem cell to be in the range of 0.00383 to 0.0131. Starting from the non-mutant growth rate, $g_0$, growth rates are assumed to increase linearly for each subsequent mutant population, with the slope given by $\Delta_g = \frac{g_k - g_0}{k}$. Ranges of stem cell number and mutation rate are retained as in the "bad luck" model.

#### The context-independent selection (CIS) case

While the clonal expansion theory introduced the notion of selective advantages to oncogenic mutants, it makes the implicit assumption that identical mutations have the same selective advantage in every individual in which they occur; stated otherwise, individuals do not differ in their propensity for mutant clonal expansion. To capture this in the CIS case, we use the same slope, $\Delta_g$ for all individuals in the simulation.

#### The context-dependent selection (CDS) case

As argued earlier, it is becoming increasingly clear that the competitive outcomes of identical mutations can depend strongly on the micro-environmental context in which cell competition occurs. In order for selection on mutants to be context-dependent in our model, we randomize the slope, $\Delta_g$ over the population from a given normal distribution. Each individual begins with the same $g_0$, but the progression of growth rates is randomized across individuals, such that individuals with large $g_i$ would progress faster towards cancer onset, while those with small, or negative values of $g_i$ would never progress to a cancerous state as the mutant gets selected against. This produces variation across individuals for cancer propensity.

### 5.3.2 Predictions from the selection models

Under the assumption of CIS, the incidence of cancer shows a strong threshold relationship with age, saturating only at 100% (Figure 5.2). The age-specific incidence declines only when the cumulative incidence approaches 100%. As with the "bad luck" model, incidence has a threshold relationship with both $n$ and $p$. Where the incidence of cancer is near the realistic range, for small values of $p$ and $n$, the late-life decrease in incidence is not observed. In this range of parameters, the log-log plot of stem cell numbers with cancer incidence gives

a straight line with slope $\geq 1$. Giving a distribution to $p$ in the realistic range reduces the sharpness of the incidence threshold with age, but saturation of incidence remains at 100% (Figure 5.3). By most of these predictions, it is difficult to resolve the CIS model from the "bad luck" model except that the log-log plot of stem cell number and life time incidence may have a slope greater than 1 in the CIS model.

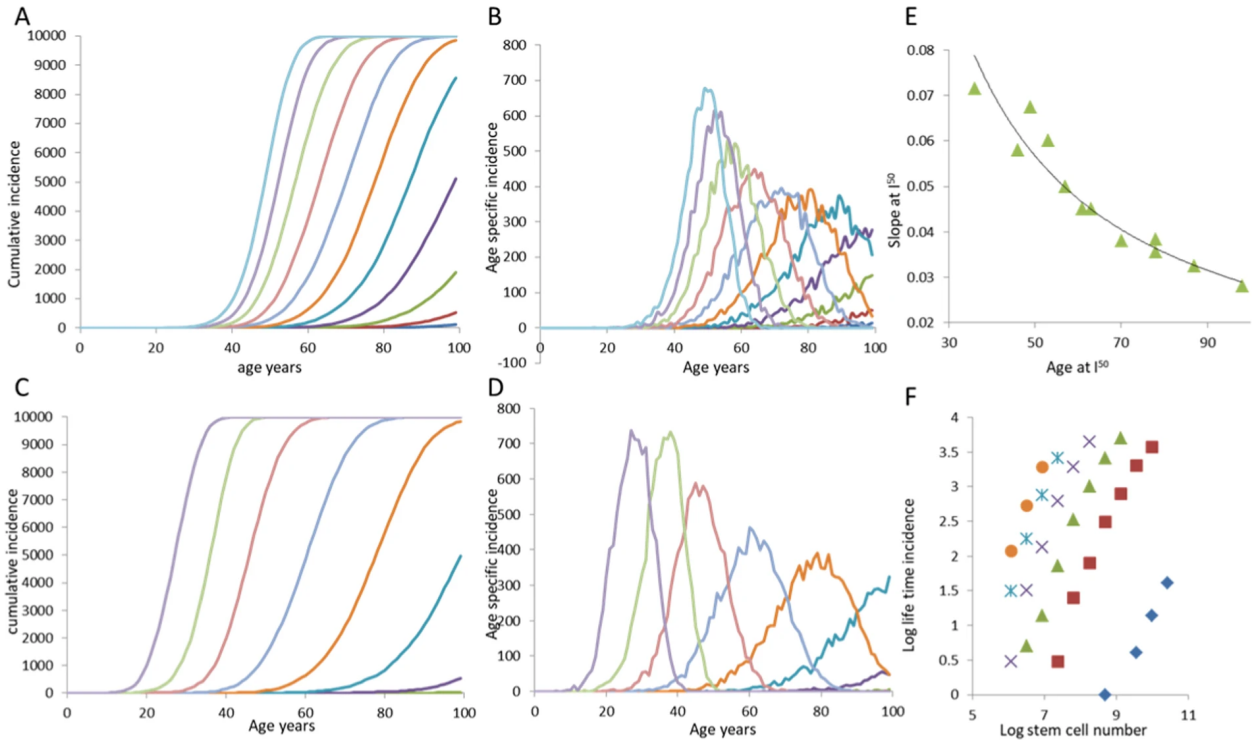



***Figure 5.2:*** **Cancer incidence predictions for the CIS model**. The nature of relationships between age, stem cell number, mutation rate and cancer incidence by the CIS model. (A) Similar to the bad luck model, there is a threshold age at which the incidence transits from near zero to near 100%. Saturation of cumulative incidence happens only when the incidence approaches 100%. The threshold age is affected by the stem cell number (A,B) as well as the mutation rate (C,D). The threshold age and the slope at $I_{50}$ holds an inverse relationship (E) similar to the bad luck model. (F) For the parameter range in which the threshold lies much beyond the lifespan, the log-log relationship between stem cell number and cumulative incidence is linear with a slope $\geq 1$, with incidence increasing with both higher $n$ and higher $p$, as with the bad luck model. For A and B, $p = 5.603 \times 10^{-9}$, $n$ ranges from (curves right to left respectively) $10^6$ to $10^{10.5}$. For C and D, $n = 1.785 \times 10^8$. $p$ ranges from (curves left to right) $10^{-6.5}$ to $10^{-10.5}$. Growth rates progress linearly in the general form, $g_i = 0.007 \times (i + 1)$, where $i = 0$ to $k$, and $k = 5$. Here, $\Delta_g = 0.007$ for all cases. Cumulative incidence at a given age refers to the incidence of cancer up to and including that age, while age-specific incidence at a given age refers to the incidence of cancer only at that age.

As opposed to the CIS model, the CDS model produces a saturating trend in cumulative incidence that saturates at a level much lower than 100% (Figure 5.4). The early saturation happens when a proportion of the population has relative growth rates zero or negative.

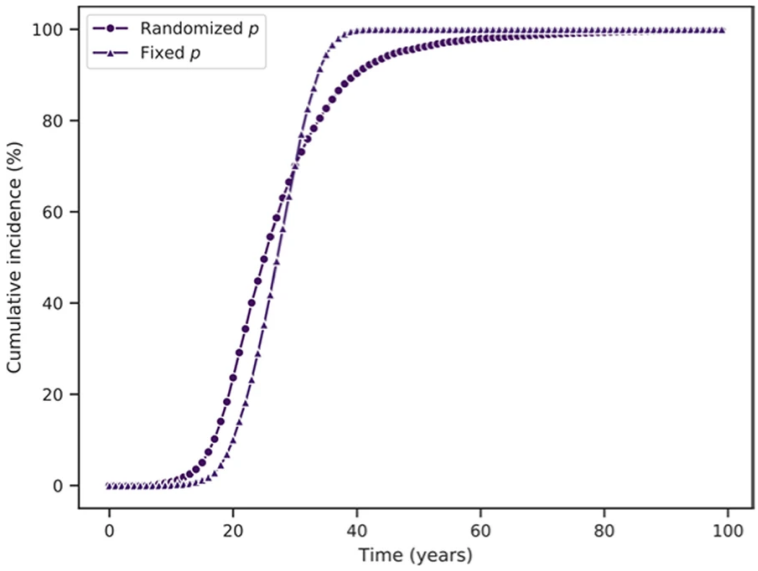



***Figure 5.3:*** **Randomizing $p$ in the CIS model**. Trends with age in cumulative incidence with randomized vs. fixed $p$ for the population in the CIS case; giving a distribution to $p$ does not reduce the saturation of incidence from 100%, but slows the transition to 100%. In the randomized case, $p$ values were drawn from a uniform distribution with range $[3.775 \times 10^{-11},\ 3.06 \times 10^{-7}]$, while for the fixed case, $p = 3.06 \times 10^{-7}$; for both, $n = 1.785 \times 10^{8}$ and $k = 5$.

This assumption is theoretically as well as empirically sound as a mutant can have positive or negative fitness (Rozhok and DeGregori, 2019). This is an important feature of the CDS model, as it allows the model to generate more realistic patterns in age-specific cancer incidence. The saturation occurs because propensity for clonal expansion varies across individuals; cancer progression occurs very quickly in some individuals, and not at all in others. Of the three models analyzed so far, only the CDS model captures a trend similar to the late-life decline observed in many cancers in humans (Figure 5.4). The stem cell number and incidence relationship on a log-log plot, in this model, can have a slope $< 1$ and often the relationship shows a saturating curve rather than a straight line. These predictions of the CDS model contrast the bad luck and CIS model and can be used as differential testable predictions. The major differential predictions of the CDS model arise from the $\Delta_g$ distribution which implies that somatic evolution of cancer in this model is mainly selection limited rather than being mutation limited.

## 5.4 Sensitivity of predictions

We relax the quantitative and qualitative assumptions of the model to test whether or not the results are artifactual, contributed by any specific assumption.

We assumed $k$ to be constant for a given cancer type, however across cancer types the estimated $k$ varies (Nunney and Muir, 2015, Martincorena, 2017). An increase in $k$ does not

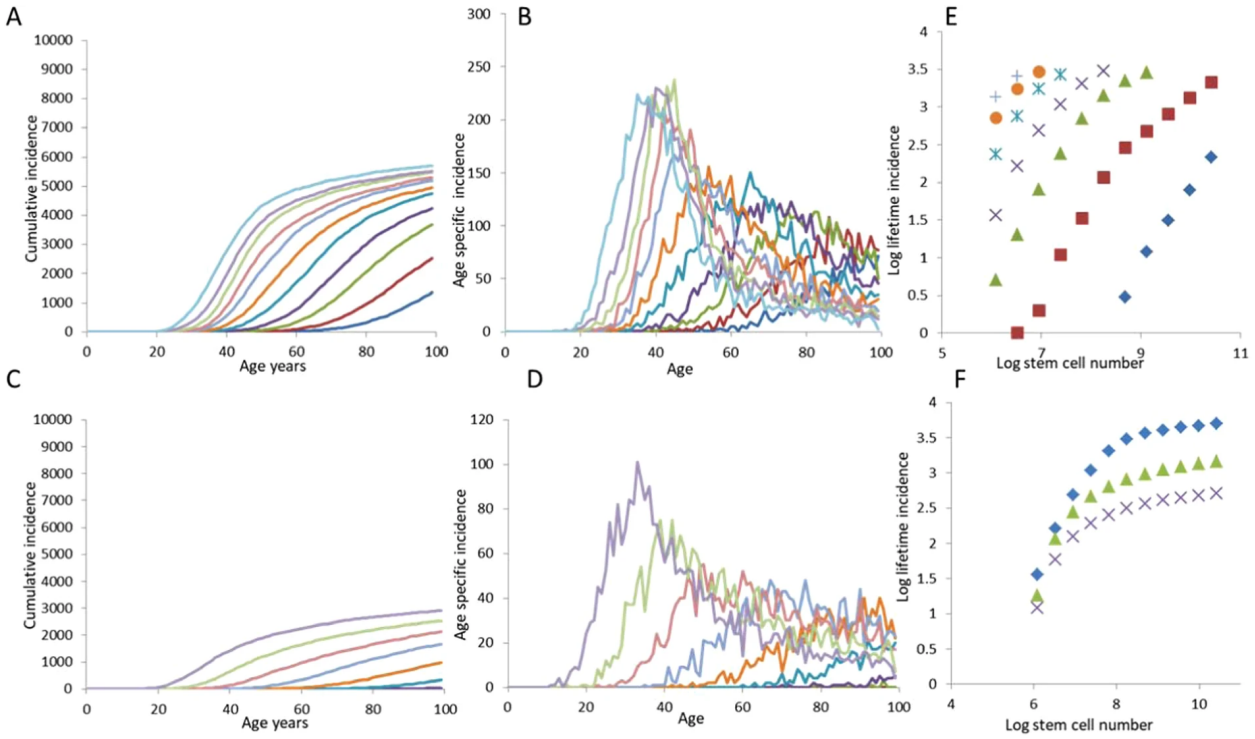



***Figure 5.4:*** **Cancer incidence predictions for the CDS model**. The nature of relationships between age, stem cell number, mutation rate and cancer incidence predicted by the CDS model. In this model the cumulative incidence can saturate much below 100% depending upon the distribution of $\Delta_g$. The age incidence curves for different values of $n$ (A,B) and different mutation rates (C,D) show an early saturating trend rather than a threshold phenomenon. The log-log plot between stem cell number and life time incidence may be saturating or linear with a slope $\leq 1$, with higher $p$ generally increasing total incidence (E). With a reduction in the mean $\Delta_g$, the cell number-incidence curve saturates early (F). All other parameter values are similar to those in Figure 5.2.

qualitatively affect the threshold phenomenon in the bad luck and CIS models but shifts the threshold age to the right. For a given threshold age, the threshold is sharper for a higher $k$. Randomizing $k$ between the range 2 to 10 (Supplementary Figures S6.3.3 to S6.3.6) increases the scatter but does not introduce any pattern by itself.

To test whether the assumption of a normal distribution for $\Delta_g$ is critical for the results, we reevaluated the predictions of the CDS model for a Gumbel and uniformly-distributed $\Delta_g$ progression. We find that the shape of the distribution does not affect the saturation of cumulative incidence, or the late-life decline in age-specific incidence predicted by the model, although the shape of the age curve changes (Supplementary Figures). We assume generally that $\Delta_g$ is constant for a given individual, but relaxing this assumption by giving a distribution to $\Delta_g$ within each individual would not change the results qualitatively. In such a case, the lowest $\Delta_g$ within an individual would be the rate-limiting step.

Somatic cell populations are potentially susceptible to mutational meltdown, and a large number of somatic mutations are in fact known to be deleterious. Earlier analysis of the effect of passenger mutations has shown that accounting for them in somatic evolution delays overall incidence of cancer in the population (McFarland et al., 2013). However, from our CIS case, we see that when total incidence in the population decreases, most cases of cancer are concentrated later in life, similar to the delay in incidence noted before (McFarland et al., 2013). This pattern is not compatible with the observed late-life decline in incidence. We expect therefore that mutational meltdown due to deleterious passenger mutations is insufficient to explain saturation due to late-life decline in incidence.

## 5.5 Differential testable predictions of the models

Summarily, the testable predictions of the three classes of models with reference to the epidemiological features that we defined in the beginning are as follows:

1. The bad luck model and the CIS model predict a threshold relationship with age. Saturation in the incidence is not predicted below 100% incidence. In CDS model, saturation at a level well below 100% is possible depending mainly on the distribution of $\Delta_g$. Nevertheless, a power law like relationship is also possible when the incidence is much below the saturation level. Different types of cancers show different shapes of age incidence curves. The power law like relationships can be explained by any of the models at certain parameters, but the saturating relationship with late life decline even at a small cumulative incidence is not described by the bad luck or the CIS model unless some extrinsic causes or statistical distributions are incorporated (Zhang, 2018).

2. The relationship between stem cell number and cancer incidence is differentially predicted by the three models. While the bad luck models expects a linear relationship with a slope = 1 on a log-log plot, the CIS expects a slope $\geq 1$. The CDS, in contrast, expects a saturating relationship or one with a decreasing slope. In the part of the curve that may be approximated to a linear relationship, the slope is $\leq 1$. Although Tomasetti and Vogelstein (Tomasetti and Vogelstein, 2015, Tomasetti et al., 2017) claim the cell number incidence relationship to be linear on a log-log scale, the slope is substantially smaller than one. Further our analysis (Supplementary Material S3) of the same data reveals that even on a log-log scale, a saturating curve describes the relationship better than a straight line. This supports the CDS model and contradicts the predictions of bad luck and CIS models.

3. The bad luck and the CIS models offer no direct explanation for non-mutagenic carcinogens, whereas the CDS model suggests that they can work by affecting the $\Delta_g$ distribution by altering microenvironmental conditions. Even carcinogenic exposures

such as ionizing radiations that were thought to affect mutagenesis alone, are now shown to act also by affecting the selective environment (Marusyk, 2009).

4. Peto's paradox is based on the assumption that the somatic evolution of cancer is mutation limited. The bad luck and CIS models are primarily mutation limited. Therefore they need an additional explanation for the Peto's paradox such as evolved anti-cancer defenses in organisms with large body sizes and longer lifespans. In CDS, on the other hand, mutations are not limiting and therefore a relationship with the body size or cell number is not expected through most of the size range. Our model is compatible with the argument that Peto's paradox can be explained by invoking selection limited cancer evolution (Nagy et al., 2007, Caulin and Maley, 2011, Noble et al., 2015, Tollis et al., 2017).

The model prediction matrix suggests that the bad luck model and the CIS model face many incompatibilities with real life data leading to rejection of these models, whereas the distribution of $\Delta_g$ in the population incorporated in the CDS model is sufficient to predict the observed patterns in the relationships between age, stem cell number and cancer incidence. Among the factors considered in the model, context dependent selection is necessary to simultaneously explain different types of shapes of the age-incidence curves, saturation at a lower incidence level, a saturation relationship between stem cell number and incidence, non-mutagenic carcinogens and Peto's paradox. While these phenomena have been separately explained by different models with different set of assumptions, simultaneous compatibility to all these phenomena is unique to context dependent selection among the models suggested so far.

## 5.6 Discussion

On the whole, the better prediction profile of the CDS model stems from the distribution of $\Delta_g$ in the population. The gradual late-life decline in incidence is because of the left-hand tail of the $\Delta_g$ distribution, where in some individuals mutants have very small, zero or negative fitness.

The role of age induced changes in the selective microenvironment is recognized by some of the prior cancer models (Rozhok et al., 2014, 2016, Rozhok and DeGregori, 2016, 2019) and supported by empirical evidence (Henry et al., 2010, Vas, 2012, Vas et al., 2012, Henry, 2015). However, these models do not address the late age saturation and decline. Incorporation of individual differences at any age class is essential to explain these patterns. Our model mainly considers individual differences in the selective environment but is compatible with the age induced changes as well. Some types of leukemia exhibit an early life peak in incidence and

a second gradual late life rise (Rozhok and DeGregori, 2019, Sallan, 2006). A combination of individual differences and age related changes in microenvironmental conditions will be able to explain such a phenomena better. The age dependent models differentiate between physiological aging and chronological aging (Rozhok and DeGregori, 2016). Within any population, individuals differ in physiological decline at a given chronological age, which necessitates incorporation of $\Delta_g$ distribution at a given chronological age in these models too. It is possible that the factors influencing the evolution of aging influence the mechanisms that determine the mean, distribution and temporal trends in $\Delta_g$ distribution. In addition, smaller animals have shorter cell generation times and faster turnover rates and as a result different species have different age-incidence curves.

The individual variation in selection could be attributed to the micro-environment in the tumor or precancerous niche, and includes all the factors that determine the selective advantage of oncogenic or precancerous mutants. Cellular interactions with the extra-cellular matrix (ECM) are known to be important in tissue homeostasis (Hansen and Bissell, 2000) and the growth of cancer cells (Cao, 2010). The tissue micro-environment is also a key player in the association between obesity and cancer risk, where obese ECM differs from normal ECM both biophysically (Druso and Fischbach, 2018) and functionally (Iyengar et al., 2016). The immune system also acts as an important mediator in cancer-related processes; a wide range of immune cells including fibroblasts and lymphocytes are known to influence cancer progression under various conditions (Hanahan and Coussens, 2012). Cancer is increasingly seen as an overactive or abnormal wound healing process (Hanahan and Weinberg, 2011) and signals of wound healing pathways might affect the growth rates of driver mutants. *In vitro*, the IGF-II concentration in culture media was found to markedly alter the selective advantage of an IGF-II over-expressing mutant in cell competition (Archetti et al., 2015). Precancerous cell lines across cancer stages also show different growth properties depending on growth factor concentration in the culture medium (Chan et al., 2014). Similarly, levels of estrogen and progesterone are known to play important roles in regulating mammary cell proliferation (Haslam and Woodward, 2001), secretion of ECM factors, tumor-stroma interactions (Woodward et al., 2000) and growth factor production (Dickson and Lippman, 1987). In intestinal tissue, colitis gave specific selective advantage to $p53$ mutants (Vermeulen, 2013). More recently, experiments have been reported in which behaviorally-enriched environments or physical exercise seemed to show cancer-suppressive effects (Cao, 2010, Rundqvist et al., 2013) that correlate with levels of particular growth factors. Studies in mice have also shown substantial variation in tumor sizes induced by identical genetic clones across individuals (Rogers, 2017). These observations together suggest that CDS factors from the microenvironment, be they genetic-epigenetic, biophysical, ECM-derived, GF-derived, hormonal or immunological, could play a limiting role in cancer progression. Identifications of the factors governing selective advantage of driver mutations needs to be a

focus of future cancer research which can potentially lead to novel lines of cancer prevention strategies.

# 6 | A mechanistic interpretation of patterns in cancer risk across biological scales

## Key contributions

This study is a collaboration between, in no particular order, Sutirth Dey (supervisor), Satavisha De (MS thesis student), and myself (PhD student).

The study has been conceptualized by Sutirth Dey and me. Satavisha and I have formulated the agent-based model used in this study, Satavisha has written the original codes for the simulations and has generated the primary data used in the study. Satavisha and I have carried out the initial analysis of the primary data, which forms Satavisha's Master's thesis (De, 2022).

For the current chapter, I have re-interpreted the primary data in a broader biological context. As part of this re-interpretation, I have re-analyzed and re-plotted the primary data in different configurations (different choices of parameter ranges and combinations, modified axes, etc.), in addition to generating new plots in the process (Figures 6.5 and 6.6). No figures have been reproduced in the current chapter as they appear in Satavisha's Master's thesis. Table 6.1 has been made by me and appears only in the current chapter, and Table 6.2 has been adapted from Table 2.2 in Satavisha's Master's thesis. I have written the following chapter from scratch with critical inputs from Sutirth Dey.

## 6.1 Introduction

Mutations are widely considered a basic substrate for cancer growth (Stratton et al., 2009, Nik-Zainal et al., 2012, Gerlinger et al., 2014, Martincorena and Campbell, 2015, Ma et al., 2018). Given that mutation is intrinsically linked to cell division, a question of fundamental importance in cancer biology is whether or not higher cell number and/or more cell division leads to an increase in the risk of cancer. The question arises essentially from the expectation that a larger number of cells and/or more frequent cell division presents greater opportunities for mutations to occur that could lead to cancerous growth (Peto et al., 1975, Peto, 2015). Given that both cell number and expected cell division frequencies vary over several orders of magnitude in biology, the scaling of cancer risk (as reflected by mutation risk) can be studied over at least four spatio-temporal scales of biological organization, and between these scales, trends in cancer risk appear to be remarkably divergent. Across species, larger and/or longer-lived organisms with potentially higher cell number and/or more lifetime cell division are not at a proportionately higher risk of cancer; in fact, the correlation between species cancer risk and body size and/or lifespan is very nearly nil (Abegglen et al., 2015, Gorbunova et al., 2014). Within a given species however, cancer risk does seem to scale with body size, as larger individuals among both humans and domesticated canines show a higher lifetime incidence of cancer (Albanes et al., 1988, Green et al., 2011, Nunney, 2018, Rafalko et al., 2023). At the sub-organismal level in humans, different tissues and organs show variation in cancer risk that can be associated with their cell number to some extent (Tomasetti and Vogelstein, 2015, Tomasetti et al., 2017), even if exposure to environmental risk factors is thought to play an important role in determining the eventual likelihood of carcinogenesis (Hochberg and Noble, 2017). Finally, trends of cancer incidence with age in the human population have been the subject of a rich pedigree of epidemiological inquiry. These studies, while foundational to the multi-stage model of cancer progression (Frank, 2007), are yet to explain the observed non-linear relationship between age and cancer incidence, which indicates that the potential for greater cell division expected from longer lifespan is not strictly predictive of cancer risk later in life (see Chapter 5 for a more detailed exploration of this idea). Given such diversity, a comprehensive understanding of carcinogenesis could be aided by the investigation of common processes that are known to act across these scales. One such possibility could stem from the fact that cancer is a competitive phenomenon between cells of varying fitnesses (Madan et al., 2022) and insofar as cell fitness pertains to cell growth rates, it can be considered an outcome of two processes: mutation and differentiation.

Cancer risk is often understood in terms of the occurrence and frequency of oncogenic mutations that increase mutant fitness relative to normal cells through a vast diversity of cellular processes (Gerlinger et al., 2014). Differentiation is essentially the establishment of a hierarchical tissue architecture that separates cells into stem and differentiated cell

pools, with a number of transit stages in between the two. This creates a gradient of division rates, such that stem cells can divide indefinitely, while transit and differentiated cells are typically limited in the number of division cycles they undergo (Pepper et al., 2007). Mutation and differentiation can also interact in a variety of ways. This includes mutations that can attenuate the rate of differentiation or induce de-differentiation, thereby enabling unrestricted cell growth (Varga and Greten, 2017). On the other hand, mutations that occur in a differentiation stage with a very limited division potential could restrict the extent to which the mutant clone can expand in the cell pool (Demeter et al., 2022). Both mutation and differentiation can therefore affect the relative frequencies of mutant and non-mutant cells in a given cell pool, and the overall risk of mutation accumulation and cancer progression is a composite function of both processes. In the context of cancer prevention, empirical evidence suggests that reducing overall mutation risk or increasing the efficiency of DNA repair can both be effective at reducing cancer risk (Tian et al., 2013, Sulak et al., 2016, Vazquez and Lynch, 2021). On the other hand, theoretical models have explored the possibility that the differentiation hierarchy could enable cancer prevention by restricting the spread of mutant clones (Werner et al., 2011, 2013, Hindersin et al., 2016, Derényi and Szöllősi, 2017, Derényi et al., 2020, Grajzel et al., 2020), although direct empirical evidence remains lacking.

One such theoretical study by Erten and Kokko (Erten and Kokko, 2020, , hereafter referred to as EK20), has recently used an individual-based model of metazoan ontogeny to arrive at optimum differentiation/division strategies that can support viable development over a range of body sizes. The EK20 model predicts that larger organisms would be expected to evolve lower rates of differentiation and more transit stages from stem to differentiated cells, effectively indicating that larger stem cell compartments could be a viable strategy to produce a larger organism. However, larger stem cell compartments could also be at a higher risk of mutation accumulation due to greater overall renewal capacity, setting the stage for a potential trade-off between viability and mutation accumulation even within ontogenetic timescales. EK20 also suggest that due to the multiplicity of parameters involved in metazoan development, an organism of a given size may be viably produced by a variety of parameter combinations. In turn, these parameter combinations could produce a range of cellular compositions for a given body size with varying levels of robustness to mutation accumulation and cancer risk.

On the one hand, EK20 has led to several interesting predictions regarding optimal ontogenetic strategies for metazoan development over a range of species sizes. On the other hand, the across-scale variation in cancer risk elaborated earlier suggests that the optimality of ontogenetic strategies inferred at the interspecific level may not be generalizable across other biological scales. Furthermore, EK20, while accounting for mutational damage-induced cell

death, do not include a biologically-relevant distribution of fitness effects (DFE) of mutations occurring over the course of ontogeny. Mutations in a somatic context fall on a spectrum of fitness, from deleterious to beneficial in context of cell growth and cancer (Loewe and Charlesworth, 2006, Eyre-Walker and Keightley, 2007), and a DFE that can reproduce such a spectrum of mutations in silico could thus lead to a holistic understanding of ontogeny and cancer risk.

The current study is an exploration of these possibilities, for which we construct an individual-based model (IBM) of the development of a cellular compartment of a specified target size at maturity from a single cell. While our framework is based on that of EK20, we forgo the genetic algorithm-based optimization strategy used by EK20 in favor of developing a mechanistic understanding of the interplay between dynamics of cell division and differentiation over the course of ontogenetic growth. Our results demonstrate that the extent of cell differentiation is a function of the ontogenetic strategy, and this in turn affects both the size of the actively-dividing cell pool and the average division frequency required to produce a mature cell compartment. We are thus able to identify a subset of ontogenetic strategies that lead to a greater frequency of cell division within a small pool of actively-dividing cells, which effectively magnifies the corresponding mutation risk for such strategies. More broadly, we are able to identify a range of strategies across target sizes that vary widely in their vulnerability to mutation accumulation, which qualitatively reflects the observed heterogeneity in the size allometries of cancer risk across biological scales. These results thus provide a theoretical framework that allows us to elucidate, across levels of biological organization, the conditions under which greater cell number does lead to higher cancer risk through mutation, and importantly, the conditions under which it does not.

## 6.2 Methods

We discuss the implementation of the model and its main findings in terms of the generic cell compartment, and then discuss the implications of our findings at each specific biological level in a later section. The model framework itself is also a part of the Master's thesis work of Satavisha De, and a description of the model therefore appears in the methodology of her Master's thesis (De, 2022). However, the model description has been substantially re-written for the current chapter.

Figure 6.1A shows a simplified scheme of the sequence of events in the simulation, and the other subpanels illustrate the chief processes and features of the model that govern the growth and composition of the cell compartment. Each individual cell is described by three attributes-fitness, differentiation stage and number of mutation. These attributes are affected by model processes including division, differentiation and mutation. We now detail

the parameterization of each of these processes following the overall sequence of events in Figure 6.1A.

### 6.2.1 Cell fitness

We begin the simulation with a generic cell compartment containing a single non-mutant cell in the least differentiated (stem) stage. This cell, upon division, or division with differentiation, can give rise to stem or differentiated cells respectively. We define $\omega$ as the cell fitness, such that cells with higher $\omega$ are more likely to divide. Since stem and non-stem cells differ in their division rates (Varga and Greten, 2017), we assume $\omega_1 = 1$, corresponding to the fitness of a stem cell in the least differentiated stage. A given stem cell must go through $T-1$ transit stages to reach the terminally-differentiated state and the transition of a cell from one transit stage to the next leads to a change in its fitness. Accordingly, we take $\omega_i$ to be the cell fitness in differentiation stage $i$, and model the effect of differentiation by assuming cell fitness to be a multiplicative function of the differentiation state as follows:

$$\omega_i = \alpha \times \omega_{i-1} \; \forall \; 1 < i \leq T-1, \tag{6.1}$$

Equation 6.1 shows that the effect of differentiation on cell fitness is mediated by $\alpha$. We define $\alpha$ as the gradient of cell fitness with differentiation and use this gradient to describe the full range of theoretical possibilities. Cell fitness can increase, decrease or remain unchanged with differentiation, and it is clear from Equation 6.1 that this corresponds to $\alpha > 1$, $\alpha < 1$, and $\alpha = 1$ respectively (Figure 6.1B).

### 6.2.2 Cell division

We use the cell fitness $\omega_i$ as the basis of division in the model. We define the divisible cell pool as $S(t) = \sum_i^{T-1} n_i$, where $n_i$ is the number of cells in differentiation stage, $i$. Cells in the stage $T$ are considered terminally-differentiated and not included in the divisible pool. In each iteration of the model, the target size of the compartment, $M$ is used to calculate required number of cells to fill the compartment, hereafter referred to as the deficit in cell number, $D(t)$. We define $D$ as the scaled difference between the target compartment size and the current number of terminally-differentiated cells in the organism, $0.05 \times [M - n_T(t)]$, where $n_T$ is the number of cells in the $T$ stage. The 0.05 scaling factor is included following EK20 to produce approximately sigmoid organismal growth (Karkach, 2006). If the deficit $D(t)$ is more than the total number of divisible cells $S(t)$, then all $S$ cells undergo division regardless of their fitness. If $D(t) < S(t)$, then a Bernoulli trial is carried out for each cell in the simulation with the probability of success defined as the relative fitness of the focal cell over the most fit cell in the pool, $\frac{\omega_{ij}}{\max(\omega)}$ for the $j$th cell in differentiation stage $i$. Among those cells that succeed in this trial, the first $D$ cells with the highest fitness are chosen to

undergo division.

### 6.2.3 Cell differentiation

Among the cells selected for division, division can occur in one of three ways with respect to the differentiation stages of the progeny cells (Figure 6.1E). When division is asymmetric, one progeny cell differentiates to the next stage while the other remains in the same stage as the parent. When division is symmetric, both progeny cells are in the same differentiation stage, and both could either differentiate to the next stage or stay in the same stage as the parent. We define two probability terms to implement these three possibilities stochastically. $P$ is defined as the probability of asymmetric cell division and $Q$ as the probability of no differentiation given that the division is symmetric. This implies that the three kinds of divisions outlined above can be expressed mathematically as: $P$, corresponding to asymmetric division, $(1-P)\times Q$, corresponding to symmetric division where both progeny cells remain in the same differentiation stage as the parent, and $(1-P)\times(1-Q)$, corresponding to symmetric division with both progeny cells differentiating into the next stage. Since the values of $P$ and $Q$ together determine how fast differentiation proceeds with cell division, we considered three different pace of differentiation in terms of combinations of and values. We call these three levels of pace of differentiation-maximum, equal and minimum-corresponding to whether cell division on average accelerates, retards, or does not affect differentiation respectively. Table 6.1 demonstrates the effective difference between the levels of pace with a numerical test case based on how many progeny cells are produced in differentiation stages $i$ and $i+1$ for every 100 parent cells dividing in differentiation stage $i$. As with cell division, we use the probabilities $P$ and $Q$ in sequential Bernoulli trials for each of the cells selected for division, first to decide if the cell divides asymmetrically, and if not, whether or not differentiation occurs. We then update the numbers of cells in the corresponding differentiation stages based on the outcomes of these trials, and adjust cell fitness according to Equation 6.1.

### 6.2.4 Mutation

The primordial stem cell at the beginning of the simulation starts with no mutations. After that, mutations occur probabilistically with every cell division event. We use a fixed mutation rate per cell per division, $\mu$, for a given simulation and as mentioned earlier, for each cell, we record the number of mutations it accumulates. For each dividing cell, we conduct a Bernoulli trial with the mutation rate and if a mutation occurs, we record its occurrence by updating the mutation count for that cell. We consider two regimes of mutation accumulation-neutral and non-neutral. Neutral mutations do not affect cell fitness, differentiation stage, or any of the differentiation rates, and their accumulation is therefore entirely driven by the overall mutation rate and the frequency of cell division. We assume that non-neutral mutations

change the value of cell fitness in its current differentiation stage and simulate this change by replacing $\omega_i$ with mutant fitness, $\omega_i'$. Based on empirical evidence on the distribution of fitness effects in the context of somatic evolution, the mutant fitness is sampled from a lognormal distribution with mean $9 \times 10^{-5}$ and standard deviation $12 \times 10^{-5}$ (Eyre-Walker et al., 2006) and the fitness for subsequent differentiation stages is re-calculated based on Equation 6.1 as $\omega'_{i+1} = \alpha \times \omega'_i$.

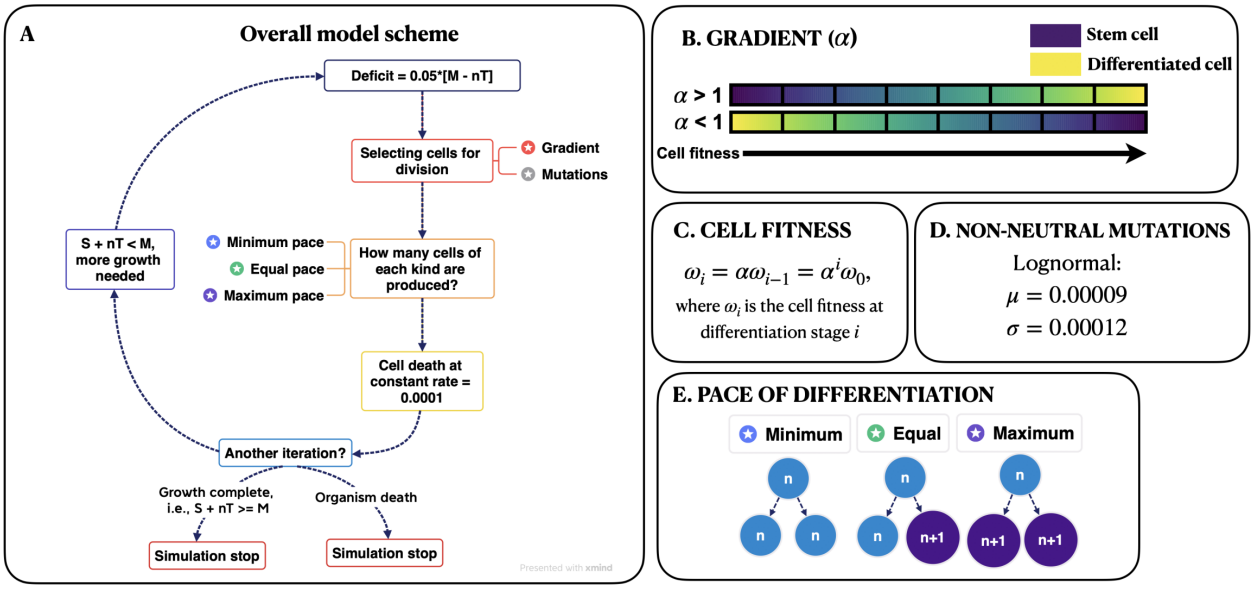



***Figure 6.1:*** **Sequence of events in the model and model components**. (A) A simplified scheme of the sequence of events in the model. (B) the effect of gradient $\alpha$ on cell fitness-for $\alpha < 1$, selection favors stem cells while for $\alpha > 1$, selection favors differentiated cells. (C) Exponential scaling of cell fitness with differentiation. (D) fitness effects for non-neutral mutations are sampled from a lognormal distribution of given mean and SD. (E) Pace of differentiation is simulated at three levels-specific parameter values and corresponding average number of differentiated cells produced per cell division are given in Table 6.1. All notations are explained in the main text.

| Pace of differentiation | $P$ | $Q$ | $(n_i, n_{i+1})$ |
|---|---|---|---|
| Minimum | 0.1 | 0.9 | $(172, 28)$ |
| Equal | 0.5 | 0.5 | $(100, 100)$ |
| Maximum | 0.1 | 0.1 | $(28, 172)$ |

***Table 6.1:*** **Numerical example of different paces of differentiation**. $P$ and $Q$ are the probabilities of asymmetric and symmetric non-differentiating division respectively, and the last column gives the average number of progeny cells in stages $i$ and $i+1$ for every 100 cells dividing in stage $i$.

### 6.2.5 Cell death

As with cell division and differentiation, cell death is simulated as a Bernoulli trial, with each cell having a constant probability of 0.0001 of dying in a given iteration of the model. Cells that succeed in the Bernoulli trial are removed from the simulation.

### 6.2.6 Termination of growth

We consider two conditions under which cell division is terminated: (1) if $S(t)+n_T(t) \geq M$, then the target size has been reached and the simulation is ended. (2) If $S(t)+n_T(t)$ is less than the target size $M$, but $S(t) = 0$, it implies that the stem cell pool has been depleted before target size could be reached and the simulation is terminated as development is considered unviable (both $S$ and $n_T$ have been defined in a previous section). It is important to note here that since $S$ is the total number of divisible cells across all stages of differentiation, target size $M$ can be reached by increasing $S$ and/or $n_T$, i.e., the number of cells in any differentiation stage from 1 to $T$. Going by condition (1) above, for a given target size, cell compartments could be composed of varying proportions of $S$ and $n_T$ even when development is viable. It also follows from our definitions that $D$ can only be reduced by an increase in the number of terminally-differentiated cells, $n_T$ (Figure 6.1A). Consequently, $n_T$ plays a key role in determining the rate at which the cell number deficit reduces during growth. Cell compartments of the same size differing in terms of $n_T$ could therefore result in different temporal dynamics of the cell number deficit. Taken together, these two features of the model allow compartment composition to vary independently of compartment size, such that cell compartments composed of different frequencies of stem, transit and differentiated cells can be produced over a range of target sizes. We consider this an essential feature of the model as it enables us to investigate the effects of compartment composition on mutation accumulation as well as how these effects scale with target size.

### 6.2.7 Simulations and parameter exploration

The following variables constitute the parameter space over which the behavior of the model will be explored: compartment size, $M$, the cell fitness gradient with differentiation, $\alpha$, pace of differentiation as defined by $P-Q$ combinations in Table 6.1, and the somatic mutation rate, $\mu$.

This parameter space is arguably vast and several simplifying assumptions are necessary for a tractable study. Due to computational constraints, we explore compartment sizes in the range of $10^3$ to $10^7$. This is a subset of the variation seen across internal organs in humans (Nunney, 1999, Tomasetti and Vogelstein, 2015), and several orders of magnitude below total body sizes across mammalian taxa (Savage et al., 2007). Within this range of

compartment size, early simulations showed that mutation rates corresponding to somatic mutation accumulation in human tissues of the order of $10^{-8}$ (Tomasetti and Vogelstein, 2015, and supplementary material thereof) resulted in very low mutation loads with very little scope for multiple mutational hits per cell. We therefore scale up mutation rates in our model to a final range of $10^{-2}$ to $10^{-6}$ per cell per division.

Empirical data for the fitness gradient $\alpha$ and pace of differentiation are rather limited, and we adopt a general approach for their exploration. As mentioned earlier, $\alpha = 1$ leads to no change in cell fitness due to differentiation and can therefore be considered a null case against which other possibilities can be compared. While it is typically assumed that differentiation usually leads to higher division rates (i.e., increasing fitness gradient), there is some empirical evidence that the fitness gradient with differentiation varies more substantially across human tissues (Bjerknes, 1985, Morrison and Kimble, 2006, Lopez-Garcia et al., 2010). We therefore explore decreasing, increasing and flat gradients, corresponding to $\alpha < 1$, $> 1$ and $= 1$ respectively, to cover a range of theoretical possibilities. The $P - Q$ combinations corresponding to the three levels of pace have been discussed in an earlier section (Table 6.1).

| Trait name | Range | Step size |
|---|---|---|
| Pace of differentiation | Minimum | N.A |
| | Equal | N.A |
| | Maximum | N.A |
| Gradient, $\alpha$ | $[0.1, 1.6]$ | 0.1 |
| Somatic mutation rate, $\mu$ | $[10^{-6}, 10^{-2}]$ | Decreasing by an order of magnitude |
| Target size, $M$ | $[10^3, 10^7]$ | Increasing by an order of magnitude |

***Table 6.2:*** **List of free parameters explored across simulations**. Detailed explanations of each parameters and the basis for their parameterization are given in the main text.

Table 6.2 summarizes the free parameters explored in this study and for each combination of these parameters, 100 replicate simulations have been run, except for the largest target size for which only 35-62 replicates could be run across levels of pace and somatic mutation rates due to computational constraints. All codes have been written in Python version 3.6 (VanderPlas, 2016) and replicate simulations run on the PARAM Brahma supercomputing facility in IISER Pune using custom SLURM scripts to parallelize replicate runs of a given parameter combination. In the interest of simplicity, we use a fixed value of $T$ for all simulations, but this is undoubtedly likely to be a parameter of interest in further study.

At the end of each simulation, we record the number of cells $n_i$ in each differentiation stage, $i$. The composition of the compartment across differentiation stages can be described by the weighted average value of differentiation stage, $\bar{\imath}$ given by:

$$\bar{\imath} = \frac{\sum_{i=1}^{T} i \cdot n_i}{\sum_{i=1}^{T} n_i} \tag{6.2}$$

We also record the number of cells $n_m$ that contain $m$ mutations, for $m$ ranging from 1 to 10. Based on this, we define the mutant clone length as the number of mutations accumulated in a given cell, such that cell that has accumulated 5 mutations over the course of a simulation is said to have reached a clone length of 5. Mutation accumulation can therefore be described by the weighted average clone length for the cell population $\overline{m}$ given by:

$$\overline{m} = \frac{\sum_{m=1}^{10} m \cdot n_m}{\sum_{m=1}^{10} n_m} \tag{6.3}$$

Finally, we also plot $M$ as the maximum value of $m$ in a given run of the model for which $n_m > 0$.

## 6.3 Results

This study seeks to investigate the variation in patterns of cancer risk allometry across multiple scales of biological organization. Using an individual-based model of the growth of a generalized cell compartment, we first examine the mechanistic basis of how the processes of cell division and differentiation respond to changes in the target size of the compartment, and then relate this to corresponding changes in the patterns of mutation accumulation across target sizes.

Model results given in the following sections develop this understanding sequentially, starting from the scaling of cell compartment composition with changing target sizes, and then elucidating the effects of cell compartment composition on the scaling of cancer risk. Finally, in the Discussion section, we consider the implications of changing target sizes in the model in terms of changing sizes of cell pools across levels of biological organization from species to organs. Based on current empirical evidence, we identify those processes and parameters that could play a mechanistic role in determining the direction of cancer risk allometry at each level.

### 6.3.1 Pace and gradient limit the effects of target size on final composition

Figure 6.2A shows the change in frequencies of cells across differentiation stages for different body sizes, gradients and all three paces of differentiation. From the first two rows, from left to right, we see that as the gradient $\alpha$ increases, the cell populations shift towards greater frequencies of the higher differentiation stages. This bias in the compartment composition is explained readily by how gradient affects the choice of cells for division; a steeper gradient, reflected either in very low or very high $\alpha$, leads to selection in favor of either relatively less differentiated or more differentiated cells respectively. This effect of gradient on the final composition is seen even more clearly in Figure 6.2B, which shows the mean differentiation stage weighted by the corresponding cell frequencies (Equation 6.2). Here, we see that as gradient increases, the weighted mean increases, following the shift in cell frequencies from lower to higher differentiated stages, with a clear transition at $\alpha = 1$. Moreover, comparing the cell frequency curves for a given target size across different levels of differentiation pace (for a given column across the rows in Figure 6.2A and B), it is clear that greater pace also increases the general push towards differentiation. The most drastic demonstration of this can be seen in the bottom row of Figure 6.2, where maximum pace swamps over all other parameters to produce organisms composed almost exclusively of terminally-differentiated cells. A significant proportion of such organisms are not developmentally viable and fail to reach the target size as they exhaust the divisible cell pool before maturity.

Figure 6.2B also shows that the target size has a clear effect on the final composition- larger target sizes produce a greater push towards differentiation and a greater proportion of differentiated cells relative to the stem cells. However, the degree to which this push is realized is dependent on both $\alpha$ and the pace of differentiation. Figure 6.2B visualizes this dependence very clearly, in which maximum pace once again serves as an extreme example of target size having no practical effect on final composition. It also shows the limiting effect of pace on the extent of differentiation with target size as larger target sizes progress to much more differentiation under equal pace than minimum pace. Likewise, the transition in body composition around $\alpha = 1$ is clearly conserved under both equal and minimum pace, albeit more clearly for smaller target sizes. We therefore surmise that while higher target size does lead to a generally higher propensity for differentiation, ontogenetic properties like pace and $\alpha$ have the potential to act as tuning knobs on how final cellular composition responds to changes in target size.

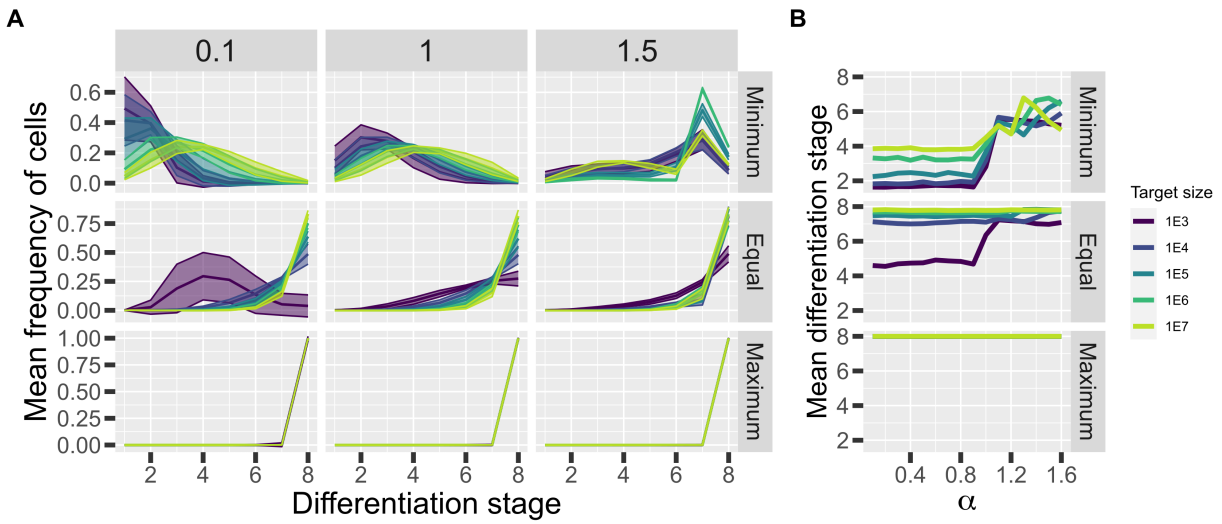



***Figure 6.2:*** **Cell compartment composition as a function of ontogenetic strategy**. (A) Relative frequencies of cells against differentiation stages across all target sizes, all three levels of pace and select values of $\alpha$; column headers give the value of $\alpha$ and row headers give the pace level; bold lines are averages over all replicate simulations and ribbons are corresponding SDs (see main text for exact number of replicates). (B) Weighted average differentiation stage against for all three pace levels as given in the row headers, across all target sizes. The calculation of the weighted average differentiation stage is explained in the main text. All values are averaged over replicate runs (see main text for exact number of replicates). Note that the y-axis scale differs across rows.

### 6.3.2 Deficit dynamics explain how gradient and pace give rise to different cellular compositions

Cell division in the model hinges entirely on the size of the deficit in each time step; a larger deficit would require a larger number of cells to be selected for division and as explained earlier, $n_T$ (number of terminally-differentiated cells) determines the size of this deficit. The previous section demonstrated that $\alpha$ and pace influence the extent of differentiation, which suggests that they could also affect how the deficit changes with time. In this section, we examine this possibility by considering how $\alpha$ and pace affect the temporal dynamics of the cell number deficit and cell division in the model. Figure 6.3A shows that the cell number deficit declines relatively faster under equal pace of differentiation than under minimum pace. This decline in cell number deficit corresponds to the trends in the number of rounds of cell division required to reach maturity-as shown in Figure 6.3B, equal pace simulations consistently undergo as many or more rounds of division than minimum pace. Likewise, we see that extreme values of $\alpha$, both very low and very high, entail more rounds of cell division than intermediate values of $\alpha$. Understanding these effects of pace and $\alpha$ requires a closer look at the workings of the model.

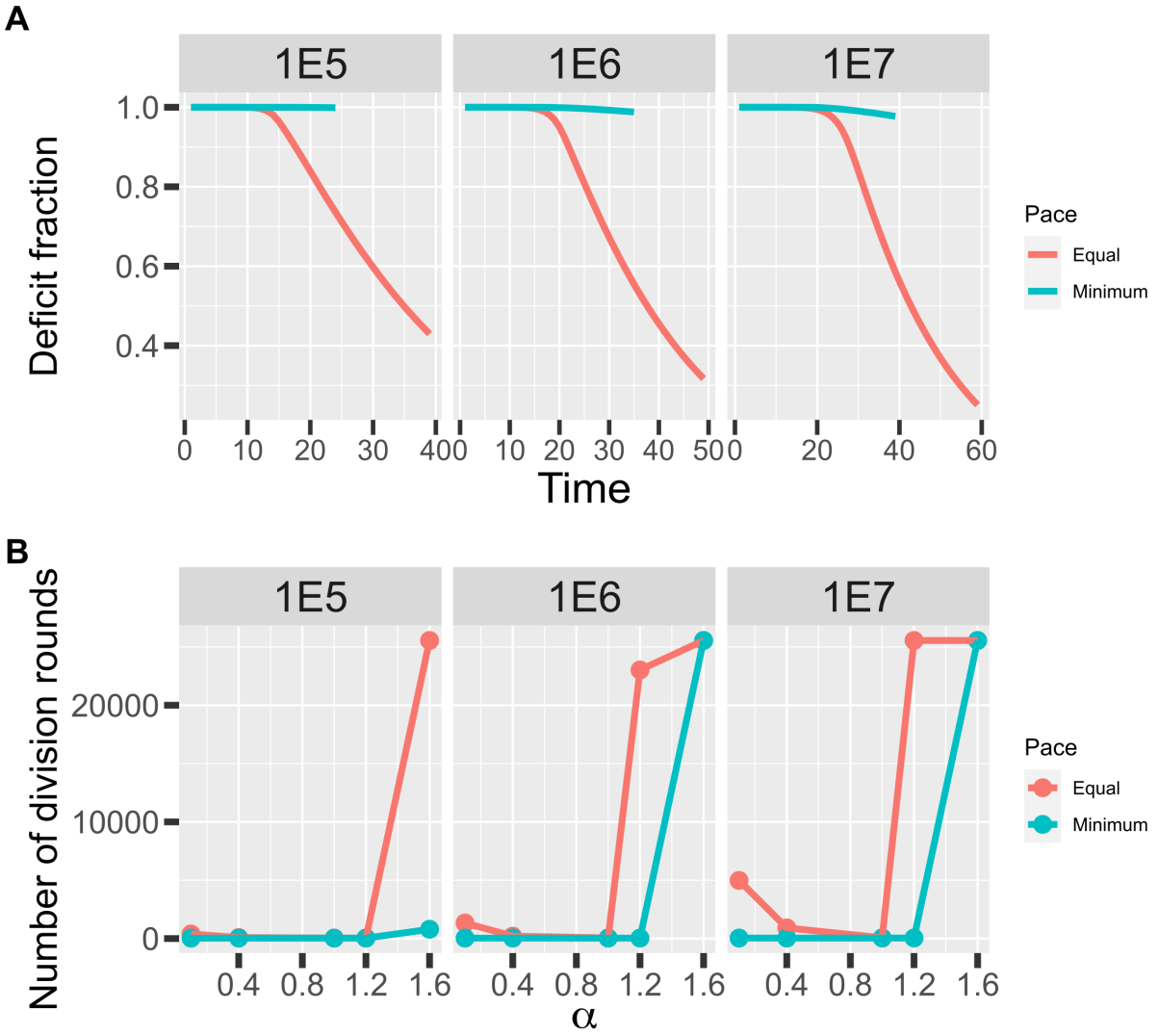



***Figure 6.3:* Conditions for tapering deficit dynamics**. (A) Deficit dynamics against time for two levels of pace and three target sizes, showing equal pace has a decreasing deficit in time whereas minimum pace maintains a higher deficit for a longer time. (B) Number of rounds of division simulated over the course of development in the model; large values of $\alpha$ and higher pace are clearly associated with more rounds of division, which in combination with a small deficit, implies a progressively smaller dividing pool of cells. For larger target sizes, number of rounds of division increases for very low values of $\alpha$ as well. All lines are averages over 9-11 replicate simulations.

Our selection algorithm begins by ranking cells by fitness, so we start with a sequence of cells arranged in decreasing order of fitness. Selection could be visualized as a "window" placed on this ordered sequence such that cells within the window are selected for further division. The deficit corresponds to the size of this window and a larger window leads to more cells being selected. Since $\alpha$ represents the scaling of cell fitness with differentiation (Figure 6.1B), it determines which differentiation stage is highest in fitness. This, in turn, determines the differentiation stage of most of the cells at the top of the fitness hierarchy

which are included in the selection window for subsequent division. Pace determines the number of further differentiated cells produced per division event, and lower pace leads to lesser overall differentiation. Consequently, when differentiation pace is low, cells in the selection window are more likely to be from a narrow range of differentiation stages if not a single differentiation stage and by corollary, higher pace leads to a more diverse group of differentiated and less differentiated cells within the selection window. Based on this logic, it can be seen that minimum pace and small values of $\alpha$ together constitute a strategy for producing preferentially less differentiated cells, whereas higher pace and higher $\alpha$ both increase the push towards differentiation. We refer to these as "low differentiation" and "high differentiation" strategies respectively, based on the net progression of differentiation in each case. As the cell number deficit is only decreased by the accumulation of terminally-differentiated cells ($n_T$), parameter combinations corresponding to high differentiation strategies are also more likely to reduce the deficit in time whereas those that restrict differentiation (low differentiation strategies) would maintain a large deficit for much of the simulated time. Furthermore, as deficit remains high for low differentiation strategies, growth happens in saltatory leaps forward and the target size is eventually reached or even exceeded with most cells being from the lower differentiation stages. On the other hand, high differentiation strategies lead to progressive reductions in deficit with time and the target size is reached almost asymptotically, with fewer and fewer cells selected in each iteration, and the target size approached in smaller steps. The parameter combination of equal pace and very low values of $\alpha$ does not strictly fall in this spectrum of high and low differentiation strategies as the two parameters act somewhat in opposition to each other. However, very low values of $\alpha$ favor less differentiated cells for division whereas equal pace leads to a higher frequency of differentiation. The selection window is placed around the lowest differentiation stages, which are progressively depleted through differentiation. The size of the divisible cell pool therefore reduces progressively, and as with the high differentiation strategies, the target size is consequently approached in smaller steps as fewer cells are available for division.

Taken together, this shows that pace and $\alpha$, by changing the way in which cells are selected for division and the extent to which cell division contributes to differentiation, can affect the cellular composition of the compartment at maturity.

### 6.3.3 A reducing deficit increases accumulation of neutral mutations

Given that mutation in our model is intrinsically linked to cell division, changes in cell division patterns could lead to altered patterns of mutation accumulation as well. This is highly relevant for cancer risk as much of cancer origin and progression is thought to occur through sequential accumulation of multiple mutations within a single cell or a small

group of cells (Frank, 2007, Calabrese and Shibata, 2010, Nunney and Muir, 2015). Figure 6.4 shows the average maximum clone length as defined in Equation 6.3 for three values of the somatic mutation rate across columns. Here, we consider both neutral mutations that do not affect cell fitness (Figure 6.4A), as well as non-neutral mutations that influence the fitness (Figure 6.4B). Sequential accumulation of neutral mutations is affected by the number of rounds of division such that, where more rounds of division are required to reach maturity, maximum clone lengths also tend to be higher on average (*cf* Figure 6.3B and Figure 6.4A). Strategies that entail more rounds of division therefore also lead to more sequential mutation accumulation. Furthermore, we see that increasing target size likewise increases the maximum clone length regardless of the differentiation strategy. This is an important observation as it represents the direct effect of the scaling of cell division with target size on the baseline risk of mutation, independent of somatic evolution driven by the fitness of mutant clones.

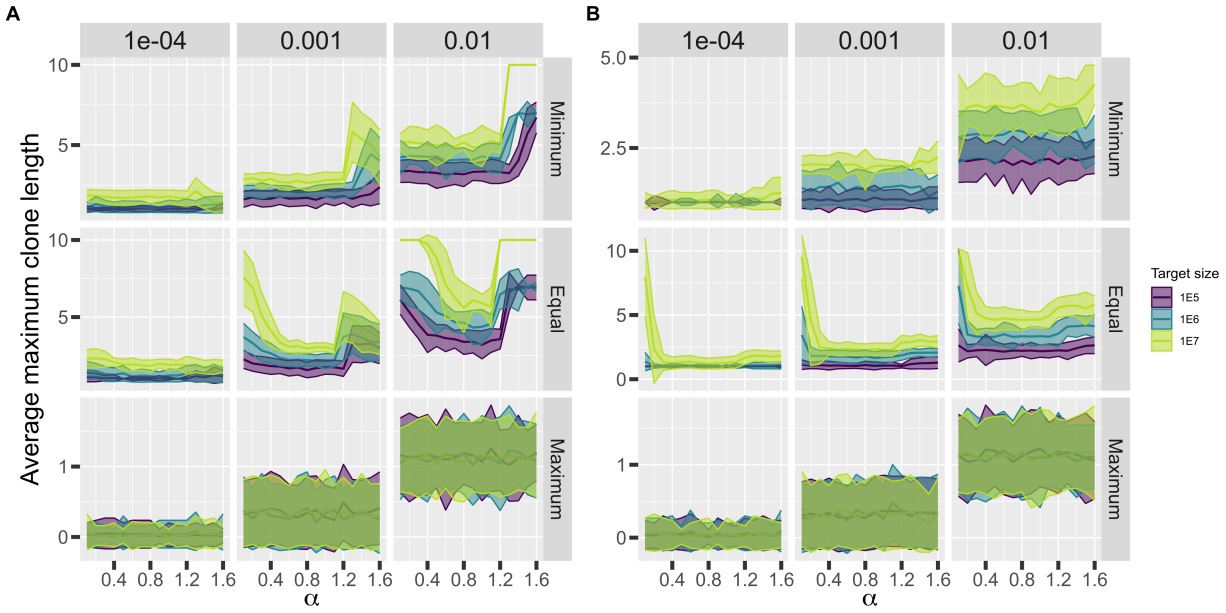



***Figure 6.4:*** **Maximum clone length corresponds to conditions of tapering deficit dynamics**. Maximum clonal length as defined in the main text against under (A) neutral and (B) non-neutral mutation accumulation. Columns are somatic mutation rates and rows are pace levels. Plotted values are averages over all replicate simulations (see main text for exact number of replicates). Note that the y-axis ranges change from (A) to (B).

We return to our earlier explanation of cell division in the model to understand these trends in the accumulation of neutral mutations. As identified earlier, high differentiation strategies, as well as equal pace with very low $\alpha$, both entail more rounds of division due to a reducing deficit and a slow approach toward the target size (Figure 6.3). As deficit declines, it also leads the selection window to shrink, so that as the target size is approached, a small set of cells with the highest fitness are selected repeatedly and undergo several rounds of division.

The higher frequency of division of a small cell pool therefore creates more opportunities for mutations over a short period of time, and particularly for multiple mutational hits in the same set of cells. On the other hand, low differentiation strategies are not associated with a declining deficit (Figure 6.3) and the selection window remains large throughout the growth period. The pool of actively dividing cells is large enough that the same cells are less likely to be selected repeatedly for division, making the sequential accumulation of multiple mutations in the same cell less likely. The maximum clone length therefore closely follows those growth strategies that produce a declining deficit.

Two more trends are important to point out in the context of mutation accumulation. First, increasing the target size and the somatic mutation rate both have qualitatively similar effects on the maximum clone length, as clone length generally increases in either case (Figure 6.4A). Second, our data also indicate that a higher clone length does not automatically imply higher mutant frequencies in the cell pool (Figure 6.6). This implies that mutation accumulation and mutant clone growth are independent processes, and while the differentiation strategy affects the risk of mutation accumulation, the actual expansion of potentially oncogenic mutations probably occurs over much longer timescales than what we have considered in our model. This is consistent with the general understanding that cancer progression involves both the accumulation and subsequent growth of oncogenic mutations (Aktipis and Nesse, 2013).

### 6.3.4 Non-neutral mutation accumulation reveals a balance between differentiation and mutation

The understanding of neutral mutation dynamics developed above can now be extended to elucidate the accumulation of non-neutral mutations. The distribution of fitness effects of mutations in our model generates largely deleterious mutations, such that large-effect beneficial mutations are relatively rare. Comparing Figures 6.4A and B, we see that in most parts of the parameter space, the trends in maximum clone length are qualitatively similar between neutral and non-neutral mutation accumulation, but the latter shows a markedly lower frequency of mutations overall with shorter clone lengths. This is consistent with the fact that deleterious mutations, which are by far more common in the model, are filtered out of the cell population by selection, which leads to fewer mutations on the whole.

However, low $\alpha$ under equal pace presents a unique region in the parameter space in terms of non-neutral mutation accumulation, as it shows particularly large maximum clone lengths (Figure 6.4B, middle row). As it has been pointed out earlier, this parameter combination leads to qualitatively similar dynamics of the cell number to high differentiation strategies, in that it leads to a progressively shrinking pool of actively-dividing cells that undergo many

more rounds of division (Figure 6.3B). Based on this similarity, patterns in non-neutral mutation accumulation under low $\alpha$ and equal pace are expected to be similar to those in high differentiation strategies. However, the data indicate that non-neutral mutation accumulation under low $\alpha$ and equal pace run contrary to this expectation, showing much larger maximum clone lengths than high differentiation strategies (*cf* Figure 6.4A and B, middle rows). This discrepancy is intriguing as two growth strategies with seemingly similar dynamics of the cell number deficit show divergent trends in the accumulation of mutations. In order to understand the cause for such a disparity, it is necessary to examine the relative frequencies of non-mutant and mutant cells for low $\alpha$ with equal pace, which clearly shows a much higher frequency of mutant cells with multiple mutations (Figure 6.6, middle rows). This arises as a consequence of the fact that under low $\alpha$ and equal pace, high-fitness mutations can potentially compensate for the reduction in cell fitness due to differentiation. Therefore, the cell number deficit can be reduced by higher frequencies of differentiated cells with a high extent of mutation. Since very low values of $\alpha$ lead to a drastic reduction in cell fitness with differentiation, multiple mutations are more likely to increase cell fitness sufficiently for them to be included in the selection window. This therefore indicates that under non-neutral mutation accumulation, the cell number deficit is reduced by different sets of cells under low $\alpha$ and equal pace than under high differentiation strategies. The former fills the deficit through the division of differentiated and multiply-mutant cells, while the latter fills the deficit through the division of differentiated, non-mutant cells. Low $\alpha$ and equal pace thus allows for antagonistic action between differentiation and mutation through their opposing effects on cell fitness. This antagonism allows the low $\alpha$ and equal pace strategy to produce divergent patterns in the accumulation of non-neutral mutations despite the dynamics of cell number deficit being qualitatively similar to that of high differentiation strategies.

### 6.3.5 Viability corresponds to depletion of the dividing cell pool

For the purposes of this study, we consider a given instance of a simulation to be viable if the total cell number reaches the target size within a margin of 10% lower or higher than the actual target size. From Figure 6.5, it is clear that under both neutral and non-neutral mutation accumulation, viability reduces consistently in regions of the parameter space corresponding to high differentiation strategies, namely high $\alpha$ and/or high pace. Figure 6.2 clearly shows that high differentiation strategies also lead to compartments with very low frequencies of the lower differentiation strategies. Particularly for the largest target sizes, this leads to an insufficient number of divisible cells for the target size to be reached.

Furthermore, viability also drops under non-neutral mutation accumulation for low values of $\alpha$, which corresponds to the unusually high clone lengths pointed out earlier (*cf* Figures

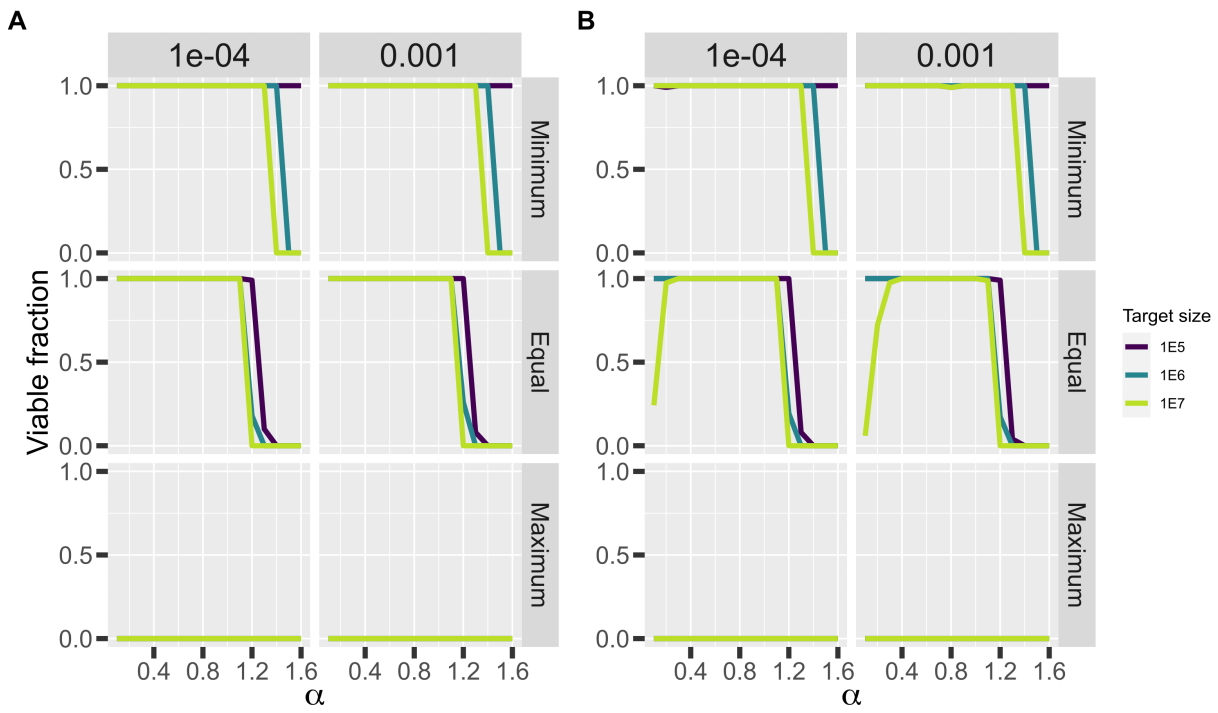



***Figure 6.5:*** **Developmental viability is related to mutation accumulation.** Fraction of viable model runs in which the target size was attained without stem cell depletion under (A) neutral and (B) non-neutral mutation accumulation. $\alpha$ is on the x-axis here, columns are somatic mutation rates and rows are pace levels. Plotted values are averages over all replicate simulations (see main text for exact number of replicates).

6.5B and 6.4B). It follows from the previous explanation that this increase in clone length excludes differentiated non-mutant cells in favor of less differentiated non-mutants and high-fitness mutants. This effectively shrinks the pool of cells available for division over the entire course of the simulation, making it more likely that divisible cells are exhausted before the target size is attained.

## 6.4 Discussion

We have shown here that while cellular composition is a composite function of $\alpha$, the pace of differentiation, and the target size, patterns in sequential mutation accumulation across these parameters can actually be understood in terms of the temporal dynamics of the cell number deficit over the course of ontogeny. More specifically, we identify so-called high differentiation strategies that are consistently associated with greater propensity for differentiation over the course of cell division (Figure 6.2) and a small cell number deficit over an extended period of time (Figure 6.3). High-differentiation strategies consequently involve the repeated selection of a small pool of high-fitness cells that undergo a disproportionately higher number of rounds of cell division. We also show that this repeated division of the same set of cells can be related directly to the sequential accumulation of several mutations

in the same cell (Figure 6.4). It bears repeating here that we take mutation accumulation as a direct proxy for cancer risk, in accordance with the current understanding of the centrality of mutation in the origin and progression of cancer. The inferences from our data regarding mutation accumulation applies to this general notion of cancer risk, such that greater mutation accumulation in some region of the parameter space indicates a higher risk of cancer for the corresponding combination of parameter values. So far, we have been discussing our findings purely in terms of how the various model parameters function to produce the patterns of cell division and mutation, and their scaling with target size. This general mechanistic understanding can now be extended to the cancer risk trends across biological scales mentioned earlier, and at each level of biological organization, we can use the model findings to speculate about mechanisms to explain the emergence of corresponding trends in cancer risk. More specifically, at each level of biological organization, we recast the known trends in cancer risk at that level in terms of the plausible action of our model parameters as applicable, in order to identify model-based mechanisms that are consistent with the cancer risk trend in question. This leads us to predictions regarding which parameters are likely to be of mechanistic importance across biological scales and therefore merit more focused investigation.

We begin at the highest organizational level, with the trends in cancer risk across species of varying body sizes. Within this biological level, cancer risk trends have been studied most extensively among mammalian species, and we therefore place our results in the context of cancer risk scaling among mammals (Abegglen et al., 2015). As mentioned earlier, empirical data regarding how cellular composition changes with body size across mammals are rather sparse. While it is known that some cell types become larger with increasing body size among birds (Kozlowski et al., 2010), and to some extent, mammals (Savage et al., 2007), it is unclear if the actual number of stem cells and/or their proportion relative to differentiated cells change systematically with body size. Results from EK20 suggest that an increase in the size of the stem cell compartment could be possible in larger species (Erten and Kokko, 2020). In the context of our model, sustaining a larger stem cell compartment requires a reduced progression of differentiation, which would be facilitated most by small values of $\alpha$. Moreover, for any given value of $\alpha$, higher pace of differentiation exhausts the stem cell compartment entirely, leaving very few cells in the lowest two differentiation stages (Figure 6.2). This indicates a potential upper limit for the pace of differentiation as body size increases across species, such that larger organisms would need to limit the rate of ontogenetic asymmetric divisions to prevent depletion of the stem cell compartment. On the other hand, we also find that $\alpha << 1$, while supporting a large stem cell compartment, also leads to significantly longer clonal lineages and higher mutant frequencies (Figures 6.4 and 6.6). Sequential mutation accumulation could therefore be the basis for a tradeoff between $\alpha$ and the pace of differentiation in determining the scaling of cellular composition with body

size. Since we consider sequential mutation accumulation a proxy for cancer, we speculate that lower pace of differentiation and intermediate values of $\alpha$ are likely to be consistent with increasing body sizes across species without a proportionate increase in cancer risk. The simple case of non-neutral mutation accumulation that we tested (Figure 6.4B) does not seem to contradict this expectation as mutant lineages become dramatically longer with body size only for equal pace and small $\alpha$.

One step below on the organizational ladder, the scaling of cancer risk with body size between individuals of the same species can also be explained based on our model results. Since $\alpha$ and the pace of differentiation are both ontogenetic variables, it is plausible to assume that for a given species, their values remain constant from one individual to the next. Based on this assumption, we see that for a given value of $\alpha$ and differentiation pace, the extent of mutation accumulation, both in terms of clonal length and to a limited degree, relative mutant frequencies, does in fact increase with increasing target size (Figures 6.4 and 6.6). Within this regime, mutation accumulation seems entirely limited by cell number and mutation rate, and increasing either one of the two increases mutation risk. The correspondence of these trends from our data with empirical observations of cancer risk is straightforward, as cancer risk is known scale with body size within a given species (Albanes et al., 1988, Green et al., 2011, Nunney, 2018, Rafalko et al., 2023). Our model data are therefore consistent with the possibility that ontogenetic parameters could be conserved between conspecifics. On the other hand, our data also suggest that larger individuals of the same species should have a greater flux towards differentiation and therefore end up with a higher proportion of more differentiated cells than smaller conspecifics. To our knowledge, empirical data have not been collected at such resolution within any species, which precludes further comments on the question.

Organ-level cancer risk arguably involves variation in almost all our model parameters, including the somatic mutation rate. Research over the past few years has shown that while much of the variation in cancer risk between organs could be ascribed to the number of lifetime stem cell divisions in each tissue (lscd, *sensu* Tomasetti and Vogelstein, 2015, Tomasetti et al., 2017), tissue-specific effects can shift the cancer risk-lscd correlation along the x-axis and/or change its slope (Noble et al., 2015). As suggested by Noble and Hochberg (Hochberg and Noble, 2017), such tissue-specific effects could stem from differential exposure to environmental risk factors leading to tissue-specific somatic mutation rates. But apart from these external influences, different tissues have different target sizes, and to some extent, organizational features like the architecture of the stem cell compartment are also known to vary across human tissues (Bjerknes, 1985, Morrison and Kimble, 2006, Lopez-Garcia et al., 2010). Our model data reflect part of such differences, such that different tissue organizations across organs could be consistent with different values for $\alpha$ and differentiation pace in the

model. On the other hand, we do not vary the number of transit stages, $T$ between the basal stem cell and differentiated cells, even though $T$ is thought to differ across tissues (Demeter et al., 2022). This is a key limitation for mutation accumulation in particular as the tapering dynamics of the cell number deficit shown in Figure 6.3 frequently occur within the highest few differentiation stages and the addition of more stages could postpone the tapering of the deficit sufficiently to allow the target size to be reached with fewer mutations. Future theoretical work would therefore need to be targeted at studying the effects of changing $T$ on ontogenetic dynamics across organs to elucidate any factors that determine cancer risk in a tissue-specific manner.

Finally, the tapering of deficit towards the end of ontogenetic growth suggests an interesting possibility for the accumulation of mutations over the course of post-maturity organismal life. It has been shown that mutations accumulate nearly linearly with age in most tissues due to cell divisions as part of baseline turnover (Blokzijl, 2016). While we do not explicitly simulate aging and lifetime somatic maintenance in our model, long-term cell turnover could be comparable to a constant small deficit over a long period with no change in any of the other parameters. Under such circumstances, unless the mutation rate itself changes, mutation accumulation would be expected to proceed more or less monotonically with further lifetime cell turnover as division continues to be concentrated within a small pool of dividing cells. A recent study has shown that somatic mutation rates across mammalian species scale negatively with lifespan but not body size (Cagan et al., 2022), which suggests that the prolonged cell turnover likely has a stronger role to play in determining mutation accumulation than a larger cell pool. *A priori*, this is consistent with the observed effects of a tapering cell number deficit on mutation accumulation in the model across target sizes. On the other hand, variables like the pace of differentiation could also change with age due to senescence-related changes in the stem cell microenvironment (Liggett and DeGregori, 2017), indicating that this is a rather simplistic view that is only a starting point for further investigations.

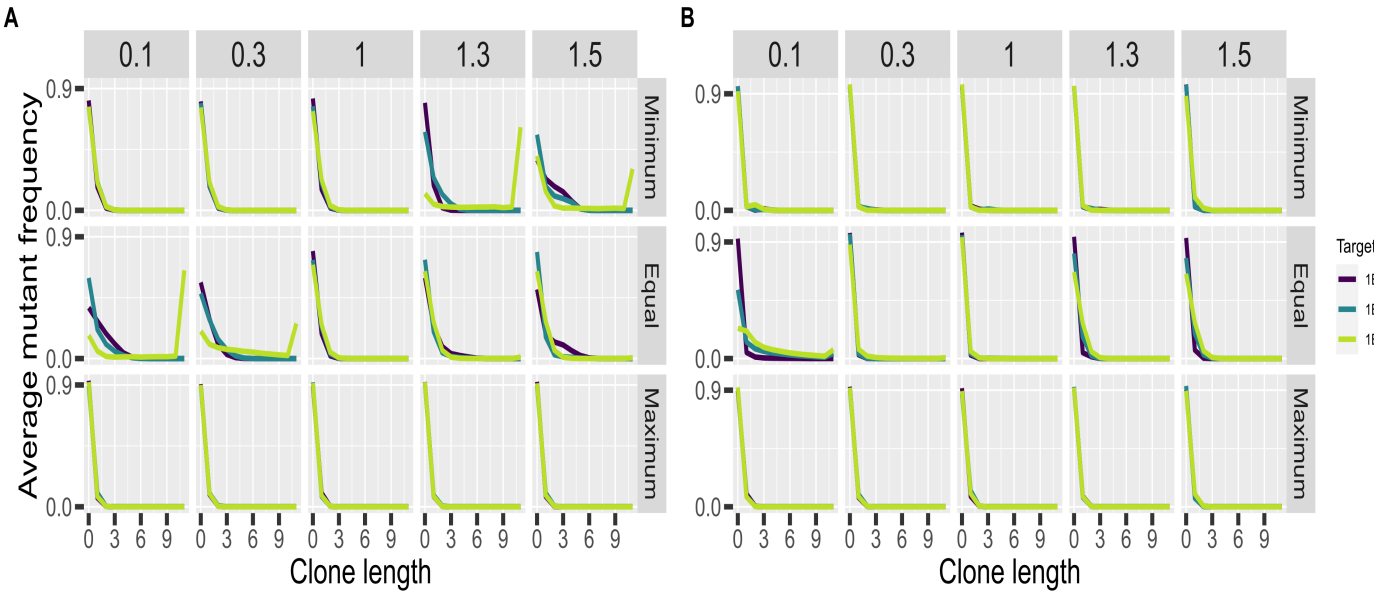



***Figure 6.6:*** **Mutant clone growth across model parameters**. Relative cell frequency against clone length, which is the number of mutations in a given cell under (A) neutral and (B) non-neutral mutation accumulation. All plots correspond to mutation rate, $\mu = 0.01$ per cell per division. As with previous figures, pace of differentiation varies across rows and across columns. Replicate numbers are the same as in Figure 6.2.

# 7 | Conclusions and future directions

Across scales of organization, ecological and evolutionary processes play key roles in determining the behavior of biological systems. Over the course of the last five chapters, two broad study systems were considered for the elucidation of the importance of ecological and evolutionary phenomena in their functioning. The first part consisted of two studies in an experimental evolution system in which lab populations of fruit flies were selected for greater dispersal, while the second part consisted of three theoretical studies that investigated several ecological and evolutionary phenomena of interest in the context of cancer.

## Dispersal evolution

Over the past several years, evidence from a wide variety of experimental systems has demonstrated the centrality of dispersal in organismal life history (Bonte and Dahirel, 2017), as well as the ecology and evolution of natural populations (Kot et al., 1996, Travis and Dytham, 1999, Lindström et al., 2011, Clobert et al., 2012). Empirical studies in the literature revolve around three central aims-documenting instances of dispersal in real-life biological systems, measuring as many features of the system as possible that could shed light on the causes and consequences of dispersal, and finally, elucidating the biological mechanisms that circumscribe the physical process of dispersal. The first two aims have been the focus of much of the mainstream investigation in dispersal over years, through both direct field observations as well as controlled mesocosm experiments (Whitmee and Orme, 2013, Stevens et al., 2013, Lustenhouwer et al., 2023), while the third aim has gained in prominence in recent years (Clobert et al., 2009, Saastamoinen and Rantala, 2013, Campos-Candela et al., 2019, DiLeo et al., 2022). Additionally, theoretical models have made significant contributions towards constructing and testing hypotheses regarding dispersal evolution (Travis and Dytham, 1999, Muller-Landau et al., 2003, Petrovskii and Morozov, 2009, Baines et al., 2020). The two mesocosm studies in this thesis on dispersal evolution in lab populations of fruit flies are rooted in the empirical background, and insofar as mesocosm experiments provide control over a large number of confounding environmental factors, they allow for a tractable study of the consequences of dispersal selection in a rigorous and robust setting. Two key insights from our data are worth highlighting again, particularly because they also point to avenues of future theoretical and empirical investigation.

First, our data clearly show that the evolution of dispersal is attenuated by the natal nutritional environment. While depletion of protein from the larval diet does not impede the evolution of higher dispersal capacity *prima facie*, we also see a distinctly diminished capacity for long-distance dispersal in the selected flies compared to unselected controls, but only when controlled for their respective average disperal distances. It has been elaborated elsewhere that this reduction in LDD could have direct consequences for local adaptation, invasion dynamics and long-term population stability based on the current understanding of

dispersal theory. However, our data also illustrate the evolution of a non-monotonic dispersal kernel when dispersal evolves under larval malnutrition, which to the best of our knowledge, is unique among existing reports of insect dispersal. While some theoretical work has attempted to explain the emergence of non-monotonic seed dispersal kernels in terms of parent-offspring conflict (Starrfelt and Kokko, 2010), both the basis of a non-monotonic dispersal kernel as well as its consequences for population dynamics in animal systems remain relatively unexplored. Whereas the emergence of monotonically-declining kernels in animals has been understood in terms of random walk processes or some variants thereof (Vinatier et al., 2011), no such conceptual framework exists currently that can generate a non-monotonic dispersal kernel in animals. How such kernel shapes could arise out of a large, heterogeneous population, and how they could further affect the evolution of the population could thus be the subject of considerable theoretical investigation.

Second, the measurement of physiological and behavioral traits in our system reveals the potential for a markedly different cost structure for dispersal evolution under larval malnutrition compared to when dispersal evolves under standard nutrition. More specifically, we see that larval malnutrition has likely led to the manifestation of costs of dispersal where none were found under standard nutrition-this is reflected in the trends in female body weight, male stress resistance, fecundity, mating characteristics and male-male aggressive behavior. It hardly needs to be emphasized that this is a subset of the whole phenotype space in *Drosophila* that could be explored in relation to dispersal, but the fact that traits considered fundamental to the organism's life history such as body weight, fecundity, and stress resistance show trends consistent with a cost of dispersal evolution signifies them as a viable starting point for the elucidation of the mechanisms of dispersal in our flies. The state-of-the-art for mechanistic elucidation largely revolves around the identification of candidate genes that could ostensibly be linked either to some aspect of dispersal directly or some dispersal-associated phenotype, if not both (Saastamoinen, 2008, Cayuela et al., 2020). While it is unclear if the identification of candidate genes is by itself sufficient for a meaningful understanding of how dispersal evolves, gene-level information can often be supplemented by independent measurements of small and large metabolites to enable the characterization of the dispersal-associated phenotype in more functional terms (Zera and Denno, 1997, Zera and Zhao, 2004, Wijerathna and Evenden, 2019, Southon et al., 2020). For our system specifically, the potential involvement of neurotransmitter signaling in the evolution of dispersal and its trait correlations has been discussed extensively, and further characterization of the phenotypic divergence between the selected and control flies would produce more leads for understanding how dispersal has evolved. While much of our discussion regarding the trait correlations with dispersal observed in our data has been set in the context of the so-called dispersal "syndromes", functional links that serve to explain how and why traits are related to dispersal probably offers a more tangible path towards a comprehensive framework for

dispersal evolution than one that relies on demonstrably labile trait correlations alone to generate predictions.

Based on the above studies, the current understanding of sexual dimorphism in the lab's dispersal selection populations remains rather nascent. Both in certain properties of the dispersal kernel, as well as certain physiological and behavioral traits, responses to dispersal selection show some signs of changing in sex-specific ways. However, no consistent pattern has yet emerged, in kernel evolution or among the correlated traits or some combination thereof, that allows us to speculate on whether and how dispersal could have evolved differently between the two sexes. It would be an understatement to say that sexual dimorphism in dispersal is a complicated matter, and there is a considerable volume of empirical and theoretical literature that has cumulatively identified several possible circumstances under which a bias in dispersal could emerge in either sex (Li and Kokko, 2019, Trochet et al., 2016). For our experimental system, the chief priority must be the characterization of the dispersal phenotype in much greater detail to identify any correspondence with expected drivers of sexual dimorphism in dispersal, such as mate and/or resource availability (Meier et al., 2011, Hovestadt et al., 2014, Stefanescu et al., 2021), or sex-specific constraints stemming from the mating system (Greenwood, 1980). Previous work has shown that selection for dispersal significantly alters the overall dispersal "syndrome" in a sex-specific manner, such that trait correlations with dispersal change between single-generation dispersers and evolved dispersers differently in either sex (Mishra et al., 2018, 2020). It would be interesting to explore if these findings hold for dispersal evolution under poor nutritional conditions, and to what extent they are consistent with prior findings. As with the nature of kernel evolution, these questions offer considerable scope for both empirical and theoretical investigation.

Notwithstanding the various implications of our results discussed so far, the lab's experimental system for dispersal selection offers scope for many other interesting perspectives. For instance, dispersal evolution has consistently been viewed for most of this thesis as the response variable to changes in the nutritional environment. But it is an equally valid line of inquiry to ask whether the fact of dispersal evolution has identifiable consequences for the organism's ability to respond to changing nutritional environments. To put it differently, do evolved dispersers respond to a change in nutrition differently from unselected controls? This is a perspective that the lab is currently well-poised to address, and experiments are ongoing that are studying physiological and life history changes in all four populations-VB, VBC, MD, and MC-under a range of different dietary conditions. Between them, studies involving these four populations have enormous potential for the investigation of dispersal evolution, nutrition, and life history.

## Ecology and evolution in cancer

The applications of ecological and evolutionary principles to understand cancer-associated phenomena are diverse and as has been demonstrated across the three studies in this thesis, can span across scales of biological organization.

Cancer growth parallels several aspects of ecological communities, in that they are heterogeneous assemblages of cells that are linked by interactions that range from mutualistic to strictly competitive (Kareva, 2011, Kotler and Brown, 2020). In the context of castration-resistant prostate cancer (CRPC), these interactions are explicitly driven by the production and consumption of the androgen, testosterone, which acts as a growth-promoting signal. Findings reported in this thesis demonstrate that while it is not a classical metabolic substrate that is converted to biomass, the effects of testosterone in mediating competition between cell types in prostate cancer can be understood using a resource-based theoretical framework. Broadly speaking, the current model is set in the context of a substantial body of work by Gatenby and colleagues over the past few years, using largely phenomenological models. Our goal here was to add processes relating to the production and consumption of resources, and investigate whether this addition produces any new insights about the prostate cancer system. In doing this, we needed to deal with a large number of moving parts and the current structure of the model represents a version that could be parameterized using the available empirical evidence. The novelty of our effort lies in its highlighting of tumor-internal processes that could be very relevant to better understanding and controlling the clinical progression of the disease. Now that the baseline behavior of this model has been established, more realistic model features can and must be incorporated in future iterations of this work.

The importance of considering resource dynamics as part of the clinical progression of prostate cancer is only just beginning to be recognized in literature (Kareva et al., 2015, Kareva and Brown, 2021) and our work indicates a possible way in which the inclusion of resource dynamics in an otherwise phenomenological model framework allows us to glean more mechanistic insights regarding the growth and progression of prostate cancer to a fully castration-resistant state. In this context, it is noteworthy that while serum markers like prostate serum antigen (PSA) have so far been used as indicators of total tumor volume (Swanson et al., 2001, Vollmer, 2010, Brady-Nicholls et al., 2020), there is a clear need for measurable indicators that reflect not just the volume of the tumor, but also its functional state in terms of the underlying processes driving tumor growth. Our work suggests that resources produced and consumed within the tumor could potentially fill this gap in the clinical monitoring of tumors by allowing for some estimation of the strength of within-tumor competition. In CRPC, such estimates of within-tumor competition acquire further

importance as they are relevant not only for the prediction of tumor growth trajectories, but also of the susceptibility of the tumor to particular kinds of therapy. It has been noted earlier that steady state composition in the model reflects testosterone dependency as well as potential resistance to testosterone-targeting treatments like abiraterone. Monitoring the state of tumor ecology is thus key for predicting both the progression of and expected therapeutic efficacy for CRPC. Furthermore, a more advanced understanding of resource dynamics in tumor ecology could also lead to the development of resource-based intervention strategies. Interventions for testosterone supply have been made in rodent models using subcutaneous implants of testosterone pellets (Umekita et al., 1996), and several modalities are already in practice for the clinical administration of testosterone in humans, particularly in the context of gender-affirmative care (Hembree et al., 2017, Coleman et al., 2022). Oxygen supply could be modified using a variety of methods targeted at increasing the recruitment of blood vessels to the tumor (Lovett et al., 2009), although this could admittedly be a somewhat coarse and messy prospect, as it is unlikely to affect the levels of oxygen alone and could allow reactive components from the tumor microenvironment to leak into systemic circulation. Attempts have also been reported to modify the pH of growing tumors using oral administration of $NaHCO_3$ in mice to control ecological tumor growth dynamics (Ibrahim-Hashim et al., 2017). Altogether, this indicates that rigorous ecological modeling could pave the way for a wide range of resource-based clinical interventions in cancer.

Finally, preliminary data from our model not included here also suggest that resources could be used in a more active manner to steer tumor growth dynamics towards a drug-sensitive state by preferentially augmenting the growth of the drug-sensitive cells. This is seen to be particularly effective at delaying the extinction of drug-sensitive cells under fluctuating treatment regimes in which periods of drug administration is interspersed with drug holidays to allow for the partial recovery of drug-sensitive cells. Fluctuating dosage regimes like adaptive therapy seek to use the biotic competition between drug-sensitive and -resistant cells to prevent the drug-resistant cells from dominating the tumor population, and in this context, external supply of resources used preferentially by the drug-sensitive cells could further support their recovery, thus helping in delaying the development of total drug resistance in the tumor. These initial findings would naturally require further exploration to develop them into realistic, clinically-applicable therapeutic regimes.

The CRPC model is entirely ecological, in that cell types did not undergo mutation to acquire further changes in their characteristics, and subsequent studies explored aspects of within-tumor evolutionary dynamics through a series of generalized tissue non-specific models. The first of these models is rooted in a human context, in which we model three possible modes of mutation accumulation within a general human tissue compartment. The model generates predictions of trends in population-level cancer incidence in humans based on the

dynamics of within-tissue somatic evolution and enables us to identify context-dependent selection on somatic mutations as a key feature of cancer progression. As has been pointed out earlier, this context dependence can stem from a wide range of individual-specific factors that all impinge on the tissue micro-environment. The findings from the model suggest that several key trends in the population-level incidence data could be explained by accounting for such inter-individual factors that could determine the susceptibility of an individual to cancer occurrence and progression. The importance of micro-environmental factors in determining the course of cancer progression has been widely recognized (Maley et al., 2017) and theories are now under investigation that suggest that temporal variations in the tissue micro-environment could be an important missing link in the current understanding of age-related trends in cancer risk (DeGregori, 2018, Rozhok and DeGregori, 2019). However, the use of epidemiological data in generating or validating hypotheses regarding the underlying mechanisms of cancer progression has rather diminished in recent years despite its rich pedigree (Armitage and Doll, 1954, Knudson, 1971), with only a few notable exceptions (Frank, 2007, Meza et al., 2008, Harding et al., 2012, Rozhok and DeGregori, 2019). The approach nevertheless remains promising, particularly as collections of epidemiological data become wider in coverage and deeper in scope (Sung et al., 2021, Ferlay et al., 2021, Cronin et al., 2022), alongside significant strides in pan-cancer multi-omics characterization (Thorsson et al., 2018, Bailey et al., 2018, Joshi et al., 2023).

In the final chapter, we consider evolutionary dynamics that could be relevant to cancer in the context of mammalian development in general. Using an individual-based model based on a general ontogenetic trajectory, we examine a range of possibilities in terms of the relationship between cell division, cell differentiation and mutation, and how they determine cell "fitness" in the context of organismal development. On the whole, while the model is agnostic of the specific tissue or species under investigation, it allows us to identify certain regions of the division-differentiation-mutation parameter space that are particularly prone to the risk of mutation accumulation and cancer. In particular, the trends reported from this model regarding the accumulation of neutral mutations are relatively new to the field and are likely to prove useful in future investigation of somatic evolution. We also see that combinations of parameter values in the model are able to recapitulate a wide range of scaling phenomena with overall cell number, allowing us to speculate on the mechanistic bases of different cancer risk allometries across levels of biological organization, from organs to species. The lack of empirical data is a considerable gap in this area of research, but the gap has been narrowed by several recent studies that have addressed a range of scaling phenomena across mammalian species (Martincorena, 2017, Sender and Milo, 2021, Cagan et al., 2022). The data from such studies, while still nascent, are a valuable starting point for further exploration of cancer risk allometry across biological scales. In particular, these studies have gathered data for a multiplicity of traits like lifespan, genome size, cell number,

cell size, nucleotide substitution rates and cumulative mutation accumulation rates within the same set of mammalian species, representing significant scope for the study of basic questions of macro-evolution including the evolution of body size, lifespan and potential cancer defense mechanisms (Tollis et al., 2017, Albuquerque et al., 2018).

# Part II

# Supplementary Material

# S1 | Modified formulation used for larval malnutrition in fruit flies

We use a modified banana-jaggery medium with the amount of dry yeast powder reduced to one-third of that of the standard food. The banana-jaggery medium is a complex form of fly feed, of which most constituent parts contribute multiple kinds of complex sugars, lipids and other micronutrients (Nutrition, 1977). Unlike other simplified media (Lee and Micchelli, 2013), it is not possible to isolate a single source of carbohydrates or protein whose levels can be changed in a factorial design to produce multiple kinds of malnutrition regimes. Yeast powder is by far the simplest and most standardized component of the food as it is obtained from the same source without the kind of batch-to-batch variation that is inevitable in natural components like bananas or jaggery. Since yeast is the primary source of protein in the banana-jaggery recipe, its reduction corresponds to reduction in the amount of total available protein by about 20%, with some auxiliary reduction in carbohydrates. Standardisation experiments showed a clear increase in time to eclosion as well as reduction in average dry body weight in both sexes (unpublished data) in unselected flies subjected to this malnutrition regime of one-third yeast powder, suggesting that it is strong enough to impose some amount of selective pressure over several generations.

| **Component** | **Standard nutrition** | **Poor nutrition** |
|---|---|---|
| Banana | 205g | 205g |
| Barley | 25g | 25g |
| Jaggery | 35g | 35g |
| Dry yeast powder | 36g | 12g |
| Agar | 12.4g | 20g |
| Ethanol (99.9%) | 22mL | 22mL |
| Water | 1.12L | 1.12L |
| Methyl paraben | 2.4g dissolved in 23mL of 99.9% ethanol | 2.4g dissolved in 23mL of 99.9% ethanol |

***Supplementary Table S1.1:*** **Recipes of banana-jaggery medium used for stock rearing and experiments**. The detailed procedure for food preparation is available at request.

# S2 | Statistical analyses for Chapter 2

## S2.1 Generation 42

### S2.1.1 Propensity

```
> options(width = 60)
> prop.glm <- glmer(cbind(MeanSuccess, MeanFail) ~ Popn + Sex + Popn:Sex +
+     (1 | Block), data = df_prop, family = "binomial")
> Anova(prop.glm, type = "III")

## Analysis of Deviance Table (Type III Wald chisquare tests)
##
## Response: cbind(MeanSuccess, MeanFail)
##                Chisq Df Pr(>Chisq)
## (Intercept)   0.2399  1     0.6242
## Popn        847.4166  1  < 2.2e-16 ***
## Sex          54.2285  1  1.785e-13 ***
## Popn:Sex    104.6150  1  < 2.2e-16 ***
## ---
## Signif. codes:
## 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
> options(width = 60)
> kable(contrast(emmeans(prop.glm, ~Popn:Sex), type = "response",
+     adjust = "tukey", method = "pairwise"), digits = 4, booktabs = T)
```

| contrast | odds.ratio | SE | df | null | z.ratio | p.value |
|---|---|---|---|---|---|---|
| MC F / MD F | 0.0937 | 0.0076 | Inf | 1 | -29.1104 | 0 |
| MC F / MC M | 0.6686 | 0.0366 | Inf | 1 | -7.3640 | 0 |
| MC F / MD M | 0.1844 | 0.0124 | Inf | 1 | -25.0991 | 0 |
| MD F / MC M | 7.1328 | 0.5816 | Inf | 1 | 24.0934 | 0 |
| MD F / MD M | 1.9673 | 0.1776 | Inf | 1 | 7.4979 | 0 |
| MC M / MD M | 0.2758 | 0.0187 | Inf | 1 | -19.0447 | 0 |

```
> options(width = 60)
> summary(cohen.d(MeanPropensity ~ Popn:Sex, data = data.frame(df_prop)))

## Extract effect sizes by groups from cohen.d.by
```

```
##             MeanPropensity  Md
## PopnforSexF            9.4 9.4
## PopnforSexM            2.8 2.8
```

### S2.1.2 Ability

```
> options(width = 60)
> ability.glm <- lmer(MeanAbility ~ Popn + Sex + Popn:Sex + (1 |
+     Block), data = df_prop)
> Anova(ability.glm, type = "III")

## Analysis of Deviance Table (Type III Wald chisquare tests)
##
## Response: MeanAbility
##               Chisq Df Pr(>Chisq)
## (Intercept) 95.7699  1  < 2.2e-16 ***
## Popn        54.0859  1  1.919e-13 ***
## Sex          0.6568  1     0.4177
## Popn:Sex     0.9679  1     0.3252
## ---
## Signif. codes:
## 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
> options(width = 60)
> kable(contrast(emmeans(ability.glm, ~Popn:Sex), adjust = "tukey",
+     method = "pairwise"), digits = 4, booktabs = T)
```

| contrast | estimate | SE | df | t.ratio | p.value |
|---|---|---|---|---|---|
| MC F - MD F | -3.6001 | 0.4895 | 9 | -7.3543 | 0.0002 |
| MC F - MC M | -0.3967 | 0.4895 | 9 | -0.8104 | 0.8481 |
| MC F - MD M | -3.3158 | 0.4895 | 9 | -6.7734 | 0.0004 |
| MD F - MC M | 3.2034 | 0.4895 | 9 | 6.5439 | 0.0005 |
| MD F - MD M | 0.2843 | 0.4895 | 9 | 0.5809 | 0.9354 |
| MC M - MD M | -2.9190 | 0.4895 | 9 | -5.9630 | 0.0010 |

```
> options(width = 60)
> summary(cohen.d(MeanAbility ~ Popn:Sex, data = data.frame(df_prop)))

## Extract effect sizes by groups from cohen.d.by
##             MeanAbility  Md
## PopnforSexF         4.2 4.2
## PopnforSexM         3.1 3.1
```

### S2.1.3 Higher moments of the kernel and 90th percentile

```
> options(width = 60)
> # Males
> wilcox.test(mc.males.SD, md.males.SD, alternative = "two.sided")

##
##  Wilcoxon rank sum exact test
##
## data:  mc.males.SD and md.males.SD
## W = 18, p-value = 0.1605
## alternative hypothesis: true location shift is not equal to 0

> # Females
> wilcox.test(mc.females.SD, md.females.SD, alternative = "two.sided")

##
##  Wilcoxon rank sum exact test
##
## data:  mc.females.SD and md.females.SD
## W = 16, p-value = 0.1049
## alternative hypothesis: true location shift is not equal to 0
```

```
> options(width = 60)
> # Males
> wilcox.test(mc.males.skew, md.males.skew, alternative = "two.sided")

##
##  Wilcoxon rank sum exact test
##
```

```
## data:  mc.males.skew and md.males.skew
## W = 64, p-value = 0.0001554
## alternative hypothesis: true location shift is not equal to 0

> # Females
> wilcox.test(mc.females.skew, md.females.skew, alternative = "two.sided")

##
##  Wilcoxon rank sum exact test
##
## data:  mc.females.skew and md.females.skew
## W = 64, p-value = 0.0001554
## alternative hypothesis: true location shift is not equal to 0
```

```
> options(width = 60)
> # Males
> wilcox.test(mc.males.kurt, md.males.kurt, alternative = "two.sided")

##
##  Wilcoxon rank sum exact test
##
## data:  mc.males.kurt and md.males.kurt
## W = 64, p-value = 0.0001554
## alternative hypothesis: true location shift is not equal to 0

> # Females
> wilcox.test(mc.females.kurt, md.females.kurt, alternative = "two.sided")

##
##  Wilcoxon rank sum exact test
##
## data:  mc.females.kurt and md.females.kurt
## W = 62, p-value = 0.0006216
## alternative hypothesis: true location shift is not equal to 0
```

```
> options(width = 60)
> # Males
> wilcox.test(mc.males.90, md.males.90, alternative = "two.sided")
```

```
## Warning in wilcox.test.default(mc.males.90, md.males.90, alternative = "two.sided"
cannot compute exact p-value with ties

##
##  Wilcoxon rank sum test with continuity correction
##
## data:  mc.males.90 and md.males.90
## W = 16.5, p-value = 0.1136
## alternative hypothesis: true location shift is not equal to 0

> # Females
> wilcox.test(mc.females.90, md.females.90, alternative = "two.sided")

## Warning in wilcox.test.default(mc.females.90, md.females.90, alternative =
"two.sided"):  cannot compute exact p-value with ties

##
##  Wilcoxon rank sum test with continuity correction
##
## data:  mc.females.90 and md.females.90
## W = 16, p-value = 0.1025
## alternative hypothesis: true location shift is not equal to 0
```

### S2.1.4 Goodness of fit indicators for Gamma function fits

| Popn | Sex | Block | D | KS.pval | ChiSq | ChiSq.pval |
|---|---|---|---|---|---|---|
| MC | F | 1 | 0.2667 | 0.2365 | 0.1678 | 1 |
| MC | F | 2 | 0.2333 | 0.3877 | 0.0942 | 1 |
| MC | F | 3 | 0.1667 | 0.8080 | 0.1558 | 1 |
| MC | F | 4 | 0.1667 | 0.8080 | 0.1502 | 1 |
| MC | M | 1 | 0.1667 | 0.7990 | 0.0533 | 1 |
| MC | M | 2 | 0.2000 | 0.5941 | 0.0892 | 1 |
| MC | M | 3 | 0.1667 | 0.8080 | 0.0951 | 1 |
| MC | M | 4 | 0.2667 | 0.2391 | 0.1178 | 1 |
| MD | F | 1 | 0.3333 | 0.0709 | 0.2877 | 1 |
| MD | F | 2 | 0.3333 | 0.0709 | 0.4481 | 1 |
| MD | F | 3 | 0.4000 | 0.0156 | 0.4007 | 1 |
| MD | F | 4 | 0.3333 | 0.0709 | 0.2066 | 1 |
| MD | M | 1 | 0.4333 | 0.0065 | 0.1951 | 1 |
| MD | M | 2 | 0.4667 | 0.0025 | 0.1926 | 1 |
| MD | M | 3 | 0.3333 | 0.0709 | 0.2229 | 1 |
| MD | M | 4 | 0.3333 | 0.0709 | 0.2190 | 1 |

### S2.1.5 Kolmogorov-Smirnov tests comparing male and female kernels

| Popn | Block | D | p |
|---|---|---|---|
| MC | 1 | 0.1333 | 0.9525 |
| MC | 2 | 0.1333 | 0.9525 |
| MC | 3 | 0.1333 | 0.9578 |
| MC | 4 | 0.1667 | 0.8080 |
| MD | 1 | 0.1667 | 0.8080 |
| MD | 2 | 0.1667 | 0.8080 |
| MD | 3 | 0.1333 | 0.9578 |
| MD | 4 | 0.1333 | 0.9578 |

### S2.1.6 All fit parameters

| Popn | Block | Sex | k | CorrectedTheta | AIC |
|---|---|---|---|---|---|
| MD | 1 | M | 1.1620 | 20.3147 | -160.4278 |
| MD | 1 | F | 1.5241 | 14.0695 | -150.1206 |
| MD | 2 | M | 0.8922 | 33.9501 | -160.0143 |
| MD | 2 | F | 1.8078 | 9.9247 | -152.6892 |
| MD | 3 | M | 1.3013 | 18.5263 | -154.9708 |
| MD | 3 | F | 1.3281 | 17.9146 | -137.3899 |
| MD | 4 | M | 1.2200 | 22.3985 | -155.2544 |
| MD | 4 | F | 1.2488 | 22.2410 | -157.1292 |
| MC | 1 | M | 1.2498 | 6.4620 | -195.2162 |
| MC | 1 | F | 1.2821 | 5.9458 | -158.0278 |
| MC | 2 | M | 1.0050 | 11.8605 | -190.9815 |
| MC | 2 | F | 0.6346 | 25.2528 | -191.3530 |
| MC | 3 | M | 1.0510 | 16.5023 | -184.0834 |
| MC | 3 | F | 1.2812 | 12.0885 | -170.3489 |
| MC | 4 | M | 0.7791 | 39.5068 | -177.9508 |
| MC | 4 | F | 0.8155 | 26.9097 | -171.7862 |

### S2.1.7 LDD-comparing scaled and unscaled means

```
> # Unscaled means (males)
> wilcox.test(mc.unscaled.males, md.unscaled.males, alternative = "two.sided")

##
##  Wilcoxon rank sum exact test
##
## data:  mc.unscaled.males and md.unscaled.males
## W = 3, p-value = 0.2
## alternative hypothesis: true location shift is not equal to 0

> # Scaled means (males)
> wilcox.test(mc.scaled.males, md.scaled.males, alternative = "two.sided")
```

```
##
##  Wilcoxon rank sum exact test
##
## data:  mc.scaled.males and md.scaled.males
## W = 16, p-value = 0.02857
## alternative hypothesis: true location shift is not equal to 0

> # Unscaled means (females)
> wilcox.test(mc.unscaled.females, md.unscaled.females,
+     alternative = "two.sided")

##
##  Wilcoxon rank sum exact test
##
## data:  mc.unscaled.females and md.unscaled.females
## W = 0, p-value = 0.02857
## alternative hypothesis: true location shift is not equal to 0

> # Scaled means (females)
> wilcox.test(mc.scaled.females, md.scaled.females, alternative = "two.sided")

##
##  Wilcoxon rank sum exact test
##
## data:  mc.scaled.females and md.scaled.females
## W = 16, p-value = 0.02857
## alternative hypothesis: true location shift is not equal to 0
```

## S2.2 Generation 66-67

### S2.2.1 Propensity

```
> options(width = 60)
> prop.glm <- glmer(cbind(MeanSuccess, MeanFail) ~ Popn +
+     Sex + Popn:Sex + (1 | Block), data = df_prop, family = "binomial")
> Anova(prop.glm, type = "III")

## Analysis of Deviance Table (Type III Wald chisquare tests)
##
```

```
## Response: cbind(MeanSuccess, MeanFail)
##                Chisq Df Pr(>Chisq)
## (Intercept)   0.1228  1  0.7260271
## Popn        861.4940  1  < 2.2e-16 ***
## Sex          11.2639  1  0.0007903 ***
## Popn:Sex     46.1537  1  1.093e-11 ***
## ---
## Signif. codes:
## 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
> options(width = 60)
> kable(contrast(emmeans(prop.glm, ~Popn:Sex), type = "response",
+     adjust = "tukey", method = "pairwise"), digits = 4, booktabs = T)
```

| contrast | odds.ratio | SE | df | null | z.ratio | p.value |
|---|---|---|---|---|---|---|
| MC F / MD F | 0.1632 | 0.0101 | Inf | 1 | -29.3512 | 0.0000 |
| MC F / MC M | 0.8440 | 0.0427 | Inf | 1 | -3.3562 | 0.0044 |
| MC F / MD M | 0.2428 | 0.0138 | Inf | 1 | -24.9337 | 0.0000 |
| MD F / MC M | 5.1726 | 0.3193 | Inf | 1 | 26.6199 | 0.0000 |
| MD F / MD M | 1.4881 | 0.0988 | Inf | 1 | 5.9834 | 0.0000 |
| MC M / MD M | 0.2877 | 0.0163 | Inf | 1 | -21.9579 | 0.0000 |

```
> options(width = 60)
> summary(cohen.d(MeanPropensity ~ Popn:Sex, data = data.frame(df_prop)))

## Extract effect sizes by groups from cohen.d.by
##             MeanPropensity  Md
## PopnforSexF            5.0 5.0
## PopnforSexM            6.8 6.8
```

### S2.2.2 Ability

```
> options(width = 60)
> ability.glm <- lmer(MeanAbility ~ Popn + Sex + Popn:Sex + (1 |
+     Block), data = df_prop)
> Anova(ability.glm, type = "III")
```

```
## Analysis of Deviance Table (Type III Wald chisquare tests)
##
## Response: MeanAbility
##                Chisq Df Pr(>Chisq)
## (Intercept) 141.5072  1    < 2e-16 ***
## Popn        178.6676  1    < 2e-16 ***
## Sex           3.7222  1    0.05369 .
## Popn:Sex      2.4147  1    0.12020
## ---
## Signif. codes:
## 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
> options(width = 60)
> kable(contrast(emmeans(ability.glm, ~Popn:Sex), adjust = "tukey",
+     method = "pairwise"), digits = 4, booktabs = T)
```

| contrast | estimate | SE | df | t.ratio | p.value |
|---|---|---|---|---|---|
| MC F - MD F | -15.2617 | 1.1418 | 9 | -13.3667 | 0.0000 |
| MC F - MC M | -2.2028 | 1.1418 | 9 | -1.9293 | 0.2820 |
| MC F - MD M | -14.9554 | 1.1418 | 9 | -13.0984 | 0.0000 |
| MD F - MC M | 13.0589 | 1.1418 | 9 | 11.4374 | 0.0000 |
| MD F - MD M | 0.3063 | 1.1418 | 9 | 0.2683 | 0.9928 |
| MC M - MD M | -12.7526 | 1.1418 | 9 | -11.1691 | 0.0000 |

```
> options(width = 60)
> summary(cohen.d(MeanAbility ~ Popn:Sex, data = data.frame(df_prop)))

## Extract effect sizes by groups from cohen.d.by
##             MeanAbility  Md
## PopnforSexF         8.9 8.9
## PopnforSexM         7.2 7.2
```

### S2.2.3 Higher moments of the kernel and 90th percentile

```
> options(width = 60)
> # Males
> wilcox.test(mc.males.SD, md.males.SD, alternative = "two.sided")

##
##  Wilcoxon rank sum exact test
##
## data:  mc.males.SD and md.males.SD
## W = 18, p-value = 0.1605
## alternative hypothesis: true location shift is not equal to 0

> # Females
> wilcox.test(mc.females.SD, md.females.SD, alternative = "two.sided")

##
##  Wilcoxon rank sum exact test
##
## data:  mc.females.SD and md.females.SD
## W = 16, p-value = 0.1049
## alternative hypothesis: true location shift is not equal to 0
```

```
> options(width = 60)
> # Males
> wilcox.test(mc.males.skew, md.males.skew, alternative = "two.sided")

##
##  Wilcoxon rank sum exact test
##
## data:  mc.males.skew and md.males.skew
## W = 64, p-value = 0.0001554
## alternative hypothesis: true location shift is not equal to 0

> # Females
> wilcox.test(mc.females.skew, md.females.skew, alternative = "two.sided")

##
##  Wilcoxon rank sum exact test
##
## data:  mc.females.skew and md.females.skew
## W = 64, p-value = 0.0001554
## alternative hypothesis: true location shift is not equal to 0
```

```
> options(width = 60)
> # Males
> wilcox.test(mc.males.kurt, md.males.kurt, alternative = "two.sided")

##
##  Wilcoxon rank sum exact test
##
## data:  mc.males.kurt and md.males.kurt
## W = 64, p-value = 0.0001554
## alternative hypothesis: true location shift is not equal to 0

> # Females
> wilcox.test(mc.females.kurt, md.females.kurt, alternative = "two.sided")

##
##  Wilcoxon rank sum exact test
##
## data:  mc.females.kurt and md.females.kurt
## W = 62, p-value = 0.0006216
## alternative hypothesis: true location shift is not equal to 0
```

```
> options(width = 60)
> # Males
> wilcox.test(mc.males.90, md.males.90, alternative = "two.sided")

## Warning in wilcox.test.default(mc.males.90, md.males.90, alternative = "two.sided"):
cannot compute exact p-value with ties

##
##  Wilcoxon rank sum test with continuity correction
##
## data:  mc.males.90 and md.males.90
## W = 16.5, p-value = 0.1136
## alternative hypothesis: true location shift is not equal to 0

> # Females
> wilcox.test(mc.females.90, md.females.90, alternative = "two.sided")

## Warning in wilcox.test.default(mc.females.90, md.females.90, alternative =
"two.sided"):  cannot compute exact p-value with ties
```

```
##
##  Wilcoxon rank sum test with continuity correction
##
## data:  mc.females.90 and md.females.90
## W = 16, p-value = 0.1025
## alternative hypothesis: true location shift is not equal to 0
```

### S2.2.4 Goodness of fit indicators for Gamma function fits

| Popn | Sex | Block | D | KS.pval | ChiSq | ChiSq.pval |
|---|---|---|---|---|---|---|
| MC | F | 1 | 0.4000 | 0.0165 | 0.0609 | 1 |
| MC | F | 2 | 0.3667 | 0.0354 | 0.1182 | 1 |
| MC | F | 3 | 0.3333 | 0.0713 | 0.0712 | 1 |
| MC | F | 4 | 0.3000 | 0.1344 | 0.0910 | 1 |
| MC | M | 1 | 0.3333 | 0.0713 | 0.1416 | 1 |
| MC | M | 2 | 0.3333 | 0.0713 | 0.0676 | 1 |
| MC | M | 3 | 0.2667 | 0.2365 | 0.0884 | 1 |
| MC | M | 4 | 0.2667 | 0.2365 | 0.0958 | 1 |
| MD | F | 1 | 0.1667 | 0.8080 | 0.0734 | 1 |
| MD | F | 2 | 0.1000 | 0.9988 | 0.4781 | 1 |
| MD | F | 3 | 0.1333 | 0.9578 | 0.2854 | 1 |
| MD | F | 4 | 0.1667 | 0.7990 | 0.6156 | 1 |
| MD | M | 1 | 0.2000 | 0.5941 | 0.0626 | 1 |
| MD | M | 2 | 0.2333 | 0.3929 | 0.0627 | 1 |
| MD | M | 3 | 0.1333 | 0.9578 | 0.0652 | 1 |
| MD | M | 4 | 0.3000 | 0.1350 | 0.1300 | 1 |

### S2.2.5 Kolmogorov-Smirnov tests comparing male and female kernels

| Popn | Block | D | p |
|---|---|---|---|
| MC | 1 | 0.1333 | 0.9525 |
| MC | 2 | 0.1333 | 0.9525 |
| MC | 3 | 0.1333 | 0.9525 |
| MC | 4 | 0.2000 | 0.5860 |
| MD | 1 | 0.1333 | 0.9578 |
| MD | 2 | 0.3000 | 0.1350 |
| MD | 3 | 0.1333 | 0.9578 |
| MD | 4 | 0.1667 | 0.7990 |

### S2.2.6 All fit parameters

| Popn | Block | Sex | k | CorrectedTheta | AIC |
|---|---|---|---|---|---|
| MD | 1 | M | 1.4357 | 15.9615 | -196.7267 |
| MD | 1 | F | 1.6427 | 14.0495 | -195.2812 |
| MD | 2 | M | 1.8762 | 15.1271 | -207.2036 |
| MD | 2 | F | 3.2613 | 7.5930 | -194.5058 |
| MD | 3 | M | 1.6430 | 12.4556 | -200.0810 |
| MD | 3 | F | 2.6627 | 9.3660 | -186.6777 |
| MD | 4 | M | 1.3602 | 18.5576 | -174.4135 |
| MD | 4 | F | 2.5281 | 12.0092 | -171.6101 |
| MC | 1 | M | 1.6075 | 6.2371 | -164.7455 |
| MC | 1 | F | 1.1496 | 4.8748 | -208.2715 |
| MC | 2 | M | 1.2409 | 6.9074 | -205.8391 |
| MC | 2 | F | 1.5171 | 6.3052 | -185.2442 |
| MC | 3 | M | 1.2446 | 7.7327 | -192.4412 |
| MC | 3 | F | 1.0997 | 7.3442 | -210.3603 |
| MC | 4 | M | 1.2877 | 9.0952 | -193.0083 |
| MC | 4 | F | 1.1031 | 8.2906 | -190.3597 |

### S2.2.7 LDD-comparing scaled and unscaled means

```
> # Unscaled means (males)
> wilcox.test(mc.unscaled.males, md.unscaled.males, alternative = "two.sided")

##
##  Wilcoxon rank sum exact test
##
## data:  mc.unscaled.males and md.unscaled.males
## W = 0, p-value = 0.02857
## alternative hypothesis: true location shift is not equal to 0

> # Scaled means (males)
> wilcox.test(mc.scaled.males, md.scaled.males, alternative = "two.sided")

##
##  Wilcoxon rank sum exact test
##
## data:  mc.scaled.males and md.scaled.males
## W = 16, p-value = 0.02857
## alternative hypothesis: true location shift is not equal to 0

> # Unscaled means (females)
> wilcox.test(mc.unscaled.females, md.unscaled.females,
+     alternative = "two.sided")

##
##  Wilcoxon rank sum exact test
##
## data:  mc.unscaled.females and md.unscaled.females
## W = 0, p-value = 0.02857
## alternative hypothesis: true location shift is not equal to 0

> # Scaled means (females)
> wilcox.test(mc.scaled.females, md.scaled.females, alternative = "two.sided")

##
##  Wilcoxon rank sum exact test
##
## data:  mc.scaled.females and md.scaled.females
## W = 16, p-value = 0.02857
## alternative hypothesis: true location shift is not equal to 0
```

# S2.3 Generation 93

## S2.3.1 Propensity

```
> options(width = 60)
> prop.glm <- glmer(cbind(MeanDisperser, MeanSource) ~
+     Popn + Sex + Popn:Sex + (1 | Block), data = df.mean,
+     family = "binomial")
> Anova(prop.glm, type = "III")

## Analysis of Deviance Table (Type III Wald chisquare tests)
##
## Response: cbind(MeanDisperser, MeanSource)
##                Chisq Df Pr(>Chisq)
## (Intercept)   33.871  1  5.890e-09 ***
## Popn        1539.402  1  < 2.2e-16 ***
## Sex           19.202  1  1.176e-05 ***
## Popn:Sex      11.213  1  0.0008122 ***
## ---
## Signif. codes:
## 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
> options(width = 60)
> kable(contrast(emmeans(prop.glm, ~Popn:Sex), type = "response",
+     adjust = "tukey", method = "pairwise"), digits = 4, booktabs = T)
```

| contrast | odds.ratio | SE | df | null | z.ratio | p.value |
|---|---|---|---|---|---|---|
| MC F / MD F | 0.0632 | 0.0044 | Inf | 1 | -39.2352 | 0e+00 |
| MC F / MC M | 1.3198 | 0.0836 | Inf | 1 | 4.3820 | 1e-04 |
| MC F / MD M | 0.1149 | 0.0073 | Inf | 1 | -34.1404 | 0e+00 |
| MD F / MC M | 20.8886 | 1.5185 | Inf | 1 | 41.8077 | 0e+00 |
| MD F / MD M | 1.8190 | 0.1307 | Inf | 1 | 8.3241 | 0e+00 |
| MC M / MD M | 0.0871 | 0.0057 | Inf | 1 | -37.0297 | 0e+00 |

```
> options(width = 60)
> cohen.d(MeanPropensity ~ Popn:Sex, data = data.frame(df.mean))
```

```
## Call: NULL
##
##  Group levels =  PopnforSexF
## Cohen d statistic of difference between two means
##                lower effect upper
## MeanPropensity   2.6    6.4    10
##
## Multivariate (Mahalanobis) distance between groups
## [1] 6.4
## r equivalent of difference between two means
## MeanPropensity
##           0.95
##
##  Group levels =  PopnforSexM
## Cohen d statistic of difference between two means
##                lower effect upper
## MeanPropensity   4.2    9.8    15
##
## Multivariate (Mahalanobis) distance between groups
## [1] 9.8
## r equivalent of difference between two means
## MeanPropensity
##           0.98
##
## Use summary for more compact output
```

# S3 | Statistical analyses for Chapter 3

For each assay, details of statistical analyses are given in three parts:

1. The code used to fit the GLM is reproduced, followed by the analysis of deviance table from the corresponding model object. All inferences of statistical significance are based on the test statistics in this table.

2. Pairwise comparisons for each trait from the corresponding model object.

3. Cohen's $d$ for each dataset

In this study, data from the male-male aggression experiment were compared using a two-sample Wilcoxon's signed rank test, and therefore does not involve analyses of the type given above.

## S3.1 Physiological traits

### S3.1.1 Dry body weight

```
> options(width = 60)
> m <- lmer(Weight ~ Popn * Sex + (1 | Block), data = bw_final)
> Anova(m, type = "III")

## Analysis of Deviance Table (Type III Wald chisquare tests)
##
## Response: Weight
##                Chisq Df Pr(>Chisq)
## (Intercept)  685.345  1  < 2.2e-16 ***
## Popn          78.117  1  < 2.2e-16 ***
## Sex         1139.782  1  < 2.2e-16 ***
## Popn:Sex      30.112  1  4.079e-08 ***
## ---
## Signif. codes:
## 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
> options(width = 60)
> kable(pairs(emmeans(m, ~Popn:Sex), simple = "Popn", reverse = TRUE),
+     digits = 4, booktabs = T)
```

| contrast | Sex | estimate | SE | df | t.ratio | p.value |
|---|---|---|---|---|---|---|
| MD - MC | F | -0.0295 | 0.0033 | 153 | -8.8384 | 0.0000 |
| MD - MC | M | -0.0036 | 0.0033 | 153 | -1.0780 | 0.2827 |

```
> options(width = 60)
> summary(cohen.d(Weight ~ Popn:Sex, data = data.frame(bw_final)))

## Extract effect sizes by groups from cohen.d.by
##              Weight   Md
## PopnforSexF  -1.04 1.04
## PopnforSexM  -0.19 0.19
```

### S3.1.2 Fecundity

```
> options(width = 60)
> fm <- glmer(Fecundity ~ Popn + (1 | Block), family = poisson,
+     data = fecundity)
> Anova(fm, type = "III")

## Analysis of Deviance Table (Type III Wald chisquare tests)
##
## Response: Fecundity
##                Chisq Df Pr(>Chisq)
## (Intercept) 543.7968  1    < 2e-16 ***
## Popn          4.1821  1    0.04085 *
## ---
## Signif. codes:
## 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
> options(width = 60)
> kable(pairs(emmeans(m, ~Popn, simple = "Popn"), type = "response",
+     reverse = TRUE), digits = 4, booktabs = T)

## NOTE: Results may be misleading due to involvement in interactions
```

| contrast | estimate | SE | df | t.ratio | p.value |
|---|---|---|---|---|---|
| MD - MC | -0.0166 | 0.0024 | 153 | -7.0119 | 0 |

```
> options(width = 60)
> summary(cohen.d(Fecundity ~ Popn, data = data.frame(fecundity)))

## Extract effect sizes from cohen.d
## cohen.d(x = Fecundity ~ Popn, data = data.frame(fecundity))
##
## Multivariate (Mahalanobis) distance between groups 0.06
##  r equivalent for each variable
## Fecundity
##     -0.03
```

### S3.1.3 Time to eclosion

#### All four blocks

```
> options(width = 60)
> dvmean1 <- lmer(DevTime ~ Popn * Sex + (1 | Block), data = dev_average)
> Anova(dvmean1, type = "III")

## Analysis of Deviance Table (Type III Wald chisquare tests)
##
## Response: DevTime
##                 Chisq Df Pr(>Chisq)
## (Intercept) 8530.3797  1     <2e-16 ***
## Popn           0.6516  1     0.4196
## Sex            1.8152  1     0.1779
## Popn:Sex       1.5290  1     0.2163
## ---
## Signif. codes:
## 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
> options(width = 60)
> kable(pairs(emmeans(dvmean1, ~Popn:Sex), simple = "Popn", reverse = TRUE),
+     digits = 4, booktabs = T)
```

| contrast | Sex | estimate | SE | df | t.ratio | p.value |
|---|---|---|---|---|---|---|
| MD - MC | F | -0.9507 | 1.1778 | 153 | -0.8072 | 0.4208 |
| MD - MC | M | -3.0102 | 1.1778 | 153 | -2.5559 | 0.0116 |

```
> options(width = 60)
> summary(cohen.d(DevTime ~ Popn:Sex, data = data.frame(dev_average)))

## Extract effect sizes by groups from cohen.d.by
##             DevTime   Md
## PopnforSexF   -0.15 0.15
## PopnforSexM   -0.41 0.41
```

### Excluding block 4 data

Visual inspection of the data indicated that block 4 could be an outlier. The analyses were therefore repeated, now excluding data from block 4.

```
> options(width = 60)
> moddev1 <- lmer(DevTime ~ Popn * Sex + (1 | Block), data = dev_average.mod)
> Anova(moddev1, type = "III")

## Analysis of Deviance Table (Type III Wald chisquare tests)
##
## Response: DevTime
##                  Chisq Df Pr(>Chisq)
## (Intercept) 30617.0150  1  < 2.2e-16 ***
## Popn           19.6544  1  9.279e-06 ***
## Sex             5.5091  1   0.018917 *
## Popn:Sex        8.8958  1   0.002858 **
## ---
## Signif. codes:
## 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
> options(width = 60)
> kable(pairs(emmeans(moddev1, ~Popn:Sex), simple = "Popn", reverse = TRUE),
+     digits = 4, booktabs = T)
```

| contrast | Sex | estimate | SE | df | t.ratio | p.value |
|---|---|---|---|---|---|---|
| MD - MC | F | -3.3442 | 0.7563 | 133.9613 | -4.4217 | 0 |
| MD - MC | M | -6.4243 | 0.7563 | 133.9613 | -8.4942 | 0 |

```
> options(width = 60)
> summary(cohen.d(DevTime ~ Popn:Sex, data = data.frame(dev_average.mod)))

## Extract effect sizes by groups from cohen.d.by
##             DevTime  Md
## PopnforSexF    -1.3 1.3
## PopnforSexM    -1.9 1.9
```

### S3.1.4 Time to death by desiccation

```
> options(width = 60)
> dmean1 <- lmer(DesTime ~ Popn * Sex + (1 | Block), data = desi_average)
> Anova(dmean1, type = "III")

## Analysis of Deviance Table (Type III Wald chisquare tests)
##
## Response: DesTime
##                Chisq Df Pr(>Chisq)
## (Intercept) 727.3205  1    < 2e-16 ***
## Popn          0.8578  1    0.35436
## Sex         109.9969  1    < 2e-16 ***
## Popn:Sex      5.4107  1    0.02001 *
## ---
## Signif. codes:
## 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
> options(width = 60)
> kable(pairs(emmeans(dmean1, ~Popn:Sex), simple = "Popn", reverse = TRUE),
+     digits = 4, booktabs = T)
```

| contrast | Sex | estimate | SE | df | t.ratio | p.value |
|---|---|---|---|---|---|---|
| MD - MC | F | 0.5150 | 0.5561 | 152.0000 | 0.9262 | 0.3558 |
| MD - MC | M | -1.3202 | 0.5597 | 152.0056 | -2.3588 | 0.0196 |

```
> options(width = 60)
> summary(cohen.d(DesTime ~ Popn:Sex, data = data.frame(desi_average)))
```

```
## Extract effect sizes by groups from cohen.d.by
##             DesTime   Md
## PopnforSexF    0.16 0.16
## PopnforSexM   -0.61 0.61
```

# S3.2 Behavioral traits

## S3.2.1 Locomotor activity

```
> options(width = 60)
> act_lm <- lmer(Activity ~ Popn + (1 | Block), data = dfdam)
> Anova(act_lm, type = "III")

## Analysis of Deviance Table (Type III Wald chisquare tests)
##
## Response: Activity
##              Chisq Df Pr(>Chisq)
## (Intercept) 103.07  1  < 2.2e-16 ***
## Popn        109.88  1  < 2.2e-16 ***
## ---
## Signif. codes:
## 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
> options(width = 60)
> kable(pairs(emmeans(act_lm, ~Popn), simple = "Popn", reverse = TRUE),
+     digits = 4, booktabs = T)
```

| contrast | estimate | SE | df | t.ratio | p.value |
|---|---|---|---|---|---|
| MD - MC | 71.1406 | 6.7867 | 251 | 10.4824 | 0 |

```
> summary(cohen.d(Activity ~ Popn, data = data.frame(dfdam)))

## Extract effect sizes from cohen.d
## cohen.d(x = Activity ~ Popn, data = data.frame(dfdam))
##
## Multivariate (Mahalanobis) distance between groups 1.27
##  r equivalent for each variable
## Activity
##     0.54
```

### S3.2.2 Rest duration

```
> options(width = 60)
> rest_glm <- glmer(cbind(Success, Failure) ~ Popn + (1 | Block),
+     family = binomial, data = dfdam)
> Anova(rest_glm, type = "III")

## Analysis of Deviance Table (Type III Wald chisquare tests)
##
## Response: cbind(Success, Failure)
##               Chisq Df Pr(>Chisq)
## (Intercept)   10.71  1   0.001066 **
## Popn        6352.72  1  < 2.2e-16 ***
## ---
## Signif. codes:
## 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
> options(width = 60)
> kable(pairs(emmeans(rest_glm, ~Popn, type = "response"), simple = "Popn",
+     reverse = TRUE), digits = 4, booktabs = T)
```

| contrast | odds.ratio | SE | df | null | z.ratio | p.value |
|---|---|---|---|---|---|---|
| MD / MC | 0.1312 | 0.0033 | Inf | 1 | -79.704 | 0 |

```
> summary(cohen.d(Rest ~ Popn, data = data.frame(dfdam)))

## Extract effect sizes from cohen.d
## cohen.d(x = Rest ~ Popn, data = data.frame(dfdam))
##
## Multivariate (Mahalanobis) distance between groups 0.84
##  r equivalent for each variable
##  Rest
## -0.39
```

### S3.2.3 Mating latency

```
> lt <- lmer(LatencyMinutes ~ Popn + (1 | Block), data = latency)

## boundary (singular) fit:  see help('isSingular')

> Anova(lt, type = "III")

## Analysis of Deviance Table (Type III Wald chisquare tests)
##
## Response: LatencyMinutes
##                Chisq Df Pr(>Chisq)
## (Intercept) 152.0223  1     <2e-16 ***
## Popn          0.0901  1      0.764
## ---
## Signif. codes:
## 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
> options(width = 60)
> kable(pairs(emmeans(lt, ~Popn), simple = "Popn", reverse = TRUE),
+     digits = 4, booktabs = T)
```

| contrast | estimate | SE | df | t.ratio | p.value |
|---|---|---|---|---|---|
| MD - MC | 0.4881 | 1.7084 | 91.5866 | 0.2857 | 0.7757 |

```
> options(width = 60)
> summary(cohen.d(LatencyMinutes ~ Popn, data = data.frame(latency)))

## Extract effect sizes from cohen.d
## cohen.d(x = LatencyMinutes ~ Popn, data = data.frame(latency))
##
## Multivariate (Mahalanobis) distance between groups 0.05
##  r equivalent for each variable
## LatencyMinutes
##           0.03
```

### S3.2.4 Copulation duration

```
> cd_model <- lmer(CDuration ~ Popn + (1 | Block), data = cd)
> Anova(cd_model, type = "III")

## Analysis of Deviance Table (Type III Wald chisquare tests)
##
## Response: CDuration
##                Chisq Df Pr(>Chisq)
## (Intercept) 579.7856  1    < 2e-16 ***
## Popn          5.7942  1    0.01608 *
## ---
## Signif. codes:
## 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
> options(width = 60)
> kable(pairs(emmeans(cd_model, ~Popn), simple = "Popn", reverse = TRUE),
+     digits = 4, booktabs = T)
```

| contrast | estimate | SE | df | t.ratio | p.value |
|---|---|---|---|---|---|
| MD - MC | -1.4816 | 0.6162 | 154.508 | -2.4043 | 0.0174 |

```
> options(width = 60)
> summary(cohen.d(CDuration ~ Popn, data = data.frame(cd)))

## Extract effect sizes from cohen.d
## cohen.d(x = CDuration ~ Popn, data = data.frame(cd))
##
## Multivariate (Mahalanobis) distance between groups 0.39
##  r equivalent for each variable
## CDuration
##     -0.19
```

### S3.2.5 Mating properties

```
> options(width = 60)
> mp <- glmer(cbind(Mated, Unmated) ~ Popn + (1 | Block), binomial,
+     data = mateprop)

## boundary (singular) fit:  see help('isSingular')
```

```
> Anova(mp, type = "III")

## Analysis of Deviance Table (Type III Wald chisquare tests)
##
## Response: cbind(Mated, Unmated)
##               Chisq Df Pr(>Chisq)
## (Intercept) 40.7215  1  1.755e-10 ***
## Popn         4.2797  1    0.03857 *
## ---
## Signif. codes:
## 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
> options(width = 60)
> kable(pairs(emmeans(mp, ~Popn), simple = "Popn", type = "response",
+     reverse = TRUE), digits = 4, booktabs = T)
```

| contrast | odds.ratio | SE | df | null | z.ratio | p.value |
|---|---|---|---|---|---|---|
| MD / MC | 0.4353 | 0.175 | Inf | 1 | -2.0687 | 0.0386 |

```
> options(width = 60)
> mateprop <- mateprop %>%
+     mutate(Propensity = Mated/(Mated + Unmated))
> summary(cohen.d(Propensity ~ Popn, data = data.frame(mateprop)))

## Extract effect sizes from cohen.d
## cohen.d(x = Propensity ~ Popn, data = data.frame(mateprop))
##
## Multivariate (Mahalanobis) distance between groups 1.15
##  r equivalent for each variable
## Propensity
##       -0.5
```

### S3.2.6 Male-male aggression

```
> mc <- (df_aggr %>%
+     filter(Popn == "MC"))$WinFreq
> md <- (df_aggr %>%
```

```
+     filter(Popn == "MD"))$WinFreq
> wilcox.test(mc, md, alternative = "two.sided")

## Warning in wilcox.test.default(mc, md, alternative = "two.sided"):  cannot
compute exact p-value with ties

##
##  Wilcoxon rank sum test with continuity correction
##
## data:  mc and md
## W = 12.5, p-value = 0.2454
## alternative hypothesis: true location shift is not equal to 0
```

# S4 | Parameterization of the resource consumption framework for CRPC

## S4.1 Derivation of cell growth parameters

For a small initial population and abundant resource supply, Equation 4.1 can be simplified to an exponential growth process by ignoring the quadratic growth term. Solving this for one population doubling leads to the following expression for the exponential growth rate for each cell type:

$$r_i = \frac{ln(2)}{\tau_{d,i}} + \delta_i, \tag{S4.1}$$

where $r_i$ is the cell-type specific growth rate, $\delta_i$ is the cell-type specific death rate as inferred from (Jain et al., 2011), and $\tau_{d,i}$ is the doubling time from available data of cell lines corresponding to each of our cell types (Bairoch, 2018). We take $T^+$, $T^p$ and $T^-$ to correspond to the LNCaP clone FGC (RRID: CVCL_1379), 22Rv1 (RRID: CVCL_1045) and PC-3 (RRID: CVCL_0035) cell lines respectively, and use the doubling times of these cell lines to calculate the growth rates by Equation S4.1. We assume that the effective carrying capacity for each cell type growing in the absence of other cell types under abundant resource availability to be 10000. Setting the LHS of Equation 4.1 to zero and rearranging, we obtain the following expression for $K_{i,max}$:

$$K_{i,max} = \frac{r_i}{r_i - \delta_i} y_i^* - 1 \approx \frac{r_i}{r_i - \delta_i} y_i^*, \tag{S4.2}$$

where $y_i^*$ is the steady-state population size of cell type $i \in T^p, T^+, T^-$, growing in the absence of competition from other cell types and abundant resource availability, and $y_i^*$ is set to 10000. Substituting this in Equation S4.2 allows us to calculate cell-type specific values for that would realise a maximum population size of 10000 for each cell type.

## S4.2 Normalization of resource dynamics

As oxygen and testosterone are produced and consumed at different rates, their stoichiometry is likely to be vastly different. In the interest of simplicity, we neglect considerations of stoichiometry in our model by normalising the abundance of each resource against the empirically measured steady state value of that resource in prostate tissue. Median oxygen tension in prostate cancer tissue is found to be about 2.5 mmHg (Stewart et al., 2010) and we take this to be the steady state level of oxygen in prostate cancer tissue. For a tumour volume of 1mL, this corresponds to a steady state oxygen level of 3.375 nmol. Setting the LHS of Equation 4.2 to zero, considering a population of only $T^-$ cells, normalising both $\lambda_{O_2}$ and $\mu_{O_2,i}$ against 3.375 nmol, and setting normalised equilibrium oxygen, $O_2$ to one, we obtain $p_{O_2}$ corresponding to vascularised PCa tissue from Equation S4.3 as 0.11 units per min. For limited oxygen supply, we assume $p_{O_2} = 0.06$ units per min.

$$p_{O_2} = \lambda_{O_2} O_2 + y^*_{T^-} \mu_{O_2,T^-} \tag{S4.3}$$

Studies on tumour tissue from patients undergoing ADT find a testosterone level of 3.75 pmol/g tissue in PCa (Titus et al., 2005). As with oxygen, we take this as the steady state value and for a 1mL tumour volume with tissue density 1.03 g/cm$^{-3}$ (Stamey et al., 1987), this corresponds to a steady state testosterone level of 3.64 pmol. All rate parameters are divided by this value to give consumption, production and decay rates in units of proportion of steady state resource level. For the rate of testosterone production by $T^p$ cells ($p_T$), empirical sources indicate an absolute rate of about 1100 pg per million cells per 48h (Dillard et al., 2008), which is about $3 \times 10^{-10}$ units per cell per min after normalization. Initial simulations suggested that this rate of production was too low for to support cell growth from a small initial number of cells within the simulated time. Following a brief parameter space exploration, we assume $p_T = 5 \times 10^{-7}$ units per cell per min to ensure sufficient testosterone availability. Setting Equation 4.3 to zero, taking normalized equilibrium testosterone, $T^* = 1$, and normalizing $\lambda_T$ to 3.64 pmol, we obtain $\mu_{T,T^p}$ for a steady-state $T^p$ population of 10000 as $6 \times 10^{-8}$ units per cell per min from Equation S4.4 below.

$$p_T - \mu_{T,T^p} = \frac{\lambda_T T^*}{y^*_{T^p}} = 4 \times 10^{-7} \tag{S4.4}$$

Since direct empirical sources were not available for $T^+$ testosterone consumption rate, we explore values over four orders of magnitude from $10^{-11}$ to $10^{-7}$ units per cell per min, with the additional assumption that $T^+$ consumed testosterone at a lower rate than that of $T^p$. Supplementary Figure S6.2.1 shows the time series of cell growth and resource dynamics for the highest rate of testosterone consumption by $T^+$, and we see that $T^+$ is outcompeted early in the simulation by $T^p$. Closer examination of the time series data indicates that while both cell types initially decrease in abundance due to resource limitation, $T^+$ declines more rapidly and is thus driven to extinction by $T^p$. Since this extinction is independent of the testosterone consumption rate of $T^+$, we fix this rate to be a value slightly less than the $T^p$ consumption rate. This is effectively a simplification of the $T^p - T^+$ interaction, as it precludes a potential amensal interaction by which $T^+$ could consume an excess of the testosterone produced by $T^p$. For external supply of testosterone, we add a constant supply term of 0.001 per min to Equation 4.3.

# S5 | Supplementary analyses for Chapter 5

## S5.1 Alternative distributions of $g$

As explained in the main text, we explore the effects of using alternative distributions for $\Delta_g$. Supplementary Figure S6.3.1 shows the effect of a Gumbel-dsitributed progression, and Supplementary Figure S6.3.2 shows that of a uniformly-distributed progression. We note that the different distributions produce qualitatively the same results as the normal distribution used throughout the main text.

## S5.2 Sensitivity of model predictions to $k$

Keeping with the sensitivity analysis in the main text, we test the effect of $k$ on the observed relationships of $p$ and $n$ with cancer incidence, and discuss the role of $k$ in greater detail. The time taken to cancer onset is a useful parameter in this context as it describes the temporal dynamics of mutation accumulation, while allowing for some limited inference regarding total incidence in the population. If time taken to cancer is largely short, population incidence is likely to be large in the given parameter space, and vice versa.

Broadly, we find that cancer incidence is significant only up to maximum $k = 10$. Increasing $k$ also reduces total incidence and shifts the observed time to cancer to later in life, as expected based on mutation accumulation. We also find that the magnitude of $k$ could modulate the strength of the association with $n$, and to a lesser extent, with $p$. Importantly, the effect of either $n$ and $p$ is apparent only when taken for one value of $k$ at a time. As Figure S6.3.3 shows, when time to cancer onset is pooled across all $k$ values, it appears largely independent of either $n$ and $p$.

Doing the same for the context-dependent case, we randomized $g$ as explained earlier, along with either $k$ and $n$, or $k$ and $p$. Remarkably, introducing $g$ as a random variable leads to most of the variance in time to cancer being explained by $g$, and to some extent, $n$. Again, the association between $g$ and time to cancer onset is modulated by $k$, as observed for the association with $n$ and $p$. As with Supplementary Figure S6.3.3, the effect of $g$ on time to cancer onset is only apparent when considered one value of $k$ at a time (Supplementary Figures S6.3.4 and S6.3.6).

## S5.3 The nature of the relationship between cancer incidence and cell number

As mentioned in the main text, the last few years have seen two prominent attempts by Tomasetti et al. to examine the relative contributions of spontaneous mutations, genetic

and environmental factors in cancer development (Tomasetti and Vogelstein, 2015, Tomasetti et al., 2017). In their analyses, they estimate *l*ifetime *s*tem *c*ell *d*ivisions (*lscd*) for 17 tissue types and correlate them with the incidence of cancer using data from the US, and across the world. In the 2017 paper (Tomasetti et al., 2017), they use the IACR datasets for cancer incidence that consisted of 423 databases corresponding to different countries, of which 347 had incidence data for all 17 cancer types considered here. With these data, they report a strong statistical association (Pearson's $r \approx 0.8$) between *lscd* and incidence of cancer in a given tissue type. The correlations are on a log-log scale, and on that scale, they are considerably strong. At face value, this is in line with the expected relationship if cancer arises out of largely random processes of mutagnesis, and the authors therefore view this association as clear evidence for the causal significance of spontaneous mutations in carcinogenesis. They use this association to further attribute the majority of cancer incidence to random replication errors alone.

Since all the data used in these papers are publicly available, we performed a model fitting exercise, taking a closer look at the nature of the associations. We considered the same 347 datasets, and found that while Pearson's $r$ was indeed distributed around a median of 0.8 as reported by Tomasetti et al., the slopes of the regression were distributed narrowly around a median value of 0.334 (Supplementary Figure S6.3.7). Going by the classical logic of cancer, it would mean that an average of 0.334 mutations are required to cause cancer, which is absurd. We expect a positive integer here, but get a substantially small fraction instead. This fractional slope has been identifed in passing as surprising (Nowak and Waclaw, 2017), without further discussion. On the other hand, evidence in support of a non-linear relationship between cancer risk and *lscd* has been accumulating in parallel (Noble et al., 2015), but on the whole, the causal significance of such non-linearity for cancer etiology remains to be clearly elucidated. A non-linear effect of *lscd* on cancer risk either challenges the hypothesis of all required oncogenic mutations coming together purely by chance, or points to significant gaps that must be addressed.

Residuals from a regression often offer insight on the linearity of a supposed relationship, and we use this insight to investigate the purported linearity of the cancer risk-*lscd* relationship on a log-log scale. In this case, we observed that distribution of residuals around the regression was not symmetric as would be expected based on the indication of a strong linear relationship on a log-log scale. Along the first and the last one-fourth of the *lscd* range, the points lie predominantly below the line, while in the middle half of the range, a higher fraction lie above the line, as shown by the respective frequencies of positive and negative residuals (Supplementary Table S5.1). In addition to the skewed distribution of points, a visual inspection of Supplementary Figure S6.3.7 also suggested that the linear regression itself was largely driven by the first 4 points in each dataset. In line with this

suspicion, upon removal of the first four points, the significance of regression was lost in all 347 datasets considered in this analysis. In order to test that the loss of significance was not due to reduction in sample size alone, the regression was performed with elimination of all other combinations of four points from each dataset (${}^{17}C_4 - 1 = 2379$ combinations per dataset, not including elimination of the first four). The distribution of Pearson's r values across the combinations revealed a striking difference, with most other combinations of points retaining the significance of the relationship. This is demonstrated by a large part of the former distribution lying below the r value threshold for $p < 0.05$ ($r_{threshold} \approx 0.55$), as shown in Supplementary Figure S6.3.8. There is therefore evidence to suggest that the log-linear relationship is largely due to the first four incidences being lower than the rest. Moreover, given the substantially less-skewed distribution of residuals about the saturation curve compared to the log-linear equation (Supplementary Table S5.1), there is sufficient basis to believe that a saturation equation might in fact be a better fit to the data than a fractional power curve. Such a saturation relationship between *lscd* and cancer incidence would indicate that *lscd*, and by extension, the stem cell number, has a quantitative effect on the development of cancer only up to a certain threshold, beyond which something else becomes limiting. As we argue in the main text, these limiting factors other than the cell number and/or turnover could stem from context-dependent selection imposed through the tissue microenvironment.

***Supplementary Table S5.1:*** **Number of positive and negative residuals from the linear regression line**. The residuals were calculated from linear regressions of cumulative cancer risk, log(CR), against log(*lscd*); data were obtained from IARC as described in the text. The columns correspond to the first one-fourth, middle half and last one-fourth of the *lscd* range. Notably, a substantial skew can be seen in the extremes of the range in the linear case, with more points below the straight line than above it, reflected by a greater number of negative residuals. Compared to the linear case, the skew in the distribution of residuals is significantly lesser for the saturation equation.

| Regression type | Residuals | First one-fourth | Middle half | Last one-fourth |
|---|---|---|---|---|
| Linear | Positive | 667 | 1919 | 4 |
| | Negative | 1415 | 1204 | 343 |
| Saturation | Positive | 1193 | 1585 | 220 |
| | Negative | 889 | 1538 | 127 |

# S6 | Supplementary Figures

## S6.1 Supplementary Figures for Chapter 2

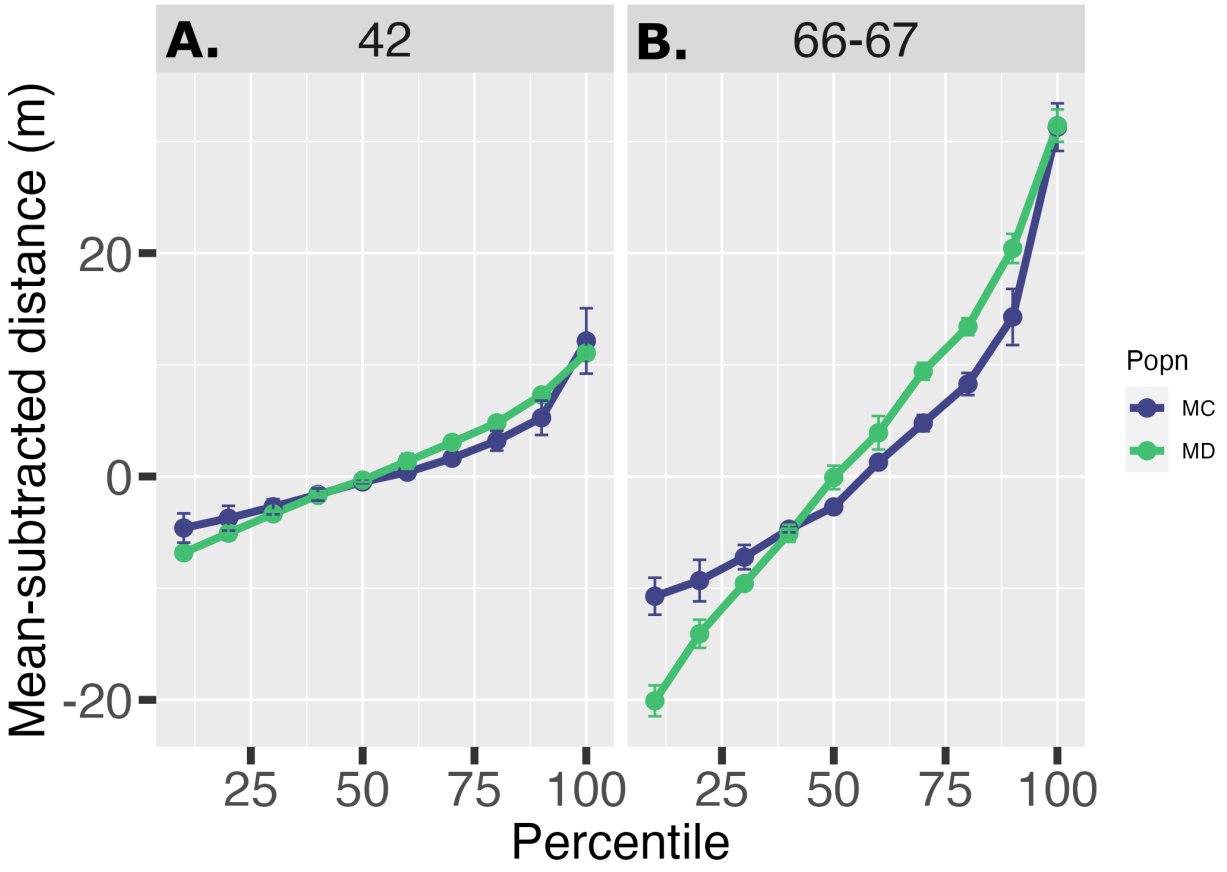



***Figure S6.1.1:*** **Percentiles of the mean-subtracted distances for both populations**. All points are ensemble means ± SD over the four blocks. From generation (A) 42 to (B) 66-67, spatial extent has increased for both MD and MC without a significant divergence in the frequency of LDDs, as shown by the apparent convergence of the distance percentiles beyond the $90^{\text{th}}$.

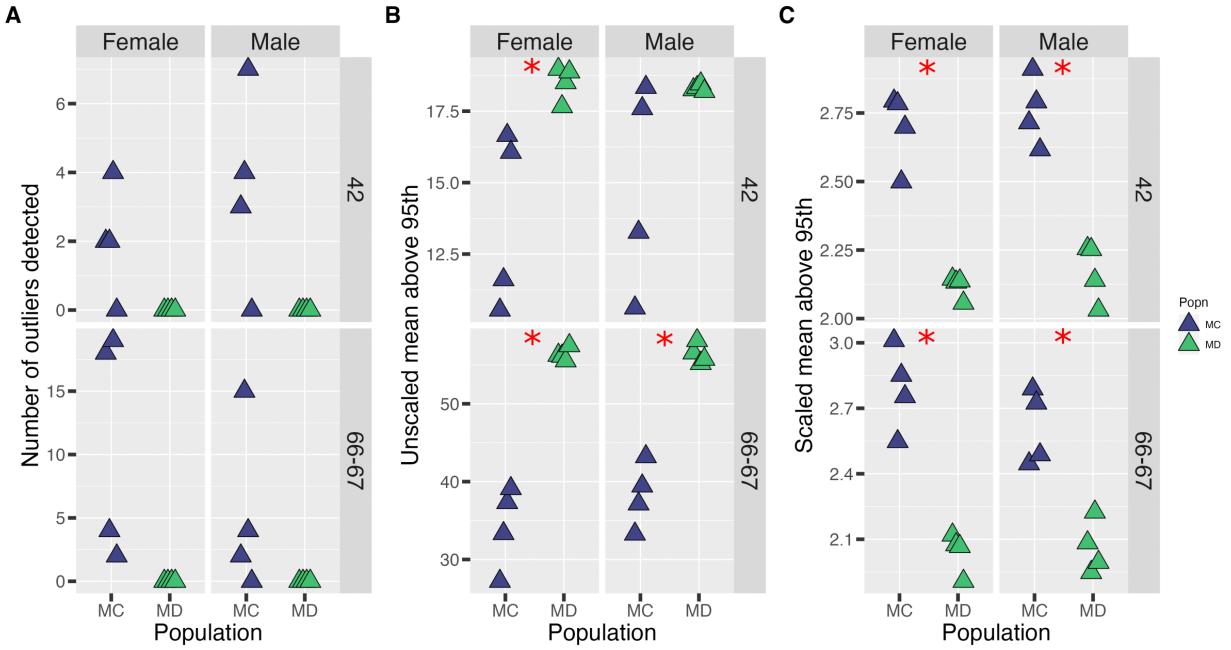



***Figure S6.1.2:*** **Comparison of multiple LDD measures across generations 42 (top row) and 66-67 (bottom row).** (A) Number of outliers detected by the Hampel identifier with a cutoff of 2.5, (B) average value of the top 5% distances traveled (distances greater than the 95$^{\text{th}}$ percentile), and (C) average of the top 5% distances, scaled by the overall average distance moved. All other specifics are the same as in Figure 2.5.

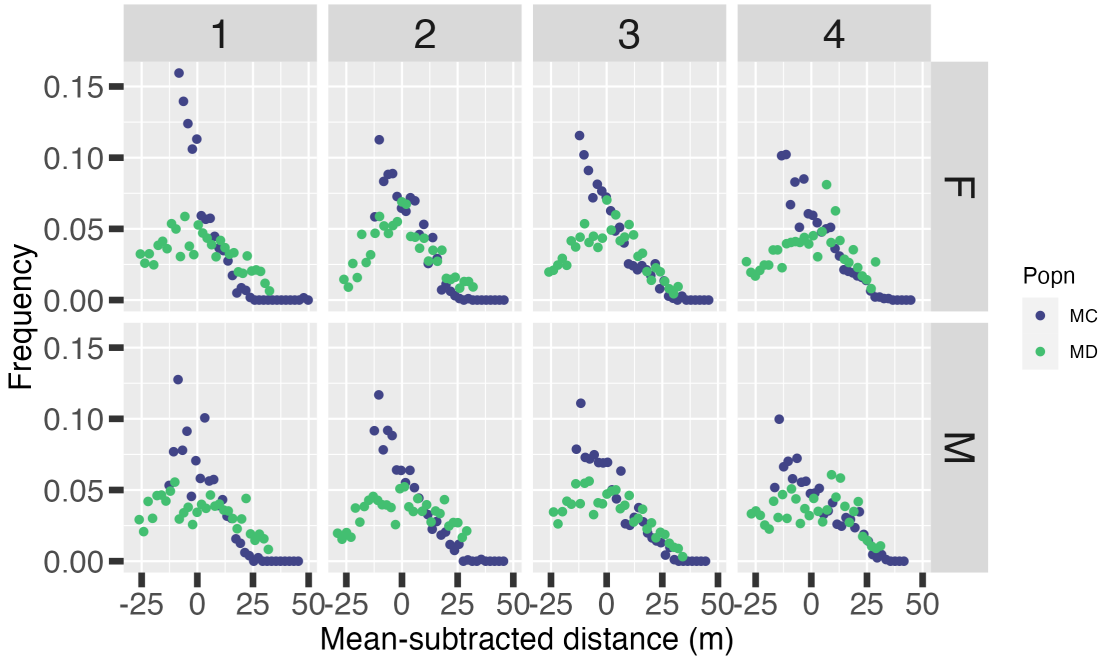



***Figure S6.1.3:*** **Dispersal kernels of both populations after mean subtraction**. From the dispersal kernel of each population, the mean of their own distribution has been subtracted. Note that the x-axis is the mean-subtracted distance and therefore is stretches on either side of zero. Points are average frequencies over within-block replicates, columns are blocks, and rows are sexes (M-male, F-female). These data are plotted only for data from generations 66-67 as the kernel shape in generation 42 is not as distinct. Note that while both distributions have moved to the left after mean subtraction, they have been shifted by different amounts corresponding to their population averages. The righthand tails appear to overlap in several cases.

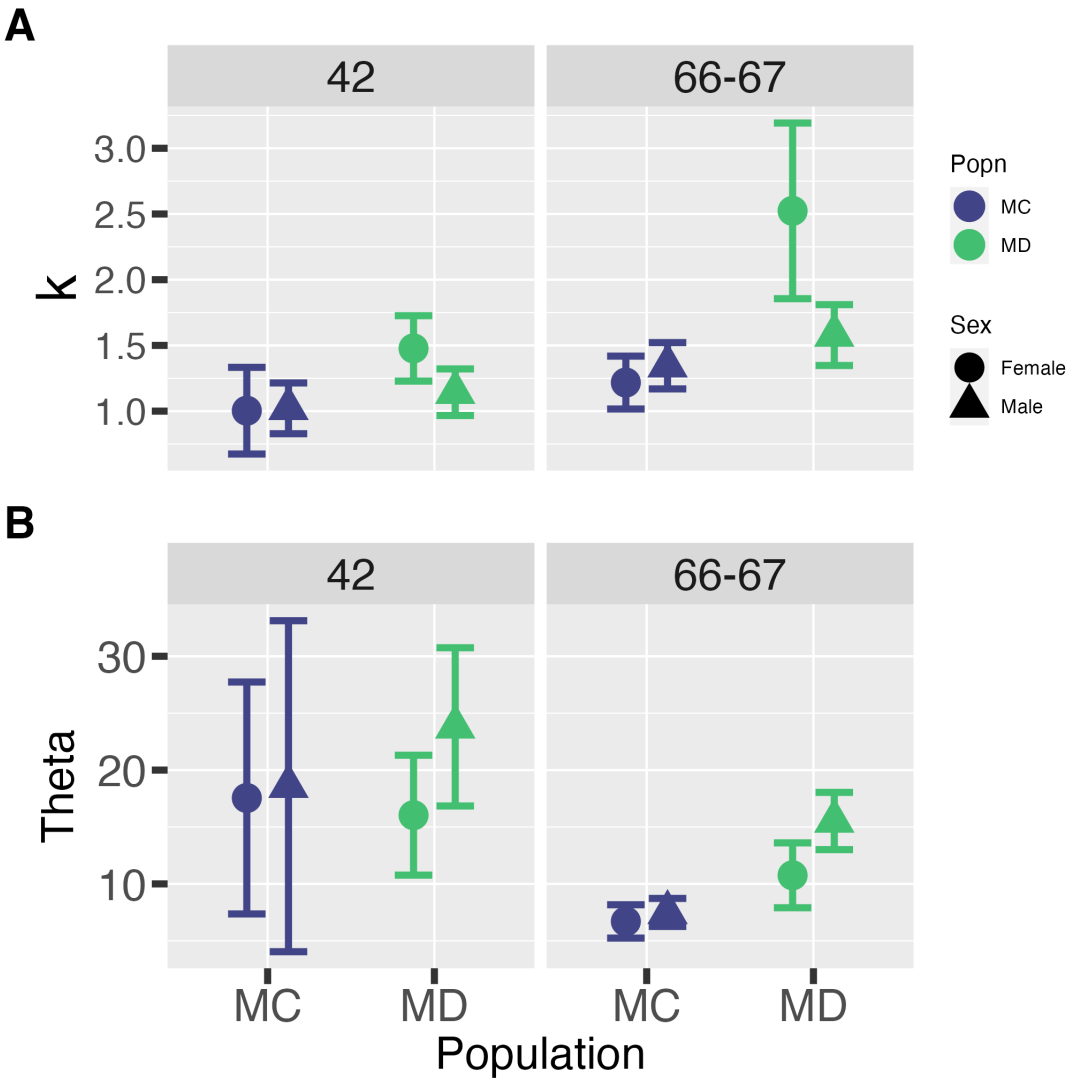



***Figure S6.1.4:*** **Model parameters as inferred by fitting the Gamma distribution function in Equation 2.1.** (A) $k$ and (B) $\theta'$. See main text for the calculation of $\theta'$. All points are ensemble means ± SD over the four blocks. While the model parameters show some sex-specific trends, block-wise Kolmogorov-Smirnov tests comparing the frequency distributions between the two sexes within each population showed no significant differences (see Supplementary Material S2 for details). It therefore remains unclear if the differences in fit parameter values between MD and MC among the sexes bear any functional significance.

## S6.2 Supplementary Figures for Chapter 4

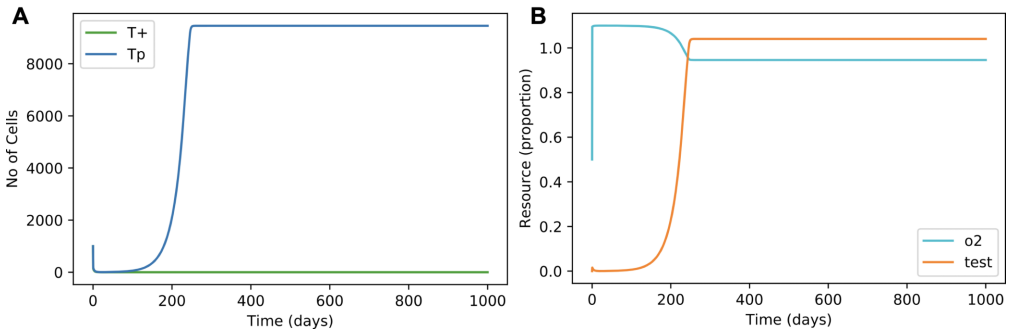



***Figure S6.2.1:*** **Fixing the testosterone consumption rate for $T^+$.** Time series of (A) $T^p$ and $T^+$ cell growth, and (B) resource dynamics in the absence of $T^-$. $T^p - T^+$ dynamics show that $T^+$ is competitively excluded by $T^p$ at steady state. Initial population size is 1000 cells of each type, $\mu_{T,T^p} = 6 \times 10^{-8}$, $\mu_{T,T^+} = 10^{-7}$, lower limit for the resource use function $f_i$ for both resources and both cell types is zero, and the upper limit is one.

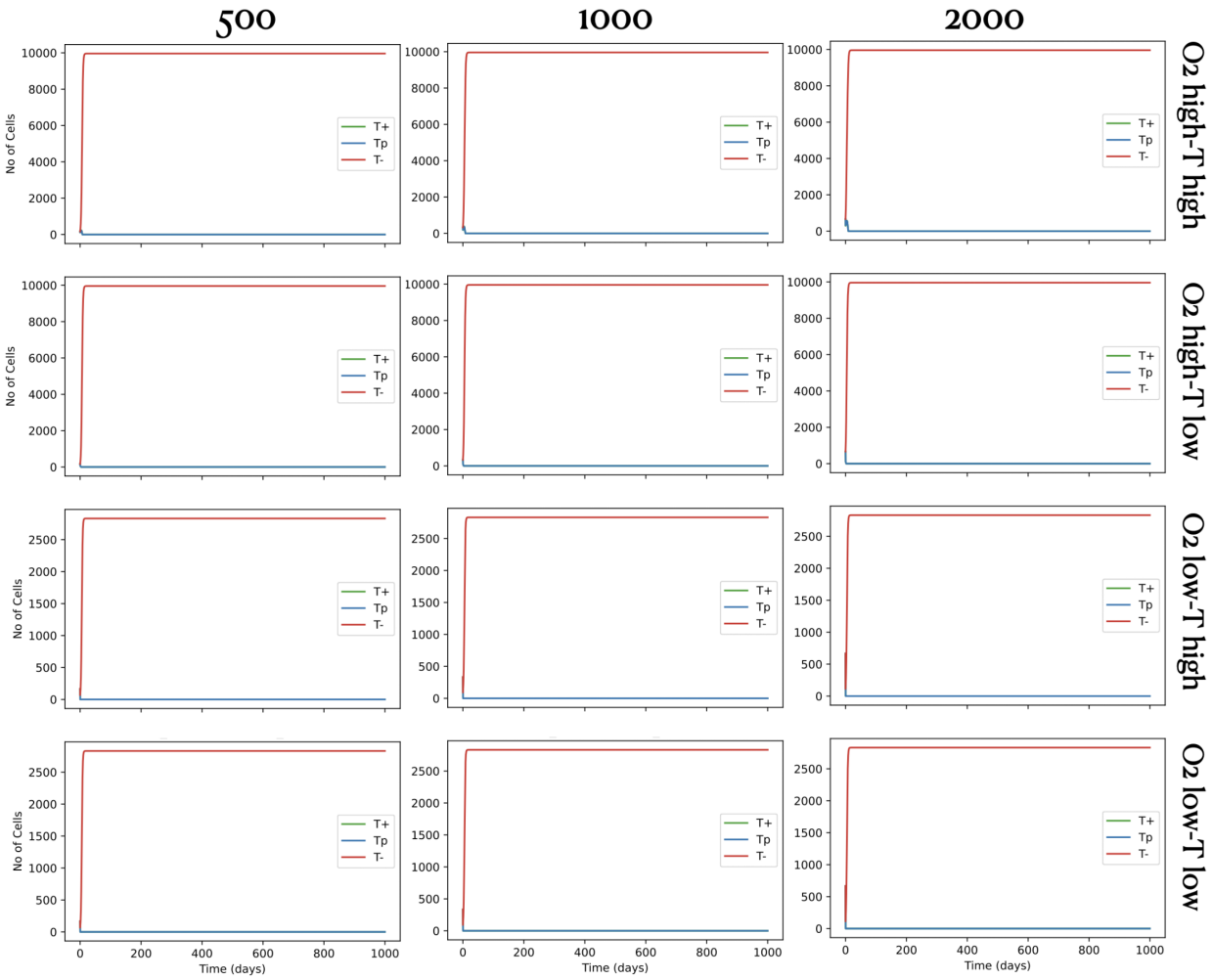



***Figure S6.2.2:*** **Cell growth time series when neither resource is supplemented externally**. Resource use efficiency combinations vary across rows and initial population size across columns. These simulations were run for 1000 days since steady state was seen to be reached within the first few days.

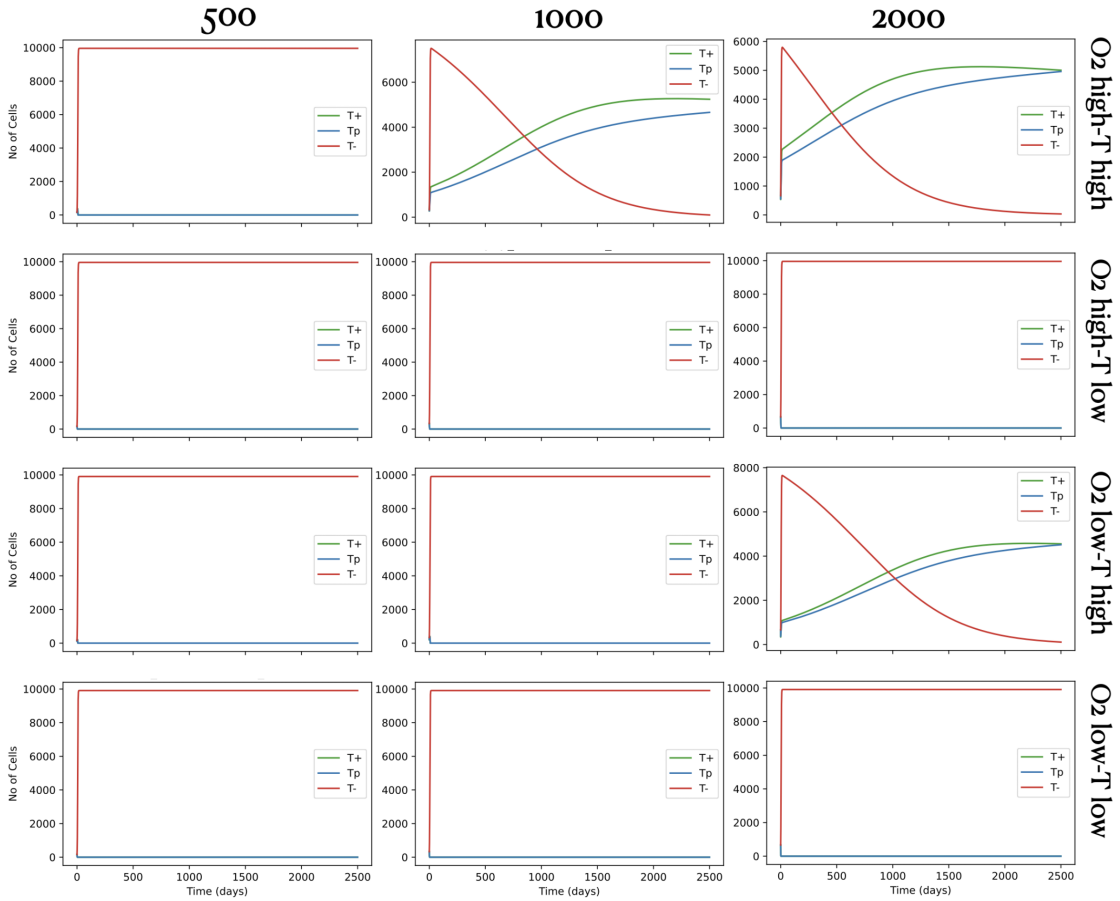



***Figure S6.2.3:*** **Cell growth time series when only oxygen is supplemented externally**. Resource use efficiency combinations vary across rows and initial population size across columns. These simulations were run for 2500 days.

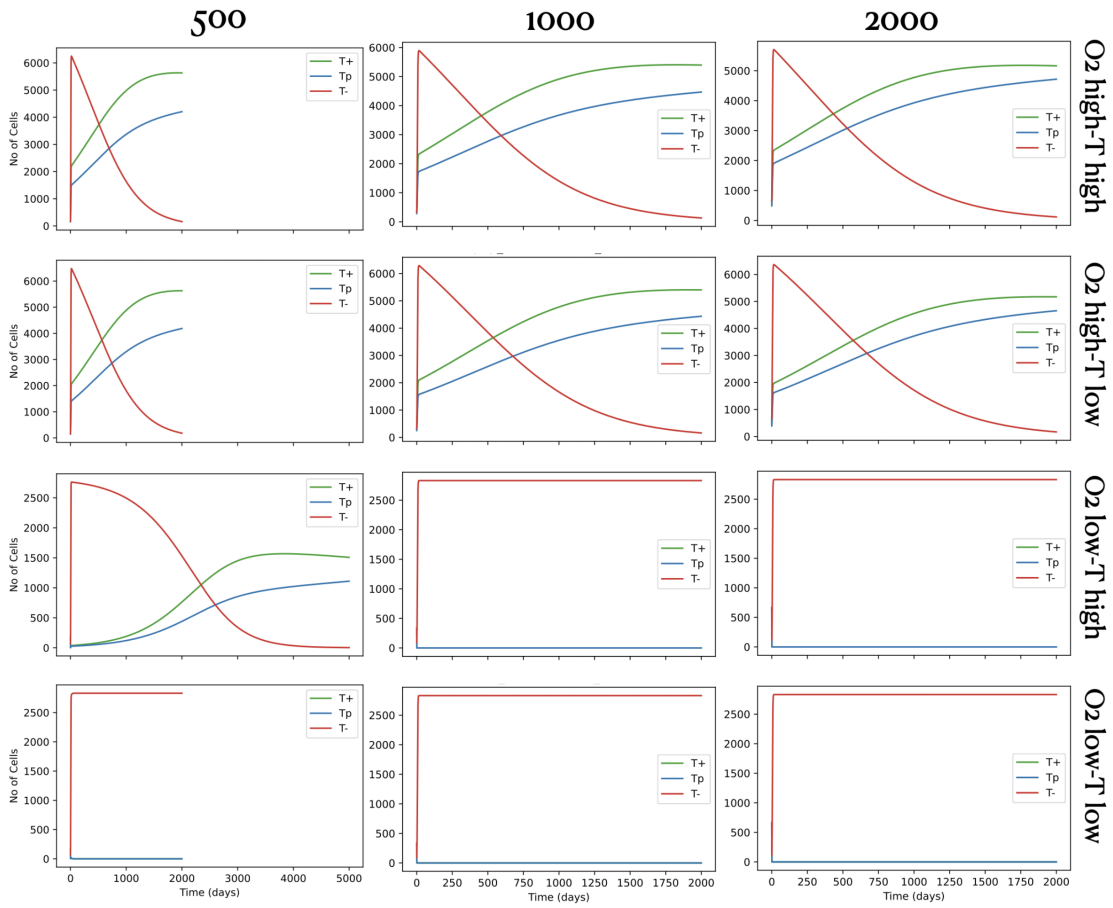



***Figure S6.2.4:*** **Cell growth time series when only testosterone is supplemented externally**. Resource use efficiency combinations vary across rows and initial population size across columns. All simulations were run for 2000 days, except for one case with initial population size 500, low use efficiency of oxygen and high use efficiency of testosterone, which was run for 5000 days for a clear plateau to emerge.

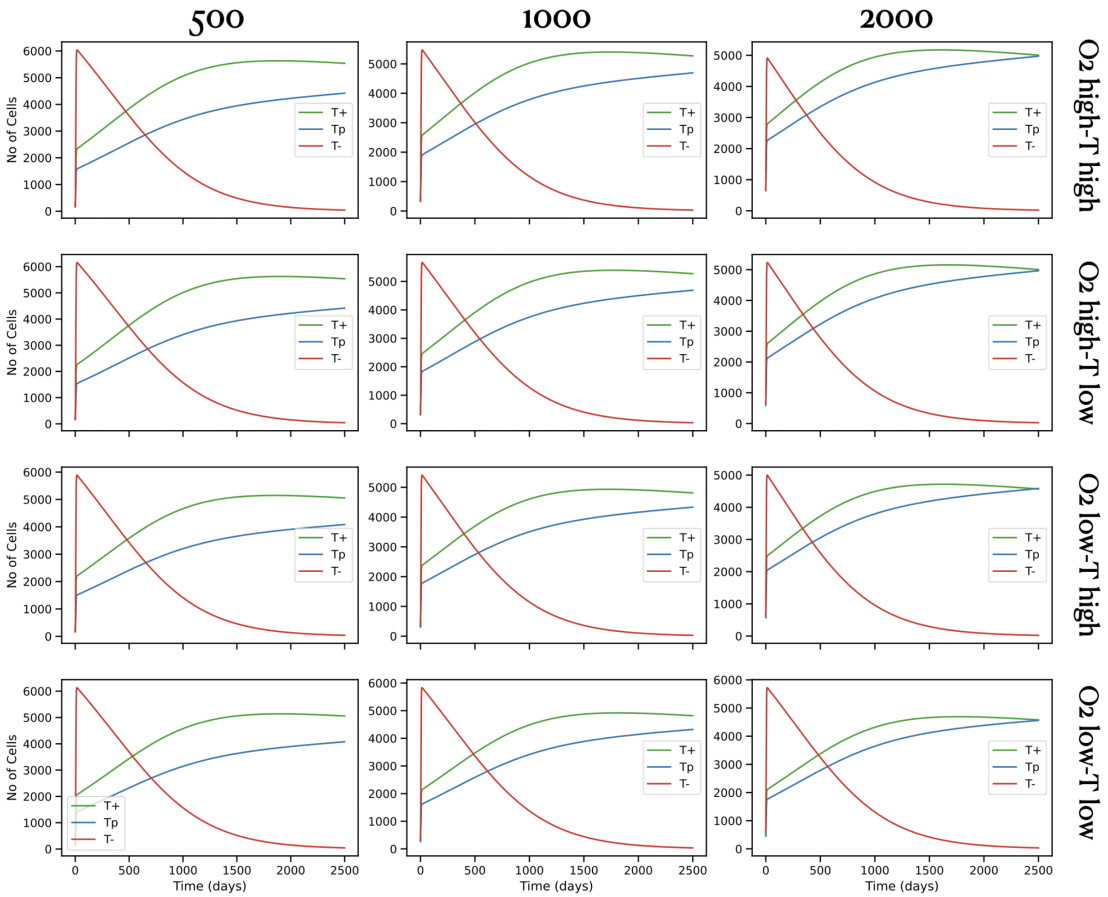



***Figure S6.2.5:*** **Cell growth time series when <u>both</u> resources are supplemented externally**. Resource use efficiency combinations vary across rows and initial population size across columns. These simulations were run for 2500 days.

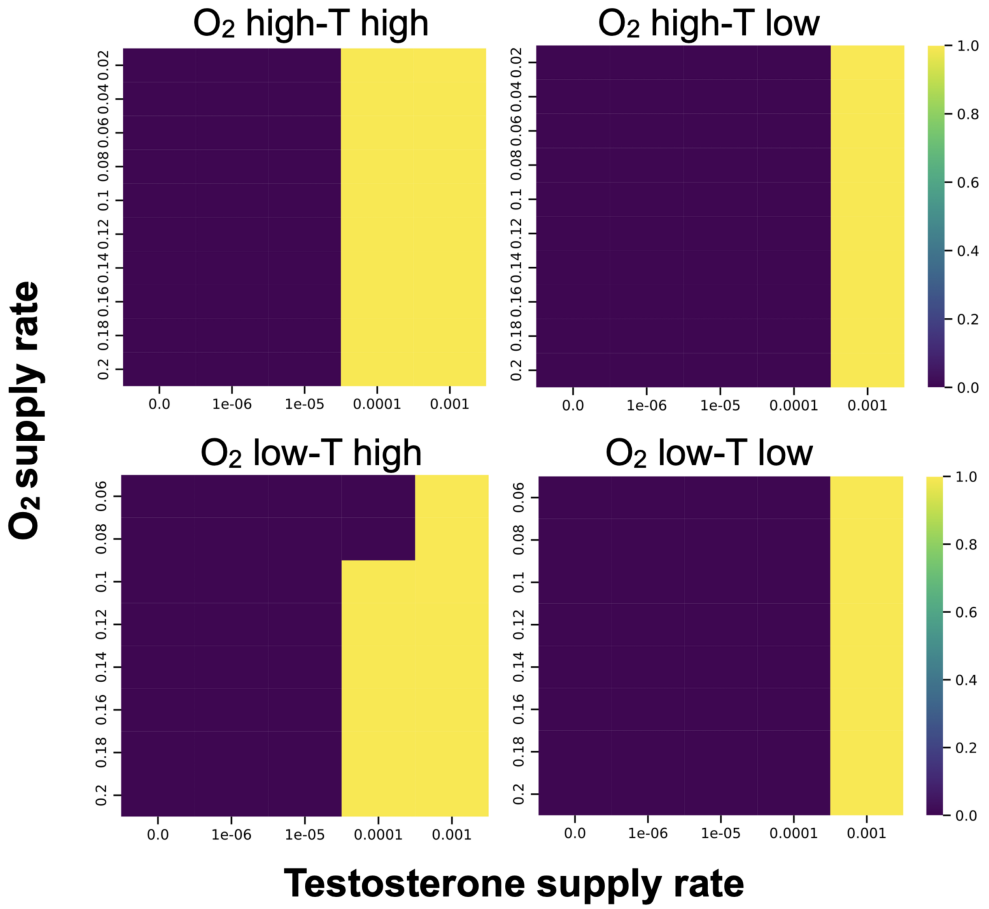



***Figure S6.2.6:*** **Overall population viability under a range of resource supply conditions**. Heat maps of steady state ratio of $[T^p + T^+]$ against total population size over a range of supply rates of oxygen and testosterone. By definition, all resource supply states where this ratio goes to zero represents a steady state population of only $T^-$. Note that under low use efficiency of oxygen, oxygen supply rates below 0.06 $\text{min}^{-1}$ could not support any cell growth at all and are therefore not shown in this plot. In all cases, simulations were initialised with equal numbers of all three cell types, with an initial population size of 500 and run for 2500 days.

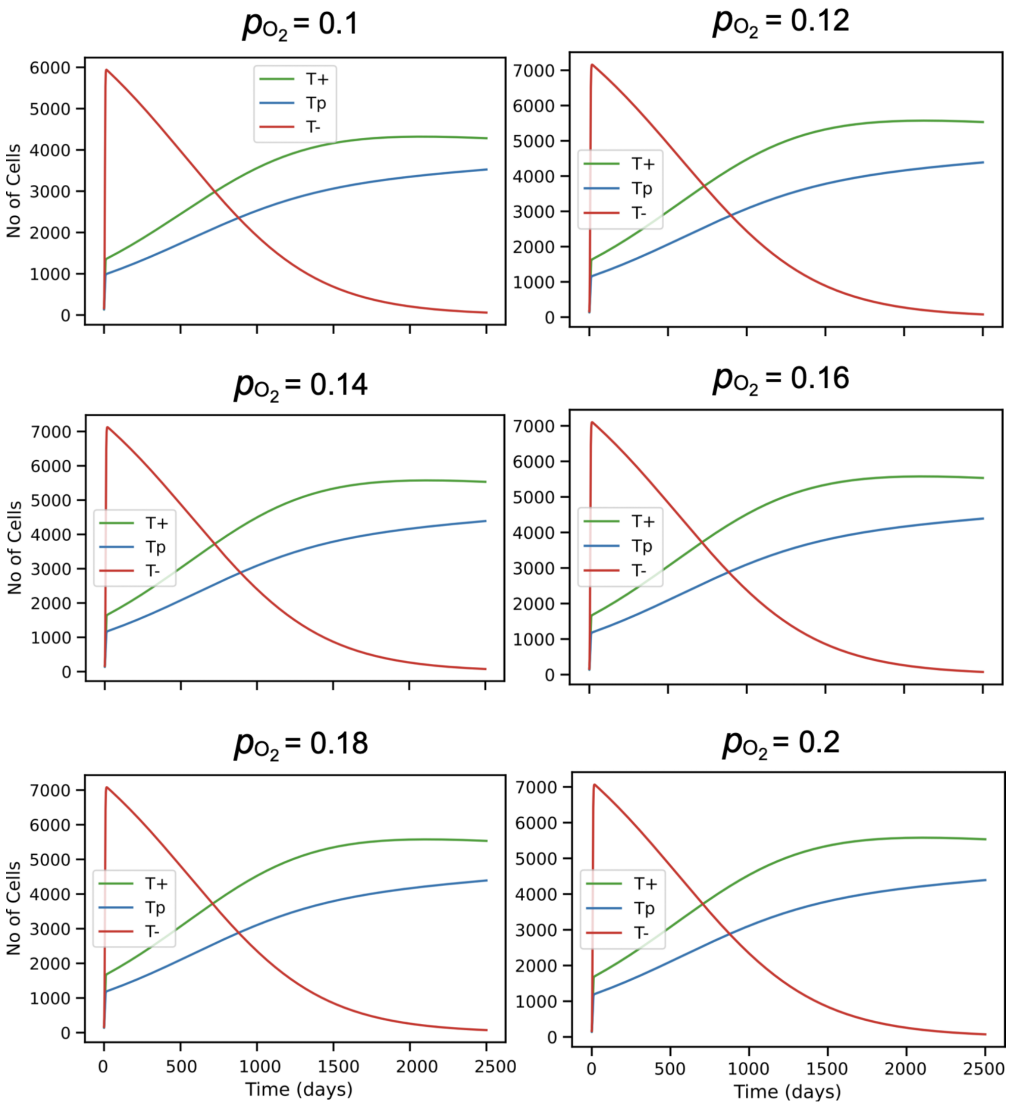



***Figure S6.2.7:*** **Time series of cell growth corresponding to Figure 4.4A in the main text for testosterone supply rate 0.0001 min$^{-1}$.** Low use efficiency of oxygen and high testosterone use efficiency, over a range of oxygen supply rates as specified in the panel titles. Initial population size was 500 with equal numbers of all three cell types.

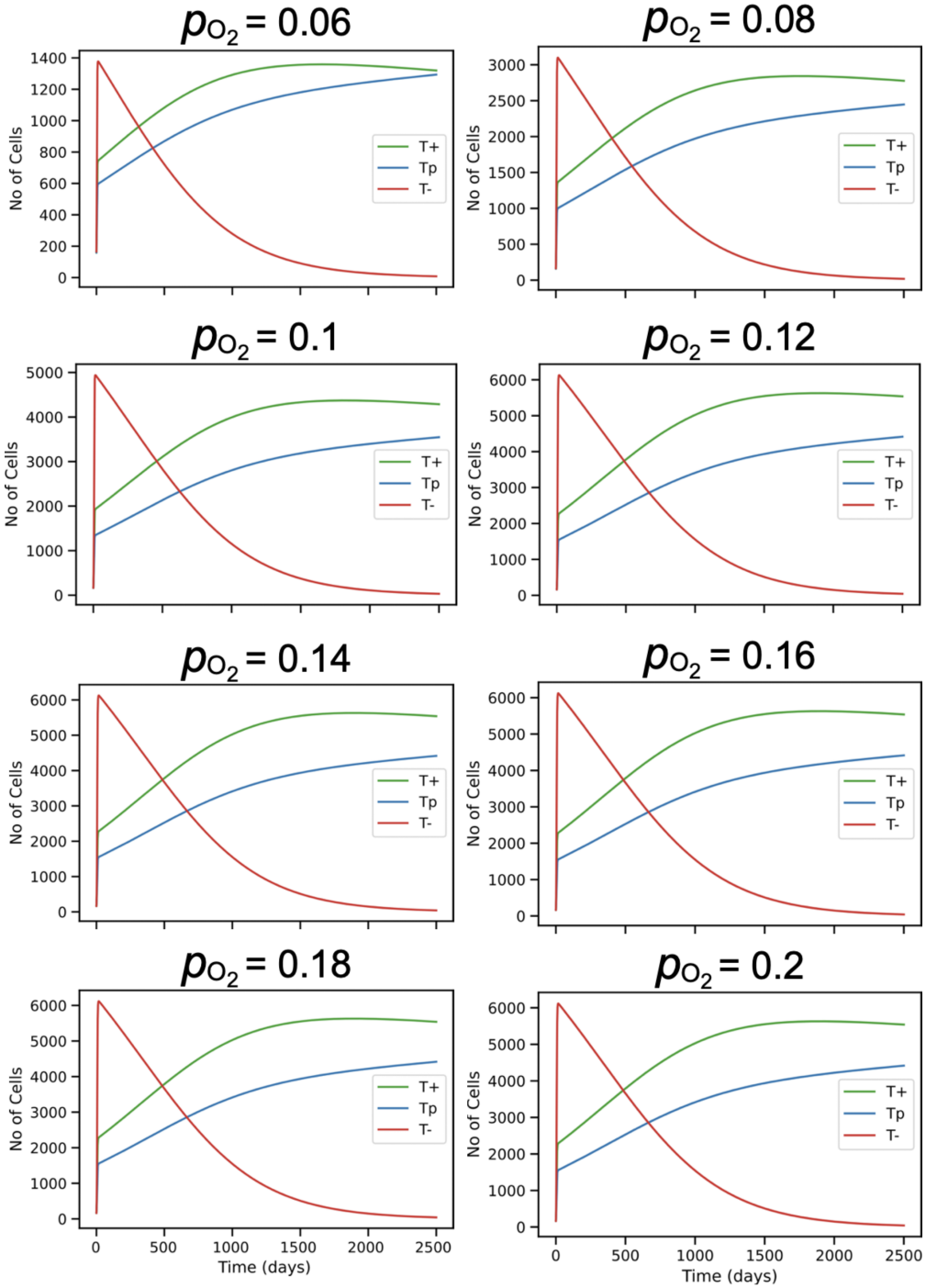



***Figure S6.2.8:*** **Time series of cell growth corresponding to Figure 4.4A in the main text for testosterone supply rate 0.001 min$^{-1}$.** Low use efficiency of oxygen and high testosterone use efficiency, over a range of oxygen supply rates as specified in the panel titles. Initial population size was 500 with equal numbers of all three cell types. Note that cell growth is now viable for lower oxygen supply rates than in the previous figure.

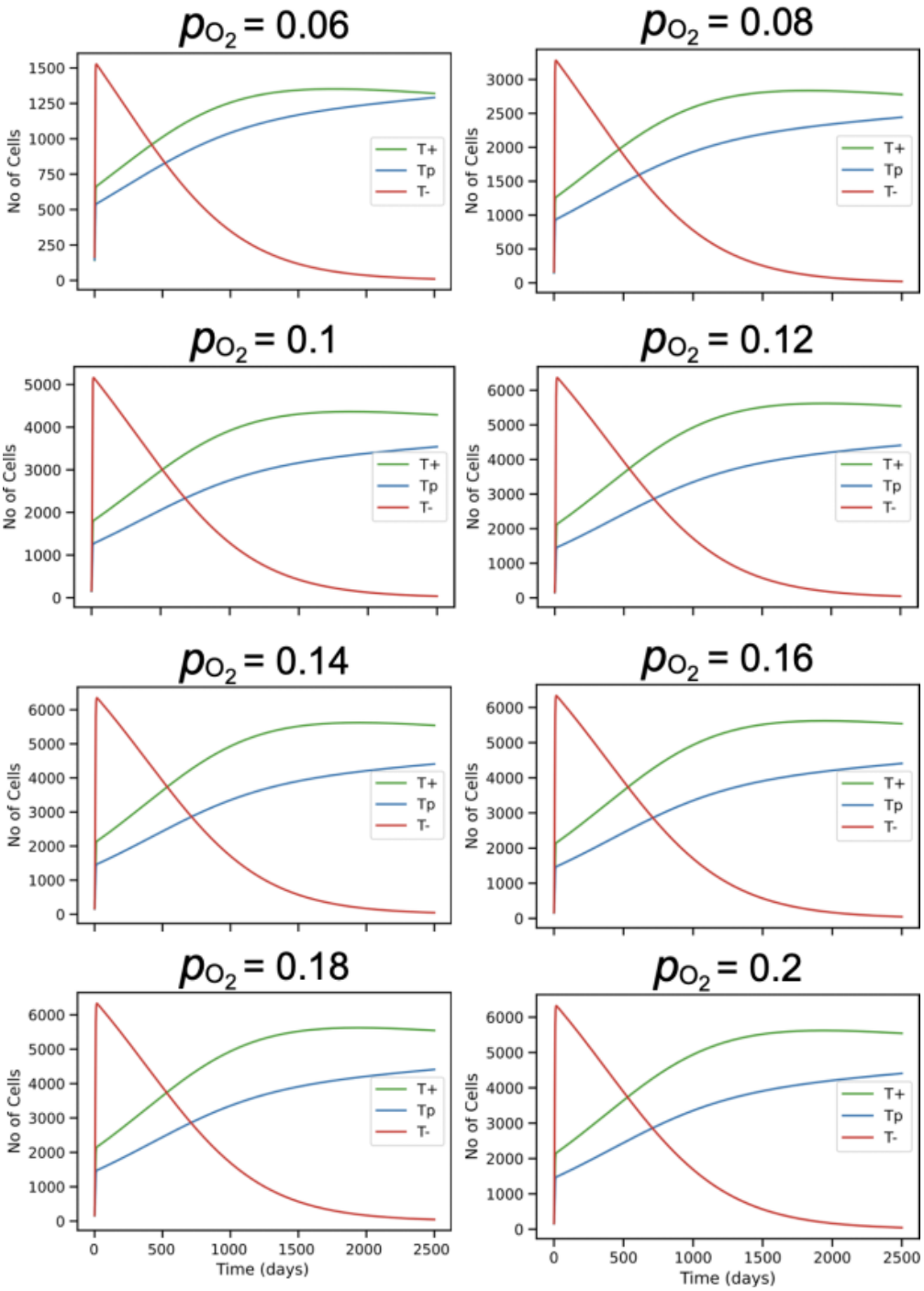



***Figure S6.2.9:*** **Time series of cell growth corresponding to Figure 4.4B in the main text**. Low use efficiency of both oxygen and testosterone, over a range of oxygen supply rates as specified in the panel titles. Initial population size was 500 with equal numbers of all three cell types, and testosterone supply rate was fixed at 0.001 $min^{-1}$.

# S6.3 Supplementary Figures for Chapter 5

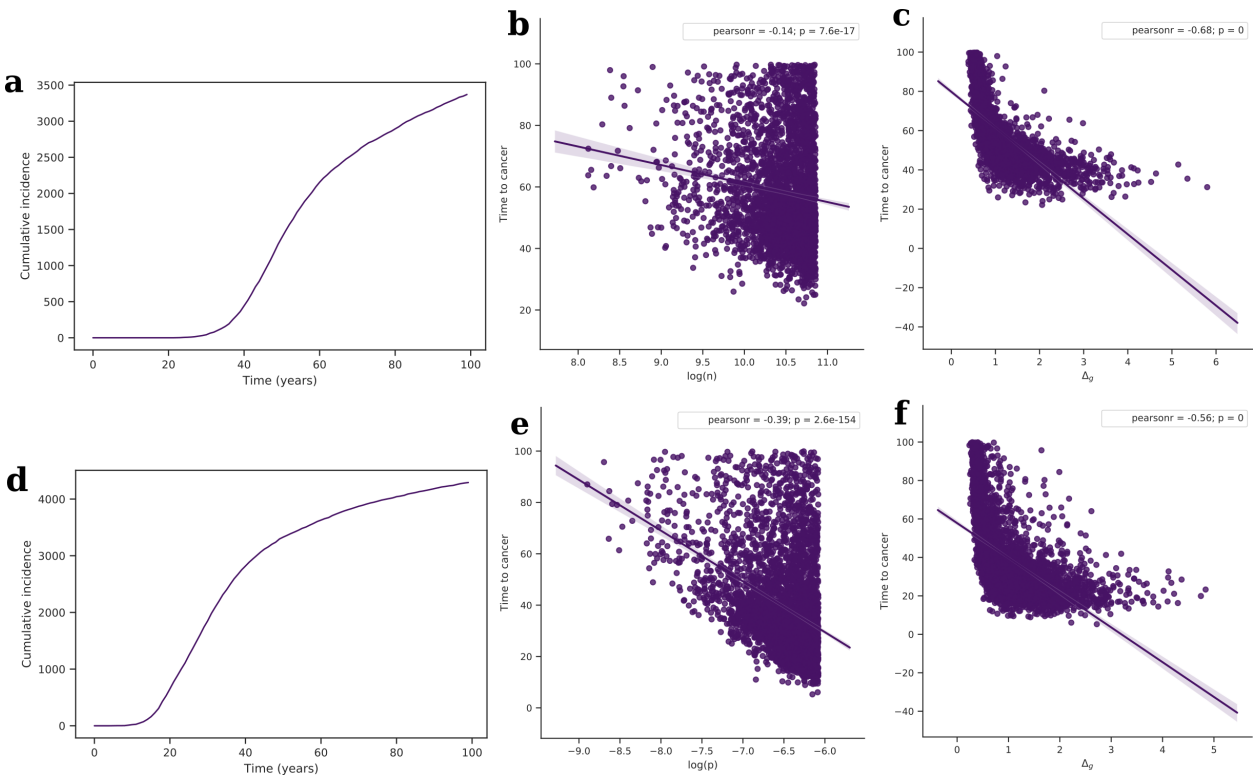



***Figure S6.3.1:*** **Alternate distributions of $g$: Gumbel**. $g$ modeled by a Gumbel-distributed random variable, $\mu$, with $\overline{\mu} = 0$ and $\sigma = 3$, co-randomizing $n$ (top row) or $p$ (bottom row), with ranges $[1.203 \times 10^6, 2.649 \times 10^{10}]$, and $[3.775 \times 10^{-11}, 3.059 \times 10^{-7}]$ respectively; (**a** and **d**) cumulative incidence (% of simulated population) vs age, and time to cancer vs (**b**) $log(n)$, (**e**) $log(p)$, and (**c** and **f**) $\Delta_g$; $\Delta_g = \frac{g_k - g0}{k}$ and $g_k = 0.007 \times \mu$. For all cases, $k = 5$.

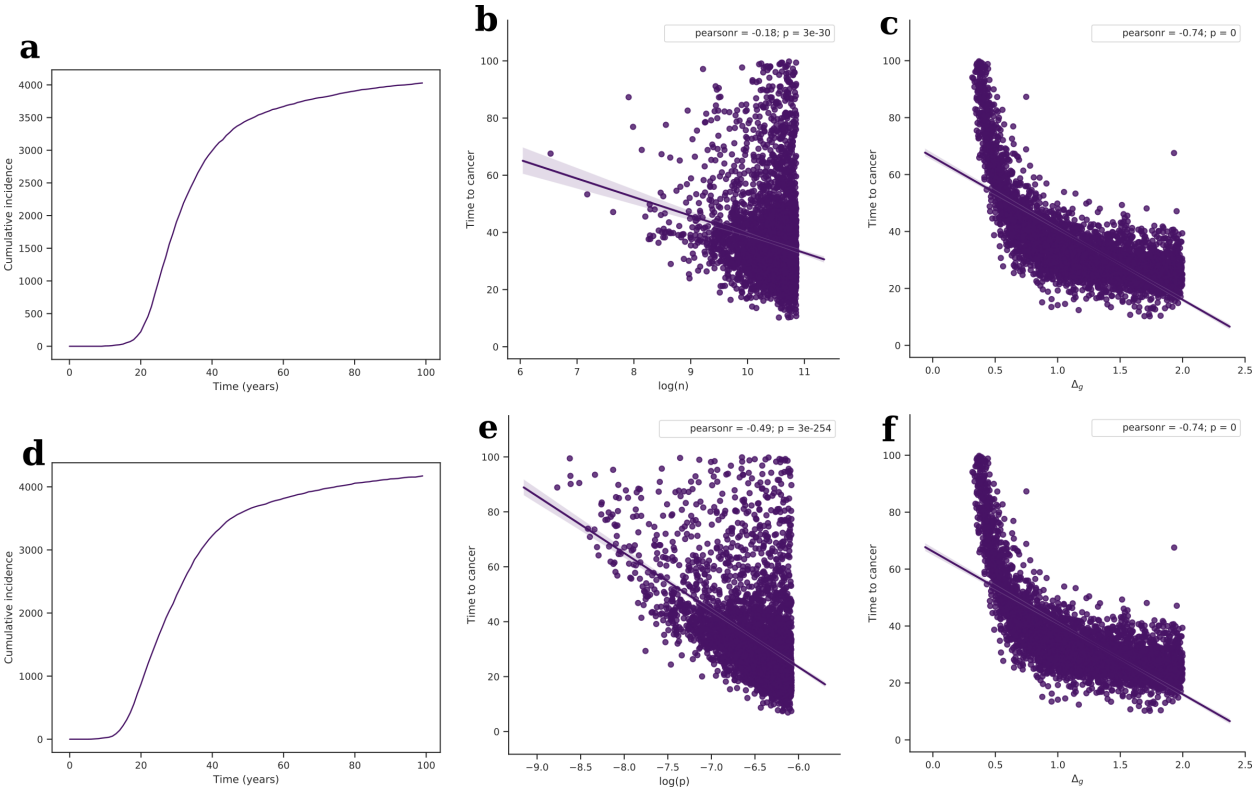



***Figure S6.3.2:*** **Alternate distributions of $g$: uniform**. $g$ modeled by a uniformly-distributed random variable, $\mu$ with range $[-10, 10]$, co-randomizing $n$ (top row) or $p$ (bottom row), with ranges as specified in Figure S6.3.1; (**a** and **d**) cumulative incidence for the simulated population vs age, and time to cancer vs (**b**) $log(n)$, (**e**) $log(p)$, and (**c** and **f**) $\Delta_g$; $\Delta_g = \frac{g_k - g_0}{k}$ and $g_k = 0.007 \times \mu$. For all cases, $k = 5$.

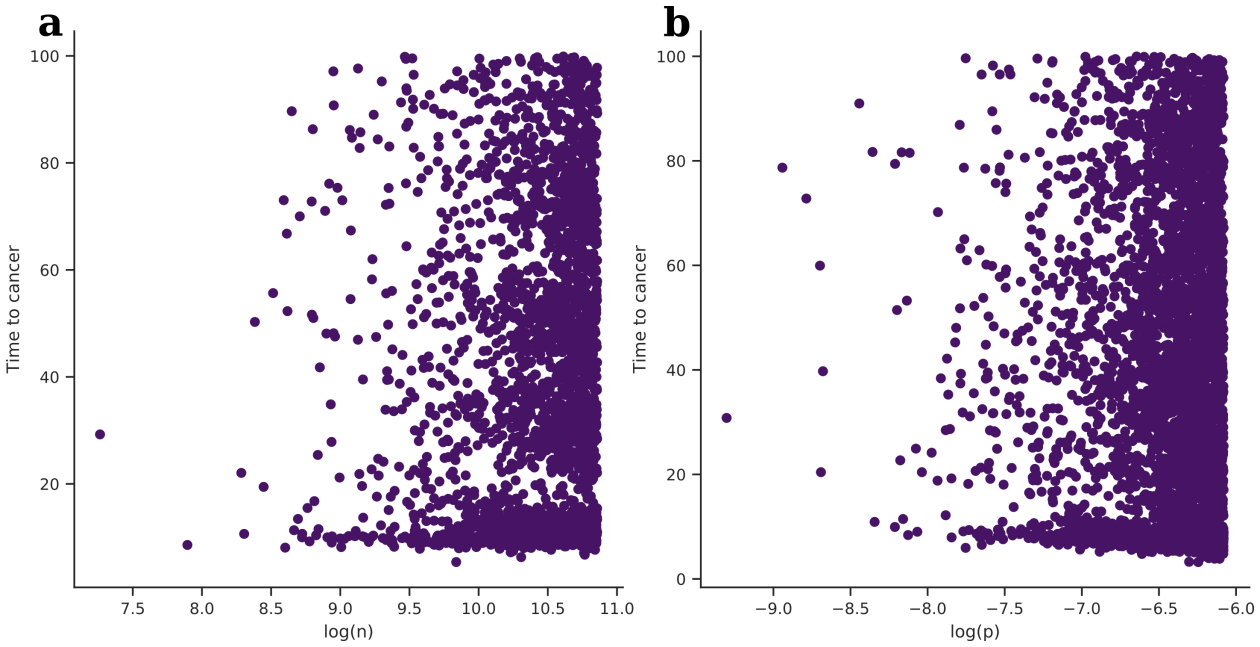



***Figure S6.3.3:*** **Effect of randomizing $k$ in the context-independent selection case**. The plots are time to cancer onset against $log(n)$ or $log(p)$, with $k$ randomized with (**a**) $n$, or (**b**) $p$; when time to cancer is pooled across values of $k$, the association between either $n$ or $p$ is practically non-existent. $k$, $n$ and $p$ were uniformly-distributed random variables with ranges $[0, 20]$, $[1.203 \times 10^{6}, 2.649 \times 10^{10}]$, and $[3.775 \times 10^{-11}, 3.059 \times 10^{-7}]$ respectively; for (**a**), $p = 5.603 \times 10^{-9}$, and for (**b**), $n = 1.785 \times 10^{8}$.

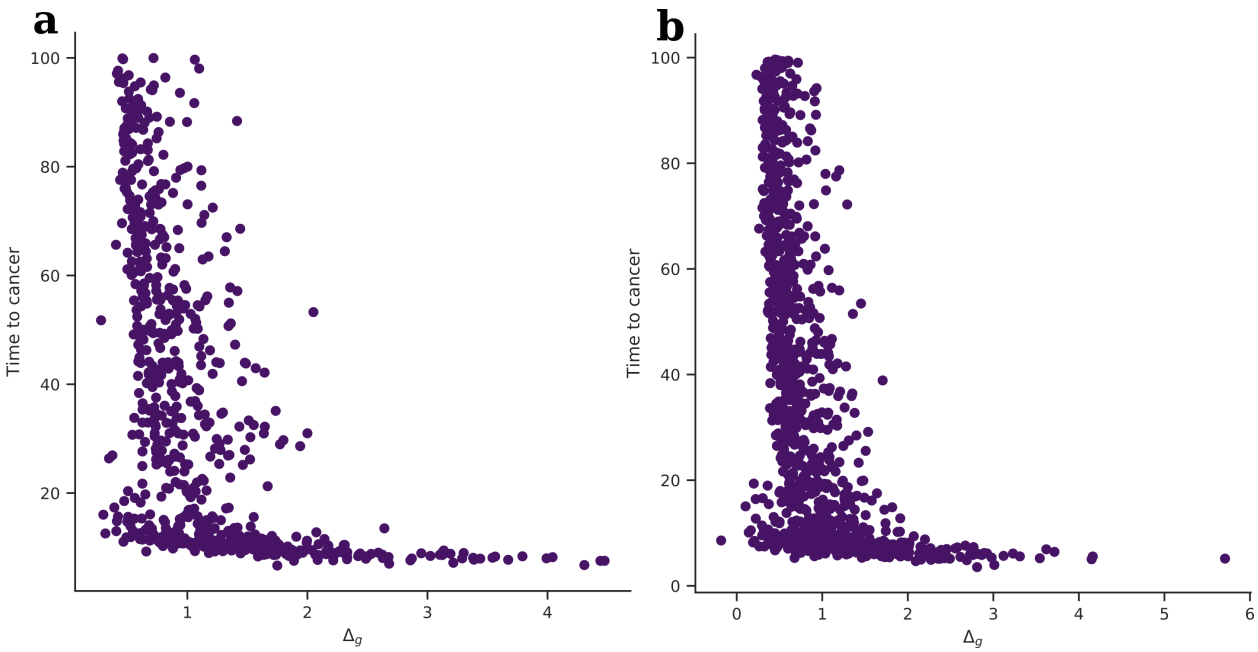



***Figure S6.3.4:*** **Effect of randomizing $k$ in the context-dependent selection case.** The plots are time to cancer onset against $\Delta_g = \frac{g_k - g_0}{k}$ and $g_k = 0.007 \times \mu$, also randomizing (**a**) $n$, or (**b**) $p$; $\Delta_g$ measures the rate of the growth rate progression in each individual, and $\mu$ is a normally-distributed random variable with $\overline{\mu} = 0$ and $\sigma = 3$. As opposed to $n$ and $p$, the association of $\Delta_g$ with time to cancer is less affected by $k$. When time to cancer is pooled across all $k$, $\Delta_g$'s effect on time to cancer appears distinctly non-linear. $k$, $n$ and $p$ were uniformly-distributed random variables with ranges $[0, 20]$, $[1.203 \times 10^6, 2.649 \times 10^{10}]$, and $[3.775 \times 10^{-11}, 3.059 \times 10^{-7}]$ respectively. For (**a**), $p = 5.603 \times 10^{-9}$, and for (**b**), $n = 1.785 \times 10^8$.

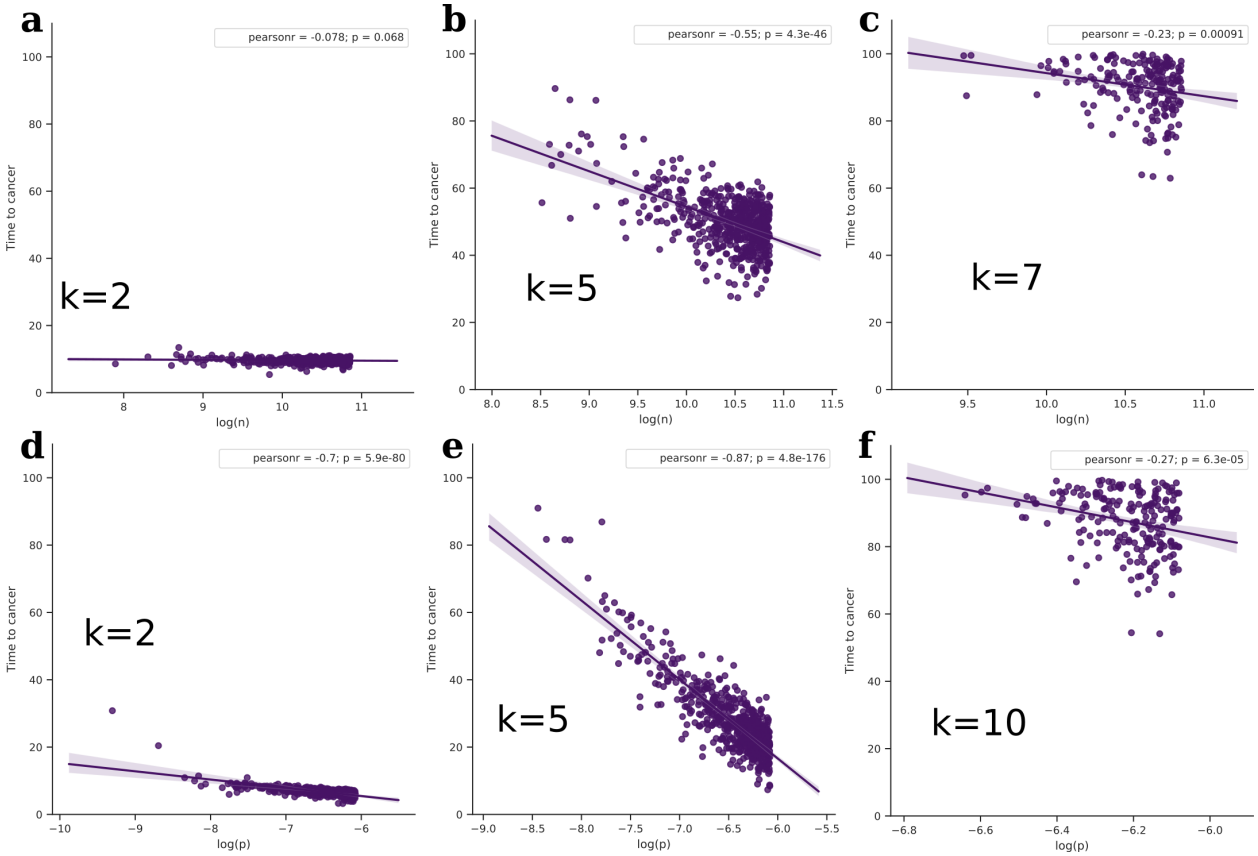



***Figure S6.3.5:*** **Effect of $k$ in the context-independent selection case**. The plots are time to cancer onset against $log(n)$ or $log(p)$, with $k$ randomized with (**a-c**) $n$, or (**d-f**) $p$; value of $k$ in the inset corresponds to the number of threshold oncogenic mutations assumed for the corresponding points. From (**a-c**), for higher threshold of oncogenic mutations, the effect of $n$ on time to cancer gets stronger, as shown by the improvement in the association. For small $k$ however, $n$ does not affect the age of cancer onset. On the other hand, $p$ has a strong effect on the time to cancer at every value of $k$ considered. $k$, $n$ and $p$ were uniformly-distributed random variables with ranges $[0, 20]$, $[1.203 \times 10^{6}, 2.649 \times 10^{10}]$, and $[3.775 \times 10^{-11}, 3.059 \times 10^{-7}]$ respectively. For (**a-c**), $p = 5.603 \times 10^{-9}$. For (**d-f**), $n = 1.785 \times 10^{8}$.

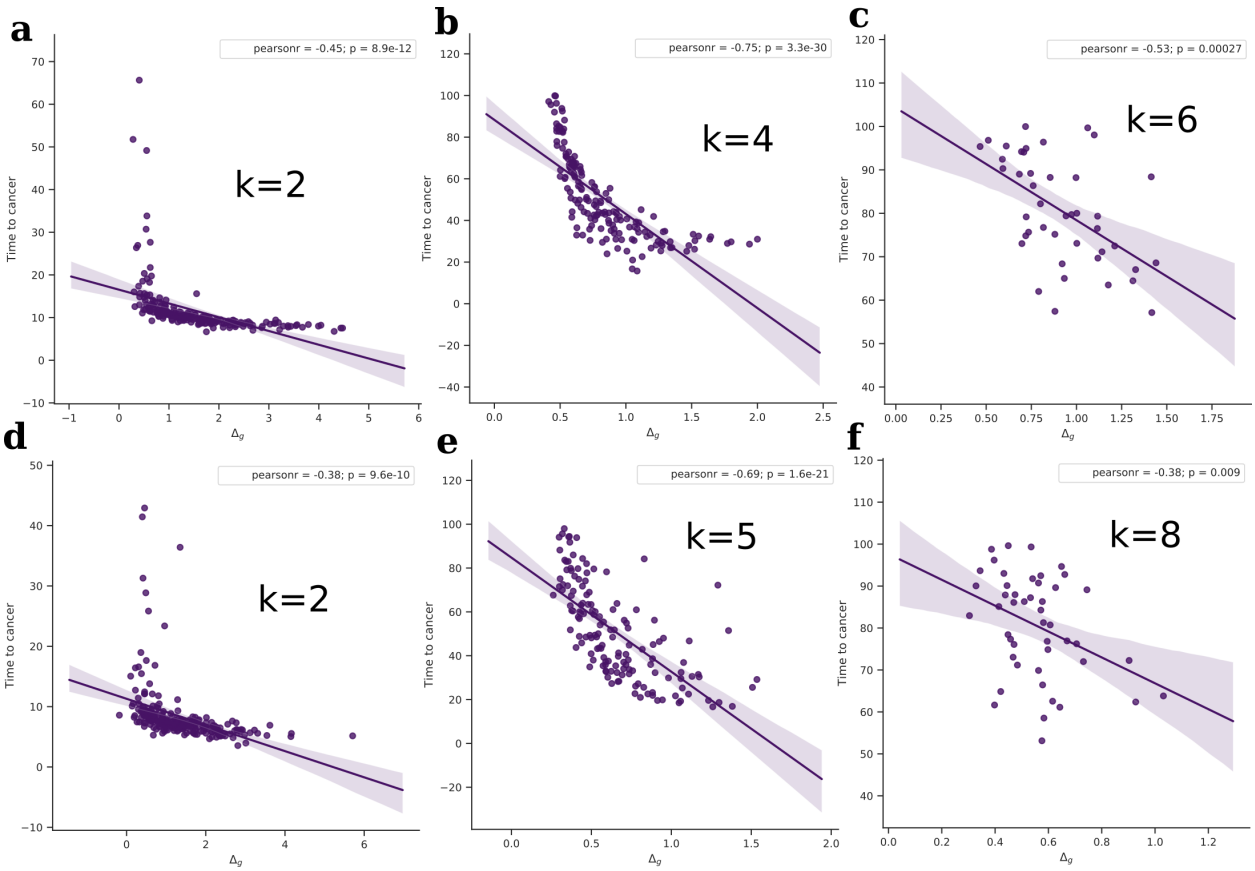



***Figure S6.3.6:*** **Effect of $k$ in the context-dependent selection case**. The plots are time to cancer onset against $\Delta_g = \frac{g_k - g_0}{k}$ as defined earlier, with $k$ randomized with (**a-c**) $n$, or (**d-f**) $p$; value of $k$ in the inset corresponds to the number of threshold oncogenic mutations assumed for the corresponding points. Compared to Supplementary Figure S6.3.5, $\Delta_g$ explains variance in time to cancer much better than either $n$ or $p$. This is true of both (**a-c**) when $n$ and $k$ are also randomized, and (**d-f**) when $p$ and $k$ are also randomized. The effect of $\Delta_g$ is nevertheless modulated by the required $k$, as reflected by the range of $\Delta_g$ for which cancer occurs; the scale of the x-axis across the figure is indicative of this effect. Ranges of $k$, $n$ and $p$, and the underlying distribution of $g$ are the same as in Supplementary Figure S6.3.5. For (**a-c**), $p = 5.603 \times 10^{-9}$. For (**d-f**), $n = 1.785 \times 10^{8}$.

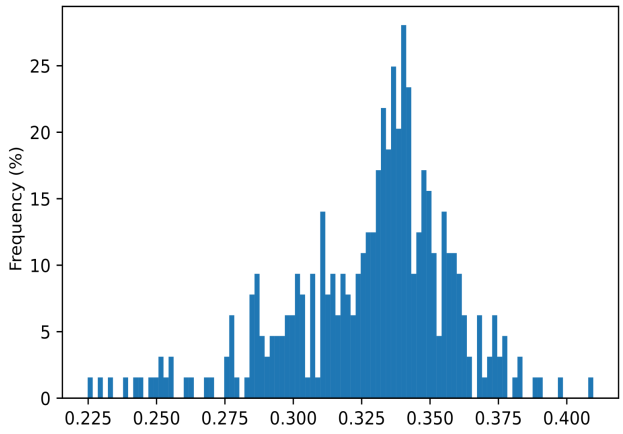



***Figure S6.3.7:*** **Distribution of slopes of the cancer risk-*lscd* regression**. Distribution of slopes of linear regression between *lscd* and cancer incidence on a log-log scale. Of the 423 datasets from corresponding national registries in the IARC database, 347 datasets had values of cumulative incidence rate for all 17 types of cancer considered. All slopes are substantially below unity, indicating non-linear relationships, with a median slope of 0.334, slopes in the range 0.225 to 0.410, and median intercept −13.630, ranged −15.630 to −11.424

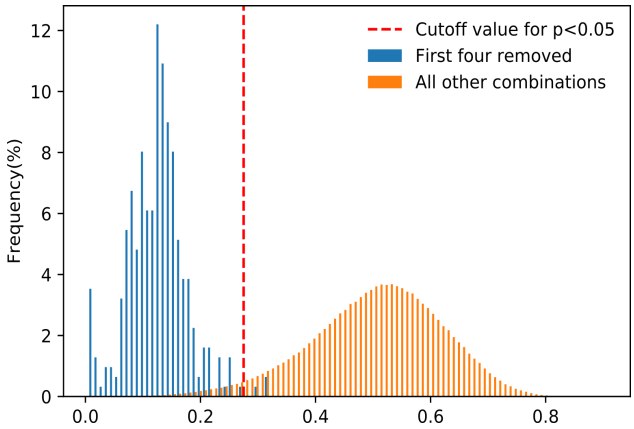



***Figure S6.3.8:*** **Distribution of Pearson's $r$ for reduced datasets**. The reduced datasets were obtained by removing any four data points from each sample of the IARC dataset, which have 17 data points correponding to the cancer types considered. Linear regression was performed against *lscd* for all possible combinations of removing four points from the sample, resulting in ${}^{17}\mathrm{C}_4 = 2380$ regression values for every sample, each now with 13 data points. The two distributions correspond to the values of Pearon's $r$ for the combination of the first four points removed, and those for all other combinations pooled respectively. The cutoff refers to the value of $r$ for which $p < 0.05$, given a sample size of 13 points. The former distribution lies predominantly below this cutoff, suggesting that the linear regression is largely driven only by the first four points. This puts into question the overall inference of linearity.

# References


Abegglen, L. M., A. F. Caulin, A. Chan, K. Lee, R. Robinson, M. S. Campbell, W. K. Kiso, D. L. Schmitt, P. J. Waddell, S. Bhaskara, S. T. Jensen, C. C. Maley, and J. D. Schiffman. 2015. Potential Mechanisms for Cancer Resistance in Elephants and Comparative Cellular Response to DNA Damage in Humans. Jama 314:1850. doi: 10.1001/jama.2015.13134.

Aktipis, C. A., and R. M. Nesse. 2013. Evolutionary foundations for cancer biology. Evolutionary Applications 6:144–159. doi: 10.1111/eva.12034.

Albanes, D., D. Y. Jones, A. Schatzkin, M. S. Micozzi, and P. R. Taylor. 1988. Adult Stature and Risk of Cancer. Cancer Research 48:1658–1662.

Albuquerque, T. A., L. Drummond do Val, A. Doherty, and J. P. de Magalhães. 2018. From humans to hydra: patterns of cancer across the tree of life. Biological Reviews 93:1715–1734. doi: 10.1111/brv.12415.

Andow, D. A. 2022. Dispersal kernel of green rice leafhopper estimated from truncated data. Population Ecology 65:111–120. doi: 10.1002/1438-390x.12141.

Archetti, M., D. A. Ferraro, and G. Christofori. 2015. Heterogeneity for IGF-II production maintained by public goods dynamics in neuroendocrine pancreatic cancer. Proc Natl Acad Sci U S A 112:1–6. doi: 10.1073/pnas.1414653112.

Armitage, P., and R. Doll. 1954. The Age Distribution of Cancer and a Multi-Stage Theory of Carcinogenesis. British journal of cancer 8:1–12.

Arnett, A. E., and N. J. Gotelli. 2003. Bergmann's rule in larval ant lions: Testing the starvation resistance hypothesis. Ecological Entomology 28:645–650. doi: 10.1111/j.1365-2311.2003.00554.x.

Arnold, P. A., P. Cassey, and C. R. White. 2017. Functional traits in red flour beetles: The dispersal phenotype is associated with leg length but not body size nor metabolic rate. Functional Ecology 31:653–661. doi: 10.1111/1365-2435.12772.

Asplen, M. K. 2018. Dispersal strategies in terrestrial insects. Current Opinion in Insect Science 27:16–20. doi: 10.1016/j.cois.2018.01.009.

B V, H. 2021. A theoretical study of the implications of resource competition for adaptive therapy of castration-resistant prostate cancer. MS Thesis. IISER Pune.

Bailey, M. H., C. Tokheim, E. Porta-Pardo, S. Sengupta, D. Bertrand, A. Weerasinghe, A. Colaprico, M. C. Wendl, J. Kim, B. Reardon, P. K.-S. Ng, K. J. Jeong, S. Cao, Z. Wang, J. Gao, Q. Gao, F. Wang, E. M. Liu, L. Mularoni, C. Rubio-Perez, N. Nagarajan, I. Cortés-Ciriano, D. C. Zhou, W.-W. Liang, J. M. Hess, V. D. Yellapantula, D. Tamborero, A. Gonzalez-Perez, C. Suphavilai, J. Y. Ko, E. Khurana, P. J. Park, E. M. V. Allen, H. Liang, and T. MC3 Working Group. 2018. Comprehensive Characterization of Cancer Driver Genes and Mutations. Cell 173:371–385.e18. doi: 10.1016/j.cell.2018.02.060.

Baines, C. B., J. M. J. Travis, S. J. McCauley, and G. Bocedi. 2020. Negative density-dependent dispersal emerges from the joint evolution of density- and body condition-dependent dispersal strategies. Evolution 74:2238–2249. doi: 10.1111/evo.14085.

Bairoch, A. 2018. The Cellosaurus, a Cell-Line Knowledge Resource. Journal of Biomolecular Techniques : JBT 29:25–38. doi: 10.7171/jbt.18-2902-002.

Bakker, K. 1959. Feeding Period, Growth, and Pupation in Larvae Ofdrosophila Melanogaster. Entomologia Experimentalis et Applicata 2:171–186. doi: 10.1111/j.1570-7458.1959.tb00432.x.

Basanta, D., J. G. Scott, M. N. Fishman, G. Ayala, S. W. Hayward, and A. R. A. Anderson. 2012. Investigating prostate cancer tumour–stroma interactions: clinical and biological insights from an evolutionary game. British Journal of Cancer 106:174–181. doi: 10.1038/bjc.2011.517.

Bateson, P., and K. N. Laland. 2013. Tinbergen's four questions: an appreciation and an update. Trends in Ecology & Evolution 28:712–718. doi: 10.1016/j.tree.2013.09.013.

Belikov, A. V. 2017. The number of key carcinogenic events can be predicted from cancer incidence. Scientific Reports 7. doi: 10.1038/s41598-017-12448-7.

Bell, J. R., D. A. Bohan, E. M. Shaw, and G. S. Weyman. 2005. Ballooning dispersal using silk: world fauna, phylogenies, genetics and models. Bulletin of Entomological Research 95:69–114. doi: 10.1079/ber2004350.

Bengtsson, B. O. 1978. Avoiding inbreeding: at what cost? Journal of Theoretical Biology 73:439–444. doi: 10.1016/0022-5193(78)90151-0.

Berwaerts, K., P. Aerts, and H. Van Dyck. 2006. On the sex-specific mechanisms of butterfly flight: Flight performance relative to flight morphology, wing kinematics, and sex in Pararge aegeria. Biological Journal of the Linnean Society 89:675–687. doi: 10.1111/j.1095-8312.2006.00699.x.

Bianchi, F. J. J. A., N. A. Schellhorn, and W. Van Der Werf. 2009. Predicting the time to colonization of the parasitoid Diadegma semiclausum: The importance of the shape of spatial dispersal kernels for biological control. Biological Control 50:267–274. doi: 10.1016/j.biocontrol.2009.04.014.

Birsoy, K., R. Possemato, F. K. Lorbeer, E. C. Bayraktar, P. Thiru, B. Yucel, T. Wang, W. W. Chen, C. B. Clish, and D. M. Sabatini. 2014. Metabolic determinants of cancer cell sensitivity to glucose limitation and biguanides. Nature 508:108–112. doi: 10.1038/nature13110.

Bitume, E. V., D. Bonte, O. Ronce, F. Bach, E. Flaven, I. Olivieri, and C. M. Nieberding. 2013. Density and genetic relatedness increase dispersal distance in a subsocial organism. Ecology Letters 16:430–437. doi: 10.1111/ele.12057.

Bitume, E. V., D. Bonte, O. Ronce, I. Olivieri, and C. M. Nieberding. 2014. Dispersal distance is influenced by parental and grand-parental density. Proceedings of the Royal Society B: Biological Sciences 281:20141061. doi: 10.1098/rspb.2014.1061.

Bjerknes, M. 1985. Assessment of the symmetry of stem-cell mitoses. Biophysical Journal 48:85–91. doi: 10.1016/s0006-3495(85)83762-0.

Blokzijl, F. 2016. Tissue-specific mutation accumulation in human adult stem cells during life. Nature 538. doi: 10.1038/nature19768.

Bonte, D., and M. Dahirel. 2017. Dispersal: a central and independent trait in life history. Oikos 126:472–479. doi: 10.1111/oik.03801.

Bonte, D., T. Hovestadt, and H.-J. Poethke. 2010. Evolution of dispersal polymorphism and local adaptation of dispersal distance in spatially structured landscapes. Oikos 119:560–566. doi: 10.1111/j.1600-0706.2009.17943.x.

Bonte, D., H. Van Dyck, J. M. Bullock, A. Coulon, M. Delgado, M. Gibbs, V. Lehouck, E. Matthysen, K. Mustin, M. Saastamoinen, N. Schtickzelle, V. M. Stevens, S. Vandewoestijne, M. Baguette, K. Barton, T. G. Benton, A. Chaput-Bardy, J. Clobert, C. Dytham, T. Hovestadt, C. M. Meier, S. C. F. Palmer, C. Turlure, and J. M. J. Travis. 2012. Costs of dispersal. Biological Reviews 87:290–312. doi: 10.1111/j.1469-185X.2011.00201.x.

Bowler, D. E., and T. G. Benton. 2005. Causes and consequences of animal dispersal strategies: relating individual behaviour to spatial dynamics. Biological Reviews 80:205–225. doi: 10.1017/s1464793104006645.

Brady-Nicholls, R., J. D. Nagy, T. A. Gerke, T. Zhang, A. Z. Wang, J. Zhang, R. A. Gatenby, and H. Enderling. 2020. Prostate-specific antigen dynamics predict individual responses to intermittent androgen deprivation. Nature Communications 11:1750. doi: 10.1038/s41467-020-15424-4.

Braendle, C., G. K. Davis, J. A. Brisson, and D. L. Stern. 2006. Wing dimorphism in aphids. Heredity 97:192–199. doi: 10.1038/sj.hdy.6800863.

Buoro, M., and S. M. Carlson. 2014. Life-history syndromes: Integrating dispersal through space and time. Ecology Letters 17:756–767. doi: 10.1111/ele.12275.

Cagan, A., A. Baez-Ortega, N. Brzozowska, F. Abascal, T. H. H. Coorens, M. A. Sanders, A. R. J. Lawson, L. M. R. Harvey, S. Bhosle, D. Jones, R. E. Alcantara, T. M. Butler, Y. Hooks, K. Roberts, E. Anderson, S. Lunn, E. Flach, S. Spiro, I. Januszczak, E. Wrigglesworth, H. Jenkins, T. Dallas, N. Masters, M. W. Perkins, R. Deaville, M. Druce, R. Bogeska, M. D. Milsom, B. Neumann, F. Gorman, F. Constantino-Casas, L. Peachey, D. Bochynska, E. S. J. Smith, M. Gerstung, P. J. Campbell, E. P. Murchison, M. R. Stratton, and I. Martincorena. 2022. Somatic mutation rates scale with lifespan across mammals. Nature 604:517–524. doi: 10.1038/s41586-022-04618-z.

Calabrese, P., and D. Shibata. 2010. A simple algebraic cancer equation: Calculating how cancers may arise with normal mutation rates. BMC Cancer 10:3. doi: 10.1186/1471-2407-10-3.

Calistro Alvarado, L. 2010. Population differences in the testosterone levels of young men are associated with prostate cancer disparities in older men. American Journal of Human Biology 22:449–455. doi: 10.1002/ajhb.21016.

Campos-Candela, A., M. Palmer, S. Balle, A. Álvarez, and J. Alós. 2019. A mechanistic theory of personality-dependent movement behaviour based on dynamic energy budgets. Ecology Letters 22:213–232. doi: 10.1111/ele.13187.

Cancer Society, A. 2023. Cancer Facts & Figures 2023. American Cancer Society .

Cao, L. 2010. Environmental and Genetic Activation of a Brain-Adipocyte BDNF/Lep- tin Axis Causes Cancer Remission and Inhibition. Cell 142. doi: 10.1016/j.cell.2010.05.029.

Carreira, S., A. Romanel, J. Goodall, E. Grist, R. Ferraldeschi, S. Miranda, D. Prandi, D. Lorente, J.-S. Frenel, C. Pezaro, A. Omlin, D. N. Rodrigues, P. Flohr, N. Tunariu, J. S. de Bono, F. Demichelis, and G. Attard. 2014. Tumor clone dynamics in lethal prostate cancer. Science Translational Medicine 6:254ra125–254ra125. doi: 10.1126/scitranslmed.3009448.

Caton, B. P., H. Fang, N. C. Manoukis, and G. R. Pallipparambil. 2022. Quantifying insect dispersal distances from trapping detections data to predict delimiting survey radii. Journal of Applied Entomology 146:203–216. doi: 10.1111/jen.12940.

Caulin, A. F., and C. C. Maley. 2011. Peto's Paradox: Evolution's prescription for cancer prevention. Trends in Ecology and Evolution 26. doi: 10.1016/j.tree.2011.01.002.

Cayuela, H., A. Besnard, J. Cote, M. Laporte, E. Bonnaire, J. Pichenot, N. Schtickzelle, A. Bellec, P. Joly, and J.-P. Léna. 2020. Anthropogenic disturbance drives dispersal syndromes, demography, and gene flow in amphibian populations. Ecological Monographs 90:e01406. doi: 10.1002/ecm.1406.

Chadov, B. F., N. B. Fedorova, and E. V. Chadova. 2015. Conditional mutations in Drosophila melanogaster: On the occasion of the 150th anniversary of G. Mendel's report in Brünn. Mutation Research/Reviews in Mutation Research 765:40–55. doi: 10.1016/j.mrrev.2015.06.001.

Chan, M. T., G. E. Lim, S. Skovsø, Y. H. C. Yang, T. Albrecht, E. U. Alejandro, C. A. Hoesli, J. M. Piret, G. L. Warnock, and J. D. Johnson. 2014. Effects of insulin on human pancreatic cancer progression modeled in vitro. BMC Cancer 14:814. doi: 10.1186/1471-2407-14-814.

Charlesworth, D., and B. Charlesworth. 1987. Inbreeding Depression and Its Evolutionary Consequences. Annual Review of Ecology and Systematics 18:237–268. doi: 10.1146/annurev.es.18.110187.001321.

Chen, S., A. Y. Lee, N. M. Bowens, R. Huber, and E. A. Kravitz. 2002. Fighting fruit flies: A model system for the study of aggression. Proceedings of the National Academy of Sciences 99:5664–5668. doi: 10.1073/pnas.082102599.

Chiu, J. C., K. H. Low, D. H. Pike, E. Yildirim, and I. Edery. 2010. Assaying Locomotor Activity to Study Circadian Rhythms and Sleep Parameters in Drosophila. JoVE (Journal of Visualized Experiments) page e2157. doi: 10.3791/2157.

Chuu, C.-P., J. M. Kokontis, R. A. Hiipakka, J. Fukuchi, H.-P. Lin, C.-Y. Lin, C. Huo, L.-C. Su, and S. Liao. 2011. Androgen suppresses proliferation of castration-resistant LNCaP 104-R2 prostate cancer cells through androgen receptor, Skp2, and c-Myc. Cancer Science 102:2022–2028. doi: 10.1111/j.1349-7006.2011.02043.x.

Clark, A. G., and W. W. Doane. 1983. Desiccation tolerance of the adipose60 mutant of Drosophila melanogaster. Hereditas 99:165–175. doi: 10.1111/j.1601-5223.1983.tb00888.x.

Clark, J. S., M. Lewis, and L. Horvath. 2001. Invasion by Extremes: Population Spread with Variation in Dispersal and Reproduction. The American Naturalist 157:537–554. doi: 10.1086/319934.

Clobert, J., M. Baguette, T. G. Benton, and J. M. Bullock. 2012. Dispersal Ecology and Evolution. Oxford University Press. ISBN: 978-0-19-960889-8.

Clobert, J., J.-F. Le Galliard, J. Cote, S. Meylan, and M. Massot. 2009. Informed dispersal, heterogeneity in animal dispersal syndromes and the dynamics of spatially structured populations. Ecology Letters 12:197–209. doi: 10.1111/j.1461-0248.2008.01267.x.

Cohen, J. 2013. Statistical Power Analysis for the Behavioral Sciences. Academic Press. ISBN: 978-1-4832-7648-9.

Coleman, E., A. E. Radix, W. P. Bouman, G. R. Brown, A. L. C. de Vries, M. B. Deutsch, R. Ettner, L. Fraser, M. Goodman, J. Green, A. B. Hancock, T. W. Johnson, D. H. Karasic, G. A. Knudson, S. F. Leibowitz, H. F. L. Meyer-Bahlburg, S. J. Monstrey, J. Motmans, L. Nahata, T. O. Nieder, S. L. Reisner, C. Richards, L. S. Schechter, V. Tangpricha, A. C. Tishelman, M. A. A. Van Trotsenburg, S. Winter, K. Ducheny, N. J. Adams, T. M. Adrián, L. R. Allen, D. Azul, H. Bagga, K. Başar, D. S. Bathory, J. J. Belinky, D. R. Berg, J. U. Berli, R. O. Bluebond-Langner, M.-B. Bouman, M. L. Bowers, P. J. Brassard, J. Byrne, L. Capitán, C. J. Cargill, J. M. Carswell, S. C. Chang, G. Chelvakumar, T. Corneil, K. B. Dalke, G. De Cuypere, E. de Vries, M. Den Heijer, A. H. Devor, C. Dhejne, A. D'Marco, E. K. Edmiston, L. Edwards-Leeper, R. Ehrbar, D. Ehrensaft, J. Eisfeld, E. Elaut, L. Erickson-Schroth, J. L. Feldman, A. D. Fisher, M. M. Garcia, L. Gijs, S. E. Green, B. P. Hall, T. L. D. Hardy, M. S. Irwig, L. A. Jacobs, A. C. Janssen, K. Johnson, D. T. Klink, B. P. C. Kreukels, L. E. Kuper, E. J. Kvach, M. A. Malouf, R. Massey, T. Mazur, C. McLachlan, S. D. Morrison, S. W. Mosser, P. M. Neira, U. Nygren, J. M. Oates, J. Obedin-Maliver, G. Pagkalos, J. Patton, N. Phanuphak, K. Rachlin, T. Reed, G. N. Rider, J. Ristori, S. Robbins-Cherry, S. A. Roberts, K. A. Rodriguez-Wallberg, S. M. Rosenthal, K. Sabir, J. D. Safer, A. I. Scheim, L. J. Seal, T. J. Sehoole, K. Spencer, C. St. Amand, T. D. Steensma, J. F. Strang, G. B. Taylor, K. Tilleman, G. G. T'Sjoen, L. N. Vala, N. M. Van Mello, J. F. Veale, J. A. Vencill, B. Vincent, L. M. Wesp, M. A. West, and J. Arcelus. 2022. Standards of Care for the Health of Transgender and Gender Diverse People, Version 8. International Journal of Transgender Health 23:S1–S259. doi: 10.1080/26895269.2022.2100644.

Cote, J., J. Clobert, T. Brodin, S. Fogarty, and A. Sih. 2010. Personality-dependent dispersal: Characterization, ontogeny and consequences for spatially structured populations. Philosophical Transactions of the Royal Society B: Biological Sciences 365:4065–4076. doi: 10.1098/rstb.2010.0176.

Crickmore, M. A., and L. B. Vosshall. 2013. Opposing Dopaminergic and GABAergic Neurons Control the Duration and Persistence of Copulation in Drosophila. Cell 155:881–893. doi: 10.1016/j.cell.2013.09.055.

Cronin, K. A., S. Scott, A. U. Firth, H. Sung, S. J. Henley, R. L. Sherman, R. L. Siegel, R. N. Anderson, B. A. Kohler, V. B. Benard, S. Negoita, C. Wiggins, W. G. Cance, and A. Jemal. 2022. Annual report to the nation on the status of cancer, part 1: National cancer statistics. Cancer 128:4251–4284. doi: 10.1002/cncr.34479.

Culig, Z., J. Hoffmann, M. Erdel, I. E. Eder, A. Hobisch, A. Hittmair, G. Bartsch, G. Utermann, M. R. Schneider, K. Parczyk, and H. Klocker. 1999. Switch from antagonist to agonist of the androgen receptor blocker bicalutamide is associated with prostate tumour progression in a new model system. British Journal of Cancer 81:242–251. doi: 10.1038/sj.bjc.6690684.

Cunningham, J. J., J. S. Brown, R. A. Gatenby, and K. Staňková. 2018. Optimal control to develop therapeutic strategies for metastatic castrate resistant prostate cancer. Journal of Theoretical Biology 459:67–78. doi: 10.1016/j.jtbi.2018.09.022.

Dahirel, M., L. Menut, and A. Ansart. 2021. Increased population density depresses activity but does not influence emigration in the snail Pomatias elegans. Journal of Zoology 313:172–181. doi: 10.1111/jzo.12846.

David, A. R., and M. R. Zimmerman. 2010. Cancer: an old disease, a new disease or something in between? Nature Reviews Cancer 10:728–733. doi: 10.1038/nrc2914.

De, S. 2022. Studying the scaling of cell division and mutation accumulation with body size using an individual-based model of organismal development. MS Thesis. IISER Pune.

DeGregori, J. 2018. Adaptive Oncogenesis: A New Understanding of How Cancer Evolves inside Us. Harvard University Press, Cambridge, MA. ISBN: 978-0-674-54539-7.

Demeter, M., I. Derényi, and G. J. Szöllősi. 2022. Trade-off between reducing mutational accumulation and increasing commitment to differentiation determines tissue organization. Nature Communications 13:1666. doi: 10.1038/s41467-022-29004-1.

Denno, R. F., G. K. Roderick, M. A. Peterson, A. F. Huberty, H. G. Dobel, M. D. Eubanks, J. E. Losey, and G. A. Langellotto. 1996. Habitat Persistence Underlies Intraspecific Variation in the Dispersal Strategies of Planthoppers. Ecological Monographs 66:389–408. doi: 10.2307/2963487.

Derényi, I., and G. J. Szöllősi. 2017. Hierarchical tissue organization as a general mechanism to limit the accumulation of somatic mutations. Nature Communications 8:14545. doi: 10.1038/ncomms14545.

Derényi, I., M. C. Demeter, and G. J. Szöllősi. 2020. Cancer Risk and the Somatic Cell Lineage Tree. URL: https://www.biorxiv.org/content/10.1101/2020.07.13.201004v1.

Dickson, R. B., and M. E. Lippman. 1987. Estrogenic Regulation of Growth and Polypeptide Growth Factor Secretion in Human Breast Carcinoma. Endocrine Reviews 8. doi: 10.1210/edrv-8-1-29.

DiLeo, M. F., E. Nonaka, A. Husby, and M. Saastamoinen. 2022. Effects of environment and genotype on dispersal differ across departure, transfer and settlement in a butterfly metapopulation. Proceedings of the Royal Society B: Biological Sciences 289:20220322. doi: 10.1098/rspb.2022.0322.

Dillard, P. R., M.-F. Lin, and S. A. Khan. 2008. Androgen-independent prostate cancer cells acquire the complete steroidogenic potential of synthesizing testosterone from cholesterol. Molecular and Cellular Endocrinology 295:115–120. doi: 10.1016/j.mce.2008.08.013.

Dobzhansky, T., and S. Wright. 1943. Genetics of natural populations x: Dispersion rates in *Drosophila Pseudoobscura*. Genetics 28:304–340. doi: 10.1093/genetics/28.4.304.

Druso, J. E., and C. Fischbach. 2018. Biophysical Properties of Extracellular Matrix: Linking Obesity and Cancer. Trends in Cancer 4. doi: 10.1016/j.trecan.2018.02.001.

Duckworth, R. A., and A. V. Badyaev. 2007. Coupling of dispersal and aggression facilitates the rapid range expansion of a passerine bird. Proceedings of the National Academy of Sciences 104:15017–15022. doi: 10.1073/pnas.0706174104.

Edelaar, P., A. M. Siepielski, and J. Clobert. 2008. Matching Habitat Choice Causes Directed Gene Flow: A Neglected Dimension in Evolution and Ecology. Evolution 62:2462–2472. doi: 10.1111/j.1558-5646.2008.00459.x.

Entling, M. H., K. Stämpfli, and O. Ovaskainen. 2011. Increased propensity for aerial dispersal in disturbed habitats due to intraspecific variation and species turnover. Oikos 120:1099–1109. doi: 10.1111/j.1600-0706.2010.19186.x.

Erten, E. Y., and H. Kokko. 2020. From zygote to a multicellular soma: Body size affects optimal growth strategies under cancer risk. Evolutionary Applications 13:1593–1604. doi: 10.1111/eva.12969.

Eyre-Walker, A., and P. D. Keightley. 2007. The distribution of fitness effects of new mutations. Nature Reviews Genetics 8:610–618. doi: 10.1038/nrg2146.

Eyre-Walker, A., M. Woolfit, and T. Phelps. 2006. The Distribution of Fitness Effects of New Deleterious Amino Acid Mutations in Humans. Genetics 173:891–900. doi: 10.1534/genetics.106.057570.

Farrokhian, N., J. Maltas, M. Dinh, A. Durmaz, P. Ellsworth, M. Hitomi, E. McClure, A. Marusyk, A. Kaznatcheev, and J. G. Scott. 2022. Measuring competitive exclusion in non–small cell lung cancer. Science Advances 8:eabm7212. doi: 10.1126/sciadv.abm7212.

Fehr, D., H. Veeraraghavan, A. Wibmer, T. Gondo, K. Matsumoto, H. A. Vargas, E. Sala, H. Hricak, and J. O. Deasy. 2015. Automatic classification of prostate cancer Gleason scores from multiparametric magnetic resonance images. Proceedings of the National Academy of Sciences 112:E6265–e6273. doi: 10.1073/pnas.1505935112.

Ferlay, J., M. Colombet, I. Soerjomataram, D. M. Parkin, M. Piñeros, A. Znaor, and F. Bray. 2021. Cancer statistics for the year 2020: An overview. International Journal of Cancer 149:778–789. doi: 10.1002/ijc.33588.

Fontana, F., M. Anselmi, and P. Limonta. 2022. Molecular mechanisms and genetic alterations in prostate cancer: From diagnosis to targeted therapy. Cancer Letters 534:215619. doi: 10.1016/j.canlet.2022.215619.

Fragoso, J. M. V., K. M. Silvius, and J. A. Correa. 2003. Long-Distance Seed Dispersal by Tapirs Increases Seed Survival and Aggregates Tropical Trees. Ecology 84:1998–2006. doi: 10.1890/01-0621.

Frank, S. A. 2007. Dynamics of Cancer. Incidence, Inheritance, and Evolution. Princeton University Press. ISBN: 978-0-691-13366-9.

Fronhofer, E. A., H. Joachim Poethke, and U. Dieckmann. 2015. Evolution of dispersal distance: Maternal investment leads to bimodal dispersal kernels. Journal of Theoretical Biology 365:270–279. doi: 10.1016/j.jtbi.2014.10.024.

Fronhofer, E. A., J. M. Stelz, E. Lutz, H. J. Poethke, and D. Bonte. 2014. Spatially correlated extinctions select for less emigration but larger dispersal distances in the spider mite Tetranychus urticae. Evolution 68:1838–1844. doi: 10.1111/evo.12339.

Gallaher, J. A., P. M. Enriquez-Navas, K. A. Luddy, R. A. Gatenby, and A. R. Anderson. 2018. Spatial Heterogeneity and Evolutionary Dynamics Modulate Time to Recurrence in Continuous and Adaptive Cancer Therapies. Cancer Research 78:2127–2139. doi: 10.1158/0008-5472.can-17-2649.

Gatenby, R. A., A. S. Silva, R. J. Gillies, and B. R. Frieden. 2009. Adaptive therapy. Cancer Research 69:4894–4903. doi: 10.1158/0008-5472.can-08-3658.

Gedye, C., and V. Navani. 2022. Find the path of least resistance: Adaptive therapy to delay treatment failure and improve outcomes. Biochimica et Biophysica Acta (BBA) - Reviews on Cancer 1877:188681. doi: 10.1016/j.bbcan.2022.188681.

Gerlinger, M., N. McGranahan, S. M. Dewhurst, R. A. Burrell, I. Tomlinson, and C. Swanton. 2014. Cancer: Evolution Within a Lifetime. Annual Review of Genetics 48:215–236. doi: 10.1146/annurev-genet-120213-092314.

Gerstung, M., C. Jolly, I. Leshchiner, S. C. Dentro, S. Gonzalez, D. Rosebrock, T. J. Mitchell, Y. Rubanova, P. Anur, K. Yu, M. Tarabichi, A. Deshwar, J. Wintersinger, K. Kleinheinz, I. Vázquez-García, K. Haase, L. Jerman, S. Sengupta, G. Macintyre, S. Malikic, N. Donmez, D. G. Livitz, M. Cmero, J. Demeulemeester, S. Schumacher, Y. Fan, X. Yao, J. Lee, M. Schlesner, P. C. Boutros, D. D. Bowtell, H. Zhu, G. Getz, M. Imielinski, R. Beroukhim, S. C. Sahinalp, Y. Ji, M. Peifer, F. Markowetz, V. Mustonen, K. Yuan, W. Wang, Q. D. Morris, P. T. Spellman, D. C. Wedge, and P. Van Loo. 2020. The evolutionary history of 2,658 cancers. Nature 578:122–128. doi: 10.1038/s41586-019-1907-7.

Ghaffarizadeh, A., R. Heiland, S. H. Friedman, S. M. Mumenthaler, and P. Macklin. 2018. PhysiCell: An open source physics-based cell simulator for 3-D multicellular systems. PLoS Computational Biology 14:e1005991. doi: 10.1371/journal.pcbi.1005991.

Gillies, R. J., P. E. Kinahan, and H. Hricak. 2016. Radiomics: Images are more than pictures, they are data. Radiology 278:563–577. doi: 10.1148/radiol.2015151169.

Gorbunova, V., A. Seluanov, Z. Zhang, V. N. Gladyshev, and J. Vijg. 2014. Comparative genetics of longevity and cancer: insights from long-lived rodents. Nature Reviews Genetics 15:531–540. doi: 10.1038/nrg3728.

Gordetsky, J., and J. Epstein. 2016. Grading of prostatic adenocarcinoma: current state and prognostic implications. Diagnostic Pathology 11:25. doi: 10.1186/s13000-016-0478-2.

Grajzel, D., I. Derényi, and G. J. Szöllősi. 2020. A compartment size-dependent selective threshold limits mutation accumulation in hierarchical tissues. Proceedings of the National Academy of Sciences 117:1606–1611. doi: 10.1073/pnas.1913104117.

Grasso, C. S., Y.-M. Wu, D. R. Robinson, X. Cao, S. M. Dhanasekaran, A. P. Khan, M. J. Quist, X. Jing, R. J. Lonigro, J. C. Brenner, I. A. Asangani, B. Ateeq, S. Y. Chun, J. Siddiqui, L. Sam, M. Anstett, R. Mehra, J. R. Prensner, N. Palanisamy, G. A. Ryslik, F. Vandin, B. J. Raphael, L. P. Kunju, D. R. Rhodes, K. J. Pienta, A. M. Chinnaiyan, and S. A. Tomlins. 2012. The mutational landscape of lethal castration-resistant prostate cancer. Nature 487:239–243. doi: 10.1038/nature11125.

Graves, J. L., E. C. Toolson, C. Jeong, L. N. Vu, and M. R. Rose. 1992. Desiccation, Flight, Glycogen, and Postponed Senescence in Drosophila melanogaster. Physiological Zoology 65:268–286. doi: 10.1086/physzool.65.2.30158253.

Green, J., B. J. Cairns, D. Casabonne, F. L. Wright, G. Reeves, and V. Beral. 2011. Height and cancer incidence in the Million Women Study: Prospective cohort, and meta-analysis of prospective studies of height and total cancer risk. The Lancet Oncology 12:785–794. doi: 10.1016/s1470-2045(11)70154-1.

Greenwood, P. J. 1980. Mating systems, philopatry and dispersal in birds and mammals. Animal Behaviour 28:1140–1162. doi: 10.1016/s0003-3472(80)80103-5.

Gregory, C. W., R. T. Johnson, J. L. Mohler, F. S. French, and E. M. Wilson. 2001. Androgen receptor stabilization in recurrent prostate cancer is associated with hypersensitivity to low androgen. Cancer Research 61:2892–2898.

Grover, J. P. 1997. Resource Competition. Springer Science & Business Media. ISBN: 978-0-412-74930-8.

Guerra, P. A., and G. S. Pollack. 2010. Colonists and desperadoes: Different fighting strategies in wing-dimorphic male Texas field crickets. Animal Behaviour 79:1087–1093. doi: 10.1016/j.anbehav.2010.02.002.

Guttal, V., F. Bartumeus, G. Hartvigsen, and A. L. Nevai. 2011. Retention Time Variability as a Mechanism for Animal Mediated Long-Distance Dispersal. Plos One 6:e28447. doi: 10.1371/journal.pone.0028447.

Hail, N., P. Chen, and L. R. Bushman. 2010. Teriflunomide (Leflunomide) Promotes Cytostatic, Antioxidant, and Apoptotic Effects in Transformed Prostate Epithelial Cells: Evidence Supporting a Role for Teriflunomide in Prostate Cancer Chemoprevention. Neoplasia 12:464–475. doi: 10.1593/neo.10168.

Hamilton, W. D., and R. M. May. 1977. Dispersal in stable habitats. Nature 269:578–581. doi: 10.1038/269578a0.

Hanahan, D., and L. M. Coussens. 2012. Accessories to the Crime: Functions of Cells Recruited to the Tumor Microenvironment. Cancer Cell 21. doi: 10.1016/j.ccr.2012.02.022.

Hanahan, D., and R. Weinberg. 2000. The hallmarks of cancer. Cell 100:57–70.

Hanahan, D., and R. A. Weinberg. 2011. Hallmarks of Cancer: The Next Generation. Cell 144:646–674. doi: 10.1016/j.cell.2011.02.013.

Hansen, E., and A. F. Read. 2020. Cancer therapy: Attempt cure or manage drug resistance? Evolutionary Applications 13:1660–1672. doi: 10.1111/eva.12994.

Hansen, E., R. J. Woods, and A. F. Read. 2017. How to Use a Chemotherapeutic Agent When Resistance to It Threatens the Patient. PLOS Biology 15:e2001110. doi: 10.1371/journal.pbio.2001110.

Hansen, R., and M. J. Bissell. 2000. Tissue architecture and breast cancer: The role of extracellu- lar matrix and steroid hormones. Endocrine Related Cancer 7. doi: 10.1677/erc.0.0070095.

Hanski, I., M. Saastamoinen, and O. Ovaskainen. 2006. Dispersal-related life-history trade-offs in a butterfly metapopulation. Journal of Animal Ecology 75:91–100. doi: 10.1111/j.1365-2656.2005.01024.x.

Hao, D., L. Wang, and L.-j. Di. 2016. Distinct mutation accumulation rates among tissues determine the variation in cancer risk. Scientific Reports 6:19458. doi: 10.1038/srep19458.

Harding, C., F. Pompei, and R. Wilson. 2012. Peak and decline in cancer incidence, mortality, and prevalence at old ages. Cancer 118:1371–1386. doi: 10.1002/cncr.26376.

Haslam, S. Z., and T. L. Woodward. 2001. Tumour-stromal interactions Reciprocal regulation of extracellular matrix proteins and ovarian steroid activity in the mammary gland. Breast Cancer Research 3. doi: 10.1186/bcr324.

Hembree, W. C., P. T. Cohen-Kettenis, L. Gooren, S. E. Hannema, W. J. Meyer, M. H. Murad, S. M. Rosenthal, J. D. Safer, V. Tangpricha, and G. G. T'Sjoen. 2017. Endocrine Treatment of Gender-Dysphoric/Gender-Incongruent Persons: An Endocrine Society Clinical Practice Guideline. The Journal of Clinical Endocrinology and Metabolism 102:3869–3903. doi: 10.1210/jc.2017-01658.

Hendricks, J. C., S. M. Finn, K. A. Panckeri, J. Chavkin, J. A. Williams, A. Sehgal, and A. I. Pack. 2000. Rest in Drosophila Is a Sleep-like State. Neuron 25:129–138. doi: 10.1016/s0896-6273(00)80877-6.

Henry, C. J. 2015. Aging-associated inflammation promotes selection for adaptive oncogenic events in B cell progenitors. The Journal of Clinical Investigation 125. doi: 10.1172/jci83024.

Henry, C. J., A. Marusyk, V. Zaberezhnyy, B. Adane, and J. DeGregori. 2010. Declining lymphoid progenitor fitness promotes aging-associated leukemogenesis. Proceedings of the National Academy of Sciences 107. doi: 10.1073/pnas.1005486107.

Higgins, S. I., and M. L. Cain. 2002. Spatially realistic plant metapopulation models and the colonization–competition trade-off. Journal of Ecology 90:616–626. doi: 10.1046/j.1365-2745.2002.00694.x.

Hill, J., C. Thomas, and O. Lewis. 1996. Effects of Habitat Patch Size and Isolation on Dispersal by Hesperia comma Butterflies: Implications for Metapopulation Structure. The Journal of Animal Ecology 65:725. doi: 10.2307/5671.

Hindersin, L., B. Werner, D. Dingli, and A. Traulsen. 2016. Should tissue structure suppress or amplify selection to minimize cancer risk? Biology Direct 11:1–11. doi: 10.1186/s13062-016-0140-7/figures/3.

Hochberg, M. E., and R. J. Noble. 2017. A framework for how environment contributes to cancer risk. Ecology Letters 20:117–134. doi: 10.1111/ele.12726.

Honěk, A. 1993. Intraspecific Variation in Body Size and Fecundity in Insects: A General Relationship. Oikos 66:483–492. doi: 10.2307/3544943.

Hovestadt, T., B. Binzenhöfer, P. Nowicki, and J. Settele. 2011. Do all inter-patch movements represent dispersal? A mixed kernel study of butterfly mobility in fragmented landscapes. Journal of Animal Ecology 80:1070–1077. doi: 10.1111/j.1365-2656.2011.01848.x.

Hovestadt, T., O. Mitesser, and H.-J. Poethke. 2014. Gender-Specific Emigration Decisions Sensitive to Local Male and Female Density. The American Naturalist 184:38–51. doi: 10.1086/676524.

Hughes, C. L., J. K. Hill, and C. Dytham. 2003. Evolutionary trade-offs between reproduction and dispersal in populations at expanding range boundaries. Proceedings of the Royal Society of London. Series B: Biological Sciences 270:S147–s150. doi: 10.1098/rsbl.2003.0049.

Ibrahim-Hashim, A., M. Robertson-Tessi, P. M. Enriquez-Navas, M. Damaghi, Y. Balagurunathan, J. W. Wojtkowiak, S. Russell, K. Yoonseok, M. C. Lloyd, M. M. Bui, J. S. Brown, A. R. Anderson, R. J. Gillies, and R. A. Gatenby. 2017. Defining cancer subpopulations by adaptive strategies rather than molecular properties provides novel insights into intratumoral evolution. Cancer Research 77:2242–2254. doi: 10.1158/0008-5472.CAN-16-2844.

Iyengar, N. M., A. Gucalp, A. J. Dannenberg, and C. A. Hudis. 2016. Obesity and Cancer Mechanisms: Tumor Microenvironment and Inflammation. Journal of Clinical Oncology 34. doi: 10.1200/jco.2016.67.4283.

Jain, H. V., S. K. Clinton, A. Bhinder, and A. Friedman. 2011. Mathematical modeling of prostate cancer progression in response to androgen ablation therapy. Proceedings of the National Academy of Sciences 108:19701–19706. doi: 10.1073/pnas.1115750108.

Jenkins, D. G., C. R. Brescacin, C. V. Duxbury, J. A. Elliott, J. A. Evans, K. R. Grablow, M. Hillegass, B. N. Lyon, G. A. Metzger, M. L. Olandese, D. Pepe, G. A. Silvers, H. N. Suresch, T. N. Thompson, C. M. Trexler, G. E. Williams, N. C. Williams, and S. E. Williams. 2007. Does size matter for dispersal distance? Global Ecology and Biogeography 16:415–425. doi: 10.1111/j.1466-8238.2007.00312.x.

Jones, K. L., V. A. Shegelski, N. G. Marculis, A. N. Wijerathna, and M. L. Evenden. 2019. Factors influencing dispersal by flight in bark beetles (Coleoptera: Curculionidae: Scolytinae): From genes to landscapes. Canadian Journal of Forest Research 49:1024–1041. doi: 10.1139/cjfr-2018-0304.

Joshi, S., R. Mishra, M. Kulkarni, D. A. Kelkar, K. Harikrishnan, P. Vaid, G. V, J. John, S. Tamhane, A. Deshpande, S. Badve, Chintamani, R. Mehrotra, L. Shashidhara, A. Kothari, and C. Koppiker. 2023. Proceedings of the 3rd Indian Cancer Genome Atlas Conference 2022: Biobanking to Omics: Collecting the Global Experience. JCO Global Oncology page e2200176. doi: 10.1200/go.22.00176.

Kareva, I. 2011. What Can Ecology Teach Us About Cancer? Translational Oncology 4:266–270. doi: 10.1593/tlo.11154.

Kareva, I., and J. S. Brown. 2021. Estrogen as an Essential Resource and the Coexistence of ER+ and ER– Cancer Cells. Frontiers in Ecology and Evolution 9.

Kareva, I., B. Morin, and C. Castillo-Chavez. 2015. Resource Consumption, Sustainability, and Cancer. Bulletin of Mathematical Biology 77:319–338. doi: 10.1007/s11538-014-9983-1/figures/4.

Karkach, A. S. 2006. Trajectories and models of individual growth. Demographic Research 15:347–400.

Kautz, M., R. Schopf, and M. A. Imron. 2014. Individual traits as drivers of spatial dispersal and infestation patterns in a host–bark beetle system. Ecological Modelling 273:264–276. doi: 10.1016/j.ecolmodel.2013.11.022.

Kayser, M. S., Z. Yue, and A. Sehgal. 2014. A Critical Period of Sleep for Development of Courtship Circuitry and Behavior in Drosophila. Science 344:269–274. doi: 10.1126/science.1250553.

Kempenaers, B., and M. Valcu. 2017. Breeding site sampling across the Arctic by individual males of a polygynous shorebird. Nature 541:528–531. doi: 10.1038/nature20813.

Knapp, M. 2016. Relative Importance of Sex, Pre-Starvation Body Mass and Structural Body Size in the Determination of Exceptional Starvation Resistance of Anchomenus dorsalis (Coleoptera: Carabidae). Plos One 11:e0151459. doi: 10.1371/journal.pone.0151459.

Knudson, A. G. 1971. Mutation and Cancer: Statistical Study of Retinoblastoma. Proceedings of the National Academy of Sciences 68:820–823. doi: 10.1073/pnas.68.4.820.

Koch, F. H., D. Yemshanov, R. D. Magarey, and W. D. Smith. 2012. Dispersal of Invasive Forest Insects via Recreational Firewood: A Quantitative Analysis. Journal of Economic Entomology 105:438–450. doi: 10.1603/ec11270.

Kokko, H., and A. López-Sepulcre. 2006. From Individual Dispersal to Species Ranges: Perspectives for a Changing World. Science 313:789–791. doi: 10.1126/science.1128566.

Kolss, M., R. K. Vijendravarma, G. Schwaller, and T. J. Kawecki. 2009. Life-history consequences of adaptation to larval nutritional stress in drosophila. Evolution 63:2389–2401. doi: 10.1111/j.1558-5646.2009.00718.x.

Korsten, P., T. van Overveld, F. Adriaensen, and E. Matthysen. 2013. Genetic integration of local dispersal and exploratory behaviour in a wild bird. Nature Communications 4:2362. doi: 10.1038/ncomms3362.

Kot, M., M. A. Lewis, and P. Van Den Driessche. 1996. Dispersal Data and the Spread of Invading Organisms. Source: Ecology 77:2027–2042.

Kotler, B. P., and J. S. Brown. 2020. Cancer Community Ecology. Cancer Control 27:1073274820951776. doi: 10.1177/1073274820951776.

Kozlowski, J., M. Czarnolełęski, A. François-Krassowska, S. MacIak, and T. Pis. 2010. Cell size is positively correlated between different tissues in passerine birds and amphibians, but not necessarily in mammals. Biology Letters 6:792–796. doi: 10.1098/rsbl.2010.0288.

Kravitz, E. A., and M. d. l. P. Fernandez. 2015. Aggression in Drosophila. Behavioral Neuroscience 129:549–563. doi: 10.1037/bne0000089.

Kume, K., S. Kume, S. K. Park, J. Hirsh, and F. R. Jackson. 2005. Dopamine Is a Regulator of Arousal in the Fruit Fly. Journal of Neuroscience 25:7377–7384. doi: 10.1523/jneurosci.2048-05.2005.

Kuznetsova, A., P. B. Brockhoff, and R. H. B. Christensen. 2017. lmerTest package: Tests in linear mixed effects models. Journal of Statistical Software 82:1–26. doi: 10.18637/jss.v082.i13.

Lancaster, J., and B. J. Downes. 2017. Dispersal traits may reflect dispersal distances, but dispersers may not connect populations demographically. Oecologia 184:171–182. doi: 10.1007/s00442-017-3856-x.

Latty, T. M., and M. L. Reid. 2010. Who goes first? Condition and danger dependent pioneering in a group-living bark beetle (Dendroctonus ponderosae). Behavioral Ecology and Sociobiology 64:639–646. doi: 10.1007/s00265-009-0881-8.

Lee, K. P., S. J. Simpson, F. J. Clissold, R. Brooks, J. W. O. Ballard, P. W. Taylor, N. Soran, and D. Raubenheimer. 2008. Lifespan and reproduction in Drosophila: New insights from nutritional geometry. Proceedings of the National Academy of Sciences of the United States of America 105:2498–503. doi: 10.1073/pnas.0710787105.

Lee, W.-C., and C. A. Micchelli. 2013. Development and Characterization of a Chemically Defined Food for Drosophila. Plos One 8:e67308. doi: 10.1371/journal.pone.0067308.

Lenth, R. V. 2023. emmeans: Estimated Marginal Means, aka Least-Squares Means.

Levin, S. A., S. R. Carpenter, H. C. J. Godfray, A. P. Kinzig, M. Loreau, J. B. Losos, B. Walker, D. S. Wilcove, and C. G. Morris, eds. 2009. The Princeton Guide to Ecology. Princeton University Press. ISBN: 978-0-691-15604-0.

Levin, S. A., H. C. Muller-Landau, R. Nathan, and J. Chave. 2003. The Ecology and Evolution of Seed Dispersal: A Theoretical Perspective. Annual Review of Ecology, Evolution, and Systematics 34:575–604. doi: 10.1146/annurev.ecolsys.34.011802.132428.

Li, X.-Y., and H. Kokko. 2019. Sex-biased dispersal: A review of the theory. Biological Reviews 94:721–736. doi: 10.1111/brv.12475.

Liggett, L. A., and J. DeGregori. 2017. Changing mutational and adaptive landscapes and the genesis of cancer. Biochimica et Biophysica Acta - Reviews on Cancer 1867:84–94. doi: 10.1016/j.bbcan.2017.01.005.

Lim, J., A. I. Fernandez, S. J. Hinojos, G. P. Aranda, J. James, C.-S. Seong, and K.-A. Han. 2018. The mushroom body D1 dopamine receptor controls innate courtship drive. Genes, Brain and Behavior 17:158–167. doi: 10.1111/gbb.12425.

Lindström, T., N. Håkansson, and U. Wennergren. 2011. The shape of the spatial kernel and its implications for biological invasions in patchy environments. Proceedings of the Royal Society B: Biological Sciences 278:1564–1571. doi: 10.1098/rspb.2010.1902.

Loewe, L., and B. Charlesworth. 2006. Inferring the distribution of mutational effects on fitness in Drosophila. Biology Letters 2:426–430. doi: 10.1098/rsbl.2006.0481.

Lopez-Garcia, C., A. M. Klein, B. D. Simons, and D. J. Winton. 2010. Intestinal Stem Cell Replacement Follows a Pattern of Neutral Drift. Science 330:822–825. doi: 10.1126/science.1196236.

Loponte, S., S. Lovisa, A. K. Deem, A. Carugo, and A. Viale. 2019. The Many Facets of Tumor Heterogeneity: Is Metabolism Lagging Behind? Cancers 11:1574. doi: 10.3390/cancers11101574.

Lorenz, M. W. 2007. Oogenesis-flight syndrome in crickets: Age-dependent egg production, flight performance, and biochemical composition of the flight muscles in adult female Gryllus bimaculatus. Journal of Insect Physiology 53:819–832. doi: 10.1016/j.jinsphys.2007.03.011.

Losos, J. B., S. J. Arnold, G. Bejerano, E. D. B. Iii, D. Hibbett, H. E. Hoekstra, D. P. Mindell, A. Monteiro, C. Moritz, H. A. Orr, D. A. Petrov, S. S. Renner, R. E. Ricklefs, P. S. Soltis, and T. L. Turner. 2013. Evolutionary Biology for the 21st Century. PLOS Biology 11:e1001466. doi: 10.1371/journal.pbio.1001466.

Losos, J. B., D. A. Baum, D. J. Futuyma, H. E. Hoekstra, R. E. Lenski, A. J. Moore, C. L. Peichel, D. Schluter, and M. J. Whitlock, eds. 2014. The Princeton Guide to Evolution. Princeton University Press. ISBN: 978-0-691-14977-6.

Love, C. R., S. Gautam, C. Lama, N. H. Le, and B. Dauwalder. 2023. The Drosophila dopamine 2-like receptor D2R (Dop2R) is required in the blood brain barrier for male courtship. Genes, Brain and Behavior 22:e12836. doi: 10.1111/gbb.12836.

Lovett, M., K. Lee, A. Edwards, and D. L. Kaplan. 2009. Vascularization Strategies for Tissue Engineering. Tissue Engineering. Part B, Reviews 15:353–370. doi: 10.1089/ten.teb.2009.0085.

Lustenhouwer, N., F. Moerman, F. Altermatt, R. D. Bassar, G. Bocedi, D. Bonte, S. Dey, E. A. Fronhofer, E. G. da Rocha, A. Giometto, L. T. Lancaster, R. B. Prather Jr, M. Saastamoinen, J. M. J. Travis, C. A. Urquhart, C. Weiss-Lehman, J. L. Williams, L. Börger, and D. Berger. 2023. Experimental evolution of dispersal: Unifying theory, experiments and natural systems. Journal of Animal Ecology 92:1113–1123. doi: 10.1111/1365-2656.13930.

Léna, J.-P., J. Clobert, M. de Fraipont, J. Lecomte, and G. Guyot. 1998. The relative influence of density and kinship on dispersal in the common lizard. Behavioral Ecology 9:500–507. doi: 10.1093/beheco/9.5.500.

Ma, X., Y. Y. Liu, Y. Y. Liu, L. B. Alexandrov, M. N. Edmonson, C. Gawad, X. Zhou, Y. Li, M. C. Rusch, J. Easton, R. Huether, V. Gonzalez-Pena, M. R. Wilkinson, L. C. Hermida, S. Davis, E. Sioson, S. Pounds, X. Cao, R. E. Ries, Z. Wang, X. Chen, L. Dong, S. J. Diskin, M. A. Smith, J. M. Guidry Auvil, P. S. Meltzer, C. C. Lau, E. J. Perlman, J. M. Maris, S. Meshinchi, S. P. Hunger, D. S. Gerhard, and J. Zhang. 2018. Pan-cancer genome and transcriptome analyses of 1,699 paediatric leukaemias and solid tumours. Nature 555:371–376. doi: 10.1038/nature25795.

Madan, E., A. M. Palma, V. Vudatha, J. G. Trevino, K. N. Natarajan, R. A. Winn, K. J. Won, T. A. Graham, R. Drapkin, S. A. McDonald, P. B. Fisher, and R. Gogna. 2022. Cell Competition in Carcinogenesis. Cancer Research 82:4487–4496. doi: 10.1158/0008-5472.can-22-2217.

Maggu, K., N. Ahlawat, M. Geeta Arun, A. Meena, and N. G. Prasad. 2021. Divergence of responses to variable socio-sexual environments in laboratory populations of Drosophila melanogaster evolving under altered operational sex ratios. Evolution 75:414–426. doi: 10.1111/evo.14138.

Maley, C. C., A. Aktipis, T. A. Graham, A. Sottoriva, A. M. Boddy, M. Janiszewska, A. S. Silva, M. Gerlinger, Y. Yuan, K. J. Pienta, K. S. Anderson, R. Gatenby, C. Swanton, D. Posada, C.-I. I. Wu, J. D. Schiffman, E. S. Hwang, K. Polyak, A. R. A. Anderson, J. S. Brown, M. Greaves, and D. Shibata. 2017. Classifying the evolutionary and ecological features of neoplasms. Nature Reviews Cancer 17:605–619. doi: 10.1038/nrc.2017.69.

Marigorta, U. M., L. A. Denson, J. S. Hyams, K. Mondal, J. Prince, T. D. Walters, A. Griffiths, J. D. Noe, W. V. Crandall, J. R. Rosh, D. R. Mack, R. Kellermayer, M. B. Heyman, S. S. Baker, M. C. Stephens, R. N. Baldassano, J. F. Markowitz, M.-O. Kim, M. C. Dubinsky, J. Cho, B. J. Aronow, S. Kugathasan, and G. Gibson. 2017. Transcriptional risk scores link GWAS to eQTLs and predict complications in Crohn's disease. Nature Genetics doi: 10.1038/ng.3936.

Markert, E. K., and A. Vazquez. 2015. Mathematical models of cancer metabolism. Cancer & Metabolism 3:14. doi: 10.1186/s40170-015-0140-6.

Martin, S. D., G. Coukos, R. A. Holt, and B. H. Nelson. 2015. Targeting the undruggable: immunotherapy meets personalized oncology in the genomic era. Annals of oncology : official journal of the European Society for Medical Oncology / ESMO 26:2367–74. doi: 10.1093/annonc/mdv382.

Martincorena, I. 2017. Universal Patterns of Selection in Cancer and Somatic Tissues. Cell 171. doi: 10.1016/j.cell.2017.09.042.

Martincorena, I., and P. J. Campbell. 2015. Somatic mutation in cancer and normal cells. Science 349:1483–1489. doi: 10.1126/science.aab4082.

Marusyk, A. 2009. Irradiation Alters Selection for Oncogenic Mutations in Hematopoietic Progenitors. Cancer Research 68. doi: 10.1158/0008-5472.can-09-0604.

Marusyk, A., D. P. Tabassum, P. M. Altrock, V. Almendro, F. Michor, and K. Polyak. 2014. Non-cell-autonomous driving of tumour growth supports sub-clonal heterogeneity. Nature 514:54–58. doi: 10.1038/nature13556.

Massot, M., J. Clobert, and R. Ferrière. 2008. Climate warming, dispersal inhibition and extinction risk. Global Change Biology 14:461–469. doi: 10.1111/j.1365-2486.2007.01514.x.

Matter, S. F., and J. Roland. 2002. An experimental examination of the effects of habitat quality on the dispersal and local abundance of the butterfly Parnassius smintheus. Ecological Entomology 27:308–316. doi: 10.1046/j.1365-2311.2002.00407.x.

McFarland, C. D., K. S. Korolev, G. V. Kryukov, S. R. Sunyaev, and L. A. Mirny. 2013. Impact of deleterious passenger mutations on cancer progression. Proceedings of the National Academy of Sciences 110. doi: 10.1073/pnas.1213968110.

McPeek, M. A., and R. D. Holt. 1992. The Evolution of Dispersal in Spatially and Temporally Varying Environments. The American Naturalist 140:1010–1027. doi: 10.1086/285453.

Meier, C. M., J. Starrfelt, and H. Kokko. 2011. Mate limitation causes sexes to coevolve towards more similar dispersal kernels. Oikos 120:1459–1468. doi: 10.1111/j.1600-0706.2011.19487.x.

Merckx, T., and H. Van Dyck. 2007. Habitat fragmentation affects habitat-finding ability of the speckled wood butterfly, Pararge aegeria L. Animal Behaviour 74:1029–1037. doi: 10.1016/j.anbehav.2006.12.020.

Meylan, S., J. Belliure, J. Clobert, and M. de Fraipont. 2002. Stress and Body Condition as Prenatal and Postnatal Determinants of Dispersal in the Common Lizard (Lacerta vivipara). Hormones and Behavior 42:319–326. doi: 10.1006/hbeh.2002.1827.

Meza, R., J. Jeon, S. H. Moolgavkar, and E. G. Luebeck. 2008. Age-specific incidence of cancer: Phases, transitions, and biological implications. Proceedings of the National Academy of Sciences 105:16284–16289. doi: 10.1073/pnas.0801151105.

Mina, M. 2017. Cancer Cell - 2017 Conditional Selection of Genomic Alterations Dictates Cancer Evolution and Oncogenic Dependencies. Cancer Cell 32. doi: 10.1016/j.ccell.2017.06.010.

Mishra, A., S. Tung, V. R. Shree Sruti, S. Srivathsa, and S. Dey. 2020. Mate-finding dispersal reduces local mate limitation and sex bias in dispersal. Journal of Animal Ecology 89:2089–2098. doi: 10.1111/1365-2656.13278.

Mishra, A., S. Tung, P. M. Shreenidhi, M. Aamir Sadiq, V. R. Shree Sruti, P. P. Chakraborty, and S. Dey. 2018. Sex differences in dispersal syndrome are modulated by environment and evolution. Philosophical Transactions of the Royal Society B: Biological Sciences 373:20170428. doi: 10.1098/rstb.2017.0428.

Mishra, A., S. Tung, V. R. S. Sruti, P. M. Shreenidhi, and S. Dey. 2022. Desiccation Stress Acts as Cause as well as Cost of Dispersal in Drosophila melanogaster. The American Naturalist 199:E111–e123. doi: 10.1086/718641.

Mohler, J. L., C. W. Gregory, O. H. Ford, III, D. Kim, C. M. Weaver, P. Petrusz, E. M. Wilson, and F. S. French. 2004. The Androgen Axis in Recurrent Prostate Cancer. Clinical Cancer Research 10:440–448. doi: 10.1158/1078-0432.ccr-1146-03.

Mole, S., and A. J. Zera. 1993. Differential allocation of resources underlies the dispersal-reproduction trade-off in the wing-dimorphic cricket, Gryllus rubens. Oecologia 93:121–127. doi: 10.1007/bf00321201.

Molofsky, J., and J.-B. Ferdy. 2005. Extinction dynamics in experimental metapopulations. Proceedings of the National Academy of Sciences of the United States of America 102:3726–3731. doi: 10.1073/pnas.0404576102.

Montgomery, R. B., E. A. Mostaghel, R. Vessella, D. L. Hess, T. F. Kalhorn, C. S. Higano, L. D. True, and P. S. Nelson. 2008. Maintenance of Intratumoral Androgens in Metastatic Prostate Cancer: A Mechanism for Castration-Resistant Tumor Growth. Cancer Research 68:4447–4454. doi: 10.1158/0008-5472.can-08-0249.

Montironi, R., A. Cimadamore, L. Cheng, A. Lopez-Beltran, and M. Scarpelli. 2018. Prostate cancer grading in 2018: limitations, implementations, cribriform morphology, and biological markers. The International Journal of Biological Markers 33:331–334. doi: 10.1177/1724600818781296.

Morrison, S. J., and J. Kimble. 2006. Asymmetric and symmetric stem-cell divisions in development and cancer. Nature 441:1068–1074. doi: 10.1038/nature04956.

Muller-Landau, H. C., S. A. Levin, and J. E. Keymer. 2003. Theoretical Perspectives on Evolution of Long-Distance Dispersal and the Example of Specialized Pests. Ecology 84:1957–1967. doi: 10.1890/01-0617.

Murrell, D. J., J. M. J. Travis, and C. Dytham. 2002. The evolution of dispersal distance in spatially-structured populations. Oikos 97:229–236. doi: 10.1034/j.1600-0706.2002.970209.x.

Muscarella, M. E., and J. P. O'Dwyer. 2020. Species dynamics and interactions via metabolically informed consumer-resource models. Theoretical Ecology 13:503–518. doi: 10.1007/s12080-020-00466-7.

Nagy, J. D., E. M. Victor, and J. H. Cropper. 2007. Why don't all whales have cancer? A novel hypothesis resolving Peto's paradox. Integrative and Comparative Biology 47:317–328. doi: 10.1093/icb/icm062.

Nall, A., and A. Sehgal. 2014. Monoamines and sleep in Drosophila. Behavioral Neuroscience 128:264–272. doi: 10.1037/a0036209.

Nardi, C., P. M. Fernandes, and J. M. S. Bento. 2008. Wing Polymorphism and Dispersal of Scaptocoris carvalhoi (Hemiptera: Cydnidae). Annals of the Entomological Society of America 101:551–557. doi: 10.1603/0013-8746(2008)101[551:wpados]2.0.co;2.

Nathan, R., W. M. Getz, E. Revilla, M. Holyoak, R. Kadmon, D. Saltz, and P. E. Smouse. 2008. A movement ecology paradigm for unifying organismal movement research. Proceedings of the National Academy of Sciences 105:19052–19059. doi: 10.1073/pnas.0800375105.

Neckameyer, W. S. 1998. Dopamine Modulates Female Sexual Receptivity in Drosophila Melanogaster. Journal of Neurogenetics 12:101–114. doi: 10.3109/01677069809167259.

Nesse, R. M. 2013. Tinbergen's four questions, organized: a response to Bateson and Laland. Trends in Ecology & Evolution 28:681–682. doi: 10.1016/j.tree.2013.10.008.

Nik-Zainal, S., P. Van Loo, D. C. Wedge, L. B. Alexandrov, C. D. Greenman, K. W. Lau, K. Raine, D. Jones, J. Marshall, M. Ramakrishna, A. Shlien, S. L. Cooke, J. Hinton, A. Menzies, L. A. Stebbings, C. Leroy, M. Jia, R. Rance, L. J. Mudie, S. J. Gamble, P. J. Stephens, S. McLaren, P. S. Tarpey, E. Papaemmanuil, H. R. Davies, I. Varela, D. J. McBride, G. R. Bignell, K. Leung, A. P. Butler, J. W. Teague, S. Martin, G. Jönsson, O. Mariani, S. Boyault, P. Miron, A. Fatima, A. Langerød, S. A. Aparicio, A. Tutt, A. M.

Sieuwerts, Å. Borg, G. Thomas, A. V. Salomon, A. L. Richardson, A.-L. Børresen-Dale, P. A. Futreal, M. R. Stratton, P. J. Campbell, A. Langerod, S. A. Aparicio, A. Tutt, A. M. Sieuwerts, Å. Borg, G. Thomas, A. V. Salomon, A. L. Richardson, A. L. Borresen-Dale, P. A. Futreal, M. R. Stratton, and P. J. Campbell. 2012. The life history of 21 breast cancers. Cell 149:994–1007. doi: 10.1016/j.cell.2012.04.023.

Niraula, S., A. J. Templeton, F. E. Vera-Badillo, A. M. Joshua, S. S. Sridhar, P. W. Cheung, P. M. Yip, A. Dodd, Z. Nugent, and I. F. Tannock. 2016. Study of testosterone-guided androgen deprivation therapy in management of prostate cancer: Study of Testosterone-Guided Androgen Deprivation. The Prostate 76:235–242. doi: 10.1002/pros.23117.

Nishiyama, T., and T. Hoshii. 2016. Testosterone-guided ADT for prostate cancer. Nature Reviews Urology 13:189–191. doi: 10.1038/nrurol.2016.9.

Noble, R., O. Kaltz, and M. E. Hochberg. 2015. Peto's paradox and human cancers. Philosophical Transactions of the Royal Society B: Biological Sciences 370:20150104. doi: 10.1098/rstb.2015.0104.

North, A., S. Cornell, and O. Ovaskainen. 2011. Evolutionary responses of dispersal distance to landscape structure and habitat loss. Evolution 65:1739–1751. doi: 10.1111/j.1558-5646.2011.01254.x.

Nowak, M. A., and B. Waclaw. 2017. Genes, environment, and "bad luck". Science 355:1266–1267. doi: 10.1126/science.aam9746.

Nowell, P. 1976. The clonal evolution of tumor cell populations. Science 194. doi: 10.1126/science.959840.

Nunney, L. 1999. Lineage selection and the evolution of multistage carcinogenesis. Proceedings of the Royal Society of London. Series B: Biological Sciences 266:493–498. doi: 10.1098/rspb.1999.0664.

———. 2018. Size matters: Height, cell number and a person's risk of cancer. Proceedings of the Royal Society B: Biological Sciences 285:20181743. doi: 10.1098/rspb.2018.1743.

Nunney, L., and B. Muir. 2015. Peto's paradox and the hallmarks of cancer: Constructing an evolutionary framework for understanding the incidence of cancer. Philosophical Transactions of the Royal Society B: Biological Sciences 370. doi: 10.1098/rstb.2015.0161.

Nutrition, A. I. o. 1977. Report of the American Institute of Nutrition Ad Hoc Committee on Standards for Nutritional Studies. The Journal of Nutrition 107:1340–1348. doi: 10.1093/jn/107.7.1340.

Ochocki, B. M., and T. E. X. Miller. 2017. Rapid evolution of dispersal ability makes biological invasions faster and more variable. Nature Communications 8:14315. doi: 10.1038/ncomms14315.

Ohlberger, J., G. Staaks, and F. Hölker. 2006. Swimming efficiency and the influence of morphology on swimming costs in fishes. Journal of Comparative Physiology B 176:17–25. doi: 10.1007/s00360-005-0024-0.

Osborne, C. K., K. Hobbs, and G. M. Clark. 1985. Effect of Estrogens and Antiestrogens on Growth of Human Breast Cancer Cells in Athymic Nude Mice1. Cancer Research 45:584–590.

Padfield, D., and G. Matheson. 2020. nls.multstart: Robust Non-Linear Regression using AIC Scores.

Page, S. T., D. W. Lin, E. A. Mostaghel, D. L. Hess, L. D. True, J. K. Amory, P. S. Nelson, A. M. Matsumoto, and W. J. Bremner. 2006. Persistent Intraprostatic Androgen Concentrations after Medical Castration in Healthy Men. The Journal of Clinical Endocrinology & Metabolism 91:3850–3856. doi: 10.1210/jc.2006-0968.

Pepper, J. W., K. Sprouffske, and C. C. Maley. 2007. Animal Cell Differentiation Patterns Suppress Somatic Evolution. PLOS Computational Biology 3:e250. doi: 10.1371/journal.pcbi.0030250.

Peto, H., F. J. Roe, P. N. Lee, L. Levy, and J. Clack. 1975. Cancer and ageing in mice and men. British Journal of Cancer 32:411–426. doi: 10.1038/bjc.1975.242.

Peto, R. 2015. Quantitative implications of the approximate irrelevance of mammalian body size and lifespan to lifelong cancer risk. Philosophical Transactions of the Royal Society B: Biological Sciences 370:20150198. doi: 10.1098/rstb.2015.0198.

Petrovskii, S., and A. Morozov. 2009. Dispersal in a Statistically Structured Population: Fat Tails Revisited. The American Naturalist 173:278–289. doi: 10.1086/595755.

Pietras, K., and A. Östman. 2010. Hallmarks of cancer: Interactions with the tumor stroma. Experimental Cell Research 316. doi: 10.1016/j.yexcr.2010.02.045.

Poethke, H. J., A. Gros, and T. Hovestadt. 2011. The ability of individuals to assess population density influences the evolution of emigration propensity and dispersal distance. Journal of Theoretical Biology 282:93–99. doi: 10.1016/j.jtbi.2011.05.012.

Pompei, F., M. Polkanov, and R. Wilson. 2001. The age distribution of cancer in mice. Toxicology and Industrial Health 17. doi: 10.1191/0748233701th091oa.

Prasad, N. G., and A. Joshi. 2003. What have two decades of laboratory life-history evolution studies onDrosophila melanogaster taught us? Journal of Genetics 82:45–76. doi: 10.1007/bf02715881.

R Core Team. 2021. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

Rafalko, J. M., K. M. Kruglyak, A. L. McCleary-Wheeler, V. Goyal, A. Phelps-Dunn, L. K. Wong, C. D. Warren, G. Brandstetter, M. C. Rosentel, L. DiMarzio, L. M. McLennan, A. L. O'Kell, T. A. Cohen, D. S. Grosu, J. Chibuk, D. W. Y. Tsui, I. Chorny, and A. Flory. 2023. Age at cancer diagnosis by breed, weight, sex, and cancer type in a cohort of more than 3,000 dogs: Determining the optimal age to initiate cancer screening in canine patients. Plos One 18:e0280795. doi: 10.1371/journal.pone.0280795.

Réale, D., D. Garant, M. M. Humphries, P. Bergeron, V. Careau, and P.-O. Montiglio. 2010. Personality and the emergence of the pace-of-life syndrome concept at the population level. Philosophical Transactions of the Royal Society B: Biological Sciences 365:4051–4063. doi: 10.1098/rstb.2010.0208.

Reim, E., D. Eichhorn, J. D. Roy, P. O. M. Steinhoff, and K. Fischer. 2019. Nutritional stress reduces flight performance and exploratory behavior in a butterfly. Insect Science 26:897–910. doi: 10.1111/1744-7917.12596.

Renault, D. 2020. A Review of the Phenotypic Traits Associated with Insect Dispersal Polymorphism, and Experimental Designs for Sorting out Resident and Disperser Phenotypes. Insects 11:214. doi: 10.3390/insects11040214.

Revelle, W. 2022. psych: Procedures for Psychological, Psychometric, and Personality Research. Northwestern University, Evanston, Illinois.

Robertson, F. W. 1960. The ecological genetics of growth in *Drosophila* 1. Body size and developmental time on different diets 1. Body size and developmental time on different diets. Genetical Research 1:288–304. doi: 10.1017/S0016672300000264.

Roff, D. A. 1990. Antagonistic pleiotropy and the evolution of wing dimorphism in the sand cricket, Gryllus firmus. Heredity 65:169–177. doi: 10.1038/hdy.1990.85.

Rogers, Z. N. 2017. A quantitative and multiplexed approach to uncover the fitness landscape of tumor suppression in vivo. Nature Methods 14. doi: 10.1038/nmeth.4297.

Ronce, O. 2007. How Does It Feel to Be Like a Rolling Stone? Ten Questions About Dispersal Evolution. Annual Review of Ecology, Evolution, and Systematics 38:231–253. doi: 10.1146/annurev.ecolsys.38.091206.095611.

Rousset, F., and S. Gandon. 2002. Evolution of the distribution of dispersal distance under distance-dependent cost of dispersal. Journal of Evolutionary Biology 15:515–523. doi: 10.1046/j.1420-9101.2002.00430.x.

Rozhok, A., and J. DeGregori. 2019. A generalized theory of age-dependent carcinogenesis. eLife 8:e39950. doi: 10.7554/eLife.39950.

Rozhok, A. I., and J. DeGregori. 2016. The Evolution of Lifespan and Age-Dependent Cancer Risk. Trends in Cancer 2. doi: 10.1016/j.trecan.2016.09.004.

Rozhok, A. I., J. L. Salstrom, and J. DeGregori. 2014. Stochastic modeling indicates that aging and somatic evolution in the hematopoietic system are driven by non-cell-autonomous processes. Aging (Albany NY) 6. doi: 10.18632/aging.100707.

———. 2016. Stochastic modeling reveals an evolutionary mechanism underlying elevated rates of childhood leukemia. Proceedings of the National Academy of Sciences 113. doi: 10.1073/pnas.1509333113.

Rundqvist, H., M. Augsten, A. Strömberg, E. Rullman, S. Mijwel, P. Kharaziha, T. Panaretakis, T. Gustafsson, and A. Östman. 2013. Effect of Acute Exercise on Prostate Cancer Cell Growth. PLoS ONE 8:e67579. doi: 10.1371/journal.pone.0067579.

Saastamoinen, M. 2008. Heritability of dispersal rate and other life history traits in the Glanville fritillary butterfly. Heredity 100:39–46. doi: 10.1038/sj.hdy.6801056.

Saastamoinen, M., and M. J. Rantala. 2013. Influence of Developmental Conditions on Immune Function and Dispersal-Related Traits in the Glanville Fritillary (Melitaea cinxia) Butterfly. Plos One 8:e81289. doi: 10.1371/journal.pone.0081289.

Sahoo, S., A. Mishra, H. Kaur, K. Hari, S. Muralidharan, S. Mandal, and M. K. Jolly. 2021. A mechanistic model captures the emergence and implications of non-genetic heterogeneity and reversible drug resistance in ER+ breast cancer cells. NAR Cancer 3:zcab027. doi: 10.1093/narcan/zcab027.

Sallan, S. E. 2006. Myths and Lessons from the Adult/Pediatric Interface in Acute Lymphoblastic Leukemia. Haematology 2006. doi: 10.1182/asheducation-2006.1.128.

Savage, V. M., A. P. Allen, J. H. Brown, J. F. Gillooly, A. B. Herman, W. H. Woodruff, and G. B. West. 2007. Scaling of number, size, and metabolic rate of cells with body size in mammals. Proceedings of the National Academy of Sciences 104:4718–4723. doi: 10.1073/pnas.0611235104.

Schäfer, M., and S. Werner. 2008. Cancer as an overhealing wound: An old hypothesis revisited. Nature Reviews Molecular Cell Biology 9. doi: 10.1038/nrm2455.

Schneider, C., J. Dover, and G. L. A. Fry. 2003. Movement of two grassland butterflies in the same habitat network: the role of adult resources and size of the study area. Ecological Entomology 28:219–227. doi: 10.1046/j.1365-2311.2003.00494.x.

Sender, R., and R. Milo. 2021. The distribution of cellular turnover in the human body. Nature Medicine 27:45–48. doi: 10.1038/s41591-020-01182-9.

Serrano, D., and J. L. Tella. 2003. Dispersal within a spatially structured population of lesser kestrels: the role of spatial isolation and conspecific attraction. Journal of Animal Ecology 72:400–410. doi: 10.1046/j.1365-2656.2003.00707.x.

Siegel, R. L., K. D. Miller, and A. Jemal. 2016. Cancer statistics, 2016. CA: A Cancer Journal for Clinicians 66:7–30. doi: 10.3322/caac.21332.

Sitaraman, D., Y. Aso, G. M. Rubin, and M. N. Nitabach. 2015. Control of Sleep by Dopaminergic Inputs to the Drosophila Mushroom Body. Frontiers in Neural Circuits 9.

Southon, R. J., A. N. Radford, and S. Sumner. 2020. Hormone-mediated dispersal and sexual maturation in males of the social paper wasp Polistes lanio. Journal of Experimental Biology 223:jeb226472. doi: 10.1242/jeb.226472.

Stamey, T. A., N. Yang, A. R. Hay, J. E. McNeal, F. S. Freiha, and E. Redwine. 1987. Prostate-Specific Antigen as a Serum Marker for Adenocarcinoma of the Prostate. New England Journal of Medicine 317:909–916. doi: 10.1056/nejm198710083171501.

Stanbrough, M., G. J. Bubley, K. Ross, T. R. Golub, M. A. Rubin, T. M. Penning, P. G. Febbo, and S. P. Balk. 2006. Increased Expression of Genes Converting Adrenal Androgens to Testosterone in Androgen-Independent Prostate Cancer. Cancer Research 66:2815–2825. doi: 10.1158/0008-5472.can-05-4000.

Starrfelt, J., and H. Kokko. 2010. Parent-Offspring Conflict and the Evolution of Dispersal Distance. The American Naturalist 175:38–49. doi: 10.1086/648605.

Stefanescu, C., A. Ubach, and C. Wiklund. 2021. Timing of mating, reproductive status and resource availability in relation to migration in the painted lady butterfly. Animal Behaviour 172:145–153. doi: 10.1016/j.anbehav.2020.12.013.

Stephens, P. J. 2011. Massive genomic rearrangement acquired in a single catastrophic event during cancer development. Cell 144. doi: 10.1016/j.cell.2010.11.055.

Stevens, V. M., A. Trochet, S. Blanchet, S. Moulherat, J. Clobert, and M. Baguette. 2013. Dispersal syndromes and the use of life-histories to predict dispersal. Evolutionary Applications 6:630–642. doi: 10.1111/eva.12049.

Stevens, V. M., A. Trochet, H. Van Dyck, J. Clobert, and M. Baguette. 2012. How is dispersal integrated in life histories: a quantitative analysis using butterflies. Ecology Letters 15:74–86. doi: 10.1111/j.1461-0248.2011.01709.x.

Stewart, G. D., J. A. Ross, D. B. McLaren, C. C. Parker, F. K. Habib, and A. C. Riddick. 2010. The relevance of a hypoxic tumour microenvironment in prostate cancer. BJU International 105:8–13. doi: 10.1111/j.1464-410X.2009.08921.x.

Stratton, M. R., P. J. Campbell, and P. A. Futreal. 2009. The cancer genome. Nature 458:719–724. doi: 10.1038/nature07943.

Sulak, M., L. Fong, K. Mika, S. Chigurupati, L. Yon, N. P. Mongan, R. D. Emes, and V. J. Lynch. 2016. TP53 copy number expansion is associated with the evolution of increased body size and an enhanced DNA damage response in elephants. eLife 5. doi: 10.7554/eLife.11994.

Sung, H., J. Ferlay, R. L. Siegel, M. Laversanne, I. Soerjomataram, A. Jemal, and F. Bray. 2021. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA: A Cancer Journal for Clinicians 71:209–249. doi: 10.3322/caac.21660.

Swanson, K. R., L. D. True, D. W. Lin, K. R. Buhler, R. Vessella, and J. D. Murray. 2001. A Quantitative Model for the Dynamics of Serum Prostate-Specific Antigen as a Marker for Cancerous Growth: An Explanation for a Medical Anomaly. The American Journal of Pathology 158:2195–2199. doi: 10.1016/s0002-9440(10)64691-3.

Syed, Z. A., V. Gupta, M. G. Arun, A. Dhiman, B. Nandy, and N. G. Prasad. 2020. Absence of reproduction-immunity trade-off in male Drosophila melanogaster evolving under differential sexual selection. BMC Evolutionary Biology 20:13. doi: 10.1186/s12862-019-1574-1.

Tatar, M., S. Post, and K. Yu. 2014. Nutrient control of Drosophila longevity. Trends in Endocrinology & Metabolism 25:509–517. doi: 10.1016/j.tem.2014.02.006.

Taylor, R. A. J. 1978. The relationship between density and distance of dispersing insects. Ecological Entomology 3:63–70. doi: 10.1111/j.1365-2311.1978.tb00903.x.

Tennant, R. W. 1993. A perspective on nonmutagenic mechanisms in carcinogenesis. Environmental Health Perspectives 101:231–236. doi: 10.1289/ehp.93101s3231.

Tennessen, J. M., W. E. Barry, J. Cox, and C. S. Thummel. 2014. Methods for studying metabolism in Drosophila. Methods 68:105–115. doi: 10.1016/j.ymeth.2014.02.034.

Teply, B. A., H. Wang, B. Luber, R. Sullivan, I. Rifkind, A. Bruns, A. Spitz, M. DeCarli, V. Sinibaldi, C. F. Pratz, C. Lu, J. L. Silberstein, J. Luo, M. T. Schweizer, C. G. Drake, M. A. Carducci, C. J. Paller, E. S. Antonarakis, M. A. Eisenberger, and S. R. Denmeade.

2018. Bipolar androgen therapy in men with metastatic castration-resistant prostate cancer after progression on enzalutamide: an open-label, phase 2, multicohort study. The Lancet Oncology 19:76–86. doi: 10.1016/s1470-2045(17)30906-3.

Tetteh, P. W., H. F. Farin, and H. Clevers. 2015. Plasticity within stem cell hierarchies in mammalian epithelia. Trends in Cell Biology 25:100–108. doi: 10.1016/j.tcb.2014.09.003.

Thorsson, V., D. L. Gibbs, S. D. Brown, D. Wolf, D. S. Bortone, T.-H. O. Yang, E. Porta-Pardo, G. F. Gao, C. L. Plaisier, J. A. Eddy, E. Ziv, A. C. Culhane, E. O. Paull, I. K. A. Sivakumar, A. J. Gentles, R. Malhotra, F. Farshidfar, A. Colaprico, J. S. Parker, L. E. Mose, N. S. Vo, J. Liu, Y. Liu, J. Rader, V. Dhankani, S. M. Reynolds, R. Bowlby, A. Califano, A. D. Cherniack, D. Anastassiou, D. Bedognetti, Y. Mokrab, A. M. Newman, A. Rao, K. Chen, A. Krasnitz, H. Hu, T. M. Malta, H. Noushmehr, C. S. Pedamallu, S. Bullman, A. I. Ojesina, A. Lamb, W. Zhou, H. Shen, T. K. Choueiri, J. N. Weinstein, J. Guinney, J. Saltz, R. A. Holt, C. S. Rabkin, and T. Cancer Genome Atlas Research Network. 2018. The Immune Landscape of Cancer. Immunity 48:812–830.e14. doi: 10.1016/j.immuni.2018.03.023.

Tian, X., J. Azpurua, C. Hine, A. Vaidya, M. Myakishev-Rempel, J. Ablaeva, Z. Mao, E. Nevo, V. Gorbunova, and A. Seluanov. 2013. High-molecular-mass hyaluronan mediates the cancer resistance of the naked mole rat. Nature 499:346–349. doi: 10.1038/nature12234.

Tilman, D. 1980. Resources: A Graphical-Mechanistic Approach to Competition and Predation. The American Naturalist 116:362–393. doi: 10.1086/283633.

Tinbergen, N. 1963. On aims and methods of Ethology. Zeitschrift für Tierpsychologie 20:410–433. doi: 10.1111/j.1439-0310.1963.tb01161.x.

Titus, M. A., M. J. Schell, F. B. Lih, K. B. Tomer, and J. L. Mohler. 2005. Testosterone and Dihydrotestosterone Tissue Levels in Recurrent Prostate Cancer. Clinical Cancer Research 11:4653–4657. doi: 10.1158/1078-0432.ccr-05-0525.

Tollis, M., A. M. Boddy, and C. C. Maley. 2017. Peto's Paradox: How has evolution solved the problem of cancer prevention? BMC Biology 15. doi: 10.1186/s12915-017-0401-7.

Tomasetti, C., L. Li, and B. Vogelstein. 2017. Stem cell divisions, somatic mutations, cancer etiology, and cancer prevention. Science 355:1330–1334. doi: 10.1126/science.aaf9011.

Tomasetti, C., and B. Vogelstein. 2015. Variation in cancer risk among tissues can be explained by the number of stem cell divisions. Science 347. doi: 10.1126/science.1260825.

Travis, J. M. J., and C. Dytham. 1998. The evolution of dispersal in a metapopulation: a spatially explicit, individual-based model. Proceedings of the Royal Society of London. Series B: Biological Sciences 265:17–23. doi: 10.1098/rspb.1998.0258.

———. 1999. Habitat persistence, habitat availability and the evolution of dispersal. Proceedings of the Royal Society of London. Series B: Biological Sciences 266:723–728. doi: 10.1098/rspb.1999.0696.

———. 2002. Dispersal evolution during invasions. Evolutionary Ecology Research 4:1119–1129.

Travis, J. M. J., H. S. Smith, and S. M. W. Ranwala. 2010. Towards a mechanistic understanding of dispersal evolution in plants: conservation implications. Diversity and Distributions 16:690–702. doi: 10.1111/j.1472-4642.2010.00674.x.

Trochet, A., E. A. Courtois, V. M. Stevens, M. Baguette, A. Chaine, D. S. Schmeller, J. Clobert, and J. J. Wiens. 2016. Evolution of Sex-Biased Dispersal. The Quarterly Review of Biology 91:297–320. doi: 10.1086/688097.

Tu, M.-P., and M. Tatar. 2003. Juvenile diet restriction and the aging and reproduction of adult Drosophila melanogaster. Aging Cell 2:327–333. doi: 10.1046/j.1474-9728.2003.00064.x.

Tung, S., A. Mishra, N. Gogna, M. Aamir Sadiq, P. M. Shreenidhi, V. R. Shree Sruti, K. Dorai, and S. Dey. 2018*a*. Evolution of dispersal syndrome and its corresponding metabolomic changes. Evolution 72:1890–1903. doi: 10.1111/evo.13560.

Tung, S., A. Mishra, P. M. Shreenidhi, M. A. Sadiq, S. Joshi, V. R. S. Sruti, and S. Dey. 2018*b*. Simultaneous evolution of multiple dispersal components and kernel. Oikos 127:34–44. doi: 10.1111/oik.04618.

Umekita, Y., R. A. Hiipakka, J. M. Kokontis, and S. Liao. 1996. Human prostate tumor growth in athymic mice: inhibition by androgens and stimulation by finasteride. Proceedings of the National Academy of Sciences 93:11802–11807. doi: 10.1073/pnas.93.21.11802.

Van Houtan, K. S., S. L. Pimm, J. M. Halley, R. O. Bierregaard Jr, and T. E. Lovejoy. 2007. Dispersal of Amazonian birds in continuous and fragmented forest. Ecology Letters 10:219–229. doi: 10.1111/j.1461-0248.2007.01004.x.

VanderPlas, J. 2016. Python Data Science Handbook: Essential Tools for Working with Data. O'Reilly Media, Inc. ISBN: 978-1-4919-1213-3.

Varga, J., and F. R. Greten. 2017. Cell plasticity in epithelial homeostasis and tumorigenesis. Nature Cell Biology 19:1133–1141. doi: 10.1038/ncb3611.

Vas, V. 2012. Contribution of an Aged Microenvironment to Aging-Associated Myeloproliferative Disease. PLoS One 7. doi: 10.1371/journal.pone.0031523.

Vas, V., K. Senger, K. Dorr, A. Niebel, and H. Geiger. 2012. Aging of the Microenvironment Influences Clonality in Hematopoiesis. PLoS One 7. doi: 10.1371/journal.pone.0042080.

Vazquez, J. M., and V. J. Lynch. 2021. Pervasive duplication of tumor suppressors in Afrotherians during the evolution of large bodies and reduced cancer risk. eLife 10:e65041. doi: 10.7554/eLife.65041.

Vermeulen, L. 2013. Defining Stem Cell Dynamics in Models of Intestinal Tumor Initiation. Science 342. doi: 10.1126/science.1243148.

Vijendravarma, R. K., S. Narasimha, and T. J. Kawecki. 2012. Chronic malnutrition favours smaller critical size for metamorphosis initiation in Drosophila melanogaster. Journal of Evolutionary Biology 25:288–292. doi: 10.1111/j.1420-9101.2011.02419.x.

Vinatier, F., F. Lescourret, P.-F. Duyck, O. Martin, R. Senoussi, and P. Tixier. 2011. Should I Stay or Should I Go? A Habitat-Dependent Dispersal Kernel Improves Prediction of Movement. Plos One 6:e21115. doi: 10.1371/journal.pone.0021115.

Vollmer, R. T. 2010. Dissecting the Dynamics of Serum Prostate-Specific Antigen. American Journal of Clinical Pathology 133:187–193. doi: 10.1309/ajcp3cjr3idrcfeo.

Wallace, B. 1966. On the Dispersal of Drosophila. The American Naturalist 100:551–563. doi: 10.1086/282449.

Watson, P. A., V. K. Arora, and C. L. Sawyers. 2015. Emerging mechanisms of resistance to androgen receptor inhibitors in prostate cancer. Nature Reviews Cancer 15:701–711. doi: 10.1038/nrc4016.

Weinberg, R. A. 1998. One renegade cell : how cancer begins, vol. 1. Basic Books. ISBN: 978-0-7538-0786-6.

Werner, B., D. Dingli, T. Lenaerts, J. M. Pacheco, and A. Traulsen. 2011. Dynamics of Mutant Cells in Hierarchical Organized Tissues. PLOS Computational Biology 7:e1002290. doi: 10.1371/journal.pcbi.1002290.

Werner, B., D. Dingli, and A. Traulsen. 2013. A deterministic model for the occurrence and dynamics of multiple mutations in hierarchically organized tissues. Journal of The Royal Society Interface 10:20130349–20130349. doi: 10.1098/rsif.2013.0349.

West, J., L. You, J. Zhang, R. A. Gatenby, J. S. Brown, P. K. Newton, and A. R. Anderson. 2020. Towards Multidrug Adaptive Therapy. Cancer Research 80:1578–1589. doi: 10.1158/0008-5472.can-19-2669.

Whitmee, S., and C. D. L. Orme. 2013. Predicting dispersal distance in mammals: a trait-based approach. Journal of Animal Ecology 82:211–221. doi: 10.1111/j.1365-2656.2012.02030.x.

Wibmer, A., H. Hricak, T. Gondo, K. Matsumoto, H. Veeraraghavan, D. Fehr, J. Zheng, D. Goldman, C. Moskowitz, S. W. Fine, V. E. Reuter, J. Eastham, E. Sala, and H. A. Vargas. 2015. Haralick texture analysis of prostate MRI: utility for differentiating non-cancerous prostate from prostate cancer and differentiating prostate cancers with different Gleason scores. European Radiology 25:2840–2850. doi: 10.1007/s00330-015-3701-8.

Wijerathna, A., and M. Evenden. 2019. Energy use by the mountain pine beetle (Coleoptera: Curculionidae: Scolytinae) for dispersal by flight. Physiological Entomology 44:200–208. doi: 10.1111/phen.12290.

Wilcox, R. R. 2010. Fundamentals of Modern Statistical Methods: Substantially Improving Power and Accuracy. Springer, New York, NY. ISBN: 978-1-4419-5524-1 978-1-4419-5525-8.

Woodward, T. L., J. Xie, J. L. Fendrick, and S. Z. Haslam. 2000. Proliferation of Mouse Mammary Epithelial Cells in Vitro: Interactions among Epidermal Growth Factor, Insulin-Like Growth Factor I, Ovarian Hormones, and Extracellular Matrix Proteins 1. Endocrinology 141. doi: 10.1210/endo.141.10.7701.

Wu, C.-I., H.-Y. Wang, S. Ling, and X. Lu. 2016. The Ecology and Evolution of Cancer: The Ultra-Microevolutionary Process. Annual Review of Genetics 50:347–369. doi: 10.1146/annurev-genet-112414-054842.

Yurkovic, A., O. Wang, A. C. Basu, and E. A. Kravitz. 2006. Learning and memory associated with aggression in Drosophila melanogaster. Proceedings of the National Academy of Sciences 103:17519–17524. doi: 10.1073/pnas.0608211103.

Zera, A. J., and R. F. Denno. 1997. Physiological and Ecology of Dispersal Polymorphism in Insects. Annual Review of Entomology 42:207–230. doi: 10.1146/annurev.ento.42.1.207.

Zera, A. J., and Z. Zhao. 2004. Effect of a Juvenile Hormone Analogue on Lipid Metabolism in a Wing-Polymorphic Cricket: Implications for the Endocrine-Biochemical Bases of Life-History Trade-Offs. Physiological and Biochemical Zoology 77:255–266. doi: 10.1086/383500.

Zhang, J., J. Cunningham, J. Brown, and R. Gatenby. 2022. Evolution-based mathematical models significantly prolong response to abiraterone in metastatic castrate-resistant prostate cancer and identify strategies to further improve outcomes. eLife 11:e76284. doi: 10.7554/eLife.76284.

Zhang, J., J. J. Cunningham, J. S. Brown, and R. A. Gatenby. 2017. Integrating evolutionary dynamics into treatment of metastatic castrate-resistant prostate cancer. Nature Communications 8:1816. doi: 10.1038/s41467-017-01968-5.

Zhang, S. X., D. Rogulja, and M. A. Crickmore. 2016. Dopaminergic Circuitry Underlying Mating Drive. Neuron 91:168–181. doi: 10.1016/j.neuron.2016.05.020.

Zhang, X. X. 2018. A simple 3-parameter model for cancer incidences. Scientific Reports 8. doi: 10.1038/s41598-017-17765-5.

Zheng, J. 2012. Energy metabolism of cancer: Glycolysis versus oxidative phosphorylation (review). Oncology Letters 4:1151–1157. doi: 10.3892/ol.2012.928.

Zollner, P. A., and S. L. Lima. 1999. Search Strategies for Landscape-Level Interpatch Movements. Ecology 80:1019–1030. doi: 10.1890/0012-9658(1999)080[1019:ssflli]2.0.co;2.

Zúñiga-Hernández, J. M., G. H. Olivares, P. Olguín, and A. Glavic. 2023. Low-nutrient diet in Drosophila larvae stage causes enhancement in dopamine modulation in adult brain due epigenetic imprinting. Open Biology 13:230049. doi: 10.1098/rsob.230049.

# List of publications/manuscripts under preparation

## Published work

1. **B. Vibishan**, and M. G. Watve. 2020. "Context-Dependent Selection as the Keystone in the Somatic Evolution of Cancer." Scientific Reports 10 (1): 4223. doi: 10.1038/s41598-020-61046-7

## Manuscripts under review

1. **B. Vibishan**, B. V. Harshavardhan, and S. Dey. 2023. "A Resource-Based Mechanistic Framework for Castration-Resistant Prostate Cancer (CRPC)." bioRxiv. doi: 10.1101/2023.09.19.558379

2. **B. Vibishan**, A. Malwade, S. R. Gupta, A. Koner, V. Khodake, and S. Dey. 2023. "Selection for dispersal under larval malnutrition results in a non-monotonic kernel in *Drosophila melanogaster*." bioRxiv. doi: 10.1101/2023.10.26.564216

## Manuscripts under preparation

1. **B. Vibishan**, A. Malwade, A. Koner, V. Khodake, and S. Dey. 2023. "Malnourished dispersers show sexually-dimorphic trait correlations under dispersal selection."