

Individual-level causes and population-level consequences of variation in fitness in an alpine rodent

Dissertation
zur
Erlangung der naturwissenschaftlichen Doktorwürde
(Dr. sc. nat.)
Mathematisch-naturwissenschaftlichen Fakultät
der
Universität Zürich
von

Timothée Bonnet

aus
Frankreich

Promotionskomitee

Prof. Dr. Lukas Keller (Vorsitz)
Dr. Erik Postma (Leitung der Dissertation)
Prof. Dr. Ben Sheldon
Prof. Dr. Barbara Tschirren
Prof. Dr. Arpat Ozgul
Dr. Jarrod Hadfield
Dr. Marc Kéry

Zürich, 2016

©Timothée Bonnet

timotheebonnetc@gmail.com

Photos: Timothée Bonnet

This is a L^AT_EX document using a customized version of the KOMA-Script's class
scrbook

Summary

This thesis investigates the stochastic and selective causes of variation in fitness components, and the evolutionary consequences of this variation in a wild rodent population. It shows the contemporary genetic evolution of body mass and decouples classic estimates of selection from adaptive evolution.

The heart of evolutionary biology is understanding the variation in organisms. For over 150 years, researchers have documented the causes of within-species variation and how it contributes to speciation and explains the fit between organisms and their environment. Recently, increasing concerns regarding rapid anthropogenic changes have driven renewed investigation of how wild populations adapt to environmental change. This new focus has revealed the difficulties measuring natural selection, disentangling evolution from plastic changes, and predicting evolutionary trajectories. For instance, there are few robust examples of contemporary evolution in wild populations, casting doubt on the possibility that evolution can rescue populations from rapid environmental change. In this thesis, I investigate the causes of natural selection and evolution in a wild population of snow voles (*Chionomys nivalis*). Thanks to 10 years of intensive individual-based monitoring and genotyping, knowledge of this population includes life-history, morphological data, and a high-resolution pedigree. This population is therefore among the best available worldwide to measure selection and evolution in action.

The population is nevertheless relatively small and recent publications suggest that the evolutionary potential in small populations is effectively cancelled by stochasticity in fitness components. I assess the methods used in those publications and demonstrate that the variation in fitness components is not purely stochastic. Small populations, including these snow voles, show evolutionary potential.

With collaborators, I then compare four common methodological frameworks to disentangle the contributions to phenotypic change of evolution, plasticity, and demography. We identify important discrepancies between the frameworks, partly originating from using different definitions, but also possessing intrinsically different capabilities. Among the considered frameworks only quantitative genetics can measure genetic change.

Applying methods from quantitative genetics to the snow vole population, I demonstrate that body mass evolved adaptively over the study period. I show that phenotypic estimates of selection are not predictive of genetic evolution: neither the mean selection nor its temporal variation are related to the rate of genetic evolution. This demonstrates that the dominant purely-phenotypic method used to measure selection risks measuring variation in nutritional status instead. Nevertheless, I employed quantitative genetics to identify the target of selection and obtain selection estimates in line with the observed genetic change

This thesis establishes contemporary evolution in a wild population and shows that



evolutionary responses to environmental change cannot be reliably estimated nor understood from purely-phenotypic methods; an explicit genetic approach is necessary.



Zusammenfassung

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Nullam a tellus. Aliquam commodo dui non ipsum. Duis mollis nisi id turpis. Donec quis ipsum. Curabitur sed nibh. Morbi suscipit justo quis orci. Ut massa tortor, ultricies vitae, lacinia eu, facilisis eu, nisl. Nulla mattis urna sed metus imperdiet ornare. Praesent sodales. Etiam laoreet. Mauris quam magna, sagittis et, pharetra eget, congue vitae, arcu. Fusce sollicitudin justo. Suspendisse lectus. Sed lobortis dolor quis lectus scelerisque ornare. Integer purus. Phasellus vel elit at nibh sagittis lobortis. Aliquam iaculis malesuada eros. Mauris metus

Contents

Summary	i
Zusammenfassung	iii
Chapter 1 General introduction	1
Chapter 2 – Successful by chance? The power of mixed models and neutral simulations for the detection of individual fixed heterogeneity in fitness components	17
Chapter 3 – Disentangling evolutionary, plastic and demographic processes underlying trait dynamics: A review of four frameworks	53
Chapter 4 – The stasis that wasn't: Adaptive body mass evolution is opposite to phenotypic selection in a wild rodent population	75
Chapter 5 – Fluctuating selection	77
Chapter 6 General discussion	79

Chapter 1

General introduction

One can't understand everything at once, we can't begin with perfection all at once! In order to reach perfection one must begin by being ignorant of a great deal. And if we understand things too quickly, perhaps we shan't understand them thoroughly.

— Fyodor Dostoyevsky, *The Idiot* (1868–9)

Si nous ne trouvons pas des choses agréables, nous trouverons du moins des choses nouvelles.

— Voltaire a.k.a. François-Marie Arouet, *Candide* (1759)

1.1 Variation in fitness

1.1.1 The origin of variation in evolutionary biology

The heart of evolutionary questioning is understanding variation among living beings (Lynch and Walsh 1998; Wayne and Miyamoto 2006; Kruuk, Charmantier, and Garant 2014). It is in fact its very starting point. Darwin opens his book *the Origin of Species* with two chapters describing variability in domestic and wild organisms (Darwin 1859). Building on these observations, Darwin then summarizes the evidence showing that variation within species is the fuel generating the astonishing diversity among species, and the striking fit between organisms and their environment.

These great answers immediately opened many more questions about the causes and consequences of variation, some of which remain not fully answered more than 150 years later. In particular, nineteenth century biologists struggled with the sources of variation within species, as is clear in *the Origin of Species* itself: “*Variability is governed by many unknown laws, of which correlated growth is probably the most important. Something, but how much we do not know, may be attributed to the definite action of the conditions of life. Some, perhaps a great, effect may be attributed to the increased use or disuse of parts*” (p. 31 Darwin 1859)¹. Of course, the effect of ageing was acknowledged and plastic responses to the environment were thought to be predominant (Wilkins 2009). Ageing and the environment does not always provide a satisfactory explanation for variation that appears within a population, however. Furthermore, Darwinian arguments build on the observation of this special kind of inherited variation that can

¹ Alternatively, biologists dismissed this within-species variation by considering that species were arbitrary boundaries in a set or continuum of variation. Darwin did not attempt to define species, but by explaining how they originate, he made some definitions indefensible (Wilkins 2009, pp. 129-163).

appear among siblings of a same litter, clutch, or pod, and that is subsequently transmitted from parent to offspring (Darwin 1859, Chapter 1). The late nineteenth century was utterly ignorant of the sources of inherited variation within species. Only at the beginning of the twentieth century were the laws of inheritance progressively discovered and spread to the scientific community (Dietrich 2006). Four more decades saw these laws formalized into a unified scientific theory to understand variation within populations (R. Fisher 1930), and explained at the molecular level (Oswald, MacLeod, and McCarty 1944; Watson and Crick 1953), thus closing the logical gap in Darwin's argument: Relatives resemble each other because they share similar gene versions on long strands of DNA, a molecule that is copied with high fidelity and transmitted from parent to offspring; There is variation among siblings because of the reshuffling and segregation of parental genes and, on occasions, because DNA mutates. The understanding of the causes of variation within species and populations has made terrific progresses and now fits elegantly in the broader evolutionary theory (Pigliucci and Müller 2010). Nevertheless, many aspects of the causes of variation are still to be refined or newly explored, especially in natural populations (Kruuk, Charmantier, and Garant 2014). In particular, the relative importance of genes and the environment in the wild remains studied in only a few populations of a few species, taxonomically biased, and concerns a limited set of traits (Lynch and Walsh 1998; Postma 2014).

Additional open research questions relate to the consequences of within-species variation. In particular, a lot of attention is paid to how genetic variation translates into adaptive evolution (Brookfield 2016). Any trait that possesses genetic variation can evolve, but it will evolve in an adaptive way only if the trait is subject to selection, be it artificial or natural. Selection occurs when the variation in the trait causes variation in *fitness*. Before we discuss the specificity of the causes and consequences of variation in fitness, we must introduce this difficult concept.

1.1.2 Variation in the definition of fitness

There has been a great deal written about the concept of fitness, including multiple conflicting definitions, which “*is hardly surprising as every important scientific concept is difficult to understand from first principles, as for instance the notions of space and time, or energy and force*” (p. 1358 Wagner 2010). I will not solve the question of the definition of fitness here, but I will try to make clear how the word is used in this thesis. To start with, in the past, there has been some confusion on whether fitness is a realized reproductive outcome or a propensity to reproduce (Brandon and Beatty 1984). It is now rather consensual that the concept of fitness is more useful when it is defined as a propensity, that is, as an expected value that cannot be measured directly because of stochasticity (Brandon and Beatty 1984; Price 1996; Krimbas 2004) and we will follow this consensus. Fitness has been defined at the level of the genetic lineage (e.g. Akçay and Van Cleve 2016), of the individual (e.g. Cam and Monnat 2000), of the genotype (e.g. Steiner and Tuljapurkar 2012), or of the population (e.g. Tienderen 2000). A propensity definition partly dissolves the problem of the level of the definition, since the expected reproductive outcome of a genotype is the same as the expected reproductive outcome of the individuals bearing this genotype, and the expected reproductive outcome of a population is the sum of the expectation of the

reproductive outcome of its individuals. Here, we will consider fitness at the level of individuals, because they are the unit most easily observable and the primary target of natural selection. More confusion on fitness comes from it being alternatively defined as the asymptotic number of descendants or as the contribution to the next generation (Wade 2006). Since we consider fitness at the level of individuals and because most of the work carried out is based on data covering about ten generations only, it is intuitive to consider fitness as the contribution to the next generation. Besides practical considerations, this choice allows for a clear, and conceptually crucial, distinction between selection, inheritance and evolution, that is blurred in asymptotic definitions (R. Fisher 1930; Arnold and Wade 1984). A slightly contentious point is whether fitness should be defined as an absolute number of offspring (Wade 2006) or a relative one (Rousset 2004), that is, whether “relative fitness” is a meaningful phrase or a tautological one. I feel like the relative definition is really closer to the interest of evolutionary biologists and avoids appending *relative* to every occurrence of *fitness*. Nevertheless the field massively favours the absolute definition and for the sake of consistency I attempted to yield to the convention (possibly with some inconsistencies). Finally, instead of a measure of reproductive success, relative fitness has recently been defined as the amount of information about the environment that populations accumulate by selection (Frank 2012). I see great conceptual promises in this view, that brings together an essentialist use of the word *fitness* and the scientific field of information theory. An information interpretation of fitness did not directly influence the work and is not necessary to understand it, but it might enlighten some of the results presented here, and evolutionary biology in general. To sum up, we define the fitness of an individual as its expected number of descendant in the next generation.

1.1.3 Causes and consequences of fitness variation in the wild

Why is there variation in individual fitness? This question attracted a lot of research attention, because (i) genetic variation in fitness controls the pace of evolution within a population, and because (ii) an intuitive consequence of evolution is the erosion of genetic variation in fitness, thus making the presence of genetic variation in fitness paradoxical (Jones 1987). In this thesis, we will not deal with the second point, the fundamental question of appearance and maintenance of genetic variation in fitness, but rather with its proximal sources. We will consider these proximal sources from two complementary angles.

First in a descriptive approach, one can decompose variation in fitness into components of variation, without mention of the underlying mechanisms. Apart from genetic variation, variation in fitness can also originate from variation in early-life, micro-environment (Turner 2009), or maternal effects (Wolf and Wade 2009). In addition, when working with wild sexual organisms, individual fitness as we defined it cannot be observed directly. Indeed, individuals are unique and their realized reproductive success does not equal their expected reproductive success. Therefore, researchers have to rely on fitness proxies, often realized reproductive success and survival, that contain a large stochastic component. Additive genetic variation in fitness is also the rate of evolution in fitness and sets the maximal rate of evolution (R. Fisher 1930). A variance decomposition approach is therefore useful to ascertain whether

variation in fitness proxies is all stochastic and environmental or whether it hides genetic variation in fitness. In doing so, it determines how much adaptive evolution can be expected to happen within a population.

Second, in a more mechanistic approach, one can investigate what characteristics make some individuals fitter than others, that is what traits are under natural selection². The study of natural selection in the wild took-off with the development of regression-based methods to accurately measure its strength and predict its effects (Lande 1979; Lande and Arnold 1983). Under some assumptions, the genetic change in response to selection on a trait is the product of a selection gradient and of additive genetic variation in that trait (Lush 1937). Therefore, by understanding what traits cause variation in fitness, one can predict what traits should evolve, as well as their direction and speed of evolution.

The study of natural selection and adaptive evolution in the wild is very topical in the context of unprecedented rates of environmental changes induced by human activities (Parmesan 2006). Anthropogenic changes provide the opportunity of natural experiments to evolutionary biologists (Altermatt, Ebert, and Altermatt 2016; Brookfield 2016), but also come with societal concerns and an ever increasing urge to better understand and predict how living things respond to the selective pressures imposed by environmental changes (Mc Carty 2001; Shaw and Shaw 2014). This regain of focus has highlighted the gaps in the understanding of adaptation in natural populations: it is still challenging to predict, or even understand retrospectively, how natural populations respond to selective pressures (Merilä, Sheldon, and Kruuk 2001; Tafani et al. 2013; Shaw and Shaw 2014; Brookfield 2016).

In order to study the evolutionary potential of wild populations and their response to selective pressure, it is necessary to measure genetic parameters. More specifically, one must determine whether the traits under selection are heritable, whether there is heritable variation in fitness and how what is the rate of genetic change for the traits of interest.

1.2 Measuring genetic variation

1.2.1 Looking up or down? Two philosophies

How to measure and make sense of genetic variation? For over a century, there have been two main approaches (Liedvogel, Cornwallis, and Sheldon 2012), that can grossly be traced back to the scientific controversy that opposed the Mendelians to the biometricalists (Dietrich 2006), and summarized as “bottom-up” and “top-down”. Bottom-up approaches, embodied by candidate gene and genome wide association studies, start from molecular data to infer the phenotypic effects of individual genetic loci. Top-down approaches, encompassed within quantitative genetics, attempt to decompose

²In this thesis, unless mentioned otherwise, we consider *sexual selection* as part of *natural selection* and of *selection*. Measuring sexual and natural selection separately, would certainly provide a finer understanding of the mechanisms of selection in the study population, but this was beyond the scope of this thesis. Nevertheless, the question was partly explored by García-Navas, Bonnet, Waldvogel, et al. 2016 and García-Navas, Bonnet, Bonal, et al. 2016.

phenotypic variation into genetic variation and other sources of variation, based solely on phenotypic data and on some knowledge of the relatedness between individuals (Lynch and Walsh 1998). Some pros and cons of both approaches are nicely illustrated by the confrontation of the quantitative genetics of mass with the genotyping of a candidate gene for mass. The former will be further developed in chapters 4 and 5 and we present it in a minimalist nutshell here: using a quantitative genetics *animal model* (Henderson 1950; Kruuk 2004), we estimated additive genetic variation in body mass and lifetime reproductive success. The latter is a side project of this PhD that does not appear in the other chapters, and we take the opportunity to present it below.

1.2.2 A candidate gene for body mass: insights and limits

We used a candidate gene approach (Fitzpatrick et al. 2005) to uncover the molecular mechanisms underlying variation in body mass. To date, the only candidate gene we fully analysed is an intronic region of the gene *lepr*, which codes for the receptor to leptin. Leptin is a hormone known to regulate fat metabolism, energy expenditure and food intake, including in rodents (Houseknecht et al. 1998).

We found a recessive allele (let call the recessive allele *a*, and the dominant allele *A*) associated with lighter individuals (Fig. 1.1A). Homozygotes *aa* were -2.9 g lighter (95% credibility interval [0.6; 5.1]), that is, 8% lighter than the mean. On average, during their lifetime, these *aa* individuals produced one third less offspring than the *AA* individuals (Fig. 1.1B). This strong difference in fitness is however not statistically significant, meaning that it could very well be the result of chance alone. These results suggest that some of the genetic variation in body mass is due to food intake and/or fat metabolism, which could not be sensed from the estimation of genetic variances and covariances. Based on such a strong phenotypic effect, *lepr* could be called a major locus, but how much of the genetic variation does it explain?

Knowing the effect of the three genotypes and the allele frequencies one can compute analytically the additive genetic variances associated with a bi-allelic locus (R. A. Fisher 1941; Lynch and Walsh 1998, p77). Thus, the additive genetic variances associated with *lepr* are 0.052g^2 for body mass and 0.006pup^2 for lifetime reproductive success. For both traits, *lepr* explains about 1% of the additive genetic variation as estimated from an animal model. This is rather large for a single locus given that quantitative traits loci typically explain a fraction of a percent to a few percent of additive genetic variance (V_A), when they have a large enough sample size to mitigate Beavis effect (Flint and Mackay 2009; Jensen, Szulkin, and Slate 2014). Still, 1% of V_A is not sufficient to infer the evolutionary potential of the trait. Finally, genotyping many more markers, for instance using high-throughput sequencing (Goodwin, McPhereson, and McCombie 2016), is unlikely to improve this situation in the snow vole population. Generally, very large sample sizes and high-quality genomic resources are necessary to explain a biologically proportion of genetic variances (Bloom et al. 2013; Jensen, Szulkin, and Slate 2014). For instance 183,727 individuals were necessary to find 180 QTL that jointly explained only 13% of additive genetic variation in human body height (Lango Allen et al. 2010). High-throughput sequencing can also been used in a top-down way, that does not identify causal genetic variants, but instead quantifies the phenotypic variation jointly explained by all the genotyped markers. Thus,

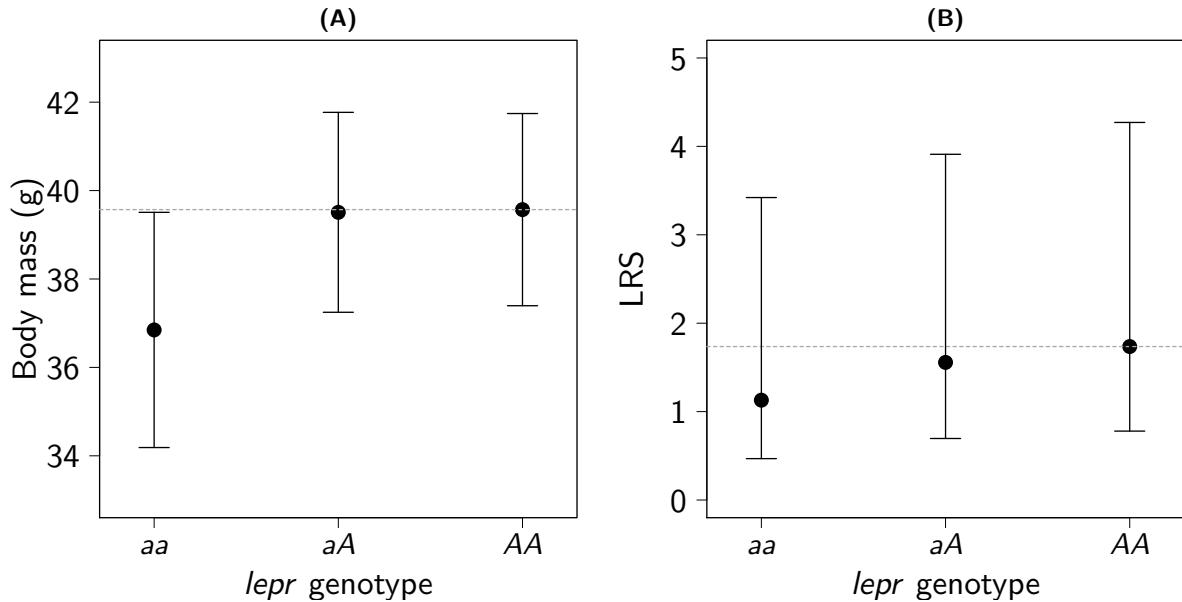


Figure 1.1: Body mass and lifetime reproductive success (LRS) as a function of *lepr* genotypes. (A) Expected body mass of snow voles bearing the three *lepr* genotypes. The expectations and 95% confidence intervals were predicted from a linear mixed model fitted to the 2311 mass measurement of 532 snow voles. The model accounted for sex, age, date of capture and their two-ways interactions, as well as year of capture and multiple measurements of the same individual. (B) Expected LRS of snow voles bearing the three *lepr* genotypes. The expectations and 95% confidence intervals were predicted from a Poisson generalized linear mixed model fitted to the LRS of 611 snow voles. The model accounted for inbreeding coefficient, year of birth and over-dispersion (using an observation-level random effect). For both panels, the dashed horizontal line projects the expected value of genotype *AA* to ease comparison with *Aa* and *aa*.

3,925 individuals and 294,831 markers were able to explain 45% of the genetic variation in human height (Yang et al. 2010). This is much better, but given knowledge on the relatedness between individuals, quantitative genetics can estimate all the genetic variation in a phenotype, without any genotyping effort.

To conclude, bottom-up approaches can better unravel the molecular mechanisms underlying phenotypes. By opening the black box of what mechanisms make up a phenotype, they can identify what is most crucial in a phenotype, how it is linked to the environment and what is the target of natural selection (Jong et al. 2014). Moreover, they contribute to building a genotype-phenotype map, a long lasting challenge in evolutionary biology (Kirschner and Gerhart 2010). On the other hand, quantitative genetics lump all the effects of individual genes and their interactions into only a few parameters, non-informative about the underlying genetic architecture (Mackay 2001; Nietlisbach and Hadfield 2015; Huang and Mackay 2016). This summarized estimation provides simple and direct measures of genetic parameters. Quantitative genetics work directly on the phenotype which is the target of selection, and the source of eco-



logical interactions. They therefore provide simple measures of genetic parameters that can directly be interpreted within the ecology of organisms. This thesis is concerned with the genetics and evolution at the level of organisms, in relation to their environment, and accordingly, most of my work relies on quantitative genetics.

1.3 This thesis

1.3.1 Objectives

In this thesis, I investigate the proximal causes of individual-level variation in fitness, and the consequences of this variation at the population level. This thesis aims at better measuring and better understanding selection and evolution in the wild. It examines the relative importance of stochasticity and selection in shaping reproductive success and survival, disentangles evolutionary from plastic changes and explores the link between selection and evolution. These questions are addressed using a combination of computer simulations and of data from the long-term individual-based monitoring of a snow vole population.

1.3.2 Snow voles in Churwalden

The snow vole (*Chionomys nivalis*, Martins 1842) is a medium-sized rodent, its adult body size ranges from 10 to 14 cm, without the tail (5 to 7.5 cm long). Contrary to a widespread intuition, snow voles are not white (Fig. 1.2). Instead, the fur colour of the upper-parts varies from light to dark taupe grey, sometimes tinted with brown or dark red. The misconception about color highlights that the species could favourably be renamed *rock vole*: it is a rock, rather than a snow, specialist (Luque-larena, López, and Gosálbez 2002) and might be associated with high elevations only because rocky areas are more widespread there. It is sparsely distributed across southern Europe and Asia Minor, from sea level up to 4000 m of elevation (Janeau and Aulagnier 1997).

Snow voles excavate burrows under the rocks, but can also use natural clefts between rocks, sometimes carrying small stones to build walls (Niederer 2008). A burrow consists of tunnels connecting chambers, one for the nest and multiple ones to stock dry plants (Janeau and Aulagnier 1997). The species is not known to hibernate and is therefore exposed to harsh winter conditions in its high-elevation range. Adult females actively defend small territories against non-relatives, and tend to form matrilineal clusters of territories, whereas adult males wander, and fight, across large overlapping home-ranges (Luque-larena, López, and Gosálbez 2004; García-Navas, Bonnet, Waldvogel, et al. 2016). The mating system is promiscuous and a same litter can be sired by multiple males. Females normally produce 1 to 4 litters of 1 to 5 pups between May and September. Juveniles generally do not reproduce in their first civil year. Although they can eat flour worms in the lab, there is no evidence that snow voles are not strictly herbivorous in the wild (Janeau and Aulagnier 1997). In the Swiss Alps, snow voles suffer predation from red foxes, stoats, various owls and corvids, and parasitism from fleas, lices and ticks (Janeau and Aulagnier 1997; Martinoli et al. 2001).





Figure 1.2: Juvenile (left picture) and adult (right picture) snow voles in their habitat in Churwalden, Switzerland. Juveniles always lack the brown hue generally found in adults. Neither adults nor juveniles are white.

The study area is located by the Churer Joch, Churwalden, in the Swiss canton Graubünden (Fig. 1.3; coordinates 46°48' N, 9°34' E), and covers about 5 ha between 1980 m and 2100 m above sea level. It consists of a west-exposed scree interspersed with small coniferous trees and with patches of alpine meadow. The study area is demarcated by extensive meadows to the south and to the north, by a coniferous forest to the west and by cliff to the east (Fig. reffig:landscape).

Another scree, called Wolfgruben, offers about 1 ha of favourable habitat, starting 300 m north-east to the monitored area. Wolfgruben was trapped in 2008 and 2013. The snow vole density was rather low, with on average five captures per night of trapping, versus 18 on the main study area. More habitat favourable to snow voles can be found 2 Km to the south. The study population is moderately isolated and receives 5 to 10 immigrants per year, on a total of 60 to 180 individuals (García-Navas, Bonnet, Waldvogel, et al. 2016).

The monitoring of this snow vole population was initiated in 2006 by Dr. Peter W. Wandeler. Dr. Erik Postma took the monitoring over in 2012, but the protocol has remained practically unchanged. This thesis contains data collected up to the year 2015. Every year from 2006 to 2016, snow voles were life-trapped multiple times between late May and early October. Traps were set during the day, opened around sunset and checked the next morning. For every snow vole capture³, we recorded sex, age, body mass, body length, tail length, date, location and signs of reproductive activity (pregnancy, lactation, swollen scrotum). In addition, all newly-captured snow voles were individually marked and genotyped for 18 microsatellites (Wandeler, Ravaiolli, and Bucher 2008). Based on the autosomal microsatellite genotypes, we reconstruct the pedigree of the population. This pedigree is the raw material for most of the work carried out during this thesis. In particular, the pedigree is used to define reproductive success, as well as to estimate the relatedness between all pairs of individuals. These two statistics are essential to estimate selection, fitness and genetic variation.

³Other species (bank voles, pine voles, wood mice, stoats, black salamanders, slugs...) were released without taking measurements.

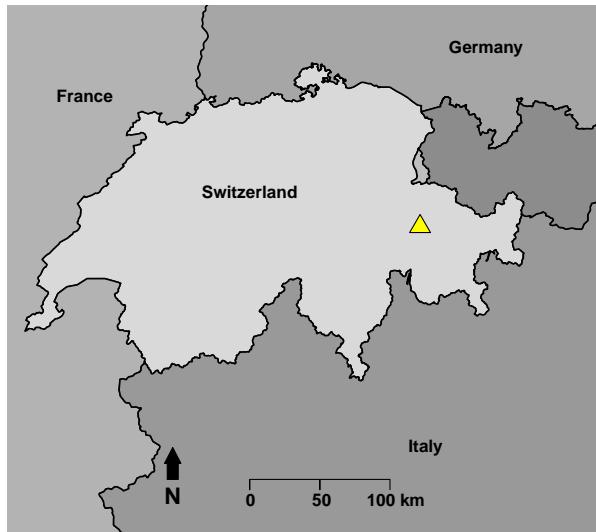


Figure 1.3: Location of the study area in Switzerland. The yellow triangle indicates the study area, with coordinates $46^{\circ}48' N$, $9^{\circ}34' E$, by Churwalden, in the Swiss canton Graubünden. Countries are filled with different shades of grey, Austria and Lichtenstein are not labelled.

1.3.3 Thesis outline

In natural populations, fitness is generally measured using individual measures of reproductive success and survival. These proxies are not fitness itself and their variation is largely stochastic, leading some authors to doubt that there is any significant variation in fitness in natural populations. Recent methodological developments appeared to support the view that variation in reproduction and survival was purely stochastic, and suggested that the potential for selection and evolution in the wild was largely over-estimated. In **chapter 2** I examine these methods and, based on computer simulations, demonstrate that they lack statistical power to detect latent variation in fitness components. Using an alternative approach we show the presence of significant variation in the propensity of reproductive success in the snow vole population, thereby showing some potential for selection and adaptive evolution in this population. We also attempt to clarify some conceptual misunderstandings between the proponents of the two methodological schools.

A similar attempt motivated **chapter 3**, where, with collaborators from different methodological schools, we review and compare four frameworks to disentangle the causes of phenotypic changes. In particular, these frameworks differ in their estimation of the relative roles of plasticity, demography and genetic change. Based on computer simulations and on mathematical comparisons, we show that the discrepancies between the frameworks primarily originates from different definitions of the com-



Figure 1.4: Distant view of the field site, taken from the west. The trapped area covers about a fifth of the width and a tenth of the height of the picture and is located in the centre. This scree is surrounded by a forest, a cliff and meadows.

ponents of change. Nevertheless, one of these frameworks, the quantitative genetics *animal model*, stands out as the only framework able to estimate genetic change and the response to selection (that is, the trans-generational consequence of variation in fitness). I relied heavily on this framework for the two next chapters.

In chapter 4, I explore the reasons of the mismatch between apparent phenotypic selection, phenotypic change and genetic change for body mass. I describe one of the first case of contemporary evolution of a quantitative trait in the wild and show that this genetic change is adaptive. Both the evolution and the selective pressure responsible for it are invisible to purely phenotypic approaches, however. Using multivariate animal models, we identify the main component of selection as juvenile viability. I then infer that the target of selection is potential adult mass in juveniles and that selection is related to a recent change in climatic conditions.

The previous chapter considered selection and evolution averaged over the whole study period, without considering their temporal dynamic within the period. The fluctuation of selection is thought to be a major determinant of the rate of evolution, and a process to consider to understand adaptation in the wild. Nevertheless, unbiased measures of the variation of selection are rare and of the coupling between variation in selection and variation in evolution has been largely ignored. Chapter 5 shows that selection fluctuates in the study population, mainly due to variation in fertility selection. The rate of adaptive evolution is, however, remarkably constant, because viability selection, the driver of body mass evolution, does not vary. In this case the fluctuation of selection is evolutionary irrelevant. These two last chapters

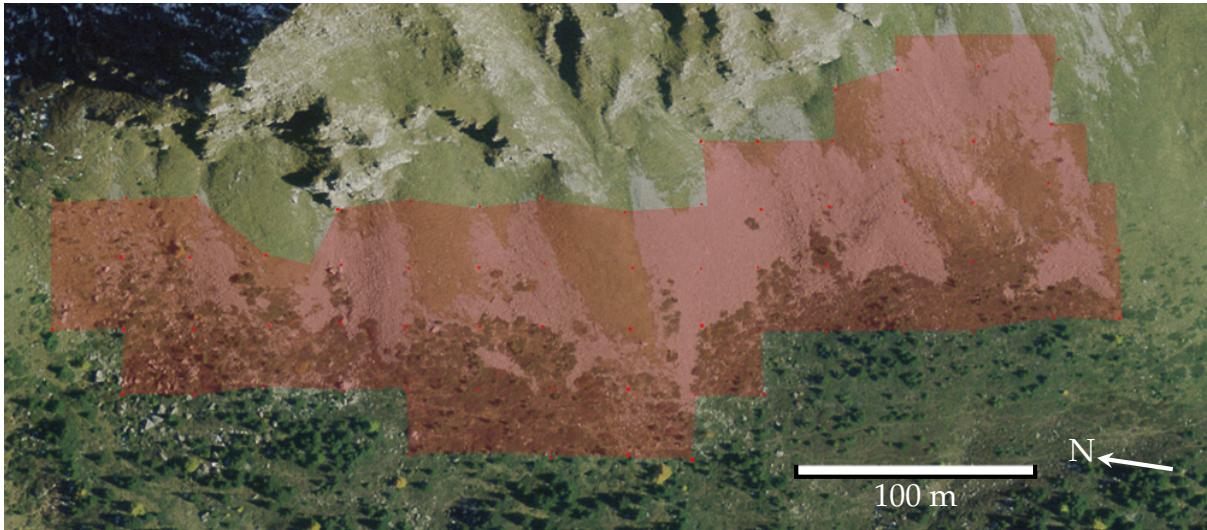


Figure 1.5: Orthophoto of the study site, from 2008. The red shading indicates the approximate area where traps are set.

highlight the dangers of relying on phenotypic estimates of selection to understand the evolutionary dynamics of natural populations.

Finally, in **chapter 6**, I summarize the progresses made during this PhD on the understanding of natural populations, and discuss some of the remaining challenges and future working directions.

References

- Akçay, E. and Van Cleve, J. 2016. "There is no fitness but fitness, and the lineage is its bearer." *Philosophical Transactions of the Royal Society B: Biological Sciences* **371** (1687): 20150085.
- Altermatt, F., Ebert, D., and Altermatt, F. 2016. "Reduced flight-to-light behaviour of moth populations exposed to long-term urban light pollution." *Biology Letters* **12**:201160111.
- Arnold, S. J. and Wade, M. J. 1984. "On the measurement of natural and sexual selection: Theory." *Evolution* **38** (4): 709–719.
- Bloom, J. S., Ehrenreich, I. M., Loo, W. T., Lite, T.-L. V., and Kruglyak, L. 2013. "Finding the sources of missing heritability in a yeast cross." *Nature* **494**:234–239.
- Brandon, R. and Beatty, J. 1984. "The propensity interpretation of 'fitness'. No interpretation is no substitute." *Philosophy of Science* **51** (2): 342–347.
- Brookfield, J. F. 2016. "Why are estimates of the strength and direction of natural selection from wild populations not congruent with observed rates of phenotypic change?" *BioEssays* **38**:1–8.

Chapter 1 General introduction

- Cam, E. and Monnat, J. Y. 2000. "Stratification based on reproductive state reveals contrasting patterns of age-related variation in demographic parameters in the kittiwake." *Oikos* **90** (3): 560–574.
- Darwin, C. 1859. *On the origin of species by means of natural selection, or the preservation of favoured races in the struggle for life*. John Murray, London.
- Dietrich, M. R. 2006. "From Mendel to molecules: A brief history of evolutionary genetics." In *Evolutionary genetics: concepts and case studies*. Edited by C. W. Fox and J. B. Wolf, 3–13. Oxford University Press, Oxford.
- Fisher, R. A. 1941. "Average excess and average effect of a gene substitution." *Annals of Eugenics* **11** (1): 53–63.
- Fisher, R. 1930. *The genetical theory of natural selection*. First. Oxford university press, Oxford, Great Britain.
- Fitzpatrick, M. J., Ben-shahar, Y., Smid, H. M., Vet, L. E. M., Robinson, G. E., and Sokolowski, M. B. 2005. "Candidate genes for behavioural ecology." *Trends in Ecology & Evolution* **20** (2): 96–104.
- Flint, J. and Mackay, T. F. C. 2009. "Genetic architecture of quantitative traits in flies, mice and humans." *Genome Research* **19**:723–733.
- Frank, S. a. 2012. "Natural selection. V. How to read the fundamental equations of evolutionary change in terms of information theory." *Journal of Evolutionary Biology* **25** (12): 2377–2396.
- García-Navas, V., Bonnet, T., Bonal, R., and Postma, E. 2016. "The role of fecundity and sexual selection in the evolution of size and sexual size dimorphism in New World and Old World voles (Rodentia: Arvicolinae)." *Oikos* **early view**.
- García-Navas, V., Bonnet, T., Waldvogel, D., Camenisch, G., and Postma, E. 2016. "Consequences of natal philopatry for reproductive success and mate choice in an Alpine rodent." *Behavioral Ecology* **27** (4): 1158–1166.
- Goodwin, S., McPherson, J. D., and McCombie, W. R. 2016. "Coming of age: ten years of next-generation sequencing technologies." *Nature Reviews Genetics* **17** (6): 333–351.
- Henderson, C. 1950. "Estimation of genetic parameters." *Annals of mathematical statistics* **21** (2): 309–310.
- Houseknecht, K. L., Baile, C. A., Matteri, R. L., and Spurlock, M. E. 1998. "The biology of leptin : a review." *Journal of animal science* **76**:1405–1420.
- Huang, W. and Mackay, T. F. 2016. "The genetic architecture of quantitative traits cannot be inferred from variance component analysis." *bioRxiv*. eprint: <http://biorxiv.org/content/early/2016/02/26/041434.full.pdf>.
- Janeau, G. and Aulagnier, S. 1997. "Snow vole - *Chionomys nivalis* (Martins 1842)." *IBEX Journal of Mountain Ecology* **4**:1–11.

- Jensen, H., Szulkin, M., and Slate, J. 2014. "Molecular quantitative genetics." In *Quantitative Genetics in the Wild*, edited by A. Charmantier, D. Garant, and L. E. B. Kruuk. Oxford: Oxford University Press.
- Jones, J. S. 1987. "The heritability of fitness: bad news for good genes?" *Trends in Ecology and Evolution* **2** (2): 35–36.
- Jong, M. a. de, Wong, S. C., Lehtonen, R., and Hanski, I. 2014. "Cytochrome P450 gene Cyp337 and heritability of fitness traits in the Glanville fritillary butterfly." *Molecular Ecology* **23** (8): 1994–2005.
- Kirschner, M. W. and Gerhart, J. C. 2010. "Facilitated variation." In *Evolution. The extended synthesis*, edited by M. Pigliucci and G. B. Müller. The MIT press, Cambridge, Massachusetts, USA.
- Krimbas, C. B. 2004. "On fitness." *Biology & Philosophy* **19** (2): 185–203.
- Kruuk, L. E. B. 2004. "Estimating genetic parameters in natural populations using the 'animal model'." *Philosophical Transactions of the Royal Society B: Biological Sciences* **359** (1446): 873–890.
- Kruuk, L. E. B., Charmantier, A., and Garant, D. 2014. "The study of quantitative genetics in wild populations." In *Quantitative Genetics in the Wild*, edited by A. Charmantier, D. Garant, and L. E. B. Kruuk. Oxford: Oxford University Press.
- Lande, R. and Arnold, S. J. 1983. "The measurement of selection on correlated characters." *Evolution* **37** (6): 1210–1226.
- Lande, R. 1979. "Quantitative genetic analysis of multivariate evolution, applied to brain: body size allometry." *Evolution*: 402–416.
- Lango Allen, H., Estrada, K., Lettre, G., Berndt, S. I., Weedon, M. N., Rivadeneira, F., Willer, C. J., et al. 2010. "Hundreds of variants clustered in genomic loci and biological pathways affect human height." *Nature* **467** (7317): 832–8.
- Liedvogel, M., Cornwallis, C. K., and Sheldon, B. C. 2012. "Integrating candidate gene and quantitative genetic approaches to understand variation in timing of breeding in wild tit populations." *Journal of evolutionary biology* **25** (5): 813–23.
- Luque-larena, J. J., López, P., and Gosálbez, J. 2002. "Microhabitat use by the snow vole *Chionomys nivalis* in alpine environments reflects rock-dwelling preferences." *Canadian Journal of Zoology* **80**:36–41.
- . 2004. "Spacing behavior and morphology predict promiscuous mating strategies in the rock- dwelling snow vole, *Chionomys nivalis*." *Canadian Journal of Zoology* **82**:1051–1060.
- Lush, J. 1937. *Animal breeding plans*. Ames, Iowa: Iowa State College Press.
- Lynch, M. and Walsh, B. 1998. *Genetics and Analysis of Quantitative Traits*. Sinauer Associates.

Chapter 1 General introduction

- Mackay, T. 2001. "The genetic architecture of quantitative traits." *Annu Rev Genet* **35**:303–309.
- Martinoli, A., Preatoni, D. G., Chiarenzi, B., Wauters, L. a., and Tosi, G. 2001. "Diet of stoats (*Mustela erminea*) in an Alpine habitat: The importance of fruit consumption in summer." *Acta Oecologica* **22** (1): 45–53.
- Mc Carty, J. P. 2001. "Ecological Consequences of Recent Climate Change." *Conservation biology* **15** (2): 320–331.
- Merilä, J., Sheldon, B. C., and Kruuk, L. E. B. 2001. "Explaining stasis: microevolutionary studies in natural populations." *Genetica* **112**:199–222.
- Niederer, A. 2008. "Das Verhalten der Schneemaus." PhD diss., Universität Basel.
- Nietlisbach, P. and Hadfield, J. D. 2015. "Heritability of heterozygosity offers a new way of understanding why dominant gene action contributes to additive genetic variance." *Evolution*: 1–16.
- Oswald, T., MacLeod, C., and McCarty, M. 1944. "Studies on the chemical nature of the substance inducing transformation of pneumococcal types: induction of transformation by a desoxyribonucleic acid fraction isolated from pneumococcus type III." *Journal of Experimental Medicine* **79** (2).
- Parmesan, C. 2006. "Ecological and evolutionary responses to recent climate change." *Annual Review of Ecology, Evolution, and Systematics*: 637–669.
- Pigliucci, M. and Müller, G. B. 2010. *Evolution. The extended synthesis*. The MIT press, Cambridge, Massachusetts, USA.
- Postma, E. 2014. "Four decades of estimating heritabilities in wild vertebrate populations: improved methods, more data better estimates?" In *Quantitative Genetics in the Wild*, edited by A. Charmentier, D. Garant, and L. E. B. Kruuk. Oxford: Oxford University Press.
- Price, P. W. 1996. *Biological evolution*. Saunders College Publishing, Philadelphia, PA.
- Rousset, F. 2004. *Genetic structure and selection in subdivided populations*. Princeton, New Jersey, USA.
- Shaw, R. G. and Shaw, F. H. 2014. "Quantitative genetic study of the adaptive process." *Heredity* **112** (1): 13–20.
- Steiner, U. K. and Tuljapurkar, S. 2012. "Neutral theory for life histories and individual variability in fitness components." *Proceedings of the National Academy of Sciences, USA* **109** (12): 4684–4689.
- Tafani, M., Cohas, A., Bonenfant, C., Gaillard, J.-M., and Allainé, D. 2013. "Decreasing litter size of marmots over time: a life history response to climate change ?" *Ecology* **94** (3): 580–586.
- Tienderen, P. H. van. 2000. "Elasticities and the link between demographic and evolutionary dynamics." *Ecology* **81** (3): 666–679.

- Turner, B. M. 2009. "Epigenetic responses to environmental change and their evolutionary implications." *Philosophical Transactions of the Royal Society B: Biological Sciences* **364** (1534): 3403–18.
- Wade, M. J. 2006. "Natural selection." In *Evolutionary genetics: concepts and case studies*. Edited by C. W. Fox and J. B. Wolf, 399–413. Oxford University Press, Oxford.
- Wagner, G. P. 2010. "The measurement theory of fitness." *Evolution* **64** (5): 1358–76.
- Wandeler, P., Ravaioli, R., and Bucher, T. B. 2008. "Microsatellite DNA markers for the snow vole (*Chionomys nivalis*)." *Mol Ecol Res* **8**:637–639.
- Watson, J. and Crick, F. 1953. "Genetical implications of the structure of deoxyribonucleic acid." *Nature* **171** (4361): 964–967.
- Wayne, M. L. and Miyamoto, M. M. 2006. "Genetic variation." In *Evolutionary genetics: concepts and case studies*. Edited by C. W. Fox and J. B. Wolf, 399–413. Oxford University Press, Oxford.
- Wilkins, J. S. 2009. *Species. A history of the idea*. Berkeley, California, USA: University of California Press.
- Wolf, J. B. and Wade, M. J. 2009. "What are maternal effects (and what are they not)?" *Philosophical Transactions of the Royal Society B: Biological Sciences* **364** (1520): 1107–15.
- Yang, J., Benyamin, B., McEvoy, B. P., Gordon, S., Henders, A. K., Nyholt, D. R., Madden, P. A., et al. 2010. "Common SNPs explain a large proportion of the heritability for human height." *Nat Gen* **42** (7): 565–569.

Chapter 2

Successful by chance? The power of mixed models and neutral simulations for the detection of individual fixed heterogeneity in fitness components

If the talents I was born with are the right ones, I may someday achieve my goal. If not, I may go through life being as stupid as I am now.

— Eiji Yoshikawa, *Musashi* (1935)

Quand on veut comprendre une chose, on se place en face d'elle, tout seul, sans secours; tout le passé du monde ne pourrait servir de rien. Et puis elle disparaît et ce qu'on a compris disparaît avec elle.

— Jean-Paul Sartre, *La nausée* (1938)

Timothée Bonnet* and Erik Postma (2016) The American Naturalist 187(1):60-74

2.1 Abstract

Heterogeneity in fitness components consists of fixed heterogeneity due to latent differences fixed throughout life (e.g. genetic variation), and dynamic heterogeneity generated by stochastic variation. Their relative magnitude is crucial for evolutionary processes, as only the former may allow for adaptation. However, the importance of fixed heterogeneity in small populations has recently been questioned. Using neutral simulations (NS), several studies failed to detect fixed heterogeneity, thus challenging previous results from mixed models (MM). To understand the causes of this discrepancy, we estimate the statistical power and false positive rate of both methods, and apply them to empirical data from a wild rodent population. While MM show high false positive rates if confounding factors are not accounted for, they have high statistical power to detect real fixed heterogeneity. In contrast, NS are also subject to high false positive rates, but always have low power. Indeed, MM analyses of the rodent population data show significant fixed heterogeneity in reproductive success, whereas NS analyses do not. We suggest that fixed heterogeneity may be more common than

is suggested by NS, and that NS are useful only if more powerful methods are not applicable and if they are complemented by a power analysis.

Keywords: *Chionomys nivalis*; individual-based model; generalized linear mixed model; simulations; snow vole; statistical power

Online enhancements: Online appendices. Data available from the Dryad Digital Repository: <http://dx.doi.org/10.5061/dryad.3cb61>.

2.2 Introduction

Within species, individual variation in lifetime reproductive success (LRS) is plentiful, with most individuals producing few or no offspring and a few individuals producing a large share of the next generation (Clutton-Brock 1988; Stearns 1992). Given their skewed and heterogeneous nature, LRS distributions are unlikely to be solely shaped by unstructured environmental stochasticity. Instead, individuals seem to differ in their probability of surviving or reproducing (Kendall et al. 2011).

Often, this individual heterogeneity in LRS is assumed to originate from latent individual differences which are fixed throughout an individual's life, i.e. that there is individual heterogeneity in frailty, quality or fitness (e.g. Vaupel, Manton, and Stallard 1979; Morris 1998; Cam and Monnat 2000). This is commonly referred to as fixed heterogeneity. Genetic variation is one source of fixed heterogeneity (e.g. Keller and Waller 2002; Ellegren and Sheldon 2008), but epigenetic, maternal and permanent environmental effects may also be important (Wolf and Wade 2009; Turner 2009). This fixed variation is usually measured retrospectively; in some cases it may have arisen at fertilization, but it may also be shaped by the environment an individual experiences throughout its life, for instance through variation in habitat choice or through gene by environment interactions. It is important to distinguish fixed heterogeneity as it is used here—that is, the repeatability of individual performance—from other sources of variation that are not due to the properties of individuals (e.g. climatic variations among years). Indeed, only fixed differences among individuals can be the target of selection and allow for adaptation, provided that these fixed differences are passed on to the next generation—be it through genes (Keller and Waller 2002), philopatry (Schauber et al. 2007) or other processes (Bonduriansky 2012).

Recent publications (Tuljapurkar, Steiner, and Orzack 2009; Steiner, Tuljapurkar, and Orzack 2010; Orzack et al. 2011; Steiner and Tuljapurkar 2012) have argued forcefully that invoking fixed differences among individuals (i.e. fixed heterogeneity) in fitness components is rarely required to explain the observed heterogeneity in LRS. Instead, they emphasize that due to the stochasticity of individual life histories, individual heterogeneity is expected even in populations of identical individuals (Caswell 2011). Indeed, if individuals take a random trajectory through the various life-history stages, and if these stages are associated with differential reproductive and survival rates, the population-level distribution of LRS may be skewed and heterogeneous. This type of heterogeneity is referred to as dynamic heterogeneity (Tuljapurkar, Steiner, and Orzack 2009). Crucially, dynamic heterogeneity originates from differences among life stages, whereas fixed heterogeneity originates from variation in the properties of individuals.

Given that most life-history traits are heritable to some degree (Mousseau and Roff 1987; Postma 2014), it is beyond doubt that some fixed heterogeneity is present in most wild populations. At the same time, the cumulative effects of individual histories on their realized lifespan and reproductive success are also unquestionable (Caswell 2011). What is subject to discussion, however, is the relative importance of fixed, versus dynamic, heterogeneity in shaping variation in LRS. Steiner and Tuljapurkar 2012 suggested that, at least in small populations, the drift generated by large life-history stochasticity is too large for fixed heterogeneity to play a significant role in shaping evolution and demography at the level of a single population. Instead, they have proposed dynamic heterogeneity as the null model to explain any observed heterogeneity. Only if this null model can be rejected should we consider an additional role for fixed heterogeneity in shaping variation in LRS or fitness components.

Tuljapurkar, Steiner, and Orzack 2009 have suggested that an appropriate tool to test for fixed heterogeneity is provided by neutral simulations (NS hereafter), which generate summary statistics describing the distribution of LRS and the pattern of life-stage transitions expected in the absence of fixed heterogeneity. These expectations can subsequently be compared to their observed counterparts to detect departures from neutrality due to the existence of fixed heterogeneity.

The application of NS to data for two sea bird populations (Steiner, Tuljapurkar, and Orzack 2010; Orzack et al. 2011), as well as to a compilation of 22 vertebrate populations (Tuljapurkar, Steiner, and Orzack 2009) has been unable to reject the null hypothesis of neutrality, leading to the conclusion that dynamic heterogeneity alone can explain the observed variation in life histories in most populations. Indeed, we are aware of only one study in which NS rejected neutrality, for one of three reproductive parameters in a roe deer population (Plard et al. 2012).

In contrast to studies relying on NS, studies employing linear mixed models (hereafter MM) commonly report evidence for fixed heterogeneity (e.g. Cam and Monnat 2000; Royle 2008; Chambert et al. 2013; Guillemain et al. 2013; Chambert, Rotella, and Higgs 2014). Interestingly, Cam et al. 2013 have provided evidence for fixed heterogeneity in a data set for which the existence of fixed heterogeneity had been dismissed based on NS (Steiner, Tuljapurkar, and Orzack 2010). However, MM and NS differ in how they deal with data: MM rely on repeated measurements of individuals, while NS use summary statistics aggregated at the population level. Compared to MM, NS are thus less data-demanding, but might be less sensitive to statistical signals at the individual level. On the other hand, aggregation might allow NS to detect effects that emerge only at the population level and are invisible to MM. More formally, the discrepancy between NS and MM suggests that they differ in either their type I (i.e. false positive) error rate, or in their type II error rate (i.e. power). For instance, the opposite conclusions reached by NS in Steiner, Tuljapurkar, and Orzack 2010 and MM in Cam et al. 2013 may be the result of the statistical power of the NS being too low, preventing the detection of fixed heterogeneity (i.e. a type II error). Alternatively, MM may have high rates of type I error, if the individual-level variances estimated by the MM are spurious, or they are unduly interpreted as the mark of fixed heterogeneity.

Applying both methods to data with known properties allows for the estimation of both types of error rates and thereby provides insight into the ability of both methods to detect fixed heterogeneity. Unfortunately however, fixed heterogeneity is the

result of latent, unobservable traits, which cannot be inferred without a modeling step (Cam et al. 2013), and it is precisely the performance of this modeling step that we investigate here. Computer simulations provide a way around this problem, as they allow one to apply methods to data sets with known underlying properties (e.g. Villemereuil, Gimenez, and Doligez 2013; Brooks, McCoy, and Bolker 2013).

Here, we simulate a series of longitudinal, individual-based, data sets through an algorithm that introduces varying amounts of fixed and dynamic heterogeneity in survival and reproduction. For illustrative purposes, these simulations are parametrized to match a population of snow voles (*Chionomys nivalis*, Martins 1842) located in the Swiss Alps. In order to assess the type I and type II error rates of both NS and MM, we subsequently analyze the simulated data sets using both methods. In a final step, we use these results to interpret the results of the application of both methods to the real snow vole data set. Figure 2.1 shows a diagram summarizing our approach. Altogether, our results highlight the lack of statistical power of NS, but at the same time emphasize that MM output should be interpreted with care. We discuss the origin of the discrepancy between NS and MM, and what this tells us about the nature of biological variability.

2.3 Material and methods

2.3.1 Data simulation

The simulation model matches the life cycle of the population of snow voles which we use in the empirical comparison of both methods. The monitoring of this population is discussed in some detail in Appendix 2.8.5. Only two age classes are modeled (non-reproducing juveniles and reproducing adults), and there are no sex-specific or spatio-temporal effects on fitness components, as the uncertainty with respect to the appropriate specification of these models would introduce an additional layer of complexity (see e.g. Cam et al. 2013). All simulated populations are monitored for 10 years. For every individual, we have perfect knowledge of survival and reproduction during the study period, but their fate beyond this period is unknown. Every year, a new cohort of 100 juveniles appears. After one year, these juveniles become adults and start reproducing. Every year, adults can reproduce once; the number of offspring produced by an individual is labeled annual reproductive success (ARS). In the real snow vole population, there is no apparent senescence in survival and the maximum age observed is four years old. Accordingly, in the simulations, adult survival probability does not vary with age until the fourth year, but all individuals still alive at that point die during the next winter. Mortality events occur after birth for juveniles and after reproduction for adults. A single sex is simulated, as the two sexes are generally analyzed separately in NS, and in MM sex differences in the mean are accounted for by fitting sex as a fixed factor.

We define a scenario as a collection of simulation parameters. For each scenario, 1000 data sets were simulated, that is 1000 putative populations with the same underlying properties. In an attempt to detect evidence for fixed heterogeneity, each data set was then analyzed using MM and NS. Note the potential for confusion between

the simulation of the data sets on the one hand, and the neutral simulation method on the other. The latter is always referred to as NS. Simulations were carried out using a C++ program (available at <https://github.com/timotheenivalis/FixDynHet>), using the pseudo-random number generator Mersenne Twister (Matsumoto and Nishimura 1998) and a command file procedure following that of IBDsim (Leblois, Estoup, and Rousset 2009). The analyses of the simulation output were all conducted in R 3.1.0 (R Core Team 2014), using the package lme4 (version 1.1-7) (Bates et al. 2015).

Due to demographic stochasticity (sensu Fox and Kendall 2002), all simulated data sets contain a baseline level of dynamic heterogeneity. Indeed, according to Tuljapurkar, Steiner, and Orzack 2009, the presence of dynamic heterogeneity results in the “scaled sequence entropy of the transition matrix between reproductive stages” (hereafter simply referred to as entropy), being greater than zero, which is always the case here. Entropy measures the rate at which the diversity of life-history trajectories increases with their length, which is due to random transitions between stages with different survival probabilities and reproductive outcomes (Tuljapurkar, Steiner, and Orzack 2009).

Beyond this baseline level of dynamic heterogeneity, heterogeneity in fitness components is introduced either as explicit fixed heterogeneity, or through a Markovian process. For the simulation of fixed heterogeneity, at birth, each individual receives a fixed quality as reproducer and survivor. These fixed qualities do not change over the course of its life. Therefore, some individuals intrinsically have a high probability to perform well, and some individuals have a high probability to perform poorly, irrespective of their past performance, as in a classic frailty model (Vaupel, Manton, and Stallard 1979). In contrast, for the simulations using a Markovian process, an individual’s probability to survive and to achieve a certain ARS is not fixed, but changes at each time step and depends solely on its ARS the time step before. Therefore, these data contain dynamic heterogeneity only. However, some of this mimics fixed heterogeneity because individual performances can persist over time. Generalized linear mixed models were used to check that the properties of the simulated data sets matched the model and the parameters used to generate them (see Appendix 2.8.1).

Simulations with explicit fixed heterogeneity At birth, every individual receives a quality as reproducer $q_{\rho,i}$, which is normally distributed with a mean of 0 and a variance equal to σ_{ρ}^2 , i.e. $q_{\rho,i} \sim \mathcal{N}(0, \sigma_{\rho}^2)$. Individuals also receive a quality as survivor $q_{\phi,i}$, with $q_{\phi,i} \sim \mathcal{N}(0, \sigma_{\phi}^2)$. These qualities are fixed for the lifetime of an individual. Because trade-offs between survival and reproduction are not considered here, the two qualities are drawn independently for each individual. The variances σ_{ρ}^2 and σ_{ϕ}^2 represent the amount of fixed heterogeneity in reproduction and survival, respectively.

If individual i is an adult at time t , its annual reproductive success, $\rho_{i,t}$, is drawn from a Poisson distribution,

$$\rho_{i,t} \sim \mathcal{P}(\exp(\log(\mu_{\rho}) + q_{\rho,i})), \quad (2.1)$$

where μ_{ρ} is the mean annual reproductive success. For an individual with $q_{\rho,i} = 0$, i.e. the average individual in a population with fixed heterogeneity, the parameter of the Poisson distribution ($\exp(\log(\mu_{\rho}) + q_{\rho,i})$) reduces to the population mean ARS (μ_{ρ}).

The qualities for reproduction ($q_{\rho,..}$) are normally distributed on the log-transformed scale of ARS.

The survival outcome of an individual i at time t , $\phi_{i,t}$, is zero (death) if the individual is four years old, and otherwise is drawn from a Bernoulli distribution:

$$\phi_{i,t} \sim \mathcal{B}(\text{logit}^{-1}(\text{logit}(\mu_\phi + j_{i,t}\beta_j) + q_{\phi,i})), \quad (2.2)$$

where $\text{logit}(p) = \log(\frac{p}{1-p})$ and its inverse function $\text{logit}^{-1}(x) = \frac{1}{1+\exp(-x)}$, where $j_{i,t}$ is a Boolean variable equal to 0 for adults and 1 for juveniles, and where β_j is the difference between the mean survival probability of juveniles and adults. For an individual with $q_{\phi,i} = 0$, the probability of survival ($\text{logit}^{-1}(\text{logit}(\mu_\phi + j_{i,t}\beta_j) + q_{\phi,i})$) reduces to $(\mu_\phi + j_{i,t}\beta_j)$, the age-specific mean survival probability. The qualities for survival ($q_{\phi,..}$) are normally distributed on the logit-transformed scale.

The mean of a log (or a logit) distribution is in general not equal to the log (or the logit) of the mean of this distribution (i.e. $\overline{\log(x)} \neq \log(\bar{x})$). Hence, Gaussian variance in individual qualities introduces a bias on the log or logit scale in the mean realized ARS and survival. If not corrected for, this bias causes the distributions of ARS and survival to deviate from their neutral expectations, which could be interpreted as evidence for fixed heterogeneity. To this end, the median individual qualities, \tilde{q}_ρ and \tilde{q}_ϕ , were iteratively modified so that the realized population means do not depend on the variances in individual qualities.

Because they are fixed for life, the individual qualities are the target of selection. Indeed, selection, i.e. the individual-level covariance between quality and relative LRS, increases with increasing variances (σ_ρ^2 and σ_ϕ^2) (Appendix 2.8.3). It could thus be argued that in response to this selection, mean latent qualities should increase and their variances decrease over time. However, here we chose not to simulate a trans-generational response to selection, as this introduces an unnecessary layer of complexity: First, a phenotypic response to selection on components of fitness is not necessarily expected. For example, environmental deterioration, which may be the result of an increase in mean competitiveness (Fisher 1958; Hadfield, Wilson, and Kruuk 2011), may mask a genetic change. Second, only the additive genetic part of the variation can respond to selection, and genetic variation may be renewed through migration, mutations and balancing selection (Fisher 1958; Charlesworth 2015). Therefore, simulating a response to selection would require much more complicated simulations and many more assumptions (e.g. an explicit genetic architecture for fitness, mechanisms to maintain genetic variation, competitive interactions). Finally, both MM and NS are blind to temporal variation, as they compute statistics averaged over the whole data set, and even if a response to selection were apparent, it would have little effect on their performance.

The simulation framework outlined above closely matches the structure of the MM later used to analyze the simulated data. Although we believe this simulation framework to be closest to biological reality, it could be argued that this may result in an overestimation of the ability of MM to deal with real data. Therefore, two alternative simulation structures not exactly matching the structure of MM were used. In the first, fixed heterogeneity was introduced on the original, rather than transformed, scale of survival probability and expected reproductive success. The results from this first al-

ternative simulation structure did not differ qualitatively from the results obtained with the standard simulation structure, so they are presented in Appendix 2.8.4. The second alternative structure considers identical individuals, that is there is no explicit fixed heterogeneity, and a Markovian process with structured transition probabilities between reproductive stages and survival probabilities (see below).

Simulations with a Markovian process Simulations were carried out as previously described, except that ARS and survival probabilities depended on their previous state and not on fixed individual qualities. This matches the structure of the NS as proposed by Tuljapurkar, Steiner, and Orzack 2009 and is referred to as the “full dynamic model” in Plard et al. 2012. Note that in this model, as shown in Plard et al. 2012, the non-random transition probabilities of the Markovian process can be interpreted either as the result of fixed heterogeneity (if successful animals have a higher than average probability of remaining successful because of their individual properties, such as genetic quality) or of dynamic heterogeneity (if the persistence of success comes from the properties of reproductive stages rather than individuals, e.g. only individuals that have a territory can reproduce and these individuals are more likely than non-reproducers to have a territory next year). Indeed, for short lived species, a Markovian process produces among-individual variance because there are only a few observations per individual, and the first outcome of a Markov chain can have a big influence on the mean individual outcome. In long-lived species, on the other hand, mean individual performances will asymptotically approach the population mean.

In these simulations, the ARS of individual i at time t , $\rho_{i,t}$, follows:

$$\begin{aligned}\rho_{i,t} &\sim \mathcal{P}(\mu_\rho); \text{ for second year individuals,} \\ \rho_{i,t} &\sim \mathcal{P}(\mu_\rho + m(\rho_{i,t-1} - \mu_\rho)); \text{ for older individuals,}\end{aligned}$$

where $\rho_{i,t-1}$ is the ARS of the focal individual the year before, μ_ρ is the mean ARS of the population and m controls the strength of the Markovian process, i.e. the degree to which current reproductive success depends on the previous reproductive success. Only positive values of m were used in order to produce an individual persistence of ARS, which may mimic latent fitness (see below).

Similarly, the survival outcome of individual i at time t , $\phi_{i,t}$, follows:

$$\begin{aligned}\phi_{i,t} &\sim \mathcal{B}(\mu_\phi + \beta_j); \text{ for juveniles} \\ \phi_{i,t} &\sim \mathcal{B}(\text{logit}^{-1}(\text{logit}(\mu_\phi) + c(\rho_{i,t-1} - \mu_\rho))); \text{ for adults,}\end{aligned}$$

where μ_ϕ is the mean adult survival, β_j is the difference between the mean survival of juveniles and adults, and c controls the correlation between reproduction and survival. Survival probability at time t depends on ARS at time $t - 1$ rather than on previous survival, as the latter is always 1 for surviving individuals. Again, only positive values of c were used to simulate persistence of the individual propensity to survive. The positive correlation between successive survival probabilities arises indirectly through the positive correlation between successive ARS, combined with the positive correlation between ARS and survival.

In the presence of allocation trade-offs between different life-history traits, or be-

tween successive expressions of the same life-history trait, negative correlations (i.e. $m < 0$) and autocorrelations (i.e. $c < 0$) could be expected. However, phenotypic correlations between life-history traits are often positive (Stearns 1992, chapter 4). This discrepancy is the result of the variance in resource acquisition, which is related to variance in latent fitness, being larger than the variance in resource allocation (Noordwijk and Jong 1986). Based on this, positive values of c and m are in line with the presence of variation in latent fitness. Indeed, a positive correlation between survival and reproduction is observed in the snow vole data (correlation between observed variation in survival and reproduction: Pearson's correlation, 0.097, 95%CI [-0.007; 0.198]. For the correlation between the latent propensities to survive and to reproduce, see Appendix 2.8.7

Simulation parameters The simulated mean survival probability from year t to year $t + 1$ was 0.4 for juveniles and 0.2 for adults (observed means in snow voles: 0.403 and 0.219, respectively). ARS, averaged over adults, was set to 3, 10 or 50 offspring. For the real snow vole population, mean ARS values of 3 (resulting in a decreasing population size) and 10 (i.e. increasing population size) are within the range observed among years (noting that we include offspring of both sexes in ARS, while we analyze vital rates for only one sex), while the value 50 aimed at confirming the direction of the trend in test performance with respect to mean ARS. The variance in individual quality, either on the original scale or on a transformed scale, σ_ϕ^2 and σ_ρ^2 , took the values 0, 0.1, 0.5, 1 or 2. In simulations without fixed heterogeneity, the m parameters took the values 0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9 or 1, while the c parameters took the values 0, 0.5 or 1. We had no a priori expectations for the heterogeneity parameters (σ_ϕ^2 , σ_ρ^2 , m and c) in the real snow vole population and thus selected the non-null values in a range from small to large relative to the mean survival and ARS.

2.3.2 Testing for fixed individual heterogeneity

Neutral Simulations (NS) NS were carried out following Tuljapurkar, Steiner, and Orzack 2009, but we used the “full stochastic model” proposed by Plard et al. 2012. Compared to the original formulation of NS, the “full stochastic model” better isolates dynamic heterogeneity by making future states independent of the current state. Thereby it removes the non-stochastic component of transition probabilities and allows testing whether “a given lifetime reproductive metric distribution is generated only by dynamic heterogeneity” (Plard et al. 2012).

Briefly, individual life histories, starting as juveniles, are simulated by producing a sequence of ARS values, with the probability of each value of ARS corresponding to its frequency in the focal data set. Mortality events, with an age-specific probability estimated from the data set, are mapped to these individual trajectories. Subsequently, properties of the resulting LRS distribution, as well as of the transition matrix between life stages, are compared between the focal data set and that obtained using NS.

Here it is crucial to highlight some differences between the NS and the way in which we simulated the data sets to which they are applied. First and foremost, in NS the propensity to reproduce and to survive is identical for all individuals and never de-

pends on the previous reproductive success. Second, in our simulations, ARS follows a Poisson distribution—all positive integers are possible values—whereas in NS, ARS are drawn from the ARS values observed in the focal data set, which can follow any distribution, and for instance may have gaps, multiple modes or extreme skewness. Third, in our simulations, mean survival probability is always 0.4 for juveniles and 0.2 for adults, while in NS these age-specific probabilities are the age-specific frequencies of survival that are realized in the focal data set. To sum up, our simulations are parametric and follow well defined distributions, while NS use empirical distributions and thereby stick to the data.

To test for a deviation from the neutral expectation, LRS distributions were compared using both Kolmogorov-Smirnov tests (used in Steiner, Tuljapurkar, and Orzack 2010) and χ^2 tests (used in Plard et al. 2012). Additionally, we calculated mean LRS, the variance in LRS, as well as the persistence of the reproductive stage transition matrix and its entropy following Plard et al. 2012. Observed values greater than the 95% quantile—or smaller than the 5% quantile in the case of entropy, because more fixed heterogeneity should decrease entropy (Tuljapurkar, Steiner, and Orzack 2009)—of the neutral distribution were considered significantly different. The proportion of data sets for which a test is significant in the absence of simulated fixed heterogeneity gives the type I error rate, whereas the proportion of data sets for which a given test is not significant in the presence of simulated fixed heterogeneity gives the type II error rate. The NS method is computationally intensive, so to minimize computational time, we used the minimal number of NS per simulated data set beyond which statistical power did not change (Appendix 2.8.2).

Mixed Models (MM) Generalized linear mixed models (GLMMs) were used to estimate the variance in reproduction and survival attributable to fixed individual heterogeneity, as well as to test for its statistical significance. Significance of the variance components was assessed using Likelihood Ratio Tests (LRT) (see e.g. Pinheiro and Bates 2000; Crainiceanu and Ruppert 2004), assuming that the statistic follows an even mixture of χ_1^2 and χ_0^2 (Self and Liang 1987). For survival, first a logistic model not allowing for individual-level heterogeneity was fitted:

$$\text{logit}(\phi_{i,t}) = \mu_\phi + \text{Age}_{i,t}, \quad (2.3)$$

where μ_ϕ denotes the intercept and $\text{Age}_{i,t}$ denotes the effect of age (juvenile or adult) of individual i at time t . In order to model individual-level heterogeneity, this model was subsequently expanded with an individual random intercept:

$$\text{logit}(\phi_{i,t}) = \mu_\phi + \text{Age}_{i,t} + z_{\phi,i}; \text{ with } z_{\phi,i} \sim \mathcal{N}(0, \hat{\sigma}_\phi^2). \quad (2.4)$$

Model (2.4) estimated the individual-level heterogeneity in survival probability, $\hat{\sigma}_\phi^2$. Moreover, a LRT comparing model (2.4) to model (2.3) tested for the significance of $\hat{\sigma}_\phi^2$.

Similarly, for ARS a first Poisson model without individual-level heterogeneity was fitted:

$$\log(\rho_{i,t}) = \mu_\rho + \text{Age}_{i,t}, \quad (2.5)$$

where μ_ρ denotes the intercept and $\text{Age}_{i,t}$ denotes the effect of age. Subsequently, an individual random intercept was included to model individual-level heterogeneity:

$$\log(\rho_{i,t}) = \mu_\rho + \text{Age}_{i,t} + z_{\rho,i}; \text{ with } z_\rho \sim \mathcal{N}(0, \hat{\sigma}_\rho^2). \quad (2.6)$$

Model (2.6) estimated the individual-level heterogeneity in reproductive ability, $\hat{\sigma}_\rho^2$. Moreover, a LRT comparing model (2.5) to model (2.6) tested for the significance of $\hat{\sigma}_\rho^2$.

In addition, for the analyses of data simulated by means of a Markovian process not including any explicit fixed heterogeneity, the models (2.5) and (2.6) were refitted while adding past reproductive success $\rho_{i,t-1}$ as a covariate. The estimated variance $\hat{\sigma}_\rho^2$ and the LRT comparing these two new models tests the significance of fixed heterogeneity while accounting for a Markovian process.

2.3.3 Analysis of the snow vole data set

A snow vole population, located in the Swiss Alps near Churwalden, at 2000m above sea level, has been monitored continuously since 2006. Analyses presented here are based on data collected until 2013. Individual recapture probability is virtually equal to 1.0, which facilitates the modeling of survival. For more information on the study site and data collection, see Appendix 2.8.5. NS were applied to the real snow vole data set exactly in the same way as to the simulated data sets, separately for males and females. For MM, starting from the models for ARS and survival used for the simulated data sets, we added sex and the sex by age interaction as additional fixed factors, as well as a random effect accounting for variation among years and an observation-level random effect. The latter accounts for overdispersion (see e.g. Atkins et al. 2013) and quantifies the overdispersion due to sources of heterogeneity not included in the model. In a second step, models also including ARS in the previous year were fitted in order to test for the presence of fixed heterogeneity after accounting for variation introduced by Markovian processes. Confidence intervals for all parameters were computed through 1000 parametric bootstraps, using the `confint` function in `lme4`. In a final step, the correlation between the propensity to survive and to reproduce was estimated using a bivariate GLMM in `MCMCglmm` (version 2.21) (Hadfield 2010). This model is detailed in Appendix 2.8.7.

2.4 Results

Mean ARS had no effect on the error rates of any test, so we merged together the the scenarios differing only by mean ARS. Therefore, all error rates are estimated based on 3000 tests (1000 data sets per scenario, times three mean ARS values).

2.4.1 Type I error rates

In the absence of simulated individual fixed heterogeneity and non-random transition probabilities between successive stages, all tests have a low rate of null-hypothesis

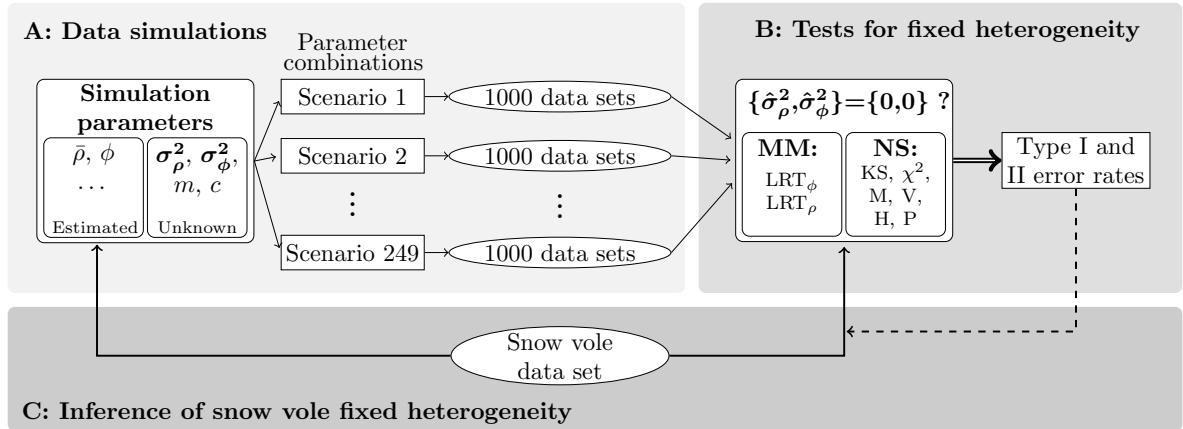


Figure 2.1: Illustration of the simulation and testing process. (A) Data simulation: The simulation model is parametrized using the life cycle and vital rates of a snow vole population, along with additional, unknown, parameters introducing fixed heterogeneity (σ_ϕ^2 and σ_ρ^2) and dynamic heterogeneity (m and c). Different combinations of these simulation parameters define 249 scenarios. For each scenario, 1000 data sets are simulated. (B) Tests for fixed heterogeneity: Each simulated data set is tested for the presence of fixed heterogeneity with both mixed models (MM) using likelihood ratio tests (LRT) on survival (ϕ) and reproduction (ρ), and neutral simulations (NS), using six different tests (see main text). Because σ_ϕ^2 and σ_ρ^2 are known for each simulated data set, we can estimate the type I and type II error rates under each scenario. (C) Analysis of the snow vole data: Both MM and NS are applied to the real snow vole data set, and the outcome is interpreted in the light of the estimated error rates of each test.

rejection (table 2.1). This means that any discrepancy between NS and MM must come from a difference in type II rather than type I error rates.

2.4.2 Type II error rates

Simulations with explicit fixed heterogeneity

Neutral simulations (NS) The Kolmogorov-Smirnov test comparing LRS distributions is significant for only one simulated data set (pertaining to the scenario $\{\sigma_\rho^2 = 1, \sigma_\phi^2 = 2, \bar{\rho} = 50\}$) out of the 72,000 data sets with explicit fixed heterogeneity on the transformed scale. For the parameter range simulated, this test has thus effectively null power. Nevertheless, p -values decrease with increasing σ_ρ^2 and σ_ϕ^2 (for $\{\sigma_\rho^2 = 0, \sigma_\phi^2 = 0, \bar{\rho}\}$ $\overline{p\text{-value}} = 0.998$, SE = 0.001; for $\{\sigma_\rho^2 = 2, \sigma_\phi^2 = 2, \bar{\rho}\}$ $\overline{p\text{-value}} = 0.776$, SE = 0.032), showing that the extremely low power is not the result of a complete calculation failure. Similar to the results of Plard et al. 2012, the χ^2 test is more powerful than the Kolmogorov-Smirnov test. Nevertheless, statistical power remains below 0.8 for moderately sized simulated variances, and its maximal value is 0.89 for the highest simulated variances (figure 2.2(A)).

Tests based on mean LRS are non-significant for all data sets and every scenario. The power of tests based on the variance in LRS increases with increasing σ_ϕ^2 , while the power peaks at intermediate values of simulated σ_ρ^2 and decreases again for higher

Table 2.1: Type I error of tests used in the MM and NS approaches, when applied to data sets without underlying fixed heterogeneity and with fully random transition probabilities

	Mixed models		Neutral simulations					
	LRT _ρ	LRT _φ	KS	χ ²	H	P	M	V
estimate	0.042	0.000	0.000	0.021	0.018	0.039	0.000	0.000
95% CI	0.039;0.054	0;0.001	0;0.001	0.016;0.027	0.014;0.023	0.033;0.047	0;0.001	0;0.001

Note: Type I error rates are estimated as the proportion of simulated data sets, generated without fixed heterogeneity nor Markovian process, for which a test provides a *p*-value below 0.05. Hence, each proportion is estimated from 3,000 tests. The 95% CI (confidence intervals) are Wilson score intervals. LRT_ρ and LRT_φ refer to the Likelihood Ratio Tests of the variance associated with the individual random intercept in reproductive success and survival, respectively. KS refers to a Kolmogorov-Smirnov test, and χ² to a χ² test, both of which compare the Lifetime Reproductive Success (LRS) distribution in a focal data set to the distribution of LRS distributions obtained through neutral simulations (NS). The four other tests are based on the distribution of values obtained by NS compared to the value in the focal data set (mean (M) and variance (V) of the LRS distribution; and entropy (H) and persistence (P) of the transition matrix between successive annual reproductive successes.

σ_ρ^2 (figure 2.2(B)). The non-monotonic shape might be the result of the simultaneous increase in both the real observed-expected difference and the sampling variance: As the simulated variances go up, the LRS distribution becomes wider and flatter. Keeping the number of NS constant, this results in a less extensive sampling of the LRS distribution and a reduced power.

Tests based on the entropy of transition matrices display a pattern that is similar to that for χ² tests, albeit with lower statistical power, this time peaking at 0.57 (figure 2.2(C)). Tests based on the persistence of transition matrices have high statistical power (≈ 0.8) for $\sigma_\rho^2 \geq 1$, while increases in σ_ϕ^2 result only in a slight increase in statistical power (figure 2.2(D)). While they reach higher statistical power than the χ² tests, they have lower power than the χ² at intermediate σ_ρ^2 values.

Mixed models (MM) In contrast to NS, the power of the likelihood ratio test for ARS (LRT_ρ) is almost perfect for $\sigma_\rho^2 \geq 0.1$. Even though fixed heterogeneity in reproduction and survival are simulated independently, the power to detect fixed heterogeneity in reproduction is marginally influenced by the value of σ_ϕ^2 (figure 2.2(E) and, more clearly, Appendix 2.8.4 figure 2.8.42.6(E)). This is because a higher variance in latent survival probability increases the proportion of individuals that reach the maximal age, which provides more successive observations of reproduction and thereby increases the power to detect variance in reproductive quality. Overall, σ_ρ^2 is slightly underestimated ($\hat{\sigma}_\rho^2 = 0.972\sigma_\rho^2$; adjusted R²=0.9997).

The LRT_φ is never significant, even for $\sigma_\phi^2 = 2$. Moreover the estimation of σ_ϕ^2 is always close to zero (average of the median values 0.029) and does not increase with increasing σ_ϕ^2 (slope and SE: -0.0016 ± 0.0006). The failure of this model illustrates the intrinsic difficulty in estimating random effects for binary traits, especially when there are few repeated measurements per individual (e.g. Albert and Anderson 1984;

Hosmer, Lemeshow, and Sturdivant 2013, chapter 9), as is the case in our short-lived simulated animals.

2.4.3 Simulations with a Markovian process

Although data sets simulated using a Markovian process do not contain explicit fixed heterogeneity, both MM and NS reject the null hypothesis of an absence of fixed heterogeneity in most of the cases (figure 2.3).

The LRT_ρ , testing for fixed heterogeneity in ARS (based on MM), rejects the null hypothesis with a high probability, except for the lowest values of c and m (figure 2.3(E)). When $m > 0$, current ARS is influenced by past ARS, which in turn introduces variance in the propensity to reproduce. When $c > 0$, current survival probability is positively influenced by current ARS. As a consequence, successful reproducers live longer, resulting in more ARS values for these individuals, which improves the ability of the MM to detect individual-level variance. The LRT_ϕ is never significant for $c = 0$, but rejects the null hypothesis at a high rate for $c \geq 0.5$, and this increases as m increases (figure 2.3(G)). This pattern was expected as c controls the correlation between survival and reproduction, and indirectly makes the probability to survive in the current time step dependent on the probability to survive in the previous time step. Increasing values of m further strengthen this correlation.

Both the Kolmogorov-Smirnov test on the LRS distribution, and the test based on mean LRS, are non-significant for any data set with Markovian process. Furthermore, the χ^2 test rejects the null hypothesis with near certainty when $c > 0$, and, when $c = 0$, with probabilities going from low to moderate with increasing m (figure 2.3(A)). Given the absence of explicit fixed heterogeneity in these data, the χ^2 test can therefore be considered to have very high type I error rates (but see the discussion). The tests based on the variance in LRS, entropy and persistence follow a similar pattern of increasing probability of null-hypothesis rejection when m and c increase, but the test based on entropy does not reach a probability higher than 0.65, while the two other tests are close to 1 for the highest values of the parameters (figures 2.3(B)-(D)).

Based on these findings, it could be argued that both MM and Plard's version of NS (Plard et al. 2012) have a very high type I error rate when the transitions between stages are structured. We examine this interpretation in more detail in the discussion. However, the rejection rate of the LRT_ρ for fixed heterogeneity in ARS is drastically reduced by the inclusion of the past ARS ($\rho_{i,t-1}$) in the two mixed models that are being compared, i.e. with and without the individual random effect (compare figure 2.3(E) and figure 2.3(F)). The type I error rate is greater than the alpha threshold of 5% only when both $m > 0.8$ and $c > 0$ (figure 2.3(F)). Moreover, the estimates of the variance in reproductive propensity are reduced by the inclusion of $\rho_{i,t-1}$ in the models: over all the scenarios, the mean is $\hat{\sigma}_\rho^2 = 0.004$, SE=0.002, with a maximal estimate of 0.144, whereas without including $\rho_{i,t-1}$, the mean is 0.050, SE=0.008, and the maximum 0.459. The former estimate is closer to zero, i.e. the individual-level variance that is explicitly simulated.

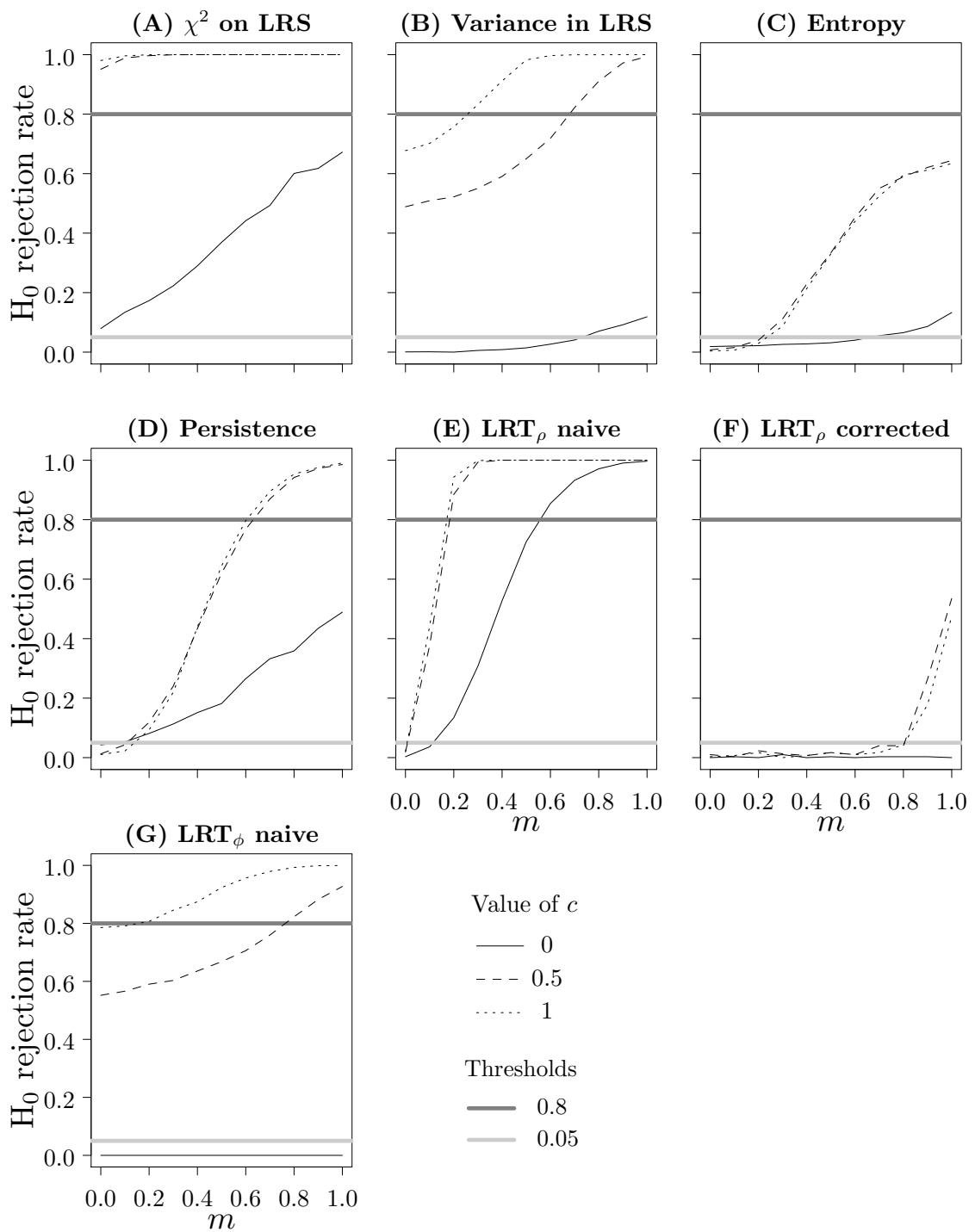


Figure 2.3: Null-hypothesis rejection rates for various methods testing for the presence of fixed heterogeneity, when none is explicitly simulated, depending on the parameter m , controlling the structure of transitions between successive annual reproductive successes, and on the parameter c , controlling the dependency between survival probability and reproductive success (see the method section “Simulations of a Markovian process” for details). The methods are: (A) a χ^2 test comparing the LRS distribution in a focal data set to the distribution of LRS distributions obtained through NS; tests based on proportion of values obtained by NS greater or equal to the value in the focal data set for (B) the variance in LRS, (C) the entropy of the transition matrix between successive annual reproductive success and (D) the persistence of this matrix; (E) a LRT for the significance of the individual random intercept in reproductive success, using models that do not account for a Markovian process, or (F) that do account for a Markovian process; (G) a LRT for the significance of the individual random intercept in survival. For survival we did not try to account for the Markovian process. Assuming that the simulated Markovian process cannot be related to fixed heterogeneity, the null-hypothesis rejection rates represent type I error rates for all values of the c and the m parameters. (A)-(D) are related to the NS framework. (E)-(G) are related to the MM framework

2.4.4 Application to the snow vole data set

Neutral simulations (NS)

For males, none of the six tests carried out within the NS framework are significant. Neither the LRS distribution, nor the transition matrix between successive values of ARS, are distinguishable from those generated using NS (table 2.2). For females, out of the six tests, two are significant: there is more persistence and more variance than expected under neutrality; and the test on mean LRS is close to being significant. However, the tests on the complete LRS distribution (Kolmogorov-Smirnov and χ^2) are far from significant (table 2.2). The latter is unsurprising as a graphical examination of the observed and the simulated neutral LRS distribution shows that the two distributions are almost indistinguishable (figure 2.4). According to the authors of the NS framework, the comparison of LRS distributions, either through a Kolmogorov-Smirnov test (in Steiner and Tuljapurkar 2012) or a χ^2 test (in Plard et al. 2012), is the gold standard when testing for the presence of fixed heterogeneity with NS (Steiner 2013, pers. comm. November 25th). Based on these NS results, there is thus no evidence for fixed heterogeneity in either of the sexes, although the results are more equivocal in females.

Mixed models (MM)

The GLMM for survival identifies significant between-years variance (5.622; 95% CI [1.133; 13.158]), but estimates a latent individual-level variance of 0 (95% CI [0; 0.248]) (see supplementary table 2.4 for all the estimates of this model).

The GLMM for ARS estimates variances among individuals (0.371; 95%CI [0.151; 0.475]) as well as among years (0.101; 95%CI [0.026; 0.452]) that are different from zero, and LRTs for both variances are highly significant. The random effect accounting for overdispersion does not significantly differ from zero, although its bootstrapped confidence interval includes positive values (table 2.5 for all the estimates of this model). When the individual random effect is not included, this overdispersion

Table 2.2: Outcomes of the various tests within the NS framework when applied to the real snow vole data set, for males and females separately

test	KS		χ^2			H p-value	P p-value	V p-value	M p-value
	D	p-value	χ^2	df	p-value				
Males	0.025	0.969	8.33	15	0.909	0.629	0.646	0.395	0.378
Females	0.030	0.902	5.50	8	0.70	0.624	0.035	0.031	0.057

Note: KS refers to the Kolmogorov-Smirnov test, and χ^2 to the χ^2 test, comparing the Lifetime Reproductive Success (LRS) distribution in a focal data set to the distribution of LRS distributions obtained through NS. The four other tests are based on the proportion of values obtained by NS greater than the value in the focal data set for the mean (M) and variance (V) of the LRS distribution, and for the entropy (H) and persistence (P) of the transition matrix between successive annual reproductive success. The *p*-values $\leq 5\%$ are shown in bold.

variance is highly significant, and the sum of squared Pearson residuals divided by the estimated residual degrees of freedom is approximately 2, while it falls to 1 with individual as a random effect. The estimation of residual degrees of freedom in GLMMs is a complex issue (Pinheiro and Bates 2000), but this approach seems to indicate that the overdispersion in the distribution is largely due to differences between individuals.

Excluding individuals reproducing for the first time, we fitted a GLMM that includes the previous reproductive success ARS_{t-1} and sex as fixed effects, and year as the only random effect. This model indicates a significant positive relationship between successive values of ARS (slope=0.0949; SE = 0.0213; *p*-value= 8×10^{-6}). Nevertheless, adding individual as a random effect greatly improved the fit of the model ($\Delta AIC = 87$; LRT: *p*-value $< 10^{-16}$), providing evidence for the existence of significant individual-level variance ($\hat{\sigma}_{id}^2 = 0.341$, bootstrapped 95% CI [0.189; 0.453]). Including ARS_{t-1} had little effect on the estimate of $\hat{\sigma}_{id}^2$ (see table 2.5), but now ARS_{t-1} no longer reached significance (slope=0.0210; SE = 0.0275; *p*-value=0.445).

Finally, the latent correlation between the propensities to survive and to reproduce was estimated as 0.32 (95% CI [-0.68;0.97]) and appears in the best model selected by DIC (see Appendix 2.8.7).

2.5 Discussion

2.5.1 Overview

Based on extensive simulations, we have shown that in the presence of fixed heterogeneity, NS have much less statistical power than MM, even when the model simulating the data does not match the structure assumed by the MM. In particular the Kolmogorov-Smirnov test, advocated in the earlier version of NS, has virtually no statistical power. In contrast, MM have low type I error rates and are not misled by the presence of dynamic heterogeneity, which in all data sets is non-zero if it is measured as entropy (Tuljapurkar, Steiner, and Orzack 2009). This finding directly contradicts the claim “[...] that random effect models will always detect unobservable fixed ef-

fектs" Steiner, Tuljapurkar, and Orzack 2010. Second, in the absence of fixed heterogeneity, Markovian transitions between successive reproductive success and survival probabilities can induce high type I error rates, both in MM and NS sensu Plard et al. 2012. However, inclusion of previous reproductive success in the MM for reproduction substantially reduces these errors. Third, when applied to a real data set for a wild population of snow voles, NS only detect ambiguous deviations from neutrality and only for females. Moreover, the main tests of the framework, based on the total distribution of LRS, fail to reject the null hypothesis in both sexes. In striking contrast, MM show strong evidence for individual latent variance in reproductive success, even when a Markovian process is accounted for. In addition, MM give some indication of the presence of individual latent variance in survival, and of a positive correlation between survival and reproduction. However, the latter two parameters are estimated with substantial uncertainty.

2.5.2 Use of simulations

Testing methods on simulated data can be difficult because the specific simulation process used can differently match the assumptions and structures of the different methods. We tried to overcome this issue by using three different simulation models. Moreover, the rejection rates of MM and NS observed in our simulations are similar to those observed when the methods are applied to real data. Indeed, in the present work we applied both methods to a snow vole data set and found that the MM approach detected individual fixed heterogeneity, while the NS approach did not detect a significant deviation from the neutral expectation. This was also the case for the other data sets to which both methods were applied (MM by Cam et al. 2013; NS by Steiner, Tuljapurkar, and Orzack 2010). On the whole we are aware of only a single case in which NS led to the rejection of neutrality (Plard et al. 2012), whereas MM commonly find evidence for significant individual fixed heterogeneity, either by estimation of positive variance components, model selection (Cam et al. 2013) or posterior predictive checks (Chambert, Rotella, and Higgs 2014). Although there is some possibility of publication bias, this pattern is consistent with our power analysis.

2.5.3 Low power of Neutral Simulations

The low power of NS probably stems from the fact that they aggregate data on vital rates, and that they do so twice: first over the lifetime of individuals, and then they aggregate individuals into population-level statistics. Thereby they first discard the repeatability of individuals, which has been shown to blur heritable differences among individuals (Vaupel 1988). Second, population-level statistics can be produced by an infinite number of different mixtures of individual types (for instance, a mean probability of 0.5 can be the result of a population consisting only of individuals with a latent probability of 0.5, or from a uniform distribution of individual probabilities between 0 and 1). Therefore, some patterns of among-individual differences are indistinguishable at the population level. Individual-level data are naturally better at identifying the causes of variation at that level (Clutton-brock and Sheldon 2010), and the ability to use non-aggregated data, for instance longitudinal information on marked

individuals, further increases this power (Brooks, McCoy, and Bolker 2013). While a method such as Plard's NS could be valuable in the absence of such data, alternative methods making use of non-aggregated information, such as MM, should be preferred whenever possible.

Importantly, within a strict null-hypothesis testing framework, the failure to reject a null hypothesis cannot be interpreted as a proof of the null hypothesis. The absence of significance in most implementations of the NS (Steiner, Tuljapurkar, and Orzack 2010; Orzack et al. 2011; Tuljapurkar, Steiner, and Orzack 2009; Plard et al. 2012) is therefore not informative with respect to the presence and the biological significance of fixed heterogeneity. The null-hypothesis testing framework can partially be relaxed by an a priori power analysis. Although comparisons of simulated data sets with and without heterogeneity were indeed presented in Steiner and Tuljapurkar 2012, there fixed heterogeneity (assumed to be genetic) was modeled as two groups of homogeneous individuals, which except for clonal organisms is biologically unrealistic. In addition, the absence of significant differences between the data sets with and without fixed heterogeneity was not interpreted as a sign of a lack of statistical power, but as evidence that fixed heterogeneity has little effect on LRS distributions.

2.5.4 Effect of Markovian transitions

When no fixed heterogeneity was explicitly simulated, both MM and NS rejected the null hypothesis that fixed heterogeneity is absent. This was to be expected for MM, given that Markovian transitions mimic individual-level variance, and MM do not model population-level transition probabilities. It is more surprising that also NS had a high rate of false positives. However, we here used the "full random model" reformulation of NS (Plard et al. 2012), and not the "full dynamic model" (Tuljapurkar, Steiner, and Orzack 2009). The latter simulates individual trajectories using a Markovian process, similar to the way data sets were simulated here, while the former simulates individual trajectories without taking into account the previous state. Hence, "full dynamic NS" would not reject the null hypothesis, and one could consider this in this case to be correct. However, as latent individual quality will necessarily produce a pattern that is consistent with a Markovian process, this formulation does not allow for a complete separation of fixed and dynamic heterogeneity (Plard et al. 2012). Observing a Markovian process is therefore in itself not informative with respect to the mechanisms shaping life histories. Hence, although they have a low type I error rate, "full dynamic NS" always have low statistical power.

We acknowledge that a Markovian process that is not due to fixed differences between individuals does mimic fixed heterogeneity, and thereby can bias estimates of between-individual variance based on full random NS and on MM. Therefore, a naive MM detects individual-level heterogeneity, irrespective of whether it is due to a population-level Markovian process or to individual-level differences. However, the type I error of MM can be substantially reduced by including previous reproductive success in the model (Rotella 2008; Cam et al. 2013). Although this is not a universal solution that accounts for all confounding factors, it highlights the flexibility of the MM framework, which allows for the incorporation of any factor that is perceived as potentially confounding based on knowledge of the study system.

2.5.5 Genetic variation as a source of fixed heterogeneity

In cases where the evidence for the presence of fixed heterogeneity is equivocal, for instance because the effects of Markovian processes and individual-level fixed differences are confounded, the use of genetic information and quantitative genetic methods has the potential to tease apart latent genetic quality from other sources of performance persistence, including stochastic transitions. Indeed, although other sources of variation may also generate fixed heterogeneity, the existence of significant additive genetic variation implies significant fixed heterogeneity, by definition determined at fertilization. Interestingly, estimates of additive genetic variation for fitness components are often large, even in small populations (for reviews see Mousseau and Roff 1987; Postma 2014). As a matter of fact, when standardized by the mean (i.e. evolvability) rather than the variance (i.e. heritability), fitness components appear to have higher additive genetic variation than other types of traits (Hansen, Pélabon, and Houle 2011; Postma 2014). In addition to our findings, this provides further support for fixed heterogeneity being more common than suggested by NS.

2.5.6 Interpretation of the snow vole results

Because they are similar in structure, our simulated data sets can shed light on the results from the analysis of the real snow vole data set. For example, it is unsurprising that the MM fails to detect individual heterogeneity in snow vole survival probabilities. The LRT_ϕ has no statistical power for simulated data sets with simulated $\sigma_\phi^2 \leq 2$, while confidence and credibility intervals indicate that the possible values of σ_ϕ^2 lay between 0 and 1 at most (supplementary tables 2.4 and 2.7). Unlike heterogeneity in individual survival probability, heterogeneity in individual reproductive success is easily detected and quantified by MM applied to simulated data sets (figure 2.2(E)). Accordingly, the analysis of the real data set identifies an individual variance in the propensity to reproduce that is significantly different from zero, and is estimated to be more than three time larger than the variance among years. Finally, given the estimate of the variance σ_ρ^2 , we can get an estimate of the statistical power of the other tests to detect fixed heterogeneity in the real snow vole data set: a significant test seems possible for the χ^2 test (figure 2.2(A)), but quite unlikely for the test based on entropy (figure 2.2(C)).

A positive correlation between individual-level variation in reproduction and survival would provide further support for fixed heterogeneity. However, as mentioned above, the estimation of individual-level variance in survival is difficult because this is a binary trait, and because due to their short lifespan there are few observations per individual. Hence there is a lot of uncertainty in the estimation of this correlation parameter. Nevertheless, the most likely values are positive (Appendix 2.8.7).

2.5.7 Fixed heterogeneity and the concept of fitness

The debate surrounding the biological significance of fixed heterogeneity appears to stem at least partly from different concepts of fitness. On the one hand, proponents of the neutral theory of life histories consider fitness to be a property of a category of

individuals, and consider variation in reproductive success among individuals to be mostly due to dynamic heterogeneity, rather than due to variation in latent individual properties (Steiner and Tuljapurkar 2012). On the other hand, researchers in the field of evolutionary ecology often see fitness as a latent property of individuals (Cam and Monnat 2000), that is, an expected value defined at the individual level that cannot be measured directly (Brandon and Beatty 1984; Price 1996; Krimbas 2004). As the mean value of a group is also the expected value of an individual belonging to this group, the two views are not fundamentally different. In sexual organisms however, each individual is unique, which makes it difficult to assign it to a hypothetical group made of identical individuals. If stochastic variation underlies most of the realized reproductive success and there are no fitness differences between individuals, as adherents of the neutral theory of life histories advocate, then it is useless to define fitness at the individual level. However, if there exists significant fixed heterogeneity, individual performances carry some information about their latent properties, for example due to their genetic makeup. In the presence of fixed heterogeneity it therefore seems useful to use an individual-level definition of fitness, differing from both group-level fitness and realized reproductive success.

2.6 Conclusions

Using extensive simulations, we have demonstrated that NS are uninformative with respect to the biological significance of fixed heterogeneity. Based on the work of Plard et al. 2012 and our power analysis, we conclude that the observation of a Markovian process in stage-transition probabilities does in itself not provide any biological insights. Within the NS framework, the full random model (Plard et al. 2012) should be preferred over the full dynamic model (Tuljapurkar, Steiner, and Orzack 2009), and the χ^2 test should be preferred over the Kolmogorov-Smirnov test. In addition, any use of NS should be complemented by an a priori power analysis, or otherwise be restricted to a strict null-hypothesis testing framework, where failure to reject the null hypothesis does not allow any conclusions regarding the null hypothesis being true, and/or the alternative hypothesis false. However, even when these improvements are included in the NS framework, we recommend that its use is restricted to data sets where individuals are not identified.

Instead, we show that MM are more powerful, but not more susceptible to type I error. Although MM can be misled by confounding factors, given a good knowledge of the biological system, it is possible to account for these confounding factors, in which case MM have a very low type I error rate.

Finally, the confrontation of our power analysis with the analysis of the real snow vole data set supports the presence of fixed heterogeneity in fitness components in this population. Further research is being carried out to identify what traits can be related to this latent heterogeneity, and how genetic and maternal effects shape these differences.

On the whole, this work supports the idea that fixed heterogeneity is more common than suggested by the studies based on NS.

2.7 Acknowledgments

We thank Arpat Ozgul, Yannis Michalakis, Benjamin M. Bolker, Gordon A. Fox and an anonymous reviewer for helpful comments on this manuscript. Thanks to Ulrich K. Steiner for discussions on an earlier version of this work and for help with the implementation of neutral simulations. Thanks also to Lukas Keller, Josh Van Buskirk and Pirmin Nietlisbach for inspiring discussions on this topic. The snow vole monitoring was authorized by the *Amt für Lebensmittelsicherheit und Tiergesundheit*, Chur, Switzerland. Funding was provided by the Swiss National Science Foundation project grant (31003A_141110).

References

- Albert, A. and Anderson, J. 1984. "On the existence of maximum likelihood estimates in logistic regression models." *Biometrika* **71** (1): 1–10.
- Atkins, D. C., Baldwin, S. A., Zheng, C., Gallop, R. J., and Neighbors, C. 2013. "A tutorial on count regression and zero-altered count models for longitudinal substance use data." *Psychology of Addictive Behaviors* **27** (1): 166–77.
- Barnett, A. G., Koper, N., Dobson, A. J., Schmiegelow, F., and Manseau, M. 2010. "Using information criteria to select the correct variance-covariance structure for longitudinal data in ecology." *Methods in Ecology and Evolution* **1** (1): 15–24.
- Bates, D., Mächler, M., Bolker, B., and Walker, S. 2015. "Fitting Linear Mixed-Effects Models Using lme4." *Journal of Statistical Software* **67** (1): 1–48.
- Bonduriansky, R. 2012. "Rethinking heredity, again." *Trends in Ecology & Evolution* **27** (6): 330–6.
- Brandon, R. and Beatty, J. 1984. "The propensity interpretation of 'fitness'. No interpretation is no substitute." *Philosophy of Science* **51** (2): 342–347.
- Brooks, M. E., McCoy, M. W., and Bolker, B. M. 2013. "A method for detecting positive growth autocorrelation without marking individuals." *PLoS One* **8** (10): e76389.
- Burnham, K. K. and Anderson, D. D. 2002. *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach* (2nd ed), 172:488. Springer-Verlag, Heidelberg.
- Cam, E., Gimenez, O., Alpizar-Jara, R., Aubry, L. M., Authier, M., Cooch, E. G., Koons, D. N., et al. 2013. "Looking for a needle in a haystack: inference about individual fitness components in a heterogeneous population." *Oikos* **122** (5): 739–753.
- Cam, E., Link, W. a., Cooch, E. G., Monnat, J.-Y., and Danchin, E. 2002. "Individual covariation in life-history traits: seeing the trees despite the forest." *The American Naturalist* **159** (1): 96–105.

- Cam, E. and Monnat, J. Y. 2000. "Stratification based on reproductive state reveals contrasting patterns of age-related variation in demographic parameters in the kittiwake." *Oikos* **90** (3): 560–574.
- Caswell, H. 2011. "Beyond R_0 : Demographic models for variability of lifetime reproductive output." *PLoS One* **6** (6): e20809.
- Chambert, T., Rotella, J. J., and Higgs, M. D. 2014. "Use of posterior predictive checks as an inferential tool for investigating individual heterogeneity in animal population vital rates." *Ecology and Evolution* **4** (8): 1389–97.
- Chambert, T., Rotella, J. J., Higgs, M. D., and Garrott, R. A. 2013. "Individual heterogeneity in reproductive rates and cost of reproduction in a long-lived vertebrate." *Ecology and Evolution* **3** (7): 2047–60.
- Charlesworth, B. 2015. "Causes of natural variation in fitness : Evidence from studies of *Drosophila* populations." *Proceedings of the National Academy of Sciences, USA* **112** (6): 1662–1669.
- Clutton-Brock, T. H. 1988. "Reproductive success in male and female red deer." In *Reproductive success. Studies of individual variation in contrasting breeding systems*. Edited by T. H. Clutton-Brock, 325–343. University of Chicago Press, Chicago / London.
- Clutton-brock, T. and Sheldon, B. C. 2010. "Individuals and populations: the role of long-term, individual-based studies of animals in ecology and evolutionary biology." *Trends in Ecology & Evolution* **25** (10): 562–573.
- Crainiceanu, C. M. and Ruppert, D. 2004. "Likelihood ratio tests in linear mixed models with one variance component." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **66** (1): 165–185.
- Ellegren, H. and Sheldon, B. C. 2008. "Genetic basis of fitness differences in natural populations." *Nature* **452**:169–175.
- Fisher, R. 1958. *The genetical theory of natural selection*. Second. Dover Publications, New York.
- Fox, G. A. and Kendall, B. E. 2002. "Demographic stochasticity and the variance reduction effect." *Ecology* **83** (7): 1928–1934.
- Guillemain, M., Green, A. J., Simon, G., and Gauthier-Clerc, M. 2013. "Individual quality persists between years: individuals retain body condition from one winter to the next in Teal." *Journal of Ornithology* **154** (4): 1007–1018.
- Hadfield, J. D. 2010. "MCMC Methods for Multi-Response Generalized Linear Mixed Models: The MCMCglmm R Package." *Journal of Statistical Software* **33** (2): 1–22. <http://www.jstatsoft.org/v33/i02/>.
- Hadfield, J. D., Wilson, A. J., and Kruuk, L. E. B. 2011. "Cryptic evolution: does environmental deterioration have a genetic basis?" *Genetics* **187** (4): 1099–113.

- Hadfield, J., Richardson, D., and Burke, T. 2006. "Towards unbiased parentage assignment: combining genetic, behavioural and spatial data in a Bayesian framework." *Mol Ecol* **15** (12): 3715–3730.
- Hansen, T. F., Pélabon, C., and Houle, D. 2011. "Heritability is not evolvability." *Evolutionary Biology* **38**:258–277.
- Haring, E., Herzig-Straschil, B., and Spitzenberger, F. 2000. "Phylogenetic analysis of Alpine voles of the *Microtus multiplex* complex using the mitochondrial control region." *J Zoolog Syst Evol Res* **38**:231–238.
- Hosmer, D. W., Lemeshow, S., and Sturdivant, R. X. 2013. *Applied logistic regression, third edition*. Hoboken, New Jersey: Wiley.
- Janeau, G. and Aulagnier, S. 1997. "Snow vole - *Chionomys nivalis* (Martins 1842)." *IBEX Journal of Mountain Ecology* **4**:1–11.
- Keller, L. and Waller, D. 2002. "Inbreeding effects in wild populations." *Trends in Ecology & Evolution* **17** (5): 19–23.
- Kendall, B. E., Fox, G. A., Fujiwara, M., and Nogaire, T. M. 2011. "Demographic heterogeneity, cohort selection, and population growth." *Ecology* **92** (10): 1985–1993.
- Krimbas, C. B. 2004. "On fitness." *Biology & Philosophy* **19** (2): 185–203.
- Leblois, R., Estoup, A., and Rousset, F. 2009. "IBD Sim: A computer program to simulate genotypic data under Isolation by Distance." *Molecular Ecology Resources* **9**:107–109.
- Marshall, T., Slate, J., Kruuk, L. E., and Pemberton, J. M. 1998. "Statistical confidence for likelihood-based paternity inference in natural populations." *Mol Ecol* **7**:639–655.
- Matsumoto, M. and Nishimura, T. 1998. "Mersenne Twister: A 623-dimensionally equidistributed uniform pseudo-random number generator." *ACM Transactions on Modeling and Computer Simulation* **8** (1): 3–30.
- Morris, D. 1998. "State-dependent optimization of litter size." *Oikos* **83**:518–528.
- Mousseau, T. A. and Roff, D. A. 1987. "Natural selection and the heritability of fitness components." *Heredity* **59** (2): 181–197.
- Noordwijk, A. J. van and Jong, G. de. 1986. "Acquisition and allocation of resources: Their influence on variation in life history tactics." *The American Naturalist* **128** (1): 137–142.
- Orzack, S. H., Steiner, U. K., Tuljapurkar, S., and Thompson, P. 2011. "Static and dynamic expression of life history traits in the northern fulmar *Fulmarus glacialis*." *Oikos* **120** (3): 369–380.
- Pinheiro, J. C. and Bates, D. M. 2000. *Mixed-Effects Models in S and S-PLUS*. Statistics and Computing Series. Springer-Verlag, New York, NY.

Chapter 2 Successful by chance?

- Plard, F., Bonenfant, C., Delorme, D., and Gaillard, J. 2012. "Modeling reproductive trajectories of roe deer females: Fixed or dynamic heterogeneity?" *Theoretical Population Biology* **82** (4): 317–328.
- Postma, E. 2014. "Four decades of estimating heritabilities in wild vertebrate populations: improved methods, more data better estimates?" In *Quantitative Genetics in the Wild*, edited by A. Charmentier, D. Garant, and L. E. B. Kruuk. Oxford: Oxford University Press.
- Price, P. W. 1996. *Biological evolution*. Saunders College Publishing, Philadelphia, PA.
- R Core Team. 2014. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. <http://www.R-project.org/>.
- Robertson, A. 1966. "A mathematical model of the culling process in dairy cattle." *Animal Production* **8**:95–108.
- Rotella, J. J. 2008. "Estimating reproductive costs with multi-state mark-recapture models, multiple observable states, and temporary emigration." In *Modeling Demographic Processes In Marked Populations. Series Environmental and Ecological Statistics, Vol 3*. Edited by D. L. Thomson, E. G. Cooch, and M. J. Conroy, 157–172. Springer.
- Royle, J. A. 2008. "Modeling individual effects in the Cormack-Jolly-Seber model: a state-space formulation." *Biometrics* **64** (2): 364–70.
- Schauber, E. M., Goodwin, B. J., Jones, C. G., and Ostfeld, R. S. 2007. "Spatial selection and inheritance: applying evolutionary concepts to population dynamics in heterogeneous space." *Ecology* **88** (5): 1112–1118.
- Self, S. G. and Liang, K. Y. 1987. "Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions." *Journal of the American Statistical Association* **82** (398): 605–610.
- Stearns, S. C. 1992. *The evolution of life histories*. Oxford University Press, Oxford.
- Steiner, U. K., Tuljapurkar, S., and Orzack, S. H. 2010. "Dynamic heterogeneity and life history variability in the kittiwake." *Journal of Animal Ecology* **79** (2): 436–44.
- Steiner, U. K. and Tuljapurkar, S. 2012. "Neutral theory for life histories and individual variability in fitness components." *Proceedings of the National Academy of Sciences, USA* **109** (12): 4684–4689.
- Tuljapurkar, S., Steiner, U. K., and Orzack, S. H. 2009. "Dynamic heterogeneity in life histories." *Ecology Letters* **12** (1): 93–106.
- Turner, B. M. 2009. "Epigenetic responses to environmental change and their evolutionary implications." *Philosophical Transactions of the Royal Society B: Biological Sciences* **364** (1534): 3403–18.
- Vaupel, J. W. 1988. "Inherited frailty and longevity." *Demography* **25** (2): 277–287.

- Vaupel, J., Manton, K., and Stallard, E. 1979. "The impact of heterogeneity in individual frailty on the dynamics of mortality." *Demography* **16** (3): 439–454.
- Villemereuil, P. de, Gimenez, O., and Doligez, B. 2013. "Comparing parent-offspring regression with frequentist and Bayesian animal models to estimate heritability in wild populations: a simulation study for Gaussian and binary traits." *Methods in Ecology and Evolution* **4** (3): 260–275.
- Wandeler, P. and Camenisch, G. 2011. "Identifying Y-chromosomal diversity by long-template PCR." *Mol Ecol Res* **11**:835–841.
- Wandeler, P., Ravaioli, R., and Bucher, T. B. 2008. "Microsatellite DNA markers for the snow vole (*Chionomys nivalis*)." *Mol Ecol Res* **8**:637–639.
- Wilberg, M. J. and Bence, J. R. 2008. "Performance of deviance information criterion model selection in statistical catch-at-age analysis." *Fisheries Research* **93** (1-2): 212–221.
- Wolf, J. B. and Wade, M. J. 2009. "What are maternal effects (and what are they not)?" *Philosophical Transactions of the Royal Society B: Biological Sciences* **364** (1520): 1107–15.

2.8 Supplementary information

2.8.1 Checking the properties of the data sets

The following Generalized Linear Models were fitted to the simulated data sets in order to test whether the data set properties matched the parameters used to generate them:

$$\text{logit}(\phi_{i,t}) = \mu_\phi + \text{Age}_{i,t}; \text{ using a binomial error structure} \quad (2.7a)$$

$$\text{logit}(\phi_{i,t}) = \mu_\phi + \rho_{i,t}; \text{ using a binomial error structure} \quad (2.7b)$$

$$\log(\rho_{i,t}) = \mu_\rho + \rho_{i,t-1}; \text{ using a quasi-Poisson error structure} \quad (2.7c)$$

$$\log(\rho_{i,t}) = \mu_\rho + \text{Age}_{i,t}; \text{ using a quasi-Poisson error structure} \quad (2.7d)$$

These were used to check that survival depended on age (2.7a), that survival depended on annual reproductive success only when that was required (2.7b), that ARS depended on previous reproductive attempts only when fixed heterogeneity for reproductive success or Markovian reproduction was simulated (2.7c) and that ARS of adults was not age-dependent (2.7d). The simulated data had all the expected properties. Furthermore, we never found a significant association between reproduction and survival. This goes against the claim made in Steiner, Tuljapurkar, and Orzack 2010 that dynamic heterogeneity alone can generate a positive association between reproduction and survival.

Instead, we argue here that the findings on which they base their claim reflects their use of reproductive stage-specific survival in their NS, and reproduction and survival being positively correlated in the source data (Cam et al. 2002). Hence, it is not the random transitions themselves that are responsible for the positive association, but the positively associated stage-specific probabilities of survival and reproduction. The origin of the latter remains unexplained, but is consistent with variation in latent fitness among individuals.

2.8.2 Optimal number of neutral simulations per data set.

The neutral simulation approach (NS) is computationally intensive: as the focal population consists of 10 cohorts of 100 individuals, performing 1000 neutral simulations (i.e. simulating 1000 hypothetical populations), requires 1,000,000 individual trajectories to be simulated for every simulated data set (and 75,000,000,000 individual trajectories for the complete study). To minimize computational time, we determined the number of neutral simulations per simulated data set beyond which statistical power did not change. Out of the six tests mentioned above, only χ^2 tests on LRS distributions are sensitive to the number of neutral simulations; while χ^2 tests based on 1000 neutral simulations differ from those based on 100 neutral simulations ($\Delta\text{power}_{1000-100}=-0.067$, $\text{se}=0.033$), the tests based on 100,000 neutral simulations do not have more statistical power than those based on 1000 neutral simulations ($\Delta\text{power}_{100,000-1000}=-0.031$, $\text{se}=0.033$), and the correlation of the statistical power

across scenarios is high ($R^2 = 0.92$). Accordingly, each simulated data set was analyzed using 1000 neutral simulations. Note that the fact that in this case statistical power plateaus already above 1000 neutral simulations is the result of the relatively short lifespan of the simulated animals, which allows for a quick exploration of all the possible individual trajectories.

2.8.3 Selection for latent quality

As outlined in the main document, we simulated fixed heterogeneity by attributing to each individual i a fixed quality for annual reproductive success ($q_{\rho,i}$) and a fixed quality as survivor ($q_{\phi,i}$). These two kind of individual qualities are normally distributed, with mean zero and variance σ_{ρ}^2 and σ_{ϕ}^2 , respectively. The selection acting on, or due to, this variation in latent individual qualities for reproduction and for survival was measured as the individual-level covariance between the qualities and a proxy for fitness (ω): relative lifetime reproductive success (Robertson 1966).

The selection coefficients increase with increasing variance in individual latent qualities, both for reproduction (figure 2.8.32.8.3) and for survival (figure 2.8.32.8.3). This confirms that the heterogeneity simulated is non-neutral.

2.8.4 Simulating fixed heterogeneity on the original scale

It could be argued that the superior statistical power of the LRT_{ρ} is the result of the simulation process used to introduce fixed heterogeneity has the same structure as the MM estimating it. To address this, additional simulations were performed in which individual reproductive success and survival probability depended on their qualities on the original scale rather than on a transformed scale. Otherwise simulations were similar to those where fixed heterogeneity was introduced on the transformed scale. To this end, the reproductive success and survival of an individual i , at time t , are drawn from

$$\rho_{i,t} \sim \mathcal{P}(\mu_{\rho} + q_{\rho,i}) \quad (2.8a)$$

$$\text{and } \phi_{i,t} \sim \mathcal{B}(\mu_{\phi} + \beta_{age} + q_{\phi,i}). \quad (2.8b)$$

Although when the variance in quality for reproduction is included on the original, non-transformed, scale, mean reproductive success (\overline{ARS}) has a dramatic negative influence on the power of the different tests, the hierarchy in the performance of the different tests does not change across the values of mean reproductive success. Therefore, we chose to present the results with pooled \overline{ARS} only (figure 2.6) Furthermore, it should be noted that although the σ_{ρ}^2 parameter values are the same in this section as in the previous one (0,0.1,0.5,1 and 2), they correspond to much smaller realized variances, as the variance is introduced on the original scale and not on a log-scale as previously. For correspondence between the variances on the two scales, see table 2.3.

Table 2.3: Realized variance on the log scale as a function of variance introduced on the original scale (σ_{ρ}^2) and mean reproductive success ($\bar{\rho}$)

ARS	σ_{ρ}^2 on original scale				
	0	0.1	0.5	1	2
3	0	0.01143	0.06649	0.16947	0.39091
10	0	0.00100	0.00506	0.01027	0.02108
50	0	0.00004	0.00020	0.00040	0.00079

Note: Each realized variance was estimated from the variance of the log of 1,000,000 draws from a normal distribution of mean $\bar{\rho}$ and variance σ_{ρ}^2 .

2.8.5 The snow vole population

A snow vole population, located in the central eastern Alps near Churwalden, Switzerland ($46^{\circ}48' N$, $9^{\circ}34' E$) at 2000m above sea level, has been monitored continuously since 2006. Analyses presented here are based on data collected until 2013. The study site consists of scree, which is the favourite habitat of the species, interspersed by patches of alpine meadows and surrounded by forest and larger meadows, which are not suitable habitats (Janeau and Aulagnier 1997). Four trapping nights are necessary for sampling the complete area. Trapping throughout the whole study area took place two (in one year), three (in three years) or five times (in four years), between late May and mid-October.

Unknown individuals were marked with a subcutaneous passive transponder (PIT, ISO transponder, Tierchip Dasmann, Tecklenburg) and an ear tissue sample was taken (maximum 2mm diameter, Thumb Type Punch, Harvard Apparatus) and stored in 90% ethanol at $-20^{\circ}C$. DNA extracted from the tissue samples was genotyped for 18 specific autosomal microsatellites developed for this population (Wandeler, Ravaoli, and Bucher 2008), and the *Sry* locus was genotyped in order to confirm the sex of all individuals. To identify cases of PIT loss as well as recaptures of juveniles initially too light for PIT injection, an identity analysis in CERVUS v.3.0 (Marshall et al. 1998) was carried out to detect resampled individuals. Parentage was assigned to all juveniles and all first-time captured adults by simultaneously reconstructing parentage and sibship using the R package MasterBayes (Hadfield, Richardson, and Burke 2006). Analyses were performed for each year separately assuming polygamy for males and females and a uniform genotyping error rate of 0.5% for all 18 loci. Parentage was assigned using a parental pool of all adults present in the examined year and the previous year. Because some rare first year individuals reproduce at the end of the season, as evidenced by the observation of pregnant and lactating first year individuals, the “juveniles” were also included in the parental pool of a second analysis excluding parent-offspring mating. Thereby eight additional parentage links could be identified. There were no inconsistencies between the reconstructed pedigree and the transmission of two sex-specific markers: a polymorphic Y-chromosome locus developed for this population (Wandeler and Camenisch 2011) and a fragment of the mitochondrial DNA control region, amplified using vole specific primers (Haring, Herzig-Straschil,

and Spitzenerger 2000). This pedigree was used to measure annual and lifetime reproductive success.

Apparent year-to-year survival could be obtained without mark-recapture modeling as the recapture probability on a given year was virtually 1: no animal was not captured in a year but captured later, and no animal was ever found to be a parent of a juvenile in a year when it had not been captured. This is not surprising since mark-recapture modeling within years estimated a between-occasion recapture probability of 0.924 (SE 0.012) for adults and of 0.814 (SE 0.030) for juveniles.

2.8.6 Univariate models of survival and reproduction in the snow vole population

The following two tables (2.4 and 2.5) present all the estimates from the univariate models used to estimate the individual-level variance in survival and reproductive propensities for the snow vole population.

Table 2.4: Estimates of coefficient of the mixed model for survival in the real snow vole data set

	Estimate	SE	p-value	Bootstrap 95% CI
Random effects:				
σ_{id}^2	0.000	-	0.500	[0;0.248]
σ_{year}^2	5.622	-	$< 10^{-16}$	[1.133;13.158]
Fixed effects:				
intercept	-1.754	0.830	0.035	[-3.393;-0.111]
age (Juvenile)	1.841	0.230	0.000	[1.369;2.411]
sex (Male)	0.306	0.295	0.300	[-0.389;0.93]
age:sex	-0.705	0.333	0.034	[-1.449;0.091]

Note: σ_{id}^2 and σ_{year}^2 refer to the variance between individuals and between years, respectively. All estimates are shown on the latent scale. The p-values for the significance of the two random effects are computed through a one-sided LRT. No standard errors (SEs) are provided for random effects. Instead, confidence intervals are computed using 1000 parametric bootstraps. The significance of the fixed effects is computed through the default Gaussian approximation provided by the package `lme4`.

Table 2.5: Estimates of coefficients of the mixed model for annual reproductive success in the real snow vole data set

	Estimate	SE	p-value	Bootstrap 95% CI
Random effects:				
σ_{obs}^2	3.3×10^{-10}	-	0.499	[0 ; 0.194]
σ_{id}^2	0.371	-	$< 10^{-16}$	[0.151 ; 0.475]
σ_{year}^2	0.101	-	$< 10^{-16}$	[0.026 ; 0.452]
Fixed effects:				
intercept	0.724	0.131	0.000	[-0.254 ; 0.266]
age (Juvenile)	-5.703	0.369	$< 10^{-16}$	[-7.425 ; -5.125]
sex (Male)	0.046	0.101	0.645	[-0.118 ; 0.200]

Note: σ_{id}^2 and σ_{year}^2 refers to the variance between individuals and between years, respectively. σ_{obs}^2 is a dummy random effect having one level per observation and used to account for potential over-dispersion in Poisson GLMMs. The p-value testing for the significance of these three random effects is computed through a one-sided likelihood ratio test. The significance of the fixed-effects is computed through the default normal approximation provided by the package lme4. Confidence intervals are computed using 1000 parametric bootstraps. The interaction between sex and age was not estimable by lme4: its inclusion produced convergence warnings and its SE was above 10^4 , without affecting other parameter estimates, and therefore it was removed from the model.

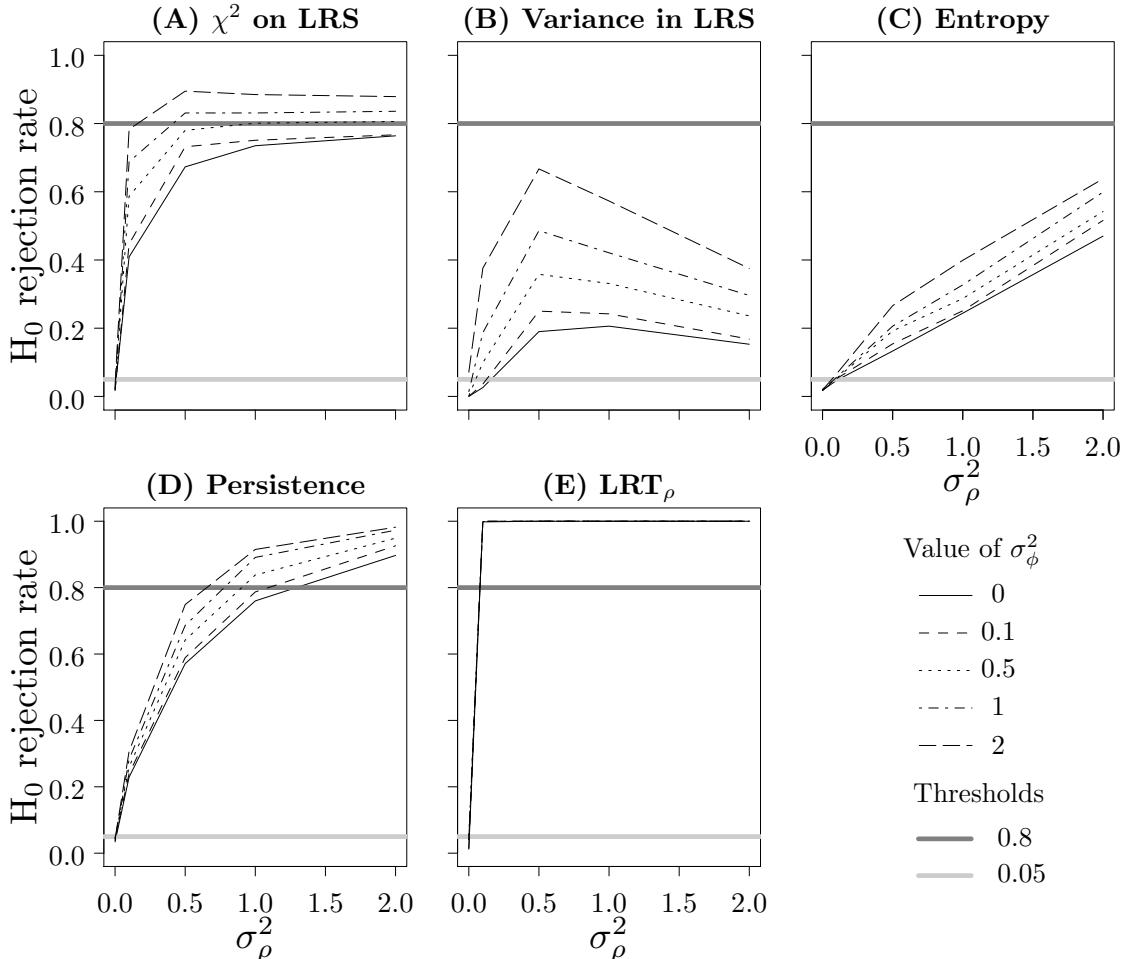


Figure 2.2: Null-hypothesis rejection rates for various methods testing for the presence of fixed heterogeneity, as a function of the variance in reproductive propensity, σ^2_ρ , and survival propensity, σ^2_ϕ , when these variances are introduced on the transformed scales. The methods are: (A) a χ^2 test comparing the LRS distribution in a focal data set to the distribution of LRS distributions obtained through the neutral simulation approach (NS); tests based on proportion of values obtained by NS greater or equal to the value in the focal data set for (B) the variance in LRS, (C) the entropy of the transition matrix between successive annual reproductive success and (D) the persistence of this matrix; (E) a LRT for the significance of the individual random intercept in reproductive success. When $\sigma^2_\rho = \sigma^2_\phi = 0$, the null-hypothesis rejection rates are equal to the type I error rates, which is expected to be 0.05 (light gray line). When $\sigma^2_\rho \neq 0$ or $\sigma^2_\phi \neq 0$, the null-hypothesis rejection rates give (1-type II error rate), i.e. statistical power. The dark gray line indicates the 0.8 threshold. (A)-(D) are related to NS, (E) is related to MM.

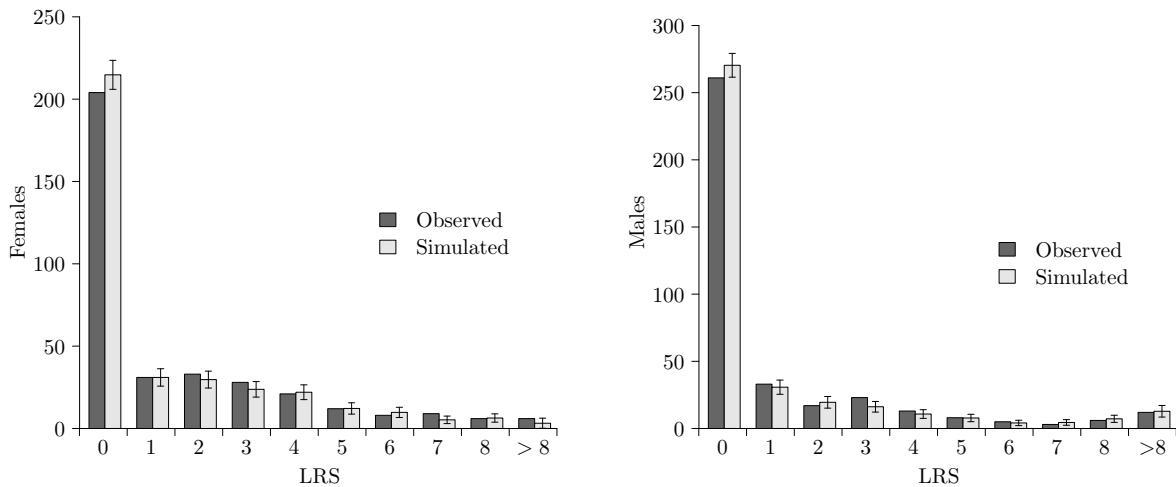


Figure 2.4: Distribution of lifetime reproductive success in the real snow vole data set, observed (dark bars) and simulated through 1000 neutral simulations (light bars with black error bars showing \pm standard deviation), for 2.4.4 females and 2.4.4 males.WRONG

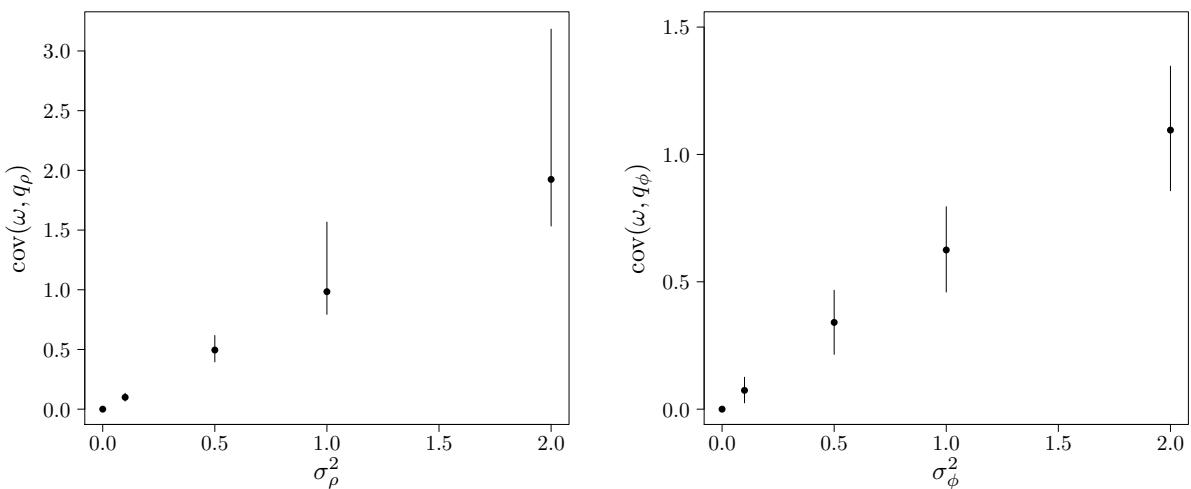


Figure 2.5: Appendix C Strength of selection on individual fixed qualities for survival and reproduction, as a function of the expected variance in these qualities. Strength of selection was measured as the individual-level covariance between the qualities and a proxy for fitness (ω): relative lifetime reproductive success; for reproduction quality and for survival quality. Vertical bars show the 95% interval of the estimate distributions.

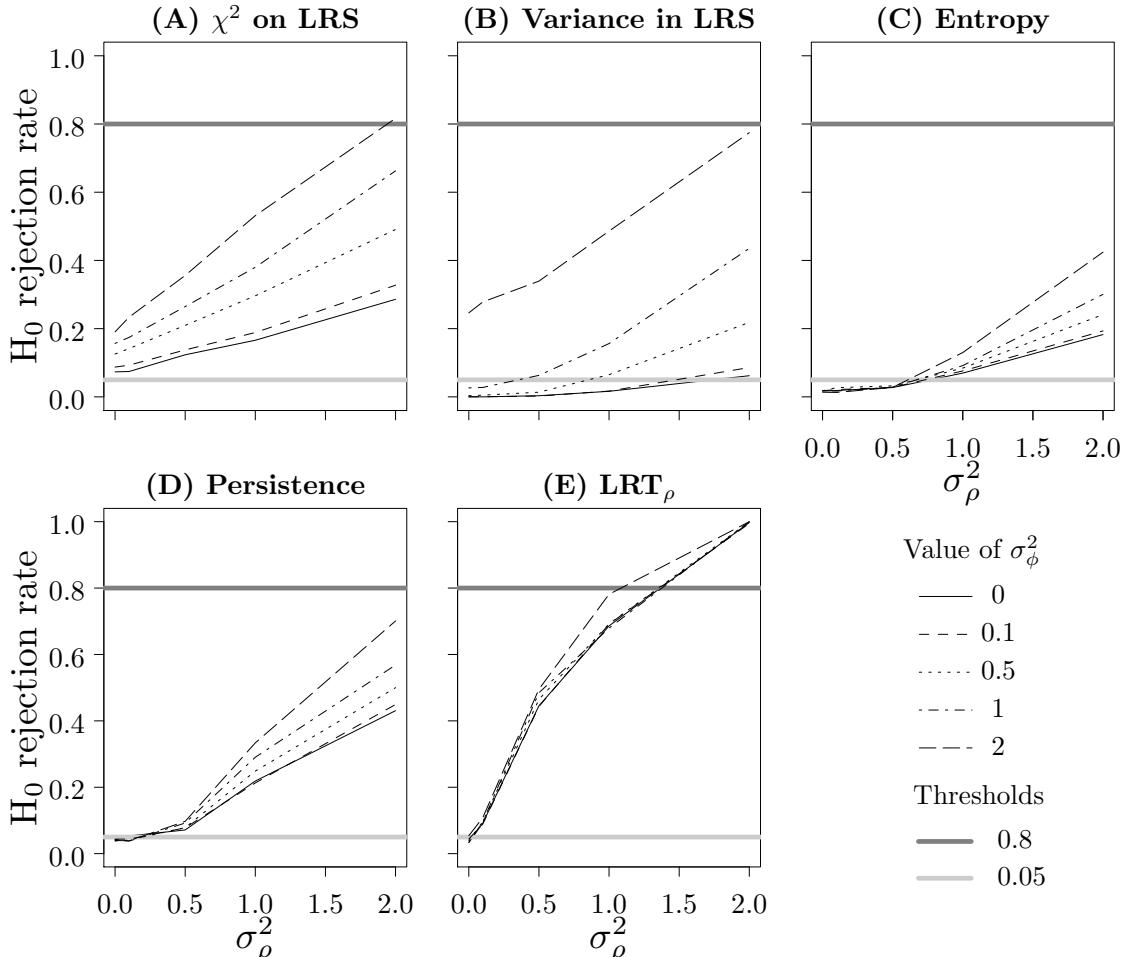


Figure 2.6: Appendix D Null-hypothesis rejection rates for various methods testing for the presence of fixed heterogeneity, depending on the variance in reproductive propensity, σ_ρ^2 , and on the variance in survival propensity, σ_ϕ^2 , when these variances are introduced on the original scales. The methods are: (A) a χ^2 test comparing the Lifetime Reproductive Success (LRS) distribution in a focal data set to the distribution of LRS distributions obtained through the neutral simulation approach (NS); tests based on proportion of values obtained by NS greater or equal to the value in the focal data set for (B) the variance in LRS, (C) the entropy of the transition matrix between successive annual reproductive success and (D) the persistence of this matrix; (E) a Likelihood Ratio Test for the significance of the individual random intercept in reproductive success. When $\sigma_\rho^2 = \sigma_\phi^2 = 0$, the null-hypothesis rejection rates are equal to the type I error rates, which is expected to be 0.05 (light gray line). When $\sigma_\rho^2 \neq 0$ or $\sigma_\phi^2 \neq 0$, the null-hypothesis rejection rates give (1-type II error rate), i.e. statistical power, which should be above 0.8 (dark gray line). (A)-(D) are related to NS, (E) is related to MM.

2.8.7 Estimation of the latent correlation between survival and reproduction

Here we provide additional details on the bivariate models to test for the latent correlation between the propensity to reproduce and the propensity to survive. See main text for more details on the univariate analyses.

In univariate models for reproduction fitted using `lme4`, neither the sex by age interaction, nor the dummy random effect controlling for overdispersion was significant. With `MCMCglmm`, the non-significance was confirmed by bivariate models using the deviance information criterion (DIC) and Bayesian credibility intervals for these two parameters. Moreover, by default `MCMCglmm` takes into account any overdispersion in a distribution assumed to be Poisson. Therefore we did not include these two explanatory variables in the final model. Posterior predictive checks revealed that the bivariate model correctly predicted the number of zeros for ARS (observed 820, predicted 807 ± 23). Moreover, the year-level covariance between survival and reproduction was estimated close to zero, and fixing it to zero improved DIC, so it was fixed to zero in the final model. Finally, the package `MCMCglmm` always includes a residual variance component for binary variables, although this variance is not estimable. We fixed this residual variance to 1, as suggested in the package course notes (<http://www.cran.r-project.org/web/packages/MCMCglmm/vignettes/CourseNotes.pdf>). This model can be written as:

$$\begin{pmatrix} \rho_{i,t} \\ \phi_{i,t} \end{pmatrix} \sim \begin{pmatrix} f_\rho \\ f_\phi \end{pmatrix} + \begin{pmatrix} \sigma_{\rho(\text{year})}^2 & 0 \\ 0 & \sigma_{\phi(\text{year})}^2 \end{pmatrix} + \begin{pmatrix} \sigma_{\rho\phi(\text{ind})}^2 & \sigma_{\rho\phi(\text{ind})} \\ \sigma_{\rho\phi(\text{ind})} & \sigma_{\phi(\text{ind})}^2 \end{pmatrix} + \begin{pmatrix} \sigma_{\rho(\text{res})}^2 & \sigma_{\rho\phi(\text{res})} \\ \sigma_{\rho\phi(\text{res})} & \sigma_{\phi(\text{res})}^2 \end{pmatrix}$$

where f_ρ and f_ϕ denote the fixed part of the model and both include an intercept, sex, age and their interaction. The σ^2 terms refer to variances and the $\sigma_{\rho\phi}$ terms refer to the covariances between ARS and survival, either at the level of years (year), of individuals (ind) or of the residuals (res).

The correlation between the individual propensity to survive and to reproduce was then calculated as $\sigma_{\rho\phi(\text{ind})} / \sqrt{\sigma_{\rho(\text{ind})}\sigma_{\phi(\text{ind})}}$. We used 1000 MCMC samples from 1,100,000 iterations with a thinning of 1000 and a burn-in of 100000. We used a non-informative parameter expanded prior. The residual variance of survival was fixed to 1, as this variance is not identifiable in binomial models. We then refitted the same model while fixing $\sigma_{\rho\phi(\text{ind})}$, $\sigma_{\rho(\text{ind})}^2$ or $\sigma_{\phi(\text{ind})}^2$ to zero, in order to compare the DIC of the two models. Although model selection on the variance-covariance random components is an active area of research (e.g. Burnham and Anderson 2002, chapter 6), the use of DIC has been shown to be robust, at least under some conditions (Wilberg and Bence 2008; Barnett et al. 2010). All models were checked by graphically assessing convergence and good mixing, and using Heidelberg stationarity tests. Moreover, thinning was sufficient to keep all auto-correlations between successive samples below 0.05.

The Bayesian bivariate model identifies variance in the ability to reproduce, $\sigma_{\rho(\text{id})}^2$. Although it is smaller than in the univariate model (table 2.7), it was still different from zero, as 97% of the posterior sample is above 0.01 and removing the random effect from the model substantially increases the DIC (table 2.6). Similar to the uni-

variate model, the estimate of the variance in the ability to survive is small, with a large uncertainty. Including this effect in a model improves (i.e. decreases) DIC in one instance (model 4 versus model 5) but not in another instance (model 2 versus model 3), see table 2.6. However, this effect appears in the best model. There is thus a large uncertainty in the estimation of variance in the ability to survive and mixed evidence for its existence. Similarly, the correlation between the two individual random effects is estimated with a large credibility interval overlapping 0 (table 2.7), and the inclusion of this parameter improves only marginally the DIC of the models (table 2.6). Nevertheless, the mode of the posterior distribution is positive and the effect is present in the best model. Altogether, these results provide limited support for the biological significance of the latent correlation between survival and reproduction.

Table 2.6: Deviance information criterion (DIC) and difference to the best model (Δ DIC), for five bivariate models of ARS and survival with different individual random effect structures

model	$\sigma_{\rho(ind)}^2$	$\sigma_{\phi(ind)}^2$	$\sigma_{\rho,\phi(ind)}$	DIC	Δ DIC
1	Yes	Yes	Yes	2554.587	0.000
2	Yes	Yes	No	2556.793	2.206
3	Yes	No	No	2556.100	1.513
4	No	Yes	No	2560.945	6.358
5	No	No	No	2564.187	9.600

Note: A “Yes” indicates that the parameter was included in the model, a “No”, that it was not. The parameters are $\sigma_{\rho(ind)}^2$, the individual-level variance in ARS; $\sigma_{\phi(ind)}^2$ the individual-level variance in survival; $\sigma_{\rho,\phi(ind)}$ the individual-level covariance between reproduction and survival. Note that it is possible to include $\sigma_{\rho,\phi}$ only when both $\sigma_{\rho(ind)}^2$ and $\sigma_{\phi(ind)}^2$ are also included in the model.

Table 2.7: Variance and correlation components for a bivariate model of survival and reproduction

	Posterior mode	95% CI
$\sigma_{\rho(ind)}^2$	0.167	$[1.4 \times 10^{-4}; 0.342]$
$\sigma_{\phi(ind)}^2$	8.9×10^{-3}	$[9.4 \times 10^{-7}; 1.048]$
$\sigma_{\rho\phi(ind)}$	0.322	$[-0.682; 0.974]$
$\sigma_{\rho(year)}^2$	0.122	$[0.030; 0.917]$
$\sigma_{\phi(year)}^2$	7.585	$[2.074; 73.123]$
$\sigma_{\rho(res)}^2$	0.230	$[1.4 \times 10^{-4}; 0.342]$
$\sigma_{\phi(res)}^2$	1	fixed
$\sigma_{\rho\phi(res)}$	0.180	$[-0.313; 0.576]$

Note: 95% CI shows 95% highest posterior density intervals.

Chapter 3

Disentangling evolutionary, plastic and demographic processes underlying trait dynamics: A review of four frameworks

“Then why do you want to know?”

“Because learning does not consist only of knowing what we must or we can do, but also of knowing what we could do and perhaps should not do.”

— Umberto Eco, *The Name of the Rose* (1954)

Koen J. van Benthem*, Marjolein Bruijning*, Timothée Bonnet*, Eelke Jongejans[†], Erik Postma[†], Arpat Ozgul[†] Accepted in Methods in Ecology and Evolution

* co-first authors

† co-last authors

3.1 abstract

1. Biologists are increasingly interested in decomposing trait dynamics into underlying processes, such as evolution, plasticity and demography. Four important frameworks that allow for such a decomposition are the quantitative genetic animal model (AM), the ‘Geber’ method (GM), the age-structured Price equation (APE), and the integral projection model (IPM). However, as these frameworks have largely been developed independently, they differ in the assumptions they make, the data they require, as well as their outcomes and interpretation.
2. Here we evaluate how each framework decomposes trait dynamics into underlying processes. To do so, we apply them to simulated data for a hypothetical animal population. Individual body size was affected by, among others, genes, maternal effects and food intake. We simulated scenarios with and without selection on body size, and with high and low heritability.
3. The APE and IPM provided similar results, as did the AM and GM, with important differences between the former and the latter. All frameworks detected

positive contributions of selection in the high but not in the low selection scenarios. However, only the AM and GM distinguished between the high and low heritability scenarios. Furthermore, the AM and GM revealed a high contribution of plasticity. The APE and IPM attributed most of the change in body size to ontogenetic growth and inheritance, where the latter captures the combined effects of plasticity, maternal effects and heritability. We show how these apparent discrepancies are mostly due to differences in aims and definitions. For example, the APE and IPM capture selection, whereas the AM and GM focus on the response to selection. Furthermore, the frameworks differ in the processes that are ascribed to plasticity and in how they take into account demography.

4. We conclude that no single framework provides the ‘true’ contributions of evolution, plasticity and demography. Instead, different research questions require different frameworks. A thorough understanding of the different definitions of their components is necessary for selecting the most appropriate framework for the question at hand, and for making biologically meaningful inferences. This work thus supports both future analysis as well as the careful interpretation of existing work.

3.2 Introduction

Understanding trait and population dynamics and how the two are intertwined is crucial for predicting population resilience and viability (e.g. Merilä and Hendry 2014). Hence, which processes shape population-level trait dynamics (i.e. changes in trait distributions over time) is a fundamental question in ecology and evolution, and one which is gaining in urgency given mounting concern regarding the consequences of anthropogenic environmental change for natural populations (e.g. Parmesan 2006).

Phenotypic trait distributions may be altered across generations by genetic (i.e. evolutionary) processes, as well as by non-genetic processes, such as phenotypic plasticity. Since the realisation that evolutionary and ecological processes may act on the same time scale, distinguishing between the role of evolution and plasticity has been the subject of a substantial body of research (Hairston et al. 2005; Gienapp et al. 2008; Post and Palkovacs 2009). To complicate matters further, changes in the demographic structure of a population may additionally shape trait distributions (Coulson and Tuljapurkar 2008). Hence, understanding and predicting trait dynamics ideally requires simultaneously taking into account all three processes (Pelletier et al. 2007; Schoener 2011).

To date, four major frameworks aiming at distinguishing between the role of evolution, phenotypic plasticity and demography have been developed: 1) The quantitative genetic framework, particularly the animal model (AM; e.g. Henderson 1950), 2) the ‘Geber method’ (GM; Hairston et al. 2005), 3) the age-structured Price equation (APE; Coulson and Tuljapurkar 2008), and 4) the application of the APE in conjunction with an integral projection model (IPM; Easterling, Ellner, and Dixon 2000; Ellner and Rees 2006; Coulson, Tuljapurkar, and Childs 2010). Several studies have tried to explicitly estimate the relative importance of evolution, plasticity and/or demography using

one of these approaches (e.g. Réale et al. 2003; Ezard, Côté, and Pelletier 2009; Ozgul et al. 2009; Rebke et al. 2010; Becks et al. 2012; Morrissey et al. 2012). Nevertheless, fully disentangling and quantifying evolutionary, ecological and demographic processes and ultimately predicting the consequential trait dynamics has proven to be problematic (Gienapp et al. 2008; Schoener 2011; Merilä and Hendry 2014). At least some of these difficulties can be attributed to the large amounts of (individual-based) long-term data required, which are often unavailable for natural populations (Clutton-brock and Sheldon 2010). However, even if sufficient data are available, synthesis of the results from the four frameworks is hampered by the fact that they have been developed largely independently of each other. As a consequence, they differ in their focus and aims, and as we show here, they define biological processes in non-equivalent ways.

Here we provide an overview of the differences, similarities and complementarity of each of these four decomposition frameworks by applying them to the same simulated datasets and comparing their outcomes. Thereby, we evaluate how they quantify the role of different ecological and evolutionary mechanisms in shaping trait dynamics under a range of biological scenarios. Together with a critical review of the theory underlying each of the frameworks, we provide comprehensive insight into their underlying assumptions, as well as the conceptual differences and similarities. This provides a much needed overview of the suitability of each framework with respect to research questions and data availability.

3.3 Applying the four frameworks

3.3.1 Data simulation

Although it comes with the loss of some biological realism, using simulated rather than empirical data enables us to evaluate the frameworks under different scenarios and allows for replication. Furthermore, simulated data do not suffer from the complications introduced by missing data. Finally, it provides a reference that aids the comparison between the results of each framework. Importantly, it is not possible to calculate “true” contributions of for example evolution without first adopting one of the frameworks and their corresponding definitions, therefore, our simulations allow only for a qualitative assessment.

Data were simulated using a two-sex individual-based model of a closed population of a hypothetical animal species, implemented in R (R Core Team 2014). Here, we provide a brief overview, while a more complete description can be found in supporting information S1. We also provide the R code on https://github.com/koenvanbenthem/Disentangling_Dynamics_IBM. We simulated a single trait, body size z . Size at birth is determined by an individual’s genotype (10 loci, with 10 alleles each and mendelian inheritance, more details in S1.1), the body size of its mother (i.e. a maternal effect as in Falconer 1965), and a stochastic component (drawn from a Gaussian distribution; S1.2). Ontogenetic growth results in an increase of body size with age. Growth rate, the proportional increase in body size, decreases with age, and is further influenced by per-capita food availability (S1.3).

Males were randomly assigned to females, who have a 50% chance of becoming reproductive after one year and whose reproductive probability increases with age. The litter size that a female produces depends on per-capita food availability, a stochastic component, and body size (S1.4). Survival probability first increases with age, but starts decreasing after year five, reflecting senescence, and is further influenced by per-capita food availability and body size. Maximum age is 30 years. Furthermore, a trade-off exists between female reproduction and survival, i.e. reproducing at time t decreases survival probability to time $t + 1$ (S1.5).

We simulated fifty time steps (years). After ten years, total food availability started to decline. Every year the available food is divided over all individuals, with some individuals randomly obtaining more than others. Individual food intake affects survival, growth and (female) reproductive success (S1.6). The first ten years were discarded from further analyses to allow the age structure to stabilize (Fig. 3.1f). The remaining data spanned 40 years (i.e. approximately 13 generations), which is comparable to the length of some of the field studies these frameworks have been applied to (Clutton-brock and Sheldon 2010).

To evaluate the behaviour of the frameworks under different circumstances, we simulated four different scenarios. First, survival and fertility selection on body size was either present (s_+) or absent (s_0). Under the s_+ scenarios, there was a positive effect of body mass on survival and on litter size for mothers. Second, the relative importance of genetic variation in shaping body size, commonly measured as heritability, was either high (h_+) or low (h_-). This was done by using either of two pre-defined genotype-phenotype maps: one with big and one with small variation in the effects of alleles. Furthermore, to keep the phenotypic variance comparable, we decreased the plastic component in birth size in the h_+ scenarios. The parameter values for each of the four scenarios (s_0h_- , s_0h_+ , s_+h_- and s_+h_+) can be found in S1.7. To evaluate the effect of stochasticity, each scenario was replicated 100 times.

Fig. 3.1 provides an illustration of some of the key characteristics of the datasets simulated under each scenario. Despite a substantial amount of stochastic variation across replicates within each scenario, clear differences in trait and population dynamics are apparent. As expected, the s_+ scenarios show a positive relation between body size and annual fitness, calculated as the sum of survival and litter size to $t + 1$, whereas the s_0 scenarios do not (Fig. 3.1e). Furthermore, the proportion of the phenotypic variance attributable to variance in the simulated genotypic values (i.e. broad-sense heritability) was ca. 0.50 in the h_+ and 0.08 in the h_- scenario.

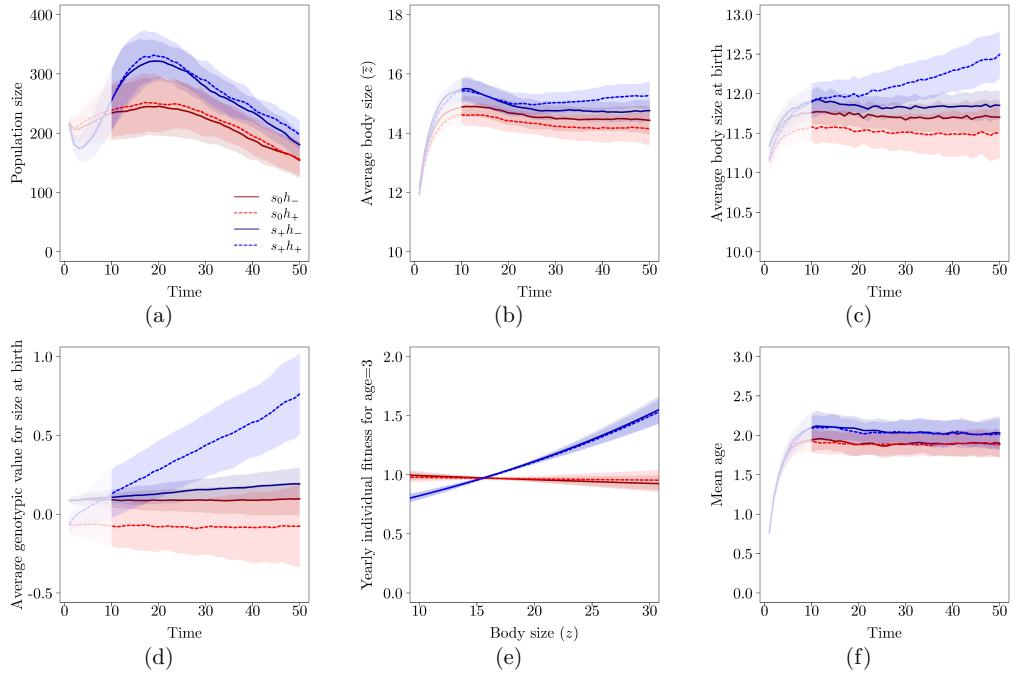


Figure 3.1: Summary of the observed population and trait dynamics of simulated datasets. (a) Trends in population size, (b) changes in mean body size, (c) mean birth size, (d) and genotypic values for body size, (e) relations between body size and yearly individual fitness (sum of survival and litter size at $t + 1$), and (f) changes in mean age. Lines indicate the averages across 100 replicates. Polygons show one standard deviation above and below the average. Red lines indicate s_0 scenarios (no viability and fertility selection), blue lines indicate s_+ scenarios (strong viability and fertility selection). Solid lines indicate h_- scenarios (low heritability), dotted lines indicate h_+ scenarios (high heritability). In a-d and f, the white polygon indicates the first 10 years, which are excluded from further analysis. In (e), lines are averaged predictions based on generalized additive models over all replicates.

Although in all scenarios population size first increased (until year 20) and then decreased (Fig. 3.1a), the population size averaged across replicates reached up to 322 and 334 individuals in scenarios s_+h_- and s_+h_+ , whereas in s_0h_- and s_0h_+ the maximum average population size was 245 and 252 individuals, respectively. Mean body size first increased rapidly, but decreased in all scenarios between the eleventh and fiftieth year (Fig. 3.1b): in s_0h_- with (mean \pm SE) -0.47 ± 0.058 [$-1.45; 0.63$ 95% interval], in s_0h_+ with -0.46 ± 0.061 [$-1.59; 0.68$], in s_+h_- with -0.75 ± 0.051 [$-1.87; 0.08$], and in s_+h_+ with -0.16 ± 0.057 [$-1.12; 0.83$]. Note that the 95% intervals, here and in the rest of the manuscript, are ranges of point estimates across replicates. They reflect the stochasticity of the simulations rather than the precision of the estimates. The standard errors for each average were calculated by dividing the standard deviation of the values of the replicates by 10 (the square root

of the number of replicates). A full power analysis of the methods is beyond the scope of this manuscript.

Contrary to average body size, genotypic values for birth size continued to increase only in scenario s_+h_+ . Here, the change in average genotypic value (across the entire population) between year 11 and year 50 was 0.62 ± 0.022 [0.23; 1.04] (Fig. 3.1d). In s_+h_- a smaller increase was observed 0.08 ± 0.0083 [-0.074; 0.24], whereas s_0h_- and s_0h_+ show on average no change in genotypic values. Correspondingly, average birth size increased only in the s_+h_+ scenario, with 0.58 ± 0.027 [0.092; 1.11], between year 11 and year 50 (Fig. 3.1c).

3.3.2 Decomposing simulated trait dynamics

Rather than providing an exhaustive overview of all methods allowing for the decomposition of trait dynamics, we have chosen to focus on four, commonly-used, frameworks. The four frameworks have different data requirements and do not yield identical results. This is illustrated in the following section, in which we analyse the simulated data using each framework.

Animal Model The animal model (AM) is a quantitative genetic method that was developed for commercial breeding (Henderson 1950, 1976), where it has been used successfully for several decades (e.g. Lynch and Walsh 1998). Only recently has it been applied to wild animal (e.g. Réale et al. 2003; Postma 2014) and plant (Stinchcombe, Simonsen, and Blows 2014) populations. For extensive explanations of the AM as applied to natural populations, see Kruuk 2004 and Wilson et al. 2009.

The AM is a linear mixed effects model that is fitted to individual-level data and assumes a quantitative genetic model, where a phenotypic trait (z) is influenced by a large number of genes with small effects (Roff 2007). The variance in z is partitioned into genetic and non-genetic sources of variation. Under the assumption that this partitioning is additive (i.e. in the absence of genotype-environment correlations and interactions), z can be written as the sum of a population mean (μ), an additive genetic effect (the breeding value, a) and a residual (environmental) value capturing plasticity (e), thus $z = \mu + a + e$. Information on the relatedness between individuals (estimated from a pedigree or genetic markers) is used as a constraint in the fit, allowing for the estimation of a . If the data allow for it, other components contributing to variation in z , such as maternal, common, and permanent environmental effects can be accounted for explicitly. This variance decomposition can be used to estimate genetic change over time—resulting from, for example, selection or genetic drift.

There are several ways to estimate evolution within the AM framework (see discussion), but here we illustrate only one. We fitted a univariate AM and quantified the change in the best linear unbiased predictors (BLUPs) for the breeding values over time (Postma 2006; Hadfield et al. 2010). We used body size as the sole response variable, and intercepts for breeding values, maternal effects, permanent environment, and year were included as random effects. Maternal and permanent environment effects were modelled by fitting maternal and individual identity, respectively. An alternative specification of the maternal effects, more in line with the simulation process, is briefly discussed further below. Age was included as a continuous fixed effect (both

as linear and quadratic terms). All fits were performed using the R-package MCMCglmm (Hadfield 2010) using inverse-Wishart priors with variance and degree of belief both set to 1. The posterior distributions were estimated based on 1,000 MCMC samples, from 50,000 iterations with a thinning interval of 40 and a burn-in of 10,000, thus ensuring that the correlation between successive samples of all parameters is below 10%.

We estimated the temporal trend in the BLUPs for all random effects. We accounted for their uncertainty following Hadfield et al. 2010 by performing a regression of the BLUPs on time for each MCMC sample of the model. This provided a posterior distribution of linear slope coefficients, estimating the change in additive genetic, maternal, and permanent environment effects per time step. More details on the fitted models are given in S2.1.

As depicted in Fig. 3.2a, in all scenarios the contributions of evolution and individual plasticity were largest, while the contributions of permanent environment and maternal effects were very small. On average, the per year change in breeding values was positive in both scenario s_+h_- (0.0013 ± 0.0003 [$-0.0038; 0.0095$]) and scenario s_+h_+ (0.014 ± 0.0007 [$0.00021; 0.029$]). Note that the large error bars in Fig. 3.2a) mostly reflect a substantial amount of variation in the rate of evolutionary change among replicates due to genetic drift, rather than the uncertainty in the point estimates. Negative contributions of individual plasticity were found, particularly in the scenarios with selection -0.02 ± 0.0013 [$-0.049; 0.0018$] and -0.019 ± 0.0013 [$-0.045; 0.0029$] for h_- and h_+ , respectively.

Despite substantial drift, we would expect the contribution of evolution averaged over replicates to be 0 in the s_0 scenarios. Instead, our model inferred a genetic decline for h_- and h_+ of -0.0057 ± 0.0005 [$-0.016; 0.0040$] and -0.0073 ± 0.0009 [$-0.024; 0.0087$], respectively. The AM therefore estimates evolution with a negative bias. The reason is a mismatch between the model structure and the simulation process. As mean size decreases with time, the maternal contributions to birth size decreases. Because we modelled maternal effects as maternal identity rather than maternal current size, this change is mistaken for evolution. We performed an additional analysis using maternal size instead of maternal identity, which strongly reduced this artefact (details and results in S2.2).

Geber method The ‘Geber method’ (GM) (Hairston et al. 2005) is a very general method that quantifies how temporal changes in various factors influence the response variable of interest. Because of this generality, the biological assumptions depend on the specific implementation. The GM may for example estimate how temporal changes in mean breeding value \bar{a} and in an environmental factor k such as food availability propagate to a population-level response variable X , such as mean trait value. Examples of its application can be found in Ellner, Geber, and Hairston 2011 and Becks et al. 2012.

Our implementation of the GM follows the analysis of fledgling mass in Ellner, Geber, and Hairston 2011. We took the average body size (\bar{z}) as the population-level response variable, and decomposed the change in \bar{z} into a contribution of the environment (\bar{k}) and a contribution of a phenotypic change in size at birth. The latter was decomposed further into an evolutionary (\bar{a}) and a plastic component (\bar{p}):

$$\frac{d\bar{z}}{dt} = \frac{\partial \bar{z}}{\partial \bar{k}} \frac{d\bar{k}}{dt} + \frac{\partial \bar{z}}{\partial \bar{a}} \frac{d\bar{a}}{dt} + \frac{\partial \bar{z}}{\partial \bar{p}} \frac{d\bar{p}}{dt} \quad (3.1)$$

For each year between years 11 and 50, we calculated the mean body size (\bar{z}), mean size at birth of newborns, the average food availability that alive individuals had access to during their life up to that moment (\bar{k}), and the mean breeding value as estimated by the AM (\bar{a}) (see above). As breeding values can not be observed directly, the application of the GM to empirical data relies on other methods such as the AM for their estimation. Finally, we calculated a plasticity term (\bar{p}), equal to the difference between the average size at birth and the average breeding value for size at birth. Thereby this term only captured plasticity in mass at birth. We fitted a linear model to estimate the effects of \bar{a} , \bar{p} and \bar{k} on \bar{z} . Using this model, together with separate linear models that describe how each of the three underlying factors changes over time, we evaluated their respective influence on \bar{z} . This procedure is described in more detail in S3.1.

The results of the GM are shown in Fig. 3.2b. The results for the evolutionary component are, as expected, nearly identical to the results of the AM. This evolutionary component is counter-acted by a decrease in food availability, as is shown by the negative ‘environmental’ contributions. The latter is largest for the s_+ scenarios, under which population size is higher (Fig. 3.1a) and per capita food availability therefore lower.

The average contributions of plasticity are more equivocal. Whereas we expected the slight reduction in maternal body size, and hence in the maternal effect, to result in a minor negative contribution of plasticity, we instead see mainly positive contributions. This is the result of the downwardly biased trend in the breeding values (as discussed above). When the analysis was repeated with the ‘true’ genotypic values from the simulations instead of the estimated breeding values, all scenarios showed negative contributions of plasticity (S3.2).

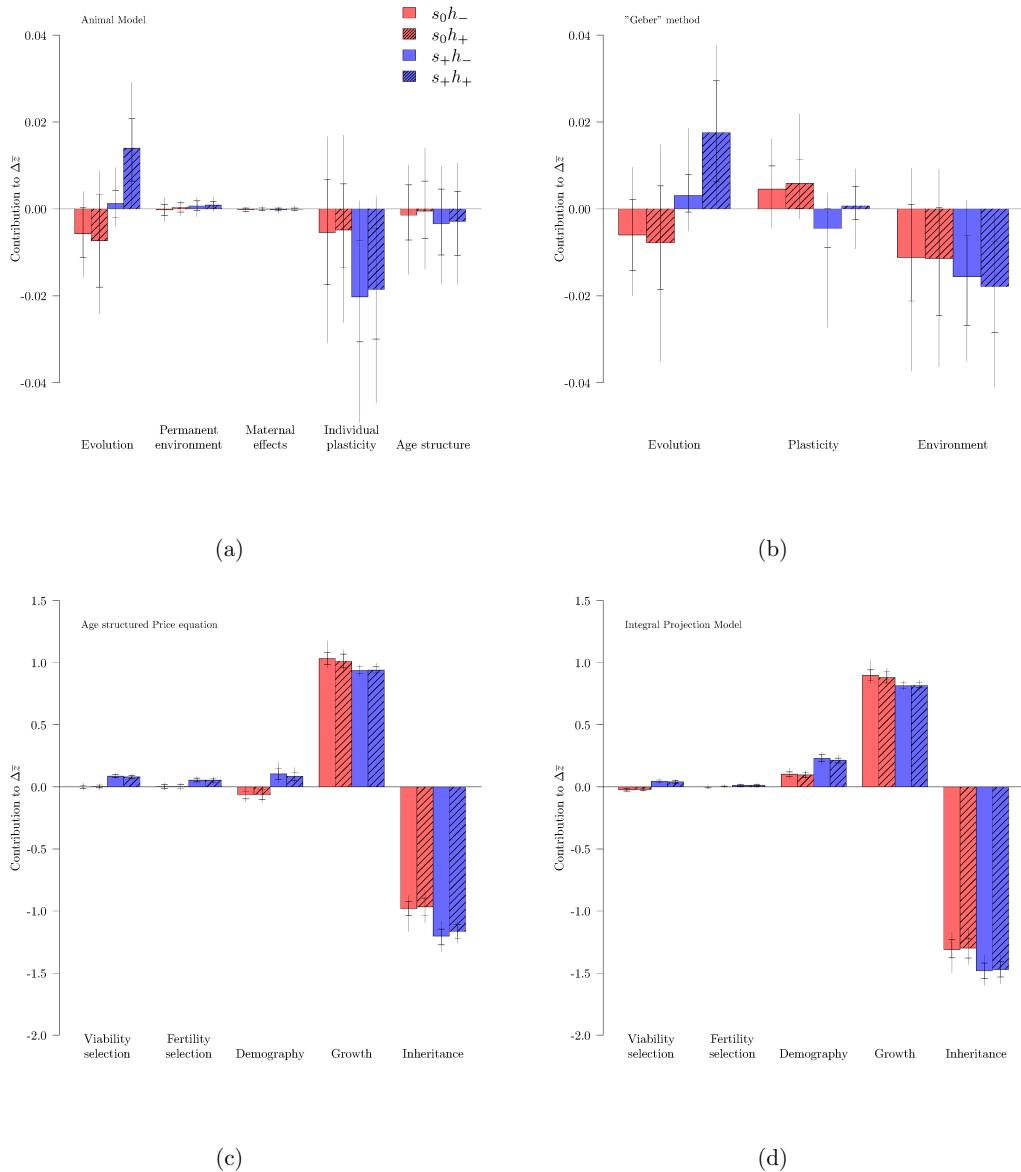


Figure 3.2: Results of the different frameworks when applied to the simulated scenarios. (a) Animal model. (b) "Geber" method. (c) Age-structured Price equation and (d) Integral projection model. In (c) and (d), demography includes changes in average body size due to the age structure, inheritance is the sum of offspring mother difference and offspring difference covariance. In (a-d), red bars indicate s_0 scenarios, blue bars indicate s_+ scenarios. Solid bars indicate h_- scenarios, and shaded bars indicate h_+ scenarios. Error bars represent the range in which 68% (error bars until horizontal lines) and 95% (entire error bars) of the contributions lie when applied to 100 replicates. The y-axis is always average contribution to mean trait change per year, although the scaling is different in (a), (b) versus (c), (d).

Age-structured Price Equation The age-structured Price equation (APE) (Coulson and Tuljapurkar 2008) is an extension of the Price equation (Price 1970). The APE does not explicitly consider genetic variation. It decomposes the change in mean trait value into seven additive components. All these contributions are either averages of, or covariances between, observable individual properties (e.g. individual survival and body size).

The two selection terms describe how selective disappearance (viability selection, VS) and selective reproduction (fertility selection, FS) alter the mean trait value. Here, VS is the covariance between z and survival, which scales with the difference in the average trait value of the whole population and the part of the population that survives to the next time step (e.g. Rebke 2012). This is referred to as the selection differential in the evolutionary literature (Robertson 1966; Lande and Arnold 1983). The contribution to the change in mean trait value due to ontogenetic development of surviving individuals is captured by the growth term. The two inheritance-related contributions were combined into one (S4.3). This combined term measures the contribution to changes in average body size due to the difference between the mother's body size (at time of giving birth) and her offspring's body size at birth (i.e. between generations). Because offspring are generally smaller than mothers, the inheritance contribution will typically be negative. This stresses that the inheritance term should not be confused with heritability, which can not be negative. Finally, the two demography contributions, here also combined into one, describe change resulting from the age structure (S4.2). The demography term arises because the other contributions are calculated per age class. This takes into account that their values depend not only on the trait value of an individual, but also on its age. The total contribution is obtained by a weighted sum of the age specific contributions.

The APE thus allows for an exact decomposition of $\Delta\bar{z}$ in discrete time into components of viability selection, fertility selection, ontogenetic growth, inheritance, and demography in populations with overlapping generations. It has been applied to a range of mammals species (Coulson and Tuljapurkar 2008; Ozgul et al. 2009, 2010; Canale et al. 2016). See S4.1 for the full equation and an explanation of the terms. Note that a stage-structured version of the Price equation has also been developed (Barfield, Holt, and Gomulkiewicz 2011).

As is commonly done in demographic analyses, we applied the APE to the female part of the population only. Under the s_0 scenarios, we find that the average VS and FS are both indistinguishable from zero (Fig. 3.2c). For the s_+ scenarios, the contribution of selection is positive, and there is no difference between the s_+h_+ and s_+h_- scenarios (VS: 0.081 ± 0.0012 [0.060; 0.10] and 0.090 ± 0.0013 [0.063; 0.11] respectively, FS: 0.054 ± 0.0015 [0.027; 0.079] and 0.055 ± 0.0015 [0.029; 0.082] respectively). Finally, the demographic contribution differs between the s_0 and s_+ scenarios, but does not differ between h_+ and h_- . This combined demography term scales with the between-age class covariance between fitness and body size (S4.2). In agreement with our simulation processes, this covariance is strong and positive, as older age classes have larger average body size, and larger individuals have higher fitness in the s_+ scenarios. The negative contribution in the s_0 scenarios is the result of a negative effect of age on survival, which in the absence of positive selection will dominate the between-age class covariance. The biggest contribution to changes in average body size comes from on-

togenetic growth. This component is slightly lower in the s_+ scenarios, due to smaller per capita food availability.

The inheritance term is more negative in the s_+ than in the s_0 scenarios. This is because in the s_+ scenarios larger mothers produce more offspring, which on average results in a larger difference between mother and offspring size: although the maternal trait value when giving birth is higher, their offspring's trait value at birth does not increase by the same amount. This leads to the average contribution of inheritance becoming more negative. Furthermore, we see that contributions from inheritance are slightly smaller (less negative) under the h_+ scenarios than under the h_- scenarios. This is because with increasing heritability, the mother-offspring difference decreases, leading to a less negative inheritance term.

Integral Projection Model The integral projection model (IPM) is a general model for projecting continuous distributions in discrete time. When describing a population, it often considers four life history processes: survival, reproduction, growth and inheritance (Ellner and Rees 2006). The dependencies of these processes on a continuous phenotypic trait z are estimated using regression models. No assumptions concerning the underlying genetics are made. Based on these regressions, the trait distribution at time $t + 1$ can be predicted from the trait distribution at time t (as well as demographic properties, such as population growth rates, e.g. Adler, Ellner, and Levine 2010; Merow et al. 2014). Over the past years, IPMs have been used to address a range of eco-evolutionary questions (e.g. Metcalf et al. 2008; Smallegange and Coulson 2013; Traill, Schindler, and Coulson 2014). While the specific decomposition we use involves applying the APE to a fitted IPM, as proposed by Coulson, Tuljapurkar, and Childs 2010, approaches using a sensitivity analysis also exist (e.g. Coulson et al. 2011; Traill, Schindler, and Coulson 2014).

An IPM was parametrized for each simulated dataset, and as we did for the APE, we only considered females. Models describing individual growth, survival and reproduction (both the probability of reproducing and the number of offspring) were fitted using generalized linear mixed models with appropriate link functions (logit for survival and reproduction probability, log for number of offspring). The contribution of inheritance was estimated as a linear regression of offspring size at birth on the size of the mother at the time of giving birth, as done in Traill, Schindler, and Coulson 2014. This differs fundamentally from heritability (h^2), where offspring size is related to the mother's size, both at the same fixed developmental stage (e.g. birth) (Chevin 2015). For all life history processes, we tested five different models: a full model containing age, size and their interaction, as well as all models nested within this full model. Furthermore, each model included a random effect for year. The model with the lowest AIC was selected and used for the IPM.

Using the selected models, a 3100×3100 matrix was parametrized (i.e. 31 age classes, 100 size classes per age class, ranging between 1 and 50) for each replicate. See S5 for more details on model fitting and the construction of the IPMs. For each IPM, we used the observed population vector at each time step (excluding the first ten years) to project the population vector to the next time step ($t + 1$). Changes in population structure, and thereby changes in \bar{z} , are decomposed into contributions from different life history processes.

We found very similar patterns as in the APE (Fig. 3.2d). Both viability and fertility selection were detected in the s_+ scenarios (VS was 0.045 ± 0.00096 [0.026; 0.063] and 0.041 ± 0.00098 [0.024; 0.060]; FS was 0.012 ± 0.00074 [0.00; 0.026] and 0.012 ± 0.00076 [-0.0044; 0.028], for h_- and h_+). In contrast, in the s_0h_- and s_0h_+ scenarios, average viability selection was -0.024 ± 0.0011 [-0.045; -0.0024] and -0.019 ± 0.0010 [-0.039; -0.00024], respectively, and fertility selection was -0.00069 ± 0.00069 [-0.014; 0.012] and 0.00068 ± 0.00059 [-0.011; 0.014]. As in the APE, the contribution of inheritance to $\Delta\bar{z}$ was large and negative in all scenarios, and was more negative in the s_+ scenarios. Furthermore, there was a consistently positive contribution of ontogenetic growth, with weaker effects in the s_+ scenarios, again due to lower per capita food availability. As in the APE, we considered both demographic terms together. This term showed positive contributions in all scenarios.

To allow for a better comparison with the other three frameworks, here we focus on the average value of $\Delta\bar{z}$, and how much various processes contribute to this. When quantifying how much of the *year-to-year variation* in $\Delta\bar{z}$ is explained by each process (as for example in Ozgul et al. 2009), the IPM and APE provide more divergent results (S6).

3.4 Discussion

We have decomposed changes in mean body size into underlying processes by applying four major frameworks to simulated data. Thereby we have shown that these frameworks differ substantially in their data requirements, which processes they consider, how these are defined, and how changes in the mean trait value are assigned to them. In the following sections we will discuss and compare the theory underlying the four frameworks, illustrated by our simulations. We will discuss the inherent differences among frameworks regarding evolution, plasticity, demography, and measures of uncertainty. These are summarised in Table 3.1. We finish by discussing each framework with respect to data availability and the research question at hand.

We have simulated scenarios with and without selection on body size, and with low and high heritability. As multiple processes influence and interact with body size, these scenarios resulted in divergent and relatively complex population and trait dynamics (Fig. 3.1). For example, in addition to genetic effects, size at birth was influenced by maternal effects and stochasticity. Moreover, ontogenetic growth was subject to both stochastic variation and a decrease in per-capita food availability. We also included a trade-off between viability and fertility. It is exactly this complexity that highlights the need for a robust framework that allows disentangling the underlying processes and quantifying their importance.

Selection and evolution

All four frameworks infer positive selection on body size in the s_+ scenarios, but not in the s_0 scenarios (Fig. 3.2). The APE and IPM detect positive viability and fertility selection in both the s_+h_+ and the s_+h_- scenarios. The AM and GM detect a strong increase in mean breeding values in the s_+h_+ scenario and a small yet positive contri-

Table 3.1: A selection of research questions and to what extent frameworks may be used to answer them, ranging from impossible without major modifications (--) to being answered by the standard formulation of the framework already (++) . AM = animal model, GM = Geber method, APE = age-structured Price equation and IPM = integral projection model. Note that scores are based on the specific application of the frameworks as we reviewed here; this involves the univariate AM, and the application of the APE to the IPM, in case of the IPM. Alternative approaches of the frameworks are mentioned in the discussion.

Question	AM	GM	APE	IPM
Does the change in trait value have a genetic basis?	++	+	--	--
Is selection acting on the trait?	+	+	++	++
Is the trait heritable?	++	±	-	-
Is the age structure responsible for the change in mean trait value?	+	±	++	++
How does individual heterogeneity affect trait value z ?	+	±	--	-
How do trait dynamics affect population dynamics?	-	+	-	++
Is an environmental change responsible for the change in mean trait value?	+	++	--	-

bution in the s_+h_- scenario. Importantly, the AM and GM estimate a genetic change (due to selection and/or drift) whereas the IPM and GPE estimate selection. This is highlighted by the fact that the AM and GM estimate a much larger contribution of evolution in the s_+h_+ compared to the s_+h_- scenario. This contrasts with the IPM and APE, where the contribution of selection is independent of the heritability.

Due to a misspecification of the maternal effects in the AM, we find a negative contribution of evolution in the s_0 scenarios. This mismatch highlights the need to adapt the model structure to the study system. Only then reliable conclusions can be drawn from the AM (see also Hadfield, Wilson, and Kruuk 2011). Indeed, we show that contributions are closer to the simulation process when we use a more appropriate specification of the maternal effects (S2.2).

Here we have chosen to quantify the contribution of evolutionary change to trait dynamics by measuring the temporal change in BLUPs for breeding value in a univariate animal model. Within a quantitative genetic framework, we could also have used the heritability estimated by the AM to apply the breeder's equation and estimate the expected response to selection. This approach has proven its effectiveness under breeding conditions, although nonlinearities in the parent-offspring regression or the trait value-fitness relationship may bias predictions (Heywood 2005). More serious difficulties arise in natural populations, where the prediction of evolution can be biased when selection acts on genetically correlated traits or when an environmental

variable dominates the covariation between traits and fitness (Rausher 1992; Morrissey, Kruuk, and Wilson 2010).

A third approach relies on a bivariate AM that estimates genetic and environmental (co)variances between a trait and a proxy for relative fitness (Lande 1979; Lynch and Walsh 2014). The additive genetic covariance is of particular interest, as following the Robertson-Price identity it provides a direct estimate of the evolutionary change (Robertson 1966; Price 1970; Lynch and Walsh 2014). Although more data demanding, this approach does not require the assumptions of the breeder's equation to be fulfilled (Morrissey et al. 2012), and avoids potentially biased trends in breeding values (Postma 2006).

Unlike the AM and GM, which quantify the change in breeding values, the APE and IPM estimate the contribution of selection, irrespective of whether this yields a genetic response. The overall contribution of selection is obtained by summing over all age-specific selection contributions. This is an attempt to remove the between-age covariation between traits and fitness (Engen, Kvalnes, and Sæther 2014), which is instead captured by the demography term. However, the age correction is not continuous, and therefore the choice of age classes determines how this total contribution of demography and selection is partitioned (see S4.4 for an example).

Most studies that have applied the APE or IPM framework to natural vertebrate populations have found a relatively small role for selection in shaping trait dynamics (e.g. Ozgul et al. 2009; Traill, Schindler, and Coulson 2014). This is in line with our application, as even in the s_+ scenarios, the contribution of the other processes was estimated to be many times larger. In the IPM, the interpretation of selection in terms of evolutionary potential critically depends on the heritability. Heritability is, however, not assessed by the IPM. Indeed, the inheritance function relates juvenile to adult (maternal) trait values, and ignores the fact that individual growth trajectories may be heritable (Chevin 2015). Alternatively, trait inheritance can be incorporated in the IPM by implementing size at birth as a fixed trait influencing offspring size (Vindenes and Langangen 2015), or by explicitly modelling the transmission of additive genetic effects within the IPM (Coulson et al. 2015; Childs, Sheldon, and Rees 2016).

Plasticity

Plasticity includes all individual-level phenotypic changes that are not attributable to genetic changes. While all four frameworks estimate a large contribution of plasticity in all scenarios, they attribute them to different biological processes. This makes it difficult to directly compare the importance of plasticity across frameworks and may potentially lead to confusion. In this section we will focus on plasticity in birth size.

We used the AM to separately estimate plasticity due to maternal and permanent environment effects (Fig. 3.2a). The contribution of maternal effects was very small. This may seem at odds with the effect of maternal adult size on offspring size at birth in our simulations, but as explained above, this was due to a mismatch between the model structure (which included a random effect of maternal identity) and the data generating process (which included an effect of maternal body size). The contribution of permanent environment was low, which is in line with the lack of a trend in the stochastic component of birth size in our simulations.

The GM captures plasticity in size at birth due to both maternal effects and stochasticity in one single term (Fig. 3.2b). Because plasticity at birth is here defined as the difference between actual birth weight and the breeding value for birth weight of an individual, by construction, the plasticity term has to compensate for the bias in estimated breeding values.

In the APE and IPM frameworks, plasticity at birth and growth are intrinsically entangled. Whereas ontogenetic growth forms the main plastic contribution to $\Delta\bar{z}$ (Figs. 3.2c and 3.2d), the body size that is attained through ontogenetic growth is only partially (through maternal effects) transmitted to the offspring. Most of the ontogenetic growth will thus be reset in the offspring: this is reflected in the strong negative contribution from inheritance (for a more detailed explanation of the inheritance terms, see S4.3.1). Also, because we applied the APE only on the female part of the population, changes in offspring body size due to selection on males (and thus fathers) will be attributed to the inheritance term.

The role of the environment

Whereas the GM defines an explicit environmental factor, in the other frameworks, the environment influences trait dynamics only indirectly through selection, plasticity and/or demography. For example, high food availability may lead to an increase in average body size through plasticity. At the same time, increased food availability may decrease competition, and thereby affect selection.

In our implementation of the GM, we defined the environment as the total food intake of an individual. Hence, the environment mainly acts through within-individual plasticity through its effect on ontogenetic growth. Importantly, the outcome of the GM depends fully on how evolution, plasticity and environment are defined. When applying the GM to field data, where not all processes are known, it is thus crucial to first identify the main drivers and attribute them to evolutionary, plastic or demographic processes.

Although in the APE and IPM effects of the environment are implicitly present in all terms, in our implementation there is no explicit quantification of this environmental effect. Although an IPM can include an environmental variable, its contribution will not be quantified by the APE when applied to that IPM. However, alternative applications of the IPM that allow exploring the effects of such an environmental variable do exist (e.g. Vindenes, Engen, and Sæther 2011). Alternatively, one can parametrize different IPMs for different environments (e.g. Ozgul et al. 2010) and use comparison methods such as life table response experiments to see how population and trait dynamics differ between these environments (Rees and Ellner 2009).

In our version of the AM, all contributions of changes in the environment, such as decreasing food availability, are captured within the residual individual plasticity term. Although not commonly done, environmental contributions can be estimated more explicitly by including additional fixed or random effects (Charmentier, Garant, and Kruuk 2014). One possibility is the inclusion of a fixed effect of food availability. Furthermore, it is possible to model interactions between the environmental variable and the additive genetic effects.

Demography

We showed how the combined demography terms in the APE scale with the covariance of age class-specific fitness and age class-specific average body size. The demography terms hence do not reflect the effect of changes in the age structure between time t and $t + 1$, but rather differences due to the existing age structure at time t . As such it provides a demographic correction of estimates of selection, similar to the one proposed by Engen, Kvalnes, and Sæther 2014.

In the AM we have quantified the demographic contribution by multiplying the slope of body size with respect to age with the predicted change in average age. This contribution is most negative in the s_+ scenarios, meaning that here a change (decrease) in the average age in the populations over time led to a decrease in the average body size in these scenarios, in agreement with the observed slight decrease in average age as shown in Fig. 3.1f.

Unexplained variation and uncertainty

Making conclusive statements regarding which factor has the largest influence on $\Delta\bar{z}$ requires a measure of the uncertainty in the estimates of each contribution. So far we have only considered the range of point estimates over the replicates, generally showing smaller ranges for APE and IPM. However, APE and IPM were estimating processes that were constant throughout replicates (e.g. selection), whereas the AM and GM were estimating quantities subject to stochasticity (e.g. genetic drift). Differences in range are thus due to the stochasticity in the simulations rather than the uncertainty in the point estimates.

While the AM allows the estimation of confidence intervals for each estimated contribution, in our implementation of the IPM, APE and GM there is no direct measure of uncertainty. For the GM, confidence intervals can be obtained using bootstrapping methods (as in Ellner, Geber, and Hairston 2011). As of yet, the lack of uncertainty quantification is a major drawback of the application of the IPM and APE. However, measures of uncertainty accompanying parameter estimates could be propagated to the decomposition, by using bootstrapping, and in the case of the IPM also by MCMC sampling.

Residual variance is explicitly quantified in the AM. The GM does evaluate the residuals of the underlying regressions, but does not include these in the final results (Ellner, Geber, and Hairston 2011). In contrast, the APE is an exact framework and hence the residual variance is zero. However, it is still subject to sampling variance. Although the IPM uses the APE, it is constructed by fitting statistical models to the data, each with their own residual term.

The AM can also account explicitly for additional sources of variation, by including the corresponding random effects (for example, we incorporated individual identity as a random effect to account for individual heterogeneity that could not be explained by additive genetic variation). IPMs can also include a random individual effect in the underlying fitted functions. This inclusion accounts for individual heterogeneity when estimating vital rates. However, although this individual heterogeneity should explicitly be propagated to the actual IPM (Vindenes and Langangen 2015), the IPM

is often parametrized with the random effect set to zero. Thereby not all individual heterogeneity is accounted for. Setting the random effect to zero might also bias the prediction because of Jensen's inequality (e.g. Fox and Kendall 2002). Individual heterogeneity can be incorporated by defining a "static trait", in addition to the continuous state variable. This static trait does not change during development, and reflects fixed individual heterogeneity caused by e.g. differences in size at birth, genetics or experienced environment (e.g. Ellner and Rees 2006; Vindenes and Langangen 2015). The role of individual heterogeneity is not captured in the GM and APE. In case of the GM, the effects of individual heterogeneity, as estimated by the AM, can be propagated to the response variable.

Conclusions and future directions

The urge for a better understanding of eco-evolutionary dynamics is reflected in the range of frameworks that have been developed over the last few years aiming at quantifying the underlying processes (Pelletier, Garant, and Hendry 2009; Schoener 2011), especially within the light of the consequences of climate change (Gienapp et al. 2008; Lavergne et al. 2010). Yet, a general, predictive framework is lacking, and applications to field data remain scarce. We have shown that the animal model (AM), 'Geber' method (GM), age-structured Price equation (APE) and integral projection model (IPM) frameworks differ in generality and data requirements. Importantly, key processes are defined and interpreted differently in the different approaches. We emphasize that one should be careful when applying one of the frameworks and interpreting the outcomes as being the "true" contributions of different processes. Indeed, we have shown that each framework has its own set of components and definitions.

All four frameworks have only recently been proposed in their current form, and are only starting to be applied to conservation-related questions. In this review we have explored the frameworks and their assumptions and limitations. Our findings are summarized in Table 3.1, where we provide an overview of which framework seems most suitable for which research question. The AM enables estimation of quantitative genetic parameters, and genetic change in particular, that cannot be estimated by the other frameworks. However, the AM, and the estimation on quantitative genetic parameters in general, is data demanding and it can be difficult to isolate confounding sources of variation when data sets are small. When individual data on reproduction, survival and growth are available, and one is interested in explicitly quantifying the contribution of within-age class selection, IPM and APE are logical choices. The AM can explicitly evaluate the effect of individual heterogeneity. Although the IPM can take this information into account as well by fitting mixed effects models, it does not evaluate its effect on trait dynamics. In contrast to the other frameworks, only the GM focuses on population-level parameters, but knowledge (or assumptions) on processes is required beforehand, i.e. it must be known what processes are shaped by evolution (or plasticity) and which by the environment.

We conclude that in isolation none of the frameworks provides a full picture. Instead, each framework answers different questions and has different data requirements. By highlighting both the similarities and the differences, we hope to have aided in the interpretation of existing work. Furthermore, we hope this work will help

researchers interested in eco-evolutionary questions in making an informed choice regarding the most suitable framework for their particular question.

Acknowledgements

We are grateful to Luis-Miguel Chevin for providing us with constructive feedback throughout the process. We also thank Robert O’Hara, Jarrod Hadfield, Loeske Kruuk and one anonymous reviewer for their help in considerably improving this manuscript. Finally, we thank Tim Coulson for comments on an earlier version of the manuscript. This work was funded by the Swiss National Science Foundation project grants (31003A_141110 and 31003A_159462/1 to EP, 31003A_146445 to AO) and an ERC starting grant (#337785 to AO).

Data Accessibility

The code for simulating the data that we used is provided as online supplementary information, and can also be found on https://github.com/koenvanbenthem/Disentangling_Dynamics_IBM.

References

- Adler, P. B., Ellner, S. P., and Levine, J. M. 2010. “Coexistence of perennial plants: an embarrassment of niches.” *Ecology Letters* **13** (8): 1019–1029.
- Barfield, M., Holt, R. D., and Gomulkiewicz, R. 2011. “Evolution in stage-structured populations.” *The American Naturalist* **177** (4): 397–409.
- Becks, L., Ellner, S. P., Jones, L. E., and Hairston, N. G. 2012. “The functional genomics of an eco-evolutionary feedback loop: linking gene expression, trait evolution, and community dynamics.” *Ecology Letters* **15** (5): 492–501.
- Canale, C. I., Ozgul, A., Allainé, D., and Cohas, A. 2016. “Differential plasticity of size and mass to environmental change in a hibernating mammal.” *Global Change Biology* **in press**.
- Charmentier, A., Garant, D., and Kruuk, L. E. B. 2014. *Quantitative Genetics in the Wild*. Oxford: Oxford University Press.
- Chevin, L.-M. 2015. “Evolution of adult size depends on genetic variance in growth trajectories: A comment on analyses of evolutionary dynamics using integral projection models.” *Methods in Ecology and Evolution* **6** (9): 981–986.
- Childs, D. Z., Sheldon, B. C., and Rees, M. 2016. “The evolution of labile traits in sex- and age-structured populations.” *Journal of Animal Ecology* **85** (2): 329–342.
- Clutton-brock, T. and Sheldon, B. C. 2010. “Individuals and populations: the role of long-term, individual-based studies of animals in ecology and evolutionary biology.” *Trends in Ecology & Evolution* **25** (10): 562–573.

- Coulson, T., MacNulty, D. R., Stahler, D. R., Wayne, R. K., Smith, D. W., et al. 2011. "Modeling effects of environmental change on wolf population dynamics, trait evolution, and life history." *Science* **334** (6060): 1275–1278.
- Coulson, T., Plard, F., Schindler, S., Ozgul, A., and Gaillard, J.-M. 2015. "Quantitative genetics meets integral projection models : unification of widely used methods from ecology and evolution." *arXiv*. <http://de.arxiv.org/abs/1509.01351v1>.
- Coulson, T. and Tuljapurkar, S. 2008. "The dynamics of a quantitative trait in an age-structured population living in a variable environment." *The American Naturalist* **172** (5): 599–612.
- Coulson, T., Tuljapurkar, S., and Childs, D. Z. 2010. "Using evolutionary demography to link life history theory, quantitative genetics and population ecology." *Journal of Animal Ecology* **79** (6): 1226–1240.
- Easterling, M. R., Ellner, S. P., and Dixon, P. M. 2000. "Size-specific sensitivity: applying a new structured population model." *Ecology* **81** (3): 694–708.
- Ellner, S. P., Geber, M. A., and Hairston, N. G. 2011. "Does rapid evolution matter? Measuring the rate of contemporary evolution and its impacts on ecological dynamics." *Ecology Letters* **14** (6): 603–614.
- Ellner, S. P. and Rees, M. 2006. "Integral projection models for species with complex demography." *The American Naturalist* **167** (3): 410–428.
- Engen, S., Kvalnes, T., and Sæther, B.-E. 2014. "Estimating phenotypic selection in age-structured populations by removing transient fluctuations." *Evolution* **68** (9): 2509–2523.
- Ezard, T. H. G., Côté, S. D., and Pelletier, F. 2009. "Eco-evolutionary dynamics: disentangling phenotypic, environmental and population fluctuations." *Philosophical Transactions of the Royal Society B: Biological Sciences* **364** (1523): 1491–1498.
- Falconer, D. 1965. "Maternal effects and selection response." In *Genetics today, Proceedings of the XI International Congress on Genetics, Vol. 3*. Edited by S. J. Geerts, 763–774. Oxford: Pergamon.
- Fox, G. A. and Kendall, B. E. 2002. "Demographic stochasticity and the variance reduction effect." *Ecology* **83** (7): 1928–1934.
- Gienapp, P., Teplitsky, C., Alho, J. S., Mills, J. A., and Merilä, J. 2008. "Climate change and evolution: disentangling environmental and genetic responses." *Molecular Ecology* **17** (1): 167–178.
- Hadfield, J. D. 2010. "MCMC Methods for Multi-Response Generalized Linear Mixed Models: The MCMCglmm R Package." *Journal of Statistical Software* **33** (2): 1–22. <http://www.jstatsoft.org/v33/i02/>.
- Hadfield, J. D., Wilson, A. J., Garant, D., Sheldon, B. C., and Kruuk, L. E. B. 2010. "The misuse of BLUP in ecology and evolution." *The American Naturalist* **175** (1): 116–125.

Chapter 3 Evolution or plasticity?

- Hadfield, J. D., Wilson, A. J., and Kruuk, L. E. B. 2011. "Cryptic evolution: does environmental deterioration have a genetic basis?" *Genetics* **187** (4): 1099–113.
- Hairston, N. G., Ellner, S. P., Geber, M. A., Yoshida, T., and Fox, J. A. 2005. "Rapid evolution and the convergence of ecological and evolutionary time." *Ecology Letters* **8** (10): 1114–1127.
- Henderson, C. 1950. "Estimation of genetic parameters." *Annals of mathematical statistics* **21** (2): 309–310.
- . 1976. "A simple method for computing the inverse of a numerator relationship matrix used in prediction of breeding values." *Biometrics* **32** (1): 69–83.
- Heywood, J. 2005. "An exact form of the breeder's equation for the evolution of a quantitative trait under natural selection." *Evolution* **59** (11): 2287–2298.
- Kruuk, L. E. B. 2004. "Estimating genetic parameters in natural populations using the 'animal model'." *Philosophical Transactions of the Royal Society B: Biological Sciences* **359** (1446): 873–890.
- Lande, R. and Arnold, S. J. 1983. "The measurement of selection on correlated characters." *Evolution* **37** (6): 1210–1226.
- Lande, R. 1979. "Quantitative genetic analysis of multivariate evolution, applied to brain: body size allometry." *Evolution*: 402–416.
- Lavergne, S., Mouquet, N., Thuiller, W., and Ronce, O. 2010. "Biodiversity and climate change: integrating evolutionary and ecological responses of species and communities." *Annual Review of Ecology, Evolution, and Systematics* **41**:321–350. <http://www.annualreviews.org/doi/pdf/10.1146/annurev-ecolsys-102209-144628>.
- Lynch, M. and Walsh, B. 1998. *Genetics and Analysis of Quantitative Traits*. Sinauer Associates.
- . 2014. *Evolution and Selection of Quantitative Traits*. Online version 30 December 2014. Sunderland, Massachusetts, USA: Sinauer.
- Merilä, J. and Hendry, A. P. 2014. "Climate change, adaptation, and phenotypic plasticity: The problem and the evidence." *Evol Appl* **7** (1): 1–14.
- Merow, C., Dahlgren, J. P., Metcalf, C. J. E., Childs, D. Z., Evans, M. E., Jongejans, E., Record, S., Rees, M., Salguero-Gómez, R., and McMahon, S. M. 2014. "Advancing population ecology with integral projection models: a practical guide." *Methods in Ecology and Evolution* **5** (2): 99–110.
- Metcalf, C. J. E., Rose, K. E., Childs, D. Z., Sheppard, A. W., Grubb, P. J., and Rees, M. 2008. "Evolution of flowering decisions in a stochastic, density-dependent environment." *Proceedings of the National Academy of Sciences* **105** (30): 10466–10470.
- Morrissey, M. B., Kruuk, L. E. B., and Wilson, A. J. 2010. "The danger of applying the breeder's equation in observational studies of natural populations." *Journal of Evolutionary Biology* **23** (11): 2277–2288. <http://www.ncbi.nlm.nih.gov/pubmed/20831731>.

- Morrissey, M. B., Parker, D. J., Korsten, P., Pemberton, J. M., Kruuk, L. E. B., and Wilson, A. J. 2012. "The prediction of adaptive evolution: empirical application of the secondary theorem of selection and comparison to the breeder's equation." *Evolution* **66** (8): 2399–2410.
- Ozgul, A., Childs, D. Z., Oli, M. K., Armitage, K. B., Blumstein, D. T., Olson, L. E., Tuljapurkar, S., and Coulson, T. 2010. "Coupled dynamics of body mass and population growth in response to environmental change." *Nature* **466** (7305): 482–485. <http://dx.doi.org/10.1038/nature09210> <http://www.nature.com/nature/journal/v466/n7305/abs/nature09210.html>.
- Ozgul, A., Tuljapurkar, S., Benton, T. G., Pemberton, J. M., Clutton-Brock, T. H., and Coulson, T. 2009. "The dynamics of phenotypic change and the shrinking sheep of St. Kilda." *Science* **325** (5939): 464–467. <http://www.ncbi.nlm.nih.gov/pubmed/19574350>.
- Parmesan, C. 2006. "Ecological and evolutionary responses to recent climate change." *Annual Review of Ecology, Evolution, and Systematics*: 637–669.
- Pelletier, F., Garant, D., and Hendry, A. 2009. "Eco-evolutionary dynamics." *Philosophical Transactions of the Royal Society B: Biological Sciences* **364** (1523): 1483–1489. <http://rstb.royalsocietypublishing.org/content/364/1523/1483.short>.
- Pelletier, F., Clutton-Brock, T., Pemberton, J., Tuljapurkar, S., and Coulson, T. 2007. "The evolutionary demography of ecological change: linking trait variation and population growth." *Science* **315** (5818): 1571–1574. <http://www.sciencemag.org/content/315/5818/1571.abstract>.
- Post, D. M. and Palkovacs, E. P. 2009. "Eco-evolutionary feedbacks in community and ecosystem ecology: interactions between the ecological theatre and the evolutionary play." *Philosophical Transactions of the Royal Society B: Biological Sciences* **364** (1523): 1629–1640. <http://rstb.royalsocietypublishing.org/content/364/1523/1629.short>.
- Postma, E. 2006. "Implications of the difference between true and predicted breeding values for the study of natural selection and micro-evolution." *Journal of Evolutionary Biology* **19**:309–320.
- Postma, E. 2014. "Four decades of estimating heritabilities in wild vertebrate populations: improved methods, more data better estimates?" In *Quantitative Genetics in the Wild*, edited by A. Charmentier, D. Garant, and L. E. B. Kruuk. Oxford: Oxford University Press.
- Price, G. R. 1970. "Selection and covariance." *Nature* **227**:520–521.
- R Core Team. 2014. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. <http://www.R-project.org/>.
- Rausher, M. D. 1992. "The measurement of selection on quantitative traits: biases due to environmental covariances between traits and fitness." *Evolution* **46** (3): 616–626.

Chapter 3 Evolution or plasticity?

- Réale, D., McAdam, A. G., Boutin, S., and Berteaux, D. 2003. "Genetic and plastic responses of a northern mammal to climate change." *Proceedings of the Royal Society of London B: Biological Sciences* **270** (1515): 591–596.
- Rebke, M. 2012. "From the Price equation to a decomposition of population change." *Journal of Ornithology* **152** (2): 555–559.
- Rebke, M., Coulson, T., Becker, P. H., and Vaupel, J. W. 2010. "Reproductive improvement and senescence in a long-lived bird." *Proceedings of the National Academy of Sciences, USA* **107** (17): 7841–7846. eprint: <http://www.pnas.org/content/107/17/7841.full.pdf>. <http://www.pnas.org/content/107/17/7841.abstract>.
- Rees, M. and Ellner, S. P. 2009. "Integral projection models for populations in temporally varying environments." *Ecological Monographs* **79** (4): 575–594.
- Robertson, A. 1966. "A mathematical model of the culling process in dairy cattle." *Animal Production* **8**:95–108.
- Roff, D. a. 2007. "A centennial celebration for quantitative genetics." *Evolution* **61** (5): 1017–1032.
- Schoener, T. W. 2011. "The newest synthesis: understanding the interplay of evolutionary and ecological dynamics." *Science* **331** (6016): 426–429. <http://www.sciencemag.org/content/331/6016/426.abstract>.
- Smallegange, I. M. and Coulson, T. 2013. "Towards a general, population-level understanding of eco-evolutionary change." *Trends in Ecology & Evolution* **28** (3): 143–148. <http://www.ncbi.nlm.nih.gov/pubmed/22944192>.
- Stinchcombe, J. R., Simonsen, A. K., and Blows, M. W. 2014. "Estimating uncertainty in multivariate responses to selection." *Evolution* **68** (4): 1188–1196.
- Traill, L. W., Schindler, S., and Coulson, T. 2014. "Demography, not inheritance, drives phenotypic change in hunted bighorn sheep." *Proceedings of the National Academy of Sciences, USA* **111** (36): 13223–13228.
- Vindenes, Y., Engen, S., and Sæther, B.-E. 2011. "Integral projection models for finite populations in a stochastic environment." *Ecology* **92** (5): 1146–1156.
- Vindenes, Y. and Langangen, Ø. 2015. "Individual heterogeneity in life histories and eco-evolutionary dynamics." *Ecology Letters* **18** (5): 417–432.
- Wilson, A. J., Réale, D., Clements, M. N., Morrissey, M. M., Postma, E., Walling, C. A., Kruuk, L. E. B., and Nussey, D. H. 2009. "An ecologist's guide to the animal model." *Journal of Animal Ecology* **79** (1): 13–26. <http://onlinelibrary.wiley.com/doi/10.1111/j.1365-2656.2009.01639.x/full>.

Chapter 4

The stasis that wasn't: Adaptive body mass evolution is opposite to phenotypic selection in a wild rodent population

Chapter 5

Fluctuating selection

All entities move and nothing remains still.

— Heraclitus cited by Plato, *Cratylus* (4-5th century BCE)

thing

— dude

Chapter 6

General discussion

It is difficult to understand the universe if you only study one planet.

— Miyamoto Musashi, *A Book of Five Rings* (circa 1645)

6.1 Overview

In this thesis, I investigated the causes and consequences of variation in fitness in a wild population. I showed that the variation in proxies for individual fitness is not purely stochastic, but is underlain by variation in latent fitness (chapitre 2). Besides, the variation in latent fitness has an additive genetic component, showing the presence of natural selection and of adaptive evolution in the snow vole population (chapter 4 and 5). I explored ways to decompose the causes of phenotypic changes and identified the animal model from quantitative genetics as a convenient tool to estimate evolution (chapter 3). Using this tool in various ways, I showed that body mass was an important contributor of variation in fitness proxies (chapter 4), but not in a consistent way over time (chapter 5). Nevertheless, body mass was a consistent contributor to variation in genetic variation for fitness (chapters 4 and 5), and therefore, body mass evolved over the study period.

Below, I will discuss further the insight brought by this thesis and the remaining challenges, in understanding the causes of phenotypic variation and the response of wild populations to environmental change.

6.2 The causes of phenotypic variation

6.2.1 Indirect genetic effects

In this thesis, I have decomposed the phenotypic variation of morphological and life-history traits into components related to additive genetic effects, maternal effects or permanent environments. This decomposition was sufficient to measure the rate of evolution of the direct genetic effects (chapter 4), that is, the direct actions of an individual's genes on its own body. Nevertheless, an individual's genes have effects reaching out beyond its body, to the environment, including other individuals (Dawkins 1982), whether it is through interactions between individuals (indirect genetic effects, e.g. maternal effects), or through the pleiotropic action of genes expressed at different life-stages (e.g. genetic conflicts).

Indirect genetic effects are an emerging question in evolutionary quantitative genetics (**OUHUIHOU**) and could be an important component shaping selection and evolution in the snow vole population. Indeed, genes within an individual are likely to affect the phenotype of another individual during at least two types of situations. Snow voles tend to inherit territories from mother to daughter and to form clusters of territories with sisters (García-Navas et al. 2016), but the presence of kin might suppress reproduction in subordinate females. Moreover, as in all placental mammals, maternal effects on offspring phenotypes are prevalent from pregnancy to weaning. Maternal effects have been studied extensively (Wolf and Wade 2009), but only recently have researchers started to measure the genetic variance in maternal effects (**ejfkj**). These genetic maternal effects could provide extra evolutionary potential in addition to that of direct genetic variation (McGlothin and Galloway 2014; Mcfarlane et al. 2015). Preliminary analyses showed the presence of additive genetic maternal effects for body mass in the snow vole population. Genetic maternal effects could be subject to selection and could evolve adaptively, and a full account of body-mass evolution should probably measure this evolution in addition to that of direct additive genetic effects. In chapter 4, maternal genetic effects are not explicitly modelled, and their evolution is assigned to phenotypic plasticity.

Thus, genetic conflicts, that is, genetic trade-off between traits expressed in different individuals (Trivers 1974), are thought to be a major constraint on evolution. For four decades, the existence of genetic parent-offspring remained untested, despite suggestive evidence from behavioural studies (Kölliker et al. 2015). Kölliker et al. 2015 demonstrated that a genetic trade-off between offspring number and offspring size constrains the evolution of size in earrings. Rollinson and Rowe 2015 then presented qualitative evidence suggesting that this constraint is widespread among animals and could be a general explanation for the evolutionary stasis of size. In chapter 4 we briefly explored the possibility that a genetic conflict constrains the evolution of body mass, and found suggestive evidence that it is not the case. The snow vole study system is not an ideal to test this hypothesis, however. First, we do not capture all juveniles—some die or emigrate before their first year—and cannot measure litter size accurately. Because mass is under selection in juveniles, selective disappearance is likely to blur the trade-off signal (Hadfield, Wilson, and Kruuk 2011). Second, the size-number genetic trade-off is best described as an explanation of evolutionary stasis of size or mass, but mass is evolving in the snow vole population, making it more difficult to formulate an expectation for the genetic covariance between mass and litter size. Finally, it is in theory possible to measure the genetic trade-off using quantitative genetics, but nor the exact model to fit nor the modelling tools are published yet (Hadfield 2012; Rollinson and Rowe 2015). An experimental approach remains the only option to quantitatively test for a size-number genetic trade-off (Kölliker et al. 2015), and such an approach appears impossible in a wild population such as Churwalden's snow voles.

6.2.2 Molecular basis of genetic variation

Improve QG estimates by solving problem of realized relatedness and immigrants being treated as part of the base population.

GWAS limited in small populations, but combination with a pedigree improves power greatly. Find causal genetic variants responsible for evolution cf scrub jays.

6.2.3 Origin and maintenance of variation in fitness

Still, our study is a snapshot, too short to expect significant loss of genetic variation. What would happen if the observed selective pressures would carry on? Fluctuating selection might contribute to the maintenance of genetic variation New mutations certainly contribute, but probably not sufficient in general (charlesworth). Population structure.

6.3 The response of wild populations to environmental change

6.3.1 Measuring selection in the wild

For over 150 years, natural selection has been known to cause the match between organisms and their environment, and biologists have attempted to understand its causes and mechanisms. The principle of natural selection is very simple: in a given environment, individuals with a phenotype that favours survival and fertility contribute more to the next generation. Therefore, it can be surprising to see how slowly the understanding of natural selection has developed, and how difficult its study remains. For most of the 20th century, the main brake to progresses was the lack of an unified framework to quantify selection in natural populations (Wade 2006). Such a framework progressively emerged, starting with the appearance of covariance-based methods (Robertson 1966; Price 1970) which efficiently measure the total effect of selection. The most influential breakdown was the popularization of regression-based methods (Lande 1979; Lande and Arnold 1983) which measures the proportional effect of selection per unit of phenotypic variation, and allows to decompose selection into the direct and indirect effects of selection on multiple traits (Broodie III, Moore, and Janzen 1995). Since then, these methods have provided thousands of estimates of selection in natural populations (Kingsolver et al. 2001; Stinchcombe et al. 2008; Kingsolver et al. 2012), showing several general patterns. For instance, directional selection is stronger and more common than suggested by early evolutionists, whereas stabilizing selection appears to be rare, while fertility selection is generally stronger than viability selection (Kingsolver et al. 2012). The abundance of estimates of selection should not be mistaken for a good understanding of natural selection, however. The estimation of selection through regression-methods faces at least three challenges that might severely hamper their significance and explain the general absence of response to selection (Merilä, Sheldon, and Kruuk 2001; Brookfield 2016).

First, to obtain an unbiased measure of selection, fitness should be regressed on the trait of interest. Fitness is rarely observable directly, and fitness proxies must be used instead. Many estimates of selection are computed on fitness components, for instance fertility and survival (Kingsolver et al. 2012). In this case, the estimation of selection can be biased in the presence of a trade-off between fitness components (Thompson

et al. 2011; Kingsolver et al. 2012; Brookfield 2016). Fortunately, this bias appears to be minor in general, with the exception of body mass (Kingsolver and Diamond 2011). For the empirical part of this thesis (chapter 4 and 5), I used fitness proxies that attempted to include all fitness components in order to avoid such biases. Thus, I used lifetime reproductive success when measuring selection within a generation, and annual reproductive success plus twice survival when measuring selection within a year. These fitness proxies are imperfect since we do not capture all juveniles and a trade-off between early juvenile survival and reproduction could bias the selection estimation. Still, the estimation of evolution using Price equation (that is, selection on the genotype) and that using the trend in BLUPs for breeding values (that is, not using any information about selection nor fitness) agree qualitatively (chapter 4), suggesting that the generational proxy for fitness is adequate.

Second, it is possible to estimate the total effect of selection on a trait with selection differentials, but it is much more difficult to disentangle the causal selective effect of a trait from the indirect selection due to other traits. In theory, it is possible to disentangle direct and indirect selection by including all the traits under selection in the analysis (Lande and Arnold 1983). In natural populations, however, it is impossible to know *a priori* what traits are under selection, and often it is impossible to measure all relevant traits (Brookfield 2016; Hadfield 2008). Furthermore, as more traits are included in a selection analysis, the statistical power to detect significant selection on any one trait decreases (Mitchell-Olds and Shaw 1987). This thesis did not bring I did detect significant indirect selection on body mass, but genetic correlations between the traits considered were such that the prediction of evolution was not affected by the inclusion of indirect selection (chapter 4). Only three traits were tested, however, and we cannot exclude that body mass is not under any direct selective pressure. The evolution of body mass could be driven by selection on an unmeasured trait. Nevertheless, this problem is irrelevant to the measures of total selection and evolution, on which chapters 4 and 5 rely.

Third, covariance-based and regression-based methods to estimate phenotypic selection essentially measure the statistical association between traits and relative fitness. Selection must however be a causal association, be it direct or indirect. If the association is entirely mediated by an environmental covariance between traits and fitness, there is no selection and no possibility of genetic response to selection (Price and Liou 1989; Rausher 1992). Body mass, the main trait analysed in this thesis, is likely to be very sensitive to this source of bias. Indeed, a favourable environment—for instance food rich and lacking parasites—is likely to lead to larger mass, high survival, and high fertility. Accordingly, phenotypic estimates of natural selection on mass and size are overwhelmingly positive (Blanckenhorn 2000; Kingsolver et al. 2012) and an excess of environmental covariance does underlie the apparent selection on mass in the snow voles (chapter 4). A solution to the problem is the experimental manipulation of the trait of interest. This can break the link between phenotype and individual quality and reveals the causal action of phenotype on fitness components (e.g. Tinbergen and Sanz 2004; Tscharren and Richner 2006). Still, experimental manipulation is no without its own limitations. Thus, manipulations are work intensive, time consuming and must be thought carefully in order to manipulate the trait of interest without affecting any other trait. Moreover, manipulations cannot easily be applied to all traits.

The approach has been widely used to study selection on brood size, but it is not clear to me how one could manipulate body mass in a controlled way (that is, without accidentally affecting other traits). My approach to the challenge of environmental covariation has been to use quantitative genetics to identify the target of natural selection (chapter 4). After having shown on-going adaptive evolution, I decomposed phenotypic selection into an additive genetic and an environmental component, for various fitness components. I found that only juvenile viability selection showed an additive genetic component, and according to the Robertson-Price identity, was the source of adaptive evolution. Understanding the mechanism of this selection and measuring its strength was then a matter of hypothesis testing. This approach could be used on other systems provided the presence of adaptive evolution. Nonetheless, it requires sufficient phenotypic and relatedness data to fit bivariate animal models. In addition, in the snow vole a single fitness drove evolution, but multiple fitness components could be involved, thus complicating the analysis. Finally, identifying the right fitness component(s) does not guarantee that the phenotypic mechanism of selection can be identified. A good understanding of the biological system will be necessary to formulate a reasonable hypothesis for the cause of selection. The testability of this hypothesis will also depend on data availability and quality, and will be subject to the limits of hypothesis testing approaches: the rate of false positive is equal to the significance level chosen, and a correlation does not prove causation.

6.3.2 Genetic response to environmental change

Predictions based on phenotypic estimates of selection (chapter 2) (chapter 3)

6.3.3 Demographic response to environmental change

Within-population variation in fitness is relevant to the demographic response to environmental changes, either through the direct effects of phenotypic variation (including genetic variation) on demographic rates (Kendall et al. 2011; Vindenes and Langangen 2015; Plard et al. 2016), or indirectly, following genetic changes (Chevin, Lande, and Mace 2010; Turcotte, Reznick, and Hare 2011; Schiffrers et al. 2013).

6.4 Conclusion

References

- Blanckenhorn, W. 2000. "The evolution of body size: what keeps organisms small?" *Quarterly Review of Biology* 75 (4): 385–407.
- Broodie III, E. D., Moore, A. J., and Janzen, F. J. 1995. "Visualizing and quantifying natural selection." *Trends in Ecology & Evolution* 10 (8): 313–318.
- Brookfield, J. F. 2016. "Why are estimates of the strength and direction of natural selection from wild populations not congruent with observed rates of phenotypic change?" *BioEssays* 38:1–8.

Chapter 6 General discussion

- Chevin, L.-M., Lande, R., and Mace, G. M. 2010. "Adaptation, plasticity, and extinction in a changing environment: towards a predictive theory." *PLoS Biology* **8** (4): e1000357.
- Dawkins, R. 1982. *The extended phenotype*. Oxford University Press, Oxford, U.K.
- García-Navas, V., Bonnet, T., Waldvogel, D., Camenisch, G., and Postma, E. 2016. "Consequences of natal philopatry for reproductive success and mate choice in an Alpine rodent." *Behavioral Ecology* **27** (4): 1158–1166.
- Hadfield, J. D. 2008. "Estimating evolutionary parameters when viability selection is operating." *Proceedings of the Royal Society of London B: Biological Sciences* **275** (1635): 723–734.
- . 2012. "The quantitative genetic theory of parental effects." In *The evolution of parental care*, 267–284. Oxford Univ. Press, Oxford, U. K.
- Hadfield, J. D., Wilson, A. J., and Kruuk, L. E. B. 2011. "Cryptic evolution: does environmental deterioration have a genetic basis?" *Genetics* **187** (4): 1099–113.
- Kendall, B. E., Fox, G. A., Fujiwara, M., and Nogeire, T. M. 2011. "Demographic heterogeneity, cohort selection, and population growth." *Ecology* **92** (10): 1985–1993.
- Kingsolver, J. G., Hoekstra, J. M., Berrigan, D., Vignieri, S. N., Hill, C. E., Hoang, A., Gilbert, P., and Beerli, P. 2001. "The strength of phenotypic selection in natural populations." *The American Naturalist* **157** (3): 245–261.
- Kingsolver, J. G. and Diamond, S. E. 2011. "Phenotypic selection in natural populations: what limits directional selection?" *The American Naturalist* **177** (3): 346–57.
- Kingsolver, J. G., Diamond, S. E., Siepielski, A. M., and Carlson, S. M. 2012. "Synthetic analyses of phenotypic selection in natural populations: lessons, limitations and future directions." *Evol Ecol* **26** (5): 1101–1118.
- Kölliker, M., Boos, S., Wong, J. W., Röllin, L., Stucki, D., Raveh, S., Wu, M., and Meunier, J. 2015. "Parent-offspring conflict and the genetic trade-offs shaping parental investment." *Nature Communications* **6**:6850.
- Lande, R. and Arnold, S. J. 1983. "The measurement of selection on correlated characters." *Evolution* **37** (6): 1210–1226.
- Lande, R. 1979. "Quantitative genetic analysis of multivariate evolution, applied to brain: body size allometry." *Evolution*: 402–416.
- Mcfarlane, S. E., Gorrell, J. C., Coltman, D. W., Humphries, M. M., Boutin, S., and Mcadam, A. G. 2015. "The nature of nurture in a wild mammal's fitness." *Proceedings of the Royal Society B: Biological Sciences* **282**:20142422.
- McGlothlin, J. W. and Galloway, L. F. 2014. "The contribution of maternal effects to selection response: An empirical test of competing models." *Evolution* **68** (2): 549–558.

- Merilä, J., Sheldon, B. C., and Kruuk, L. E. B. 2001. "Explaining stasis: microevolutionary studies in natural populations." *Genetica* **112**:199–222.
- Mitchell-Olds, T. and Shaw, R. G. 1987. "Regression analysis of natural selection: statistical inference and biological interpretation." *Evolution* **41** (6): 1149–1161.
- Plard, F., Gaillard, J.-M., Coulson, T., and Tuljapurkar, S. 2016. "Des différences, pourquoi? Transmission, maintenance and effects of phenotypic variance." *Journal of Animal Ecology* **85** (2): 356–370.
- Price, G. R. 1970. "Selection and covariance." *Nature* **227**:520–521.
- Price, T. and Liou, L. 1989. "Selection on clutch size in birds." *The American Naturalist* **134** (6): 950–959.
- Rausher, M. D. 1992. "The measurement of selection on quantitative traits: biases due to environmental covariances between traits and fitness." *Evolution* **46** (3): 616–626.
- Robertson, A. 1966. "A mathematical model of the culling process in dairy cattle." *Animal Production* **8**:95–108.
- Rollinson, N. and Rowe, L. 2015. "Persistent directional selection on body size and a resolution to the paradox of stasis." *Evolution* **69** (9): 2441–2451.
- Schiffers, K., Bourne, E. C., Lavergne, S., Thuiller, W., and Travis, J. M. J. 2013. "Limited evolutionary rescue of locally adapted populations facing climate change." *Philosophical Transactions of the Royal Society B: Biological Sciences* **368** (20120083): 20120083.
- Stinchcombe, J. R., Agrawal, A. F., Hohenlohe, P. a., Arnold, S. J., and Blows, M. W. 2008. "Estimating nonlinear selection gradients using quadratic regression coefficients: double or nothing?" *Evolution; international journal of organic evolution* **62** (9): 2435–40.
- Thompson, D. J., Hassall, C., Lowe, C. D., and Watts, P. C. 2011. "Field estimates of reproductive success in a model insect: behavioural surrogates are poor predictors of fitness." *Ecology Letters* **14** (9): 905–913.
- Tinbergen, J. M. and Sanz, J. J. 2004. "Strong evidence for selection for larger brood size in a great tit population." *Behavioral Ecology* **15** (4): 525–533.
- Trivers, R. L. 1974. "Parent-offspring conflict." *Am Zool* **14**:249–264.
- Tschirren, B. and Richner, H. 2006. "Parasites shape the optimal investment in immunity." *Proceedings of the Royal Society B: Biological Sciences* **273** (1595): 1773–1777.
- Turcotte, M. M., Reznick, D. N., and Hare, J. D. 2011. "Experimental assessment of the impact of rapid evolution on population dynamics." *Evolutionary Ecology Research* **13**:113–131.
- Vindenes, Y. and Langangen, Ø. 2015. "Individual heterogeneity in life histories and eco-evolutionary dynamics." *Ecology Letters* **18** (5): 417–432.

Chapter 6 General discussion

Wade, M. J. 2006. "Natural selection." In *Evolutionary genetics: concepts and case studies*. Edited by C. W. Fox and J. B. Wolf, 399–413. Oxford University Press, Oxford.

Wolf, J. B. and Wade, M. J. 2009. "What are maternal effects (and what are they not)?" *Philosophical Transactions of the Royal Society B: Biological Sciences* **364** (1520): 1107–15.