

Individual-Level Causes and Population-Level Consequences of Variation in Fitness in an Alpine Rodent

Dissertation
zur
Erlangung der naturwissenschaftlichen Doktorwürde
(Dr. sc. nat.)
vorgelegt der
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. Barbara Tschirren
Prof. Dr. Arpat Ozgul
Dr. Marc Kéry

Zürich, 2016

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

Diese Doktorarbeit untersucht die stochastischen und selektiven Ursachen der Variation in Fitnesskomponenten und deren evolutionären Konsequenzen in einer freilebenden Nagetierpopulation. Sie zeigt die gegenwärtige, genetische Evolution von Körpermasse und entkoppelt klassische Selektionsschätzungen von adaptiver Evolution.

Das Herzstück der Evolutionsbiologie liegt im Verständnis der Vielfalt von Organismen. Während über 150 Jahren haben Forscher die Ursachen von intraspezifischer Variation dokumentiert, wie sie zur Artbildung beiträgt und zum Zusammenpassen von Organismen mit ihrer Umwelt. Zunehmende Bedenken wegen der schnellen anthropogenischen Veränderungen haben in letzter Zeit eine erneute Erforschung, wie sich freilebende Populationen an Umweltveränderungen anpassen, vorangetrieben. Dieser neue Fokus offenbart die Schwierigkeiten im Messen von natürlicher Selektion, die Entflechtung von Evolution und plastischen Veränderungen und dem Vorhersagen von evolutionären Entwicklungsverläufen. Unter anderem gibt es nur wenige robuste Beispiele von gegenwärtiger Evolution in freilebenden Populationen, die, die Möglichkeit, dass Evolution Populationen bei schnellen Veränderungen der Umweltbedingungen rettet, fraglich erscheinen lässt. In dieser Doktorarbeit erforsche ich die Ursachen natürlicher Selektion und Evolution in einer freilebenden Population von Schneemäusen (*Chionomys nivalis*). Dank 10 Jahren intensivem Individuenbasiertem Monitoring und Genotypisierung, beinhaltet der Erkenntnisstand dieser Population Lebensweise, morphologische Daten und einen hochaufgelösten Stammbaum. Deswegen ist diese Population unter den besten weltweit verfügbaren um Selektion und Evolution in Aktion zu messen.

Trotzdem ist die Population ziemlich klein und neuerliche Publikationen legen nahe, dass das evolutionäre Potential in kleinen Populationen effektiv von Stochastik in Fitnesskomponenten aufgehoben wird. Ich beurteile diese Methoden, die in diesen Publikationen benutzt wurden und demonstriere, dass die Variation in Fitnesskomponenten nicht ausschliesslich stochastisch ist. Kleine Populationen, einschliesslich diese Schneemäuse, zeigen evolutionäres Potential.

Mit Kollaboratoren vergleiche ich dann vier häufig benutzte methodologische Ansätze um die Anteile von Evolution, Plastizität und Demographie an der phänotypischen Veränderung zu entflechten. Wir identifizieren wichtige Unstimmigkeiten zwischen den Ansätzen, die teilweise vom Gebrauch von unterschiedlichen Definitionen, aber auch vom Besitz von intrinsisch unterschiedlichen Fähigkeiten stammen. Unter den in Betracht gezogenen Ansätzen kann nur quantitative Genetik genetische Veränderungen messen.

Durch die Anwendung von quantitativ-genetischen Methoden an der Schneemaus-Population demonstriere ich, dass sich Körpermasse über die Studiendauer adaptiv entwickelt. Ich zeige auf, dass phänotypische Schätzungen von Selektion nicht

genetische Evolution hervorsagen: weder die durchschnittliche Selektion noch die temporale Variation hängen mit der Rate genetischer Evolution zusammen. Das legt dar, dass die dominante, rein phänotypische Methode zur Messung von Selektion stattdessen die Messung von Variation im Ernährungszustand riskiert. Dennoch habe ich quantitative Genetik zur Identifikation von Selektion verwendet und Selektionsschätzungen erhalten, die mit der beobachteten genetischen Veränderung übereinstimmen.

Diese Doktorarbeit weist gegenwärtige Evolution in einer freilebenden Population nach und zeigt dass evolutionäre Reaktionen auf Umweltveränderungen von rein phänotypischen Methoden weder zuverlässig eingeschätzt noch verstanden werden können; ein expliziter, genetischer Ansatz ist notwendig.

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	13
Chapter 3	
Disentangling evolutionary, plastic and demographic processes underlying trait dynamics: A review of four frameworks	45
Chapter 4	
The stasis that wasn't: Adaptive body mass evolution is opposite to phenotypic selection in a wild rodent population	85
Chapter 5	
Fluctuating selection but no fluctuating evolution in a wild rodent population	105
Chapter 6	
General discussion	125
Bibliography	151
Acknowledgements	153
Curriculum Vitae	159
Addresses of co-authors	167



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. / If we do not find anything very pleasant, at least we shall find something new.

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

1.1 Variation in fitness

1.1.1 The origin of variation

The heart of evolutionary questioning is understanding variation among living beings (Lynch and Walsh 1998; Wayne and Miyamoto 2006; Kruuk, Charmantier, and Garant 2014). As a matter of fact, it is 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 upon these observations, he then goes on to show that variation within species is the fuel generating the astonishing diversity among species, and the striking fit between organisms and their environment.

These fundamental insights immediately gave rise to many more questions about the causes and consequences of variation, some of which remain unresolved until today, more than 150 years later. Nineteenth century biologists particularly struggled with the origin of variation within species. For example, in *the Origin of Species* Darwin writes: “*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. Although by some the environment was thought to play a predominant role, through what would be nowadays called *phenotypic plasticity*, and the effect of ageing was acknowledged as well (Wilkins

¹Some biologists dismissed within-species variation as they considered species to be arbitrary boundaries within a continuum of variation. Although Darwin did not attempt to define species, by explaining how they originate, he made some species definitions indefensible (Wilkins 2009, pp. 129–163).

2009), these did not provide a generally satisfactory explanation for variation within populations.

While Darwinian arguments build on the observation that variation that can appear among siblings of a same litter, clutch, or pod, and that this variation may subsequently transmitted from parent to offspring (Darwin 1859, Chapter 1), the late-nineteenth century was utterly ignorant of the sources of such inherited variation among individuals within a species. It was only at the beginning of the twentieth century when the laws of inheritance were progressively being uncovered and started to spread through the scientific community (Dietrich 2006), and it took another four decades until these laws were formalized into a unified scientific theory that provided an understanding of variation within populations (R. Fisher 1930), and also on a molecular level (Oswald, MacLeod, and McCarty 1944; Watson and Crick 1953). These breakthroughs finally allowed 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.

Our understanding of the causes of variation within species and populations has made terrific progress and now fits elegantly within the broader theory of evolutionary (Pigliucci and Müller 2010). Nevertheless, many aspects of the causes of variation are still in need of refinement and further exploration, especially in natural populations (Kruuk, Charmantier, and Garant 2014). For example, the relative importance of genes and the environment in shaping variation in the wild has been studied in only a few populations of a few, taxonomically biased, species, and concerns only a limited set of traits (Lynch and Walsh 1998; Postma 2014). Furthermore, the consequences of within-species variation remains a very active field of research, trying to understand how, among others, genetic variation translates into adaptive evolution (Brookfield 2016). Although any trait that possesses genetic variation can in principle evolve, it will evolve in an adaptive manner only if subject to selection, be it artificial or natural. As selection occurs when the variation in the trait causes variation in *fitness*, before we can discuss the causes and consequences of selection in any more detail, we must first introduce this difficult concept.

1.1.2 Variation in how fitness is defined

A great deal has been written about the concept of fitness, and multiple, often conflicting, definitions of fitness have been put forward, 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 do not intend to solve the question of how to define fitness here, but I will rather try to make it clear how the word is used throughout this thesis.

First of all, there has been some debate on whether fitness is a realized reproductive outcome, or rather a propensity to reproduce (Brandon and Beatty 1984). However, the consensus is now that the concept of fitness is more useful when it is defined as a propensity, that is, as an expected value that because of stochasticity cannot be measured directly (Brandon and Beatty 1984; P. W. Price 1996; Krimbas 2004), and

here we will adhere to this idea. Furthermore, 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). Importantly, defining fitness as a propensity partly removes the problem of the level at which to define fitness, as 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 as defined at the level of individuals, because they are the unit most easily observable and the primary target of natural selection.

More confusion with respect to the concept of fitness comes from it being alternatively defined as the contribution of an individual to the next generation, or as the asymptotic number of descendants into the distant future (Wade 2006). As we here consider fitness to be property of an individual and because most of the work carried out is based on data covering only about ten generations, it is most intuitive to consider fitness to capture the contribution to the next generation. Besides practical considerations, this allows for a clear, and conceptually crucial, distinction between selection, inheritance and evolution, that is blurred when asymptotic definitions of fitness are employed (R. Fisher 1930; Arnold and Wade 1984). For example, in the simple case of a closed finite population of clonal organisms with no balancing selection, one individual currently living will eventually be the ancestor of the whole population, while all the other currently living individuals will have no descendant at all. As consequence, the asymptotic fitness of the one successful individual equals the population size, and the asymptotic fitness of every other individual is zero. Observing only the starting point and the end point perfectly measures asymptotic fitness, but it does not explain the processes by which the descendants of one individual invaded the population. Did this individual was simply lucky? Did it bear a selective advantage? Did a selective advantage appear by mutation in its descendants? To infer the relative roles of chance (drift), selection, and inheritance (mutations), one must describe the generation-to-generation changes in lineage frequencies and attempt to relate it to observable differences between lineages.

A slightly more contentious question is whether fitness should be defined as an absolute number of offspring (Wade 2006) or as relative number (Rousset 2004), that is, whether “relative fitness” is a meaningful phrase or a tautology. The relative definition avoids appending *relative* to every occurrence of *fitness*, and seems closer to the interest of evolutionary biologists. Nevertheless, many favour the absolute definition, as it has a concrete and easily interpretable meaning. For the sake of consistency, I attempted to yield to this convention as much as possible.

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 (S. a. Frank 2012). Indeed, the first and secondary theorems of natural selection can be rigorously written in terms of gain and loss of bits of information, with populations gaining information by selection, and losing it by imperfect transmission or environmental change. I see great conceptual promises in this view, as it brings together an intuitive meaning of the word *fitness* and the scientific field of information theory, with all its powerful tools and concepts. Nevertheless, despite

showing promise, this interpretation of fitness did not directly influence the work presented in this thesis, and I will therefore not go into it in any more detail here.

To summarise the above, I define fitness as the *expected* number of descendant in the *next generation* of an *individual*.

1.1.3 Causes and consequences of fitness variation in the wild

Why is there variation in individual fitness? This question attracts 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, which makes the existence of genetic variation in fitness paradoxical (Jones 1987). In this thesis, I deal with the second point, the fundamental question of appearance and maintenance of genetic variation in fitness, only briefly in chapter 5. Instead I mostly focus on the sources of variation in fitness, which is addressed in all chapters, and I will approach this from two complementary angles.

In the first, descriptive, approach, one decomposes variation in fitness, i.e. the opportunity for selection, into components of variation. In addition to genetic variation, variation in fitness may originate from variation in early-life, micro-environmental (Turner 2009), and maternal effects (Wolf and Wade 2009). Furthermore, because when working with wild sexually reproducing organisms, individual fitness as we defined it here cannot be observed directly, and their realized reproductive success does not equal their expected reproductive success. This means that researchers have to rely on fitness proxies such as number of offspring and survival, both of which contain also a stochastic component. As the additive genetic variance in fitness is equal to the rate of evolution of fitness (R. Fisher 1930), a variance decomposition approach allows us to determine how much adaptive evolution can be expected to happen within a population, and how important environmental and stochastic sources of variation are.

In the second, more mechanistic approach, one can investigate which characteristics make some individuals fitter than others, or in other words, which 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 the additive genetic variation in that trait (Lush 1937). Therefore, by understanding which traits cause variation in fitness, one can predict which traits should evolve, as well as in which direction and at what rate.

The study of natural selection and adaptive evolution in the wild is very topical in the context of the unprecedented rates of environmental changes induced by human activities (Parmesan 2006). These anthropogenic changes provide an opportunity in the form of a natural experiment to evolutionary biologists (Altermatt, Ebert, and Altermatt 2016; Brookfield 2016), but also come with societal concerns and an ever

²Unless mentioned otherwise, in this thesis I use *natural selection* and *selection* interchangeably, and I consider *sexual selection* as part of *natural selection*. Although measuring sexual selection separately would certainly provide a finer understanding of the mechanisms of selection acting in our study population, this was beyond the scope of this thesis. Nevertheless, the question was partly explored by García-Navas, Bonnet, Waldvogel, Camenisch, et al. 2016 and García-Navas, Bonnet, Bonal, et al. 2016.

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 revival of our interest in the process of adaptation in natural populations has highlighted the gaps in the understanding: Despite decades of research on the topic, it remains challenging to predict, or even to understand retrospectively, how natural populations respond to selection (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 pressures, 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 what the rate of genetic change for the traits of interest is.

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, the two main approaches can be traced back to the scientific controversy that opposed the Mendelians to the biometricians (Dietrich 2006), and can be summarized as “bottom-up” and “top-down” approaches (Liedvogel, Cornwallis, and Sheldon 2012). Bottom-up approaches, embodied by candidate gene and genome-wide association studies, aim to infer the individual genetic loci that underlie phenotypic variation. Top-down approaches, encompassed within the field of quantitative genetics, attempt to decompose phenotypic variation into genetic variation and other sources of variation, based solely on phenotypic data and knowledge of the relatedness between individuals (Lynch and Walsh 1998). Some of the pro’s and con’s of both approaches are nicely illustrated by the confrontation of the quantitative genetics of mass in snow voles (see below for detailed description of the study species and population) with genotype data for a candidate gene for mass. The former will be further developed in chapters 4 and 5, but in a nutshell, we estimated additive genetic variation in body mass and lifetime reproductive success using a quantitative genetics *animal model* (C. Henderson 1950; Kruuk 2004). The candidate gene approach was a side project of this Ph.D. project and does not appear in the other chapters. Hence we present it below in some more detail.

1.2.2 A candidate gene for body mass: insights and limits

We used a candidate gene approach (Fitzpatrick et al. 2005) to uncover a molecular mechanism underlying variation in body mass. To this end we focused on 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, also in rodents (Houseknecht et al. 1998).

We found a recessive allele (from now on referred to as *a*, whereas the dominant allele is referred to as *A*) which was associated with lower body mass (Fig. 1.1A).

Individuals homozygotes for this allele aa were on average -2.9 g lighter (95% credibility interval [0.6; 5.1]), that is, 8% lighter than the mean. Furthermore, across their lifetime, these aa individuals produced on average a third fewer offspring than the AA individuals (Fig. 1.1B). This large difference in fitness was however not statistically significant. These results thus suggest that some of the genetic variation in body mass is attributable to genetic variation in food intake and/or fat metabolism, which is something the estimation of genetic variances and covariances is unable to tell us. Although based on its large strong phenotypic effect, *lepr* could be called a locus of major effect, how much of the additive genetic variation does it explain?

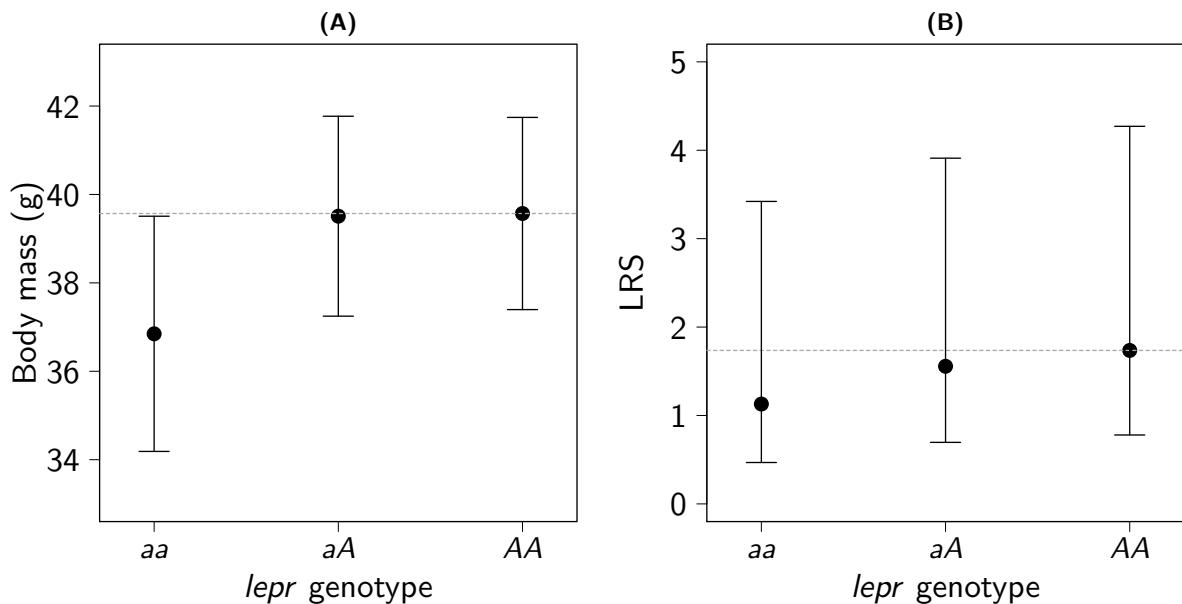


Figure 1.1: Body mass and lifetime reproductive success (LRS) as a function of *lepr* genotype. (A) Expected body mass of snow voles bearing one of the three *lepr* genotypes. The expectations and 95% confidence intervals were predicted from a linear mixed model fitted to the 2311 mass measurements of 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 one of the three *lepr* genotypes. The expectations and 95% confidence intervals were predicted from a Poisson generalized linear mixed model fitted to the lifetime reproductive success 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 .

Knowledge of both the effect of the three genotypes and the allele frequencies one can analytically compute the additive genetic variances associated with a bi-allelic locus (R. A. Fisher 1941; Lynch and Walsh 1998, p77). The additive genetic variances associated with *lepr* are 0.052g^2 for body mass and 0.006pup^2 for lifetime reproductive success. This means that 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, as those studies that have a large enough sample size to avoid biases introduced by the Beavis effect, typically find that quantitative trait loci explain a fraction of a percent to a few percent of the additive genetic variance (V_A) (Flint and Mackay 2009; Jensen, Szulkin, and Slate 2014). Nevertheless, 1% of V_A is not sufficient to infer the evolutionary potential of the trait, and genotyping many more markers, for instance using high-throughput sequencing (Goodwin, McPherson, and McCombie 2016), is unlikely to improve this situation in this small population. Generally speaking, very large sample sizes and high-quality genomic resources are necessary to explain a biologically relevant proportion of the additive genetic variance (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). Admittedly, high-throughput sequencing data can also be used in a top-down way, which does not aim to identify causal genetic variants, but instead quantifies the phenotypic variation jointly explained by all the genotyped markers. Using this approach, 3,925 individuals genotyped for 294,831 markers, (Yang et al. 2010) were able to explain 45% of the genetic variation in human height. Albeit much better, provided knowledge on the relatedness among individuals is available, quantitative genetics can estimate all the additive genetic variance, and this without any genotyping effort.

To conclude, bottom-up approaches allow unravelling the molecular mechanisms underlying phenotypic variation. By opening the black box and revealing these mechanisms, they can identify where this variation comes from, how it is linked to the environment and what the target of natural selection is (Jong et al. 2014). Moreover, they contribute to building a genotype-phenotype map, a long-lasting challenge in evolutionary biology (Kirschner and Gerhart 2010). In contrast, quantitative genetics lumps all the effects of individual genes and their interactions into a few parameters which are largely non-informative with respect to the underlying genetic architecture (Mackay 2001; Nietlisbach and Hadfield 2015; Huang and Mackay 2016). However, quantitative genetics provides a simple and direct measure of key evolutionary parameters. As they are based directly on data on the phenotype, which is the target of selection and the source of ecological interactions, they 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 causes of individual-level variation in fitness, and the consequences of this variation at the population level. This thesis aims at improving the measurement, and thereby our understanding, of 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 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 the widespread idea that snow voles are white, the fur colour of the upper-parts varies from light to dark taupe grey, sometimes tinted with brown or dark red (Fig. 1.2). Indeed, the species could probably 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 and snow 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).



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.

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 or migrate 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, Camenisch, 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, lice and ticks (Janeau and Aulagnier 1997; Martinoli et al. 2001).

The study area is located near 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. 1.4).

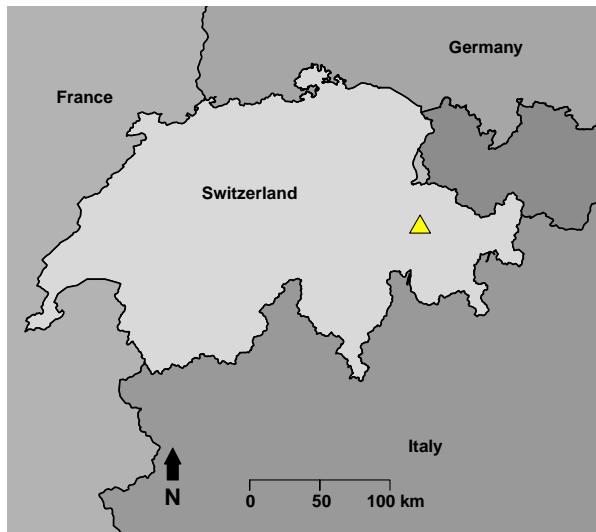


Figure 1.3: Location of the study area in Switzerland. The yellow triangle indicates the study area, with coordinates 46°48' N, 9°34' E, by Churwalden, in the Swiss canton Graubünden. Countries are filled with different shades of grey, Austria and Lichtenstein are not labelled.

Another scree offers about 1 ha of favourable habitat, starting 300 m north-east to the monitored area. This area 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, Camenisch, 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



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.

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, Ravaoli, 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 presented in this thesis. In particular, it 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.

1.3.3 Thesis outline

In natural populations, fitness is generally measured using individual measures of reproductive success and survival. Importantly, these are proxies of fitness and their variation has a large stochastic component. This has lead 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 strongly overestimated. In chapter 2 I examine these methods and, based on computer simulations, demonstrate that they lack the statistical power

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

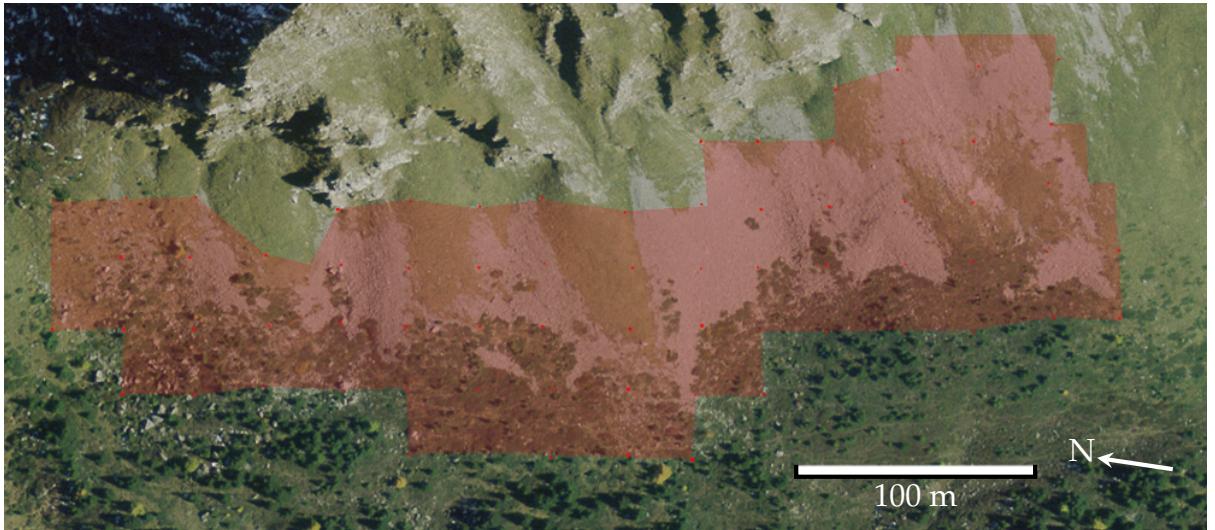


Figure 1.5: Orthophoto of the study site, from 2008. The red shading indicates the approximate area where traps are set.

to detect latent variation in fitness components. Using an alternative approach I 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. I also attempt to clarify some conceptual misunderstandings between the proponents of the two methodological schools.

In **chapter 3**, with collaborators from different methodological schools, I review and compare four frameworks that claim to be able to disentangle the causes of temporal phenotypic change. While these frameworks appear to come to different conclusions with respect to the relative roles of plasticity, demography and genetic change, based on computer simulations and mathematical comparisons, we show that these discrepancies primarily originate from different definitions of the components 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 cases of contemporary adaptive evolution of a quantitative trait in the wild. Both the evolution and the selective pressure responsible for it are invisible to purely phenotypic approaches, however. Using multivariate animal models, I 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 dynamics across the study period. Fluctuating selection is thought to be a major determinant of the rate of evolution, and a potentially important process when it comes to understanding adaptation in the

wild. Nevertheless, unbiased measures of the variation of selection are rare, and the coupling between variation in selection and variation in evolution has been largely ignored. **Chapter 5** shows that selection fluctuates in the snow vole 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 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 Ph.D. on the understanding of natural populations, and discuss some of the remaining challenges and future directions.

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. / When you wish to understand a thing, you face it, alone, without help; all the knowledge of the world could not be of any use. And then it vanishes and so does what you understood.

— 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.

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 meth-

ods to detect fixed heterogeneity. Unfortunately however, fixed heterogeneity is the result of latent, unobservable traits, which cannot be inferred without a modelling step (Cam et al. 2013), and it is precisely the performance of this modelling 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 analyse 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 modelled (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 labelled 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 analysed 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 analysed 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 IBDs im (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 (R. 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 (R. 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 analyse 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 alternative 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 between 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 analyse 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 depends 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:

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

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 modelling 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 over-dispersion (see e.g. Atkins et al. 2013) and quantifies the over-dispersion 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

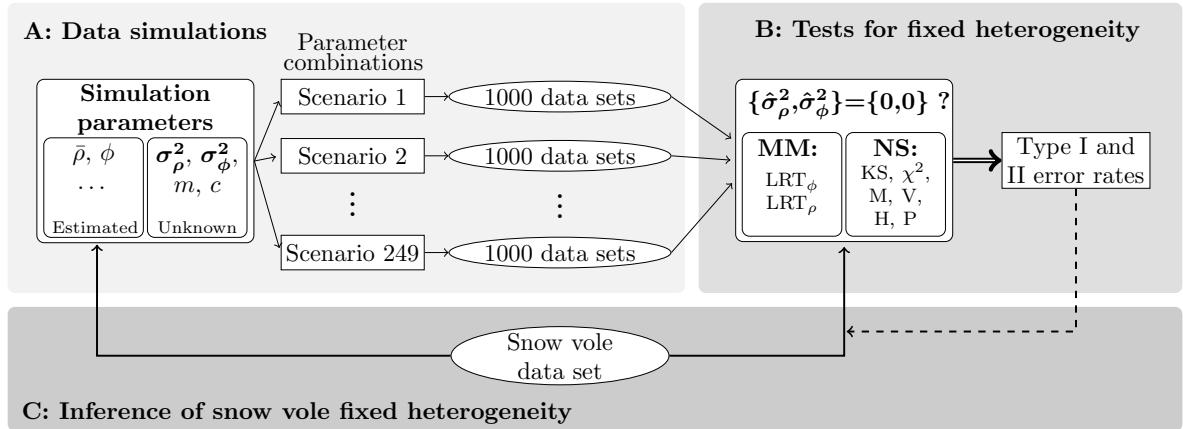


Figure 2.1: Illustration of the simulation and testing process. (A) Data simulation: The simulation model is parametrized using the life cycle and vital rates of a snow vole population, along with additional, unknown, parameters introducing fixed heterogeneity (σ_ϕ^2 and σ_ρ^2) and dynamic heterogeneity (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.

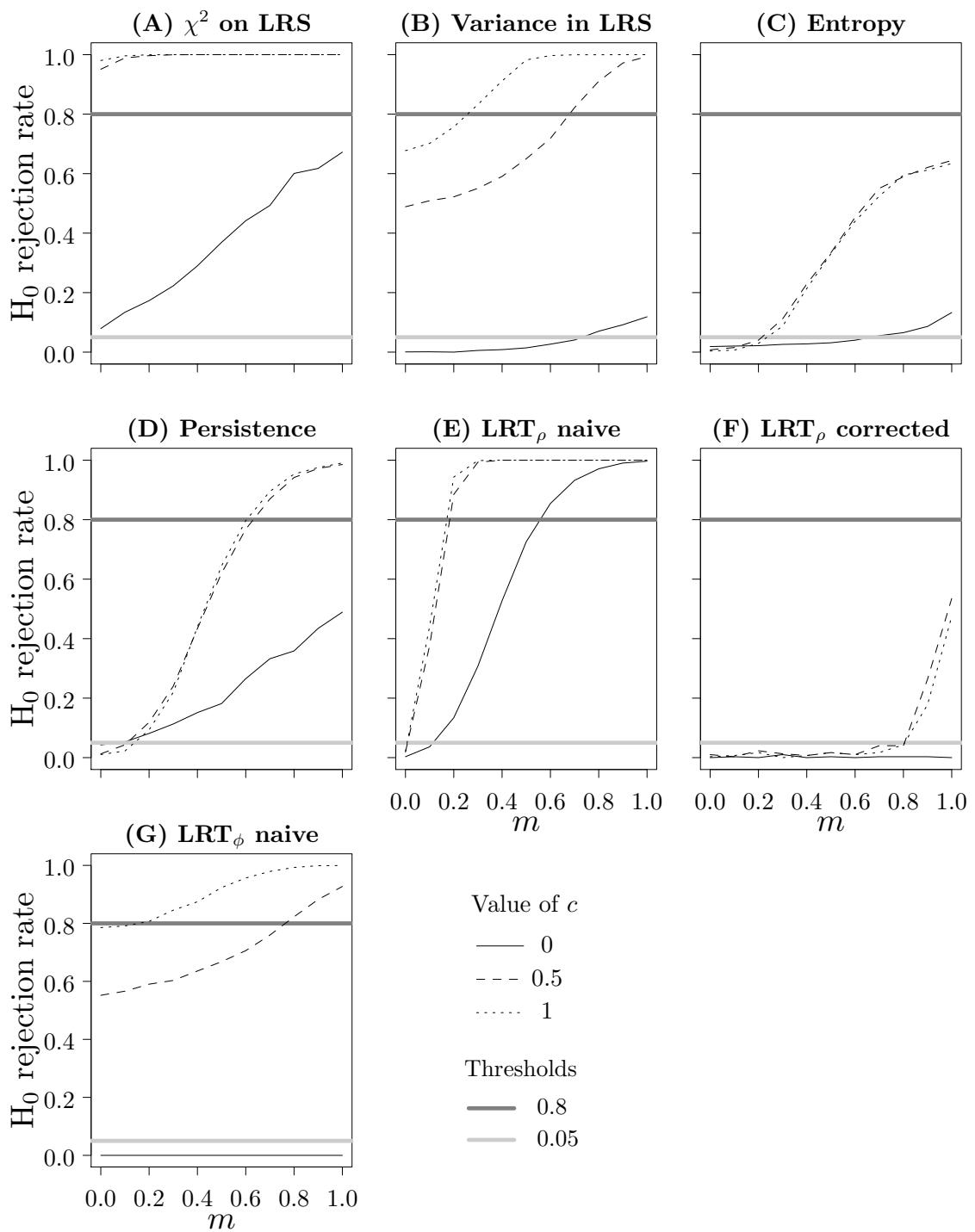


Figure 2.3: Null-hypothesis rejection rates for various methods testing for the presence of fixed heterogeneity, when none is explicitly simulated, depending on the parameter m , controlling the structure of transitions between successive annual reproductive successes, and on the parameter c , controlling 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 over-dispersion

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 over-dispersion 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 modelled 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; P. W. 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 make-up. 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).

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 2.8.3) and for survival (figure 2.8.3). 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 ($46^\circ 48' \text{ N}, 9^\circ 34' \text{ E}$) 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°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 re-sampled 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 modelling 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 modelling 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.

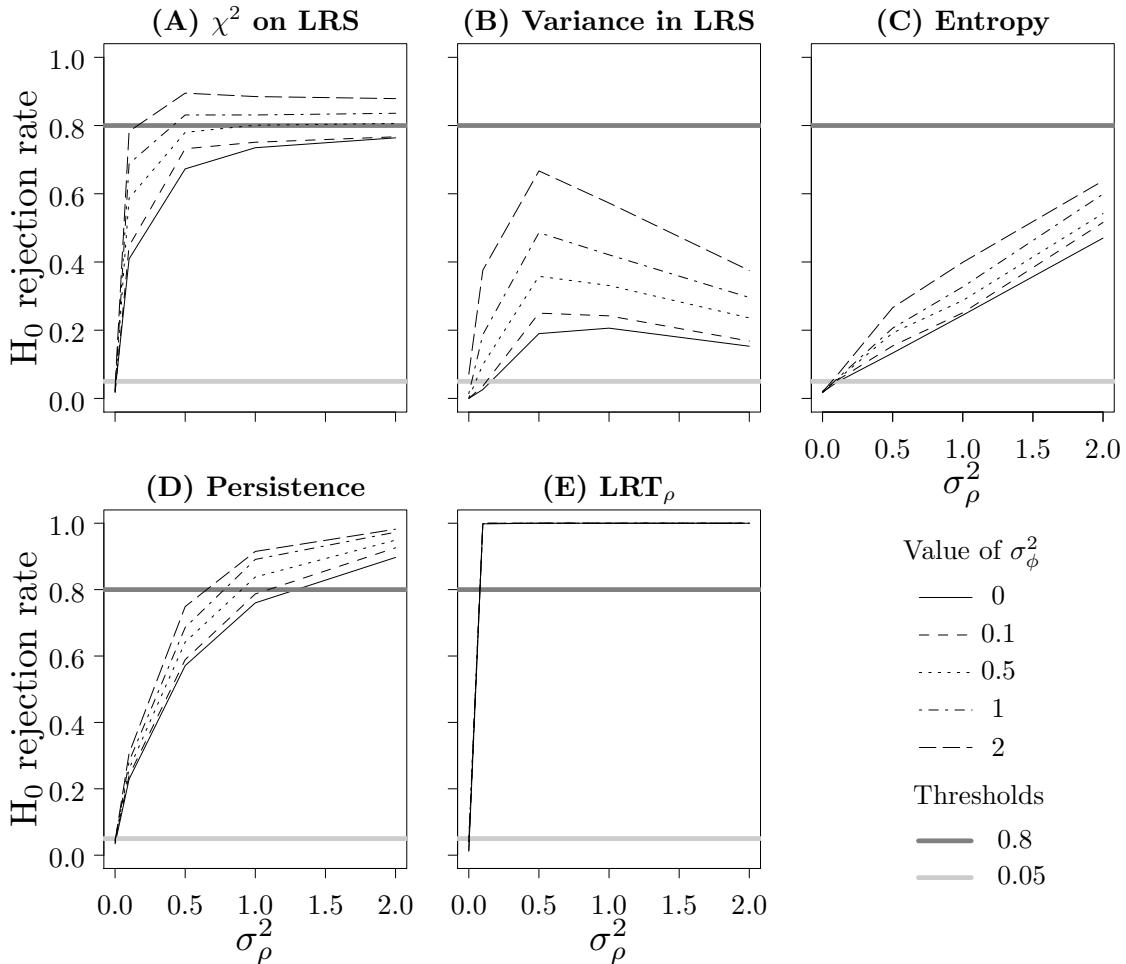


Figure 2.2: Null-hypothesis rejection rates for various methods testing for the presence of fixed heterogeneity, as a function of the variance in reproductive propensity, σ_ρ^2 , and survival propensity, σ_ϕ^2 , when these variances are introduced on the transformed scales. The methods are: (A) a χ^2 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 $\sigma_\rho^2 = \sigma_\phi^2 = 0$, the null-hypothesis rejection rates are equal to the type I error rates, which is expected to be 0.05 (light grey 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. The dark grey line indicates the 0.8 threshold. (A)-(D) are related to NS, (E) is related to MM.

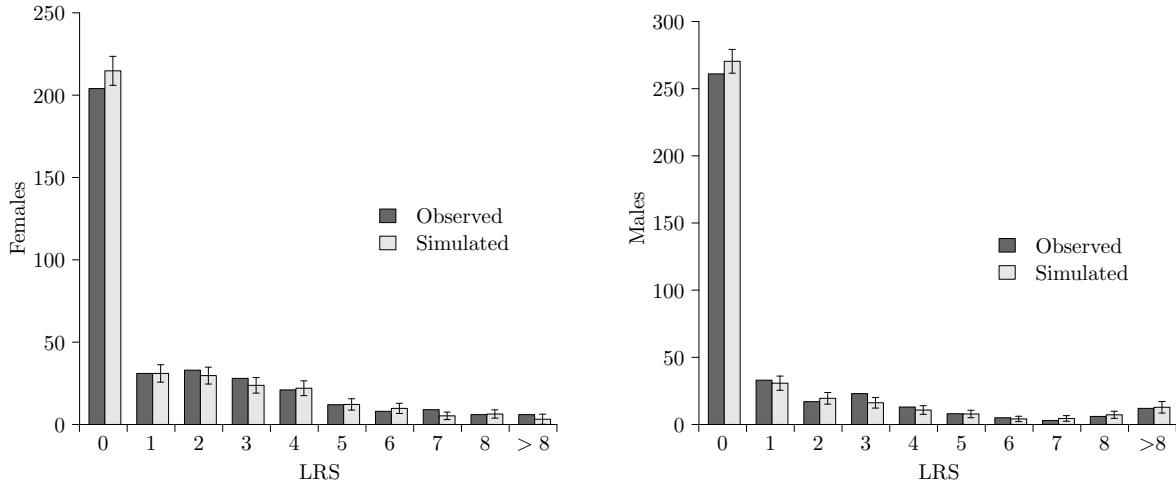


Figure 2.4: Distribution of lifetime reproductive success in the real snow vole data set, observed (dark bars) and simulated through 1000 neutral simulations (light bars with black error bars showing \pm standard deviation), for 2.4.4 females and 2.4.4 males.WRONG

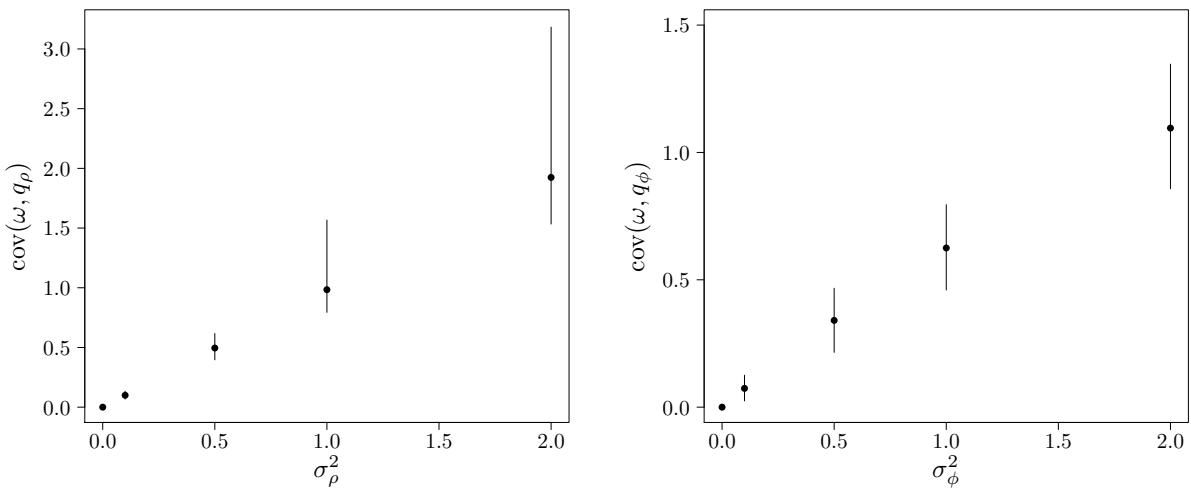


Figure 2.5: Appendix C Strength of selection on individual fixed qualities for survival and reproduction, as a function of the expected variance in these qualities. Strength of selection was measured as the individual-level covariance between the qualities and a proxy for fitness (ω): relative lifetime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

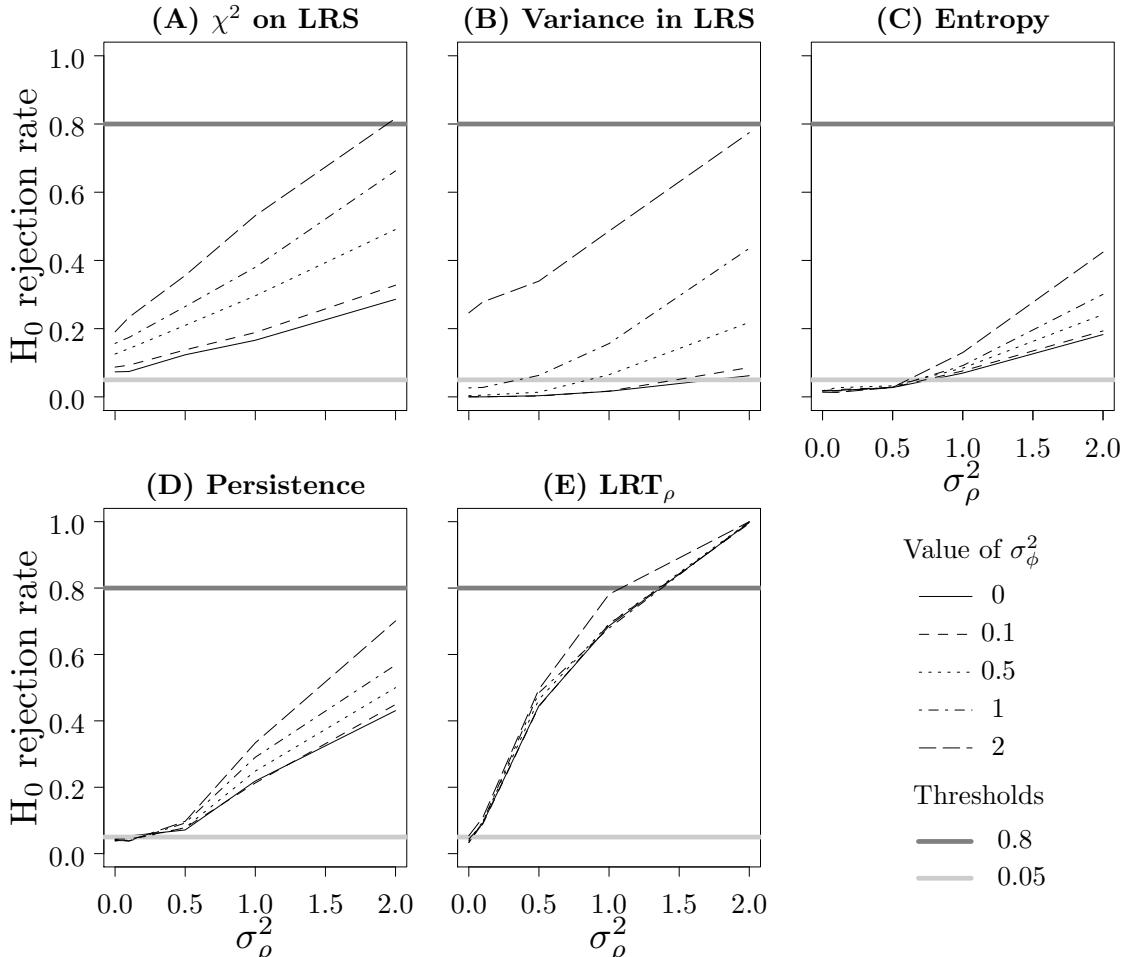


Figure 2.6: Appendix D Null-hypothesis rejection rates for various methods testing for the presence of fixed heterogeneity, depending on the variance in reproductive propensity, σ_ρ^2 , and on the variance in survival propensity, σ_ϕ^2 , when these variances are introduced on the original scales. The 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 over-dispersion 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 over-dispersion 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:

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

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 $\sigma_{\rho\phi(ind)} / \sqrt{\sigma_{\rho(ind)} \sigma_{\phi(ind)}}$. 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 $\sigma_{\rho\phi(ind)}$, $\sigma_{\rho(ind)}^2$ or $\sigma_{\phi(ind)}^2$ 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, $\sigma_{\rho(ind)}^2$. 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)

*La Nature est un temple où de vivants piliers
Laissent parfois sortir de confuses paroles;
L’homme y passe à travers des forêts de symboles
Qui l’observent avec des regards familiers.*

*Nature is a temple where living columns
Let slip from time to time uncertain words;
Man finds his way through forests of symbols
Which regard him with familiar gazes.*

— Charles Baudelaire, *Les Fleurs du Mal, Correspondances*
(1868)

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. C. 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, Parker, 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.

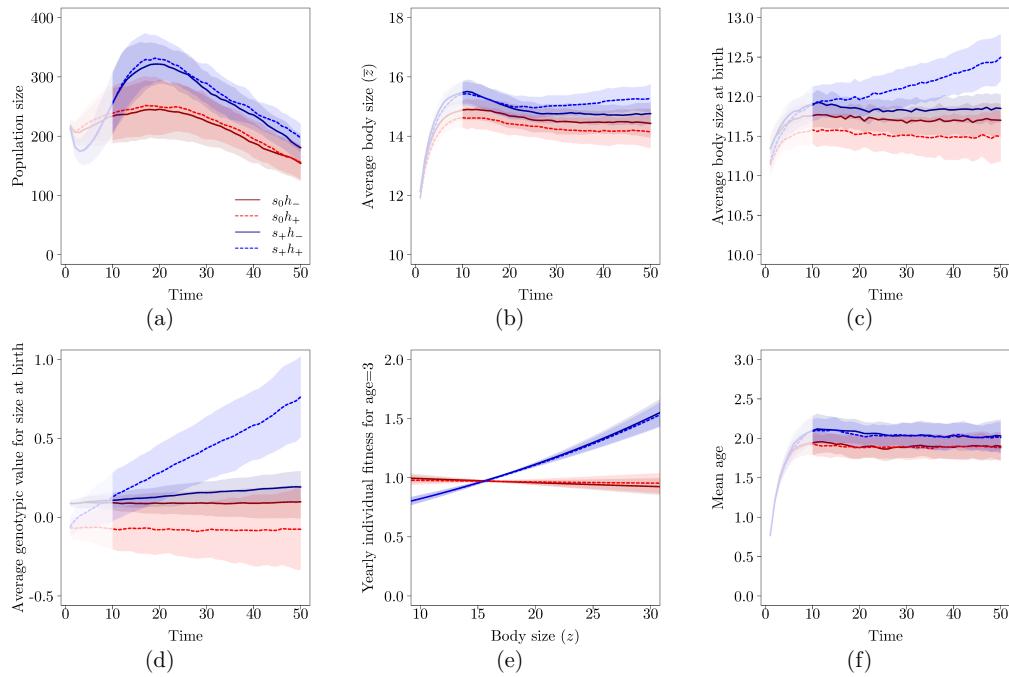


Figure 3.1: Summary of the observed population and trait dynamics of simulated datasets. (a) Trends in population size, (b) changes in mean body size, (c) mean birth size, (d) and genotypic values for body size, (e) relations between body size and yearly individual fitness (sum of survival and litter 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 ($\text{mean} \pm \text{SE}$) -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 (C. 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).

Chapter 3 Evolution or plasticity?

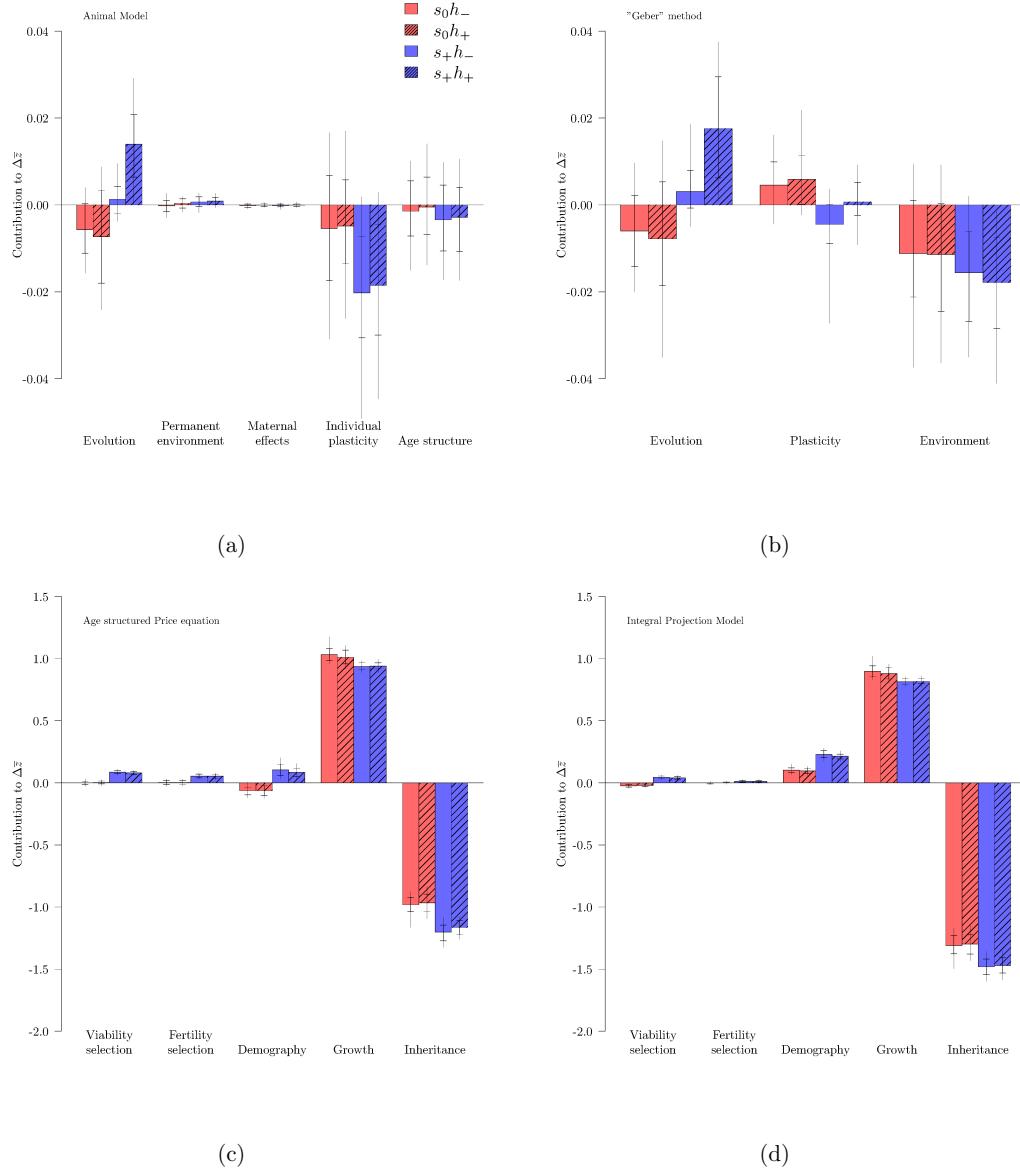


Figure 3.2: Results of the different frameworks when applied to the simulated scenarios. (a) Animal model. (b) "Geber" method. (c) Age-structured Price equation and (d) Integral projection model. In (c) and (d), demography includes changes in average body size due to the age structure, inheritance is 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 (G. R. 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 contrib-

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 non-linearities 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; G. R. 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, Parker, 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.

3.5 Details on data simulation

In the following paragraphs we provide detailed information on the data simulation of a closed population of a hypothetical animal. We used an individual-based model, containing both males and females. We started with a population of 200 1-year old individuals, and fifty time steps were simulated. In this simulated population, body size was a fitness related trait, shaped by an individual's genotype, maternal effects, the environment and its ontogenetic growth. Every year, first food is distributed among all individuals. Subsequently, each individual either survives or dies. Surviving individuals age, grow and then they potentially reproduce. Details on these processes are given below.

3.5.1 Simulation of the genotypes

Each diploid individual had a genotype for body size consisting of 10 independent genes, for each of which there are 10 alleles (leading to a total of $(10 + \frac{10 \cdot 9}{2})^{10}$ possible genotypes – i.e. per gene there are 10 ways of being homozygote and $\frac{10 \cdot 9}{2}$ ways of being heterozygote). The importance of each gene, that is the variance in the additive effects of its alleles, was drawn from a folded normal distribution. Subsequently, the effect of the 10 homozygotes for this gene was determined by drawing a number from a normal distribution with a standard deviation equal to the importance of that gene. Finally, the genotypic effects of heterozygotes was determined for all pairs of alleles per gene, by drawing a dominance value from a uniform distribution bounded by the additive effects of the two alleles. Based on these values, and assuming no epistasis, we obtained body size genotypic values for all possible genotypes (10 homozygotes plus $\frac{10 \cdot 9}{2}$ heterozygotes) by summing the additive effects across alleles and genes.

For the first cohort of 200 individuals, individual genotypes were drawn randomly from all possible genotypes, while for subsequent cohorts the genotype was inherited from the parents in a Mendelian way.

3.5.2 Birth size

Birth size of individual i (z_0^i) was determined as an intercept value β_{z0} , equal to 10 in our simulation, plus the sum of three processes: genotypic effects, maternal effects and stochasticity. An individual's genotypic value b_i was determined by the inherited genotype. Second, a fraction M of the size of the mother at the moment of reproduction (z_m^i), represented a maternal effect. This maternal effect thus depends on the phenotype of the mother, which in turn is partially determined by her genotype. This fraction M equalled 0.1 in our application. This yielded the expected birth size in absence of stochasticity:

$$\zeta_i = \beta_{z0} + b_i + M \cdot z_m^i . \quad (3.2)$$

Third, we included a random plasticity component, reflecting the experienced micro-environment, drawn from a normal distribution with standard deviation σ_{z0} ,

which equalled 0.5 or 1 depending on the scenario (Appendix 3.5.7), to obtain an individual's birth weight z_0^i .

$$z_0^i = \mathcal{N}(\zeta_i, \sigma_{z0}) \quad (3.3)$$

3.5.3 Growth

Body size increased over time due to growth, which was proportional to body size. There was no heritable variation in growth rate. For one-year-old individuals ($\alpha = 1$), mean proportional growth was μ_{growth} , and equalled 1.245. Proportional growth for individual i decreased with age (α_i), approaching 1.

$$\gamma_i = (\mu_{growth} + \alpha_i - 1) / \alpha_i \quad (3.4)$$

Age-dependent proportional growth was further influenced by individual food intake E_i , whereby low food availability decreased proportional growth:

$$g_i = 1 + (\gamma_i - 1) \cdot \left(\frac{2}{1 + e^{-c \cdot E_i}} - 1 \right). \quad (3.5)$$

Here, E_i is food availability obtained by individual i , and c is a food to growth conversion, which we set at 0.05. To include random plasticity in growth, the yearly realized proportional growth was drawn from a normal distribution, with mean g_i and a standard deviation σ_{growth}^i , the latter depending on γ_i , and calculated as

$$\sigma_{growth}^i = (\gamma_i - 1) / 2 \quad (3.6)$$

To obtain the new body size z'_i , proportional growth was multiplied with current body size z_i :

$$z'_i = \mathcal{N}(g_i, \sigma_{growth}^i) \cdot z_i. \quad (3.7)$$

Equation 3.6 implies that variance in growth decreased with age and that individuals could shrink. This led to individual variation in growth within and between years. Individual proportional growth (g_i) was not correlated across years. However, because growth was proportional, individuals that grew more in one year, on average also gained more (absolute) size in the next year.

3.5.4 Reproduction

We explicitly modelled variation in the reproductive success of females, but males were randomly assigned to females once the reproductive success of the females was

determined. Annual reproductive success of female i (F_i) was the product of the reproduction probability p_{repr}^i , and the litter size L_i .

The probability of reproduction increased with age:

$$p_{repr}^i = 1/(1 + \exp(-\alpha_i - \alpha_m)), \quad (3.8)$$

where α_i is individual age and α_m is the age at which individuals become mature, which is set to 1 in our simulations. This implies that individuals started reproducing one year after their birth with 50% probability. Whether the female i reproduces or not was simply drawn from a Bernoulli distribution:

$$\pi_i = \mathcal{B}(p_{repr}^i). \quad (3.9)$$

Expected litter size was a function of body size and food availability:

$$L_i = \exp(\log(L_0) + \beta_{L1} \cdot (z_i - z_0) + \beta_{L2} \cdot (E_i^{1/3})/10), \quad (3.10)$$

where L_0 is a baseline reproductive success and was set at 0.5, β_{L1} is the strength of the body size effect (set at 0 or 0.04, depending on the scenario; Appendix 3.5.7), β_{L2} is the effect of food availability (set at 0.1), E_i is the current food availability and z_0 is a centering parameter and was set at 15 (approximately the mean mass in the simulated population). The realized litter size was then drawn from a Poisson distribution and a value of 1 was added to make sure the litter contains at least one offspring:

$$\rho_i = \mathcal{P}(L_i) + 1. \quad (3.11)$$

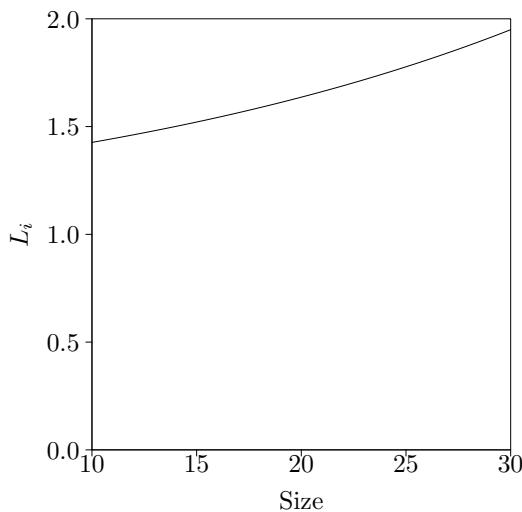


Figure 3.3: Expected reproductive outcome per year as a function of size.

Individual annual reproductive success was subsequently calculated as

$$F_i = \pi_i \cdot \rho_i. \quad (3.12)$$

3.5.5 Survival

Survival was positively affected by food availability and was a function of age, the latter through a bathtub function: survival first increased with age (α) but started decreasing after an age of 5, reflecting senescence. Thus, considering only the effect of age, the survival probability of individual i can be written as

$$\phi_{\alpha,i} = 1 - m \cdot \exp(-(\alpha_i - \alpha_s)/4) + \exp((\alpha_i - \alpha_s)\log(2)/(\alpha_m - \alpha_s)), \quad (3.13)$$

where m is the baseline mortality (0.1), α_i is the current age of i , α_m is the maximal age (30) and α_s is the age after which survival decreases again due to senescence (set at 5). Survival also varied depending on the size of the individual (z_i) and on the

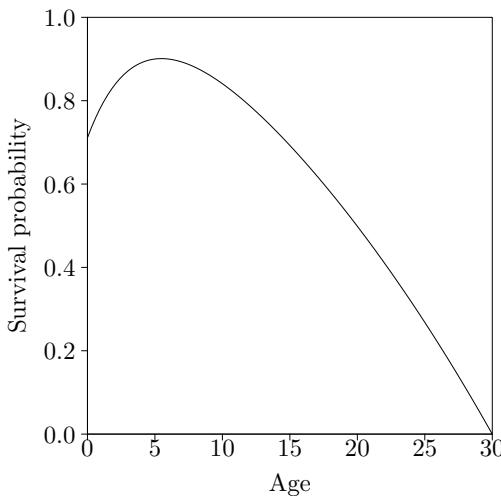


Figure 3.4: Probability of survival to the next year, as a function of age.

reproductive success during the previous year (F_i). Thus, we have:

$$\phi_{m,i} = \text{logit}^{-1}(\text{logit}(\phi_0) - \beta_\phi(z_i - 15) + F_i h_F), \quad (3.14)$$

where ϕ_0 is a baseline survival probability (0.75 in our application), β_ϕ is the linear selection gradient, which was set at either 0 or 0.2 (SI Appendix 3.5.7). h_F is the strength of the trade-off between survival and reproduction and was set at 0.01. The survival probability of the individual i is the product of the age-specific survival probability and of the size- and reproduction-specific survival probability:

$$\phi_i = \phi_{m,i} \phi_{\alpha,i}. \quad (3.15)$$

Finally, a shortage of food acquisition by the individual (E) can decrease survival. If

$E < 40$:

$$\phi'_i = 1 - (1 - \phi_i)(E/40)^{1/2}. \quad (3.16)$$

3.5.6 Food availability

Total food availability differed per year and the amount an individual obtained influenced individual growth and reproduction. The expected total amount of food available in year t started to decrease after year 10, and can be written as

$$E_\mu(t) = \begin{cases} \mu_E, & \text{if } t \leq 10, \\ \mu_E + (t - 10) \cdot \beta_E, & \text{otherwise.} \end{cases} \quad (3.17)$$

We set μ_E at 15000 and β_E is the yearly rate of decrease after year 10, and was set at -200.

To include random variation in food availability, the realized amount of food available was drawn from a folded normal distribution, whereby the variance of the Gaussian (\mathcal{N}) was proportional to the mean.

$$E_{tot}(t) = |\mathcal{N}(E_\mu(t), \sigma_E)| \quad (3.18)$$

In our simulations, σ_E was set at $E_\mu^t/10$. The total amount of food was divided over all N alive individuals, whereby some individuals obtained more than others. The amount of food that an individual obtained was assigned randomly. To do so, each year (t), a measure of successful foraging (η_i^t) was assigned to each individual. For each individual, this was a number between 0.3 and 1 from a uniform distribution, i.e.:

$$\eta_i(t) = \mathcal{U}(0.3, 1), \quad (3.19)$$

Using these, we calculated the food amount obtained by individual i at time t as

$$E_i(t) = \frac{E_{tot}(t) \cdot \eta_i(t)}{\sum_{i=1}^N \eta_i(t)}. \quad (3.20)$$

3.5.7 Scenarios

We have simulated data under four biologically different scenarios. These scenarios are presented below. For each scenario, we have ran 100 replicates, resulting in a total of 400 datasets.

Table 3.2: Simulated scenarios

Scenario	Description	Parameter value
s_0h_-	Low genetic variance	$V_A \approx 1$
	High plasticity in birth size	$\sigma_{z_0} = 1$
	No fertility selection	$\beta_{L1} = 0$
	No viability selection	$\beta_\phi = 0$
s_0h_+	High genetic variance	$V_A \approx 5$
	Low plasticity in birth size	$\sigma_{z_0} = 0.5$
	No fertility selection	$\beta_{L1} = 0$
	No viability selection	$\beta_\phi = 0$
s_+h_-	Low genetic variance	$V_A \approx 1$
	High plasticity in birth size	$\sigma_{z_0} = 1$
	High fertility selection	$\beta_{L1} = 0.04$
	High viability selection	$\beta_\phi = 0.2$
s_+h_+	High genetic variance	$V_A \approx 5$
	Low plasticity in birth size	$\sigma_{z_0} = 0.5$
	High fertility selection	$\beta_{L1} = 0.04$
	High viability selection	$\beta_\phi = 0.2$

3.5.8 Full R code

The R code that was used to generate the analysed datasets is provided as additional supplementary information on [GitHub](#) (doi: [10.5281/zenodo.59412](https://doi.org/10.5281/zenodo.59412)).

3.6 The application of the animal model

In this appendix we give details on the application of the animal model to the simulated data. We have applied a univariate animal model, using the temporal trends in BLUPs (best linear unbiased predictors) to decompose changes in \bar{z} into underlying processes.

3.6.1 Animal model using the BLUPs approach

Following Hadfield et al. 2010, we estimated the temporal change in the mean components of variation (breeding values, maternal effects, individual repeatability and within individual residual variation) by regressing their estimates on time, within a Bayesian framework. To this end we fitted a univariate animal model on the vector of size observations, z , as:

$$z = X_z b_z + D_1 a + D_2 m + D_3 p + D_4 y + I r . \quad (3.21)$$

Here X_z , D_1 , D_2 , D_3 , D_4 and I are design matrices, b_z is a matrix of fixed effects, a , m , p and y random effects accounting for the variance associated with breeding value, mothers, permanent environment and years respectively, and r represents residuals. The fixed part of the model included an intercept and the effect of age (up to second order). In addition to a random additive genetic effect, the model thus included three additional random effects: the identity of individual (accounting for any permanent environment effects), the identity of the mother and the year of measurement (Kruuk 2004).

From this model, the posterior distribution of each Best Linear Unbiased Predictors (BLUPs), that is the value of each level of the random effect, of the random effect a , m , p and r were extracted. Each posterior sample, j , of the BLUPs was regressed on time. For instance, for breeding values:

$$a_{i,j} = \mu_{a,j} + \bar{t}_i \beta_{a,j} + \epsilon_{i,j} , \quad (3.22)$$

where $a_{i,j}$ is the posterior sample j of the breeding value of the individual i , $\mu_{a,j}$ is the intercept for the j th regression, \bar{t}_i is the mean time of presence of the individual i , $\beta_{a,j}$ is the slope of the j th regression and $\epsilon_{i,j}$ is the error. Combining all the slope estimates, $\beta_{a,j}$ we thus obtained the posterior distribution of the rate of change in the random effect (Hadfield et al. 2010).

3.6.2 Alternative modelling of maternal effects

The AM described above does not match the simulation process on maternal effects. Indeed, in our simulations, offspring size at birth is influenced by the current size of their mothers (in an additive, proportional way). The effect of a given mother on her offspring size therefore changes through her life. In contrast, the AM models maternal effects as a random intercept on mother identity, which assumes that a given mother influences her offspring size by a fixed propensity throughout her life. Various formulations of maternal effects in quantitative genetic models are for instance compared in

McGlothlin and Galloway 2014. In this section, we present results from an AM that models maternal effects in a way that is more in agreement with the simulations process: maternal effects are modelled by including maternal size as a covariate, rather than maternal identity as a random effect.

Compared to the results of the AM in the main text, this alternative AM formulation estimated a less negative contribution of genetic evolution to trait dynamics, and a more negative contribution of maternal effects (Fig. B.1 shows the results of the two different AM for the scenario s_0h_-). In the other scenarios a similar effect was obtained by changing the model specification (results not shown).

The negative contribution of maternal effects makes sense in our simulations, because maternal effects are modelled by adding up 10% of mother size to the offspring birth size, and because mean size decreases with time (Fig. 1b). As a result, the effect of mother size on offspring birth size decreases with time. The scenario s_0h_- did not include selection on size and no directional genetic change was therefore expected. However, the AM estimated on average a negative contribution of evolution when maternal effects were modelled as maternal identity. This bias is divided by two when maternal effects are modelled in the same way as they were simulated. The reason why the bias does not entirely disappear probably has to do with the fact that the maternal effects we have simulated are heritable (because they are proportional to size, which is itself heritable). As such, a model matching our simulation process should also account for genetic maternal effects—in addition to phenotypic maternal effects—(McGlothlin and Galloway 2014) in order to fully disentangle the change in breeding values for size from the change in maternal effects.

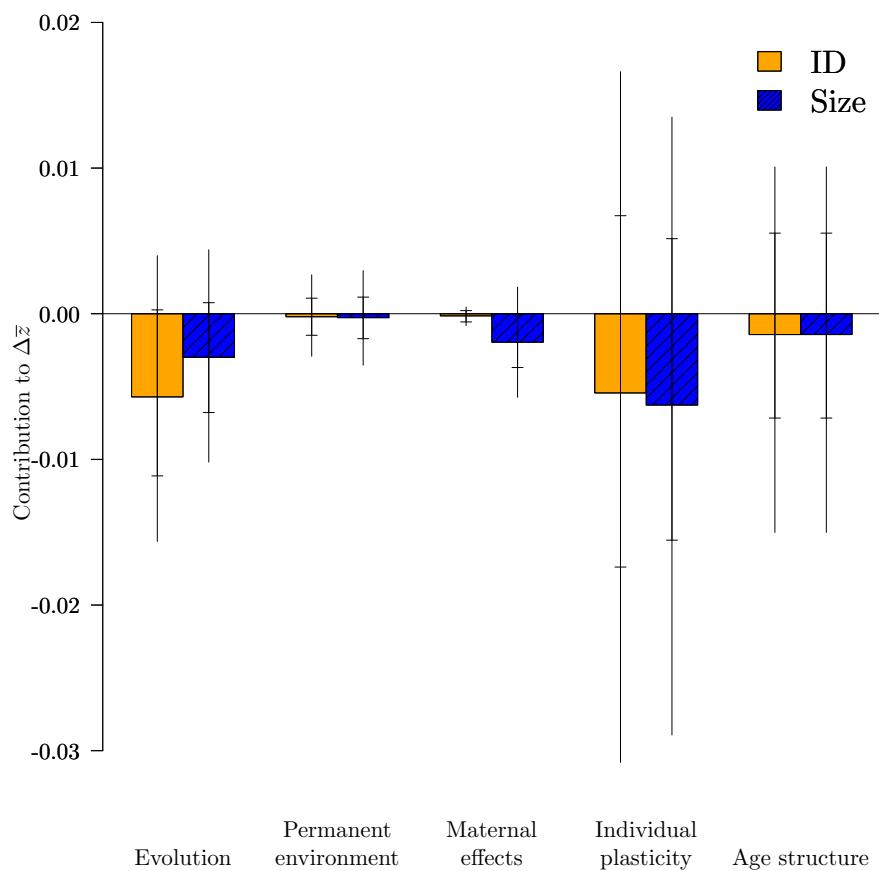


Figure 3.5: Results of two animal models, either modelling maternal effects by fitting mother identity as a random intercept (yellow plain bars), or by fitting the size of the mother at birth of the focal individual as a continuous covariate (blue striped bars). The latter model is closer to the process of data simulations, and provides a less biased estimation of genetic change (the expectation is zero for both models). 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 of the scenario s_0h_- .

3.7 The application of the ‘Geber’ method

In this section, we provide more information on the applied genotype-phenotype-environment equation (Ellner, Geber, and Hairston 2011; Hairston et al. 2005), this is the variant of the ‘Geber’ method that Ellner, Geber, and Hairston 2011 developed and forms the basis of our analysis. First, we describe the method in more detail. This framework was applied to each simulated dataset.

3.7.1 The method

In our analysis, we have followed the example described in Ellner, Geber, and Hairston 2011 on great tits to quantify the contribution of the three processes in influencing average body size \bar{z} in the population. This version depends on two steps.

First, we fitted a linear model to estimate the effects of average breeding value (\bar{a}), average plasticity at birth (\bar{p}) and average consumed food (\bar{k}) on the average trait value (\bar{z}).

We then used separate regressions of each of the three factors on time to predict their values at the beginning ($t = 11$) and at the end of the interval ($t' = 50$). Using these predicted values, together with the model for z , we estimated \hat{z} at time t ($=\hat{z}(\hat{a}_t, \hat{p}_t, \hat{k}_t)$) and at time t' ($=\hat{z}(\hat{a}_{t'}, \hat{p}_{t'}, \hat{k}_{t'})$). We then predicted the change in \hat{z} for mixed scenarios, where one or two factors were estimated at time t and the remaining factor(s) at time t' (e.g. $z(\hat{a}_t, \hat{p}_t, \hat{k}_t)$). We used all eight (2^3) combinations of the three factors at either time point and regressed the so obtained eight values of \hat{z} on three binary dummy variables (one for each factor, indicating which of the values of the corresponding factor was used for the calculation of z). The coefficients of this regression provided the relative contributions of the three factors.

3.7.2 Using genotypic values from the simulation

When genotypic values from the simulations were used instead of breeding value estimates by the AM, we obtain the result shown in Figure 3.6. We see that now the bias in the evolutionary components has disappeared and that the plasticity component is in line with the expectation as outlined in the main text (i.e. small but negative, due to a slight decrease in average maternal body size weighted by litter size over time.).

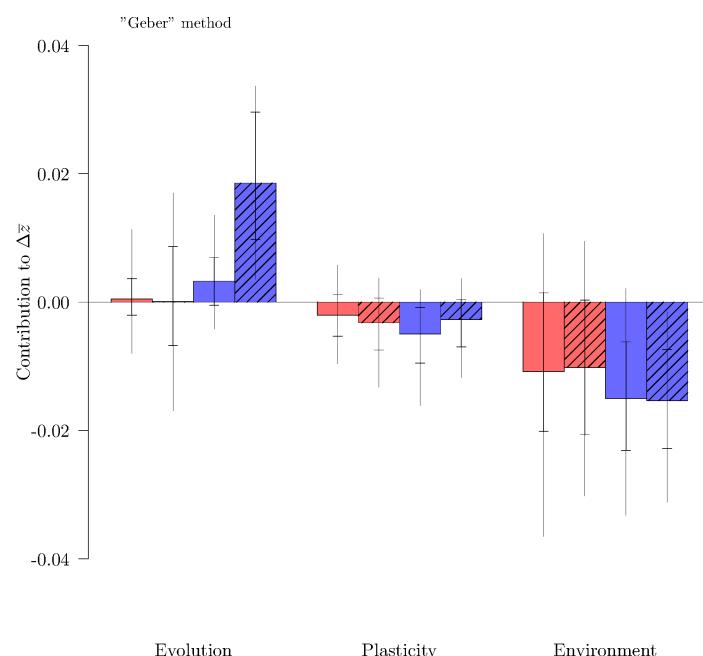


Figure 3.6: Results of the GM when based on the actual (simulated) genotypic values instead of on the estimated breeding values.

3.8 The application of the age-structured Price equation

Here we give the full age-structured Price equation as proposed by Coulson and Tuljapurkar 2008. We have applied this equation to decompose changes in z in the simulated datasets into underlying processes including demography, selection and plasticity. Subsequently, we explain how the demographic terms and the inheritance terms are interpreted. Finally, we demonstrate how the choice of age classes influences the calculated contribution of selection and demography, illustrated by a simple example. Throughout this SI we use the convention that juveniles are born into age class 1 and not in age class 0. This was done to stay closer to the original notation.

3.8.1 Full equation

The age-structured Price equation can be written as:

$$\Delta z(t) = \underbrace{\sum_{\alpha=1}^{\Omega-1} \Delta c(\alpha, t) \bar{z}(\alpha, t) - c(\Omega, t) \bar{z}(\Omega, t)}_{\text{Demo S}} + \underbrace{\sum_{\alpha=1}^{\Omega} \frac{c(\alpha, t)}{W(t)} \bar{r}(\alpha, t) \bar{z}(\alpha, t)}_{\text{Demo R}} + \sum_{\alpha=1}^{\Omega} \frac{c(\alpha, t)}{W(t)} \left(\underbrace{\text{cov}(z, s)(\alpha, t)}_{\text{VS}} + \underbrace{\text{cov}(z, r)(\alpha, t)}_{\text{FS}} \right) + \sum_{\alpha=1}^{\Omega} \frac{c(\alpha, t)}{W(t)} \left(\underbrace{\bar{r}(\alpha, t) \bar{d}(\alpha, t)}_{\text{OMD}} + \underbrace{\text{cov}(d, r)(\alpha, t)}_{\text{ODC}} + \underbrace{\bar{s}g(\alpha, t)}_{\text{Growth}} \right) \quad (3.23)$$

The equation distinguishes between seven terms: demographic changes due to survival and due to reproduction (Demo S and Demo R), viability and fertility selection (VS and FS), offspring mother difference (OMD), offspring difference covariance (ODC) and ontogenetic growth. Here, to get the full contribution, underbraces should be extended to also include the sum and the weighting factors shown. For example the full contribution of the viability selection (VS) is obtained by multiplying the viability selection per age class with the shown factor ($\frac{c(\alpha, t)}{W(t)}$) and summing the contributions of all age classes. The separate terms are defined as follows:

$c(\alpha, t)$	The proportion of individuals of age α in the population at time t
$\Delta c(\alpha, t)$	The difference between the proportion of individuals of age $\alpha + 1$ in the population at time $t + 1$ and the proportion of individuals of age α at time t . It measures how much the importance of a cohort in terms of population numbers changes from one timestep to the next.
$\bar{z}(\alpha, t)$	The average trait value for all individuals of age α at time t .
Ω	The maximum age
$\bar{W}(t)$	The population growth rate from t to $t + 1$.
$\bar{r}(\alpha, t)$	The average number of offspring that an individual of age α at time t contributes to the population at time $t + 1$.
$\text{cov}(z, s)(\alpha, t)$	The covariance between trait values and survival for all individuals of age α at time t ; viability selection.
$\text{cov}(z, r)(\alpha, t)$	The covariance between the number of offspring and individual of age α at time t contributes to the population at time $t + 1$ and the trait values of these parents; fertility selection.
$\bar{d}(\alpha, t)$	The average difference between the trait value of an individual with age α at time t and the average trait value of its offspring
$\text{cov}(d, r)(\alpha, t)$	The covariance between the number of offspring that an individual of age α at time t contributes to the population at time $t + 1$ and how much the average trait value of the offspring differs from the trait value of this individual
$\bar{s}g(\alpha, t)$	The average change in trait value for the individuals with age α at time t , regardless of survival. Non-surviving individuals have a growth of $g_i = 0$.

In the main text we do not discuss all seven terms, instead we combined the two demography terms (Demo S and Demo R) together, to obtain one joint demography term. Furthermore, the OMD and ODC were combined, to obtain one inheritance term. We are thus left with five terms. The reason for doing so is that this eases the interpretation.

3.8.2 Interpreting the demographic terms

The interpretation of the demographic terms becomes more straightforward when we combine the two demographic terms (we shall call the combined term D_{tot}) and rewrite them.

$$D_{tot} = \sum_{\alpha=1}^{\Omega-1} \Delta c(\alpha, t) \bar{z}(\alpha, t) - c(\Omega, t) \bar{z}(\Omega, t) + \sum_{\alpha=1}^{\Omega} \frac{c(\alpha, t)}{\bar{W}(t)} \bar{r}(\alpha, t) \bar{z}(\alpha, t) \quad (3.24)$$

We start by first writing out $\Delta c(\alpha, t)$.

$$D_{tot} = \sum_{\alpha=1}^{\Omega-1} (c(\alpha + 1, t + 1) - c(\alpha, t)) \bar{z}(\alpha, t) - c(\Omega, t) \bar{z}(\Omega, t) + \sum_{\alpha=1}^{\Omega} \frac{c(\alpha, t)}{\bar{W}(t)} \bar{r}(\alpha, t) \bar{z}(\alpha, t) \quad (3.25)$$

Now we re-arrange the terms, to find:

$$D_{tot} = \sum_{\alpha=1}^{\Omega-1} c(\alpha+1, t+1) \bar{z}(\alpha, t) + \sum_{\alpha=1}^{\Omega} \frac{c(\alpha, t)}{\bar{W}(t)} \bar{r}(\alpha, t) \bar{z}(\alpha, t) - \sum_{\alpha=1}^{\Omega} c(\alpha, t) \bar{z}(\alpha, t) \quad (3.26)$$

We notice that, since Ω is the maximum age in the population, we know that $c(\Omega+1, t) = 0$. Hence, we can safely change the upper limit of the first sum on the right side of the previous equation.

$$D_{tot} = \sum_{\alpha=1}^{\Omega} c(\alpha+1, t+1) \bar{z}(\alpha, t) + \sum_{\alpha=1}^{\Omega} \frac{c(\alpha, t)}{\bar{W}(t)} \bar{r}(\alpha, t) \bar{z}(\alpha, t) - \sum_{\alpha=1}^{\Omega} c(\alpha, t) \bar{z}(\alpha, t) \quad (3.27)$$

Next we realize that $c(\alpha, t) = \frac{N(\alpha, t)}{N(t)}$ and that $\bar{W}(t) = \frac{N(t+1)}{N(t)}$ to obtain:

$$D_{tot} = \sum_{\alpha=1}^{\Omega} c(\alpha+1, t+1) \bar{z}(\alpha, t) + \sum_{\alpha=1}^{\Omega} \frac{N(\alpha, t)}{N(t+1)} \bar{r}(\alpha, t) \bar{z}(\alpha, t) - \sum_{\alpha=1}^{\Omega} c(\alpha, t) \bar{z}(\alpha, t) \quad (3.28)$$

By combining the sums, we get:

$$D_{tot} = \sum_{\alpha=1}^{\Omega} \left(c(\alpha+1, t+1) + \frac{N(\alpha, t) \bar{r}(\alpha, t)}{N(t+1)} - c(\alpha, t) \right) \bar{z}(\alpha, t) \quad (3.29)$$

We have now obtained a term $N(\alpha, t) \bar{r}(\alpha, t)$: the number of individuals in age class α times the average number of offspring for individuals from this age class. This number is then divided by $N(t+1)$. Overall, this term thus calculates which fraction of the population at time $t+1$ will be newborn offspring from individuals of age α at time t . We will call this term $c_{\alpha}(1, t+1)$.

$$D_{tot} = \sum_{\alpha=1}^{\Omega} (c(\alpha+1, t+1) + c_{\alpha}(1, t+1) - c(\alpha, t)) \bar{z}(\alpha, t) \quad (3.30)$$

Now we notice that the term $c(\alpha+1, t+1) + c_{\alpha}(1, t+1) - c(\alpha, t)$ corresponds to the difference of on one hand the fraction of the population at $t+1$ that stems from a given cohort (with age α) through both survival and reproduction and on the other hand the proportional size of that cohort at time t . We denote this differential contribution as $\Delta f(\alpha, t)$. For instance, if currently a fraction of 0.2 of the population belongs to age class 3 and in the next time step, through survival and reproduction, a fraction of 0.3 of the population stems for these age 3 individuals, $\Delta f(3, t)$ would be 0.1. We thus write:

$$D_{tot} = \sum_{\alpha=1}^{\Omega} \Delta f(\alpha, t) \bar{z}(\alpha, t) \quad (3.31)$$

Furthermore, we know that the covariance between these terms over age classes is:

$$\text{cov}(\Delta f(\alpha, t), \bar{z}(\alpha, t))(t) = E(\Delta f(\alpha, t) \bar{z}(\alpha, t)) - E(\Delta f(\alpha, t)) E(\bar{z}(\alpha, t)) \quad (3.32)$$

Here, E refers to the average over age classes. That is:

$$E(b(\alpha, t)) = \frac{1}{\Omega} \sum_{\alpha=1}^{\omega} b(\alpha, t) \quad (3.33)$$

We know, however, that $E(\Delta f(\alpha, t)) = 0$; if the differential contribution of one age class increases, the differential contribution of another has to decrease, which becomes more clear when we write the terms out again:

$$\begin{aligned} E(\Delta f(\alpha, t)) &= \frac{1}{\Omega} \sum_{\alpha=1}^{\Omega} (c(\alpha + 1, t + 1) + c_{\alpha}(1, t + 1) - c(\alpha, t)) \\ &= \frac{1}{\Omega} \sum_{\alpha=1}^{\Omega} (c(\alpha + 1, t + 1) + c_{\alpha}(1, t + 1)) - \sum_{\alpha=1}^{\Omega} c(\alpha, t) \\ &= 1 - 1 = 0 \end{aligned} \quad (3.34)$$

Feeding this into equation 3.32, we find:

$$\text{cov}(\Delta f(\alpha, t), \bar{z}(\alpha, t))(t) = E(\Delta f(\alpha, t)\bar{z}(\alpha, t)) = \frac{1}{\Omega} \sum_{\alpha=1}^{\Omega} \Delta f(\alpha, t)\bar{z}(\alpha, t) \quad (3.35)$$

We now use this result to rewrite equation 3.31

$$D_{tot} = \Omega \cdot \text{cov}(\Delta f(\alpha, t), \bar{z}(\alpha, t))(t) \quad (3.36)$$

Here, the covariance is taken to be the covariance over age classes. This equation provides us with an interpretation of the demographic terms: they measure how much the average body size of a cohort covaries with its differential contribution to the next time step. This term is thus some sort of ‘between age class selection’. Furthermore, it scales linearly with the number of age classes.

3.8.3 Combining the OMD and ODC terms

We have combined the OMD and the ODC term, into one ‘inheritance’ term. The sum of these terms can be written as:

$$\text{OMD} + \text{ODC} = \sum_{\alpha=1}^{\Omega} \frac{c(\alpha, t)}{\bar{W}(t)} \left(\bar{r}(\alpha, t)\bar{d}(\alpha, t) + \text{cov}(d, r)(\alpha, t) \right) \quad (3.37)$$

Here, the ODC term evaluates the covariance of litter size and the difference in body size between the mother and the offspring among individuals within an age class. Consider a population where bigger females, within an age class, produce more, but not bigger offspring. The fact that the offspring are not bigger, will lead to a more negative mother offspring for these individuals (i.e. the mother is bigger, the offspring is of the same size, hence the difference between the mother and the offspring will be bigger). This term will then be negative. The ODC term combines both the between

age class covariance (the same effect, but among age classes) as well as an overall average effect (in general offspring will be smaller than the parents). These effects are thus rather subtle and we have chosen not to treat them explicitly in the main text. Instead, using the identity:

$$\text{cov}(d, r)(\alpha, t) = \overline{d \cdot r}(\alpha, t) - \bar{d}(\alpha, t)\bar{r}(\alpha, t) \quad (3.38)$$

we find an expression for the total inheritance

$$\text{Inheritance} = \text{OMD} + \text{ODC} = \sum_{\alpha=1}^{\Omega} \frac{c(\alpha, t)}{\bar{W}(t)} \overline{d \cdot r}(\alpha, t) \quad (3.39)$$

This term simply contains all effects due to offspring being smaller than their mothers and we therefore call it inheritance.

Interpretation of the inheritance term

As shown the inheritance term is the sum of the OMD and the ODC term. The OMD term consists of a complex interaction between demography (if the average number of offspring increases, the mean age of the population decreases), plasticity (changes in offspring body size due to maternal effects and other environmentally-induced effects) and inheritance (affecting the average mother offspring difference). The ODC term takes into account how litter size co-varies with the differences between the offspring's and the mother's body size. Such a covariance may arise either directly (new-borns being smaller in bigger litters (e.g. Speakman 2008)), or indirectly (for example through body size, if heavier mothers have bigger litters, as well as a larger, more negative, mother offspring difference). Our simulation contains no explicit trade-off between litter size and mother offspring difference and the observed covariance is thus an indirect covariance. The inheritance term thus combines demography, maternal effects, ontogenetic development and stochasticity. Thereby they include different plastic processes, including both between-individual and within-individual plasticity.

3.8.4 Demographic contribution and choice of time steps

In this section, we demonstrate how the choice of age classes directly influences the estimated contribution of selection and of demography, illustrated by a simple example using made up data.

The APE quantifies selection within age classes and takes the sum over all age classes to estimate the total contribution from selection:

$$\sum_{\alpha=1}^{\Omega} \frac{c(\alpha, t)}{\bar{W}(t)} (\text{cov}(z, s)(\alpha, t) + \text{cov}(z, r)(\alpha, t)). \quad (3.40)$$

The sum of these terms is in general not equal to the contribution of total selection ($\text{cov}(z, s)(t) + \text{cov}(z, r)(t)$). The difference between the total contribution of the age-specific selection and the total contribution due to selection depends on the choice of

age classes and will be more pronounced when smaller age classes are chosen (imagine the most extreme scenario in which each individual is in a separate age class: no selection within age classes will occur).

For this example, we only consider survival (s_i) and a trait value (say body mass, z_i), assuming that individuals cannot grow, nor can they reproduce. This reduces the age-structured price equation to:

$$\Delta \bar{z} = \sum_{\alpha=1}^{\Omega-1} \Delta c(\alpha, t) \bar{z}(\alpha, t) - c(\Omega, t) \bar{z}(\Omega, t) + \sum_{\alpha=1}^{\Omega} \frac{c(\alpha, t)}{\bar{W}(t)} \text{cov}(s, z)(\alpha, t) \quad (3.41)$$

We consider a population of 10 individuals, each born in a different month. Measurement of the population takes place on the first day of each month. The column birth indicates in which month the individual was first seen, the column death indicates in which month the individual was found dead.

ID	birth	death	mass
1	1	11	7.70
2	2	11	5.10
3	3	20	6.00
4	4	16	5.30
5	6	19	6.60
6	8	22	7.70
7	9	22	6.20
8	10	28	6.10
9	11	23	7.90
10	12	25	7.50

The question we ask is how the average body mass changes from month 13 (that is january of the second year) to month 25 (january of the next year). In month 13 all individuals except number 1 and 2 are still alive. In month 25 only individual 8 is still alive. Over this timespan we note that:

$$\Delta \bar{z} = \bar{z}(25) - \bar{z}(13) = 6.1 - 6.6625 = -0.5625 \quad (3.42)$$

Using a 1 year timestep

On a one year basis, all individuals are in the same age group. On this basis we call month 13 the start of the year $t = 1$ and month 25 the start of the year $t = 25$. Hence, we see that (s indicating survival to the next year):

	mass	age	s
3	6.00	1	0
4	5.30	1	0
5	6.60	1	0
6	7.70	1	0
7	6.20	1	0
8	6.10	1	1
9	7.90	1	0
10	7.50	1	0

We set $\Omega = 2$ (in years) since it is the maximum age that individuals can reach over this time step. Now the age structured price equation further reduces to:

$$\Delta \bar{z} = (c(2,2) - c(1,1))\bar{z}(1,1) - c(2,1)\bar{z}(2,1) + \frac{c(1,1)}{\bar{W}(1)}\text{cov}(s,z)(1,1) + \frac{c(2,1)}{\bar{W}(1)}\text{cov}(s,z)(2,1) \quad (3.43)$$

Now we note that:

$$\begin{aligned} c(1,1) &= c(2,2) = 1 \\ c(2,1) &= 0 \\ \bar{W}(1) &= \frac{1}{8} \\ \text{cov}(s,z)(1,1) &= \bar{s}\bar{z}(1,1) - \bar{z}(1,1)\bar{s}(1,1) = -0.070\dots \end{aligned}$$

Plugging these numbers into the age structured price equation it is easy to see that the only contribution to $\Delta \bar{Z}$ comes from the survival selection term.

$$\Delta \bar{z} = \frac{c(1,1)}{\bar{W}(1)}\text{cov}(s,z)(1,1) = 8 \cdot -0.070\dots = -0.05625 \quad (3.44)$$

We can thus conclude that the change in trait value between month 13 and month 25 is completely caused by survival selection.

Using a 1 month timestep

When using a timestep with a length of 1 month, every single individual will have a unique age. Because of this, $\text{cov}(z,s)(\alpha, t)$ has to be 0 for every single age class. All the changes in Δz just have to arise because of the other terms in the equation and we can thus—also—conclude that the change in body mass between month 13 and month 25 is entirely due to demography.

3.9 The application of the integral projection model

In this section we will give details on the application of the integral projection model (IPM) framework in order to decompose changes in mean body size into underlying processes (Coulson, Tuljapurkar, and Childs 2010), using the simulated data. First, we give details on the constructed IPM, and second, all functions which were fitted to the data are shown.

3.9.1 Construction of the IPM

An IPM was constructed to describe population dynamics in discrete time whereby body size z was used as a continuous state variable. We built an age-structured IPM, considering only females. In total, the IPM consisted of four kernels, which were all made a function of age α and body size z : 1) A survival kernel $S(\alpha, z)$, describing yearly survival probabilities for individuals of size z and age α . 2) A growth kernel $G(z'|\alpha, z)$ describing probabilities for individuals of size z and age α at time t , to obtain size z' at time $t + 1$. 3) A reproduction kernel $R(\alpha, z)$, describing yearly reproductive success for a female of size z , as the product of the reproduction probability function ($p_{repr}(\alpha, z)$) and the number of offspring produced function $f_{littersize}(\alpha, z)$. 4) An inheritance function $D(z'|\alpha, z)$, describing the probability of offspring having size z' at time $t + 1$, given that their mother has age α and size z at time t . These four kernels together form the IPM:

$$n(t+1, \alpha', z') = \begin{cases} \sum_{\alpha=0}^{30} \int R(\alpha, z) D(z'|\alpha, z) n(t, \alpha, z) dz, & \text{if } \alpha' = 0 \\ \int S(\alpha' - 1, z) G(z'|\alpha' - 1, z) n(t, \alpha' - 1, z) dz, & \text{else} \end{cases} \quad (3.45)$$

The IPM was discretized into a 3100×3100 matrix (i.e. 100 size classes per age class) and we followed methods described by Coulson, Tuljapurkar, and Childs 2010 to perform the decomposition. More details on the construction of IPMs can be found in e.g. Ellner and Rees 2006 and Merow et al. 2014.

3.9.2 Fitted vital rate functions

Vital rates were fitted using generalized linear mixed models, whereby year was included as random effect. For all vital rates, we tested different models including age and body size as fixed effects influencing the intercept, and an age*size interaction. For each vital rate, we applied model selection and used the model with the lowest AIC. Below, we show for each vital rate the most complex model, including the age*size interaction.

Yearly survival probability was estimated using mixed logistic regression, on binomial data describing survival to $t + 1$.

$$S(\alpha, z) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 \cdot z + \beta_2 \cdot \alpha + \beta_3 \cdot \alpha \cdot z + \epsilon_{year1} + \epsilon_{res1})}} \quad (3.46)$$

Annual growth was estimated using linear mixed effects models. Here, $g(\alpha, z)$ is size at $t + 1$.

$$g(\alpha, z) = \beta_4 + \beta_5 \cdot z + \beta_6 \cdot \alpha + \beta_7 \cdot \alpha \cdot z + \epsilon_{year2} + \epsilon_{res2} \quad (3.47)$$

The standard deviation of the residual error term ϵ_{res2} was used to obtain the growth kernel. Here we used the normal distribution density function $\mathcal{N}(x|\mu, \sigma)$, that returns the probability density at point x of a normal distribution with mean μ and standard deviation σ :

$$G(z'|\alpha, z) = \mathcal{N}(z'|g(\alpha, z), \sigma_{res2}) \quad (3.48)$$

The reproduction probability function was fitted using mixed effects logistic regression, on binomial data describing reproduction to $t + 1$ (0 = not reproductive, 1 = reproductive).

$$p_{repr}(\alpha, z) = \frac{1}{1 + e^{-(\beta_8 + \beta_9 \cdot z + \beta_{10} \cdot \alpha + \beta_{11} \cdot \alpha \cdot z + \epsilon_{year3} + \epsilon_{res3})}} \quad (3.49)$$

Litter size was estimated using linear mixed models, with a log link function. Here, we included data on litter size for all reproductive females, and performed the regression on these numbers subtracted by one. In the IPM, $f_{littersize}(\alpha, z)$ was divided by two to include only female offspring.

$$f_{littersize}(\alpha, z) = e^{\beta_{12} + \beta_{13} \cdot z + \beta_{14} \cdot \alpha + \beta_{15} \cdot \alpha \cdot z + \epsilon_{year4} + \epsilon_{res4}} \quad (3.50)$$

Finally, the inheritance function was fitted, relating offspring size to maternal size (at the moment they reproduce):

$$d(\alpha, z) = \beta_{16} + \beta_{17} \cdot z + \beta_{18} \cdot \alpha + \beta_{19} \cdot \alpha \cdot z + \epsilon_{year5} + \epsilon_{res5} \quad (3.51)$$

The standard deviation of the residual error term ϵ_{res5} was used to obtain the offspring size distribution, where \mathcal{N} is again the normal distribution probability function:

$$D(z'|\alpha, z) = \mathcal{N}(z'|d(\alpha, z), \sigma_{res5}) \quad (3.52)$$

Using the fitted relations shown above, we parametrized an IPM for a median year. To do so, for each vital rate, we calculated the linear part of the model, using the estimated fixed effects. To avoid bias due to Jensen's inequality (e.g. Fox and Kendall 2002), we sampled 10,000 times from a normal distribution with $\mu=0$ and a standard deviation equal to the estimated random effect (ϵ_{year}), and added this to the fixed part. We subsequently applied the appropriate link function and averaged the outcomes. Vital rates were combined into the IPM according to Eq. 3.45.

3.10 Variance decomposition in the APE and IPM framework

In the main text, we limited ourselves to a discussion of the absolute contributions of the different terms of the APE and IPM to the changes in mean trait value. Another focus of interest is to analyse which constituents underlie the biggest variance in this change. In the application of age-structured Price equation (APE) by Ozgul et al. 2009 on the sheep population on St. Kilda, both the absolute contributions to $\Delta\bar{z}$ over time and a variance partitioning of $\Delta\bar{z}$ into contributions were regarded. This partitioning aims at finding which processes cause *changes* in $\Delta\bar{z}$ and is thus not directly comparable with contributions to changes in \bar{z} (or: $\Delta\bar{z}$ itself). We have performed a variance partitioning for both the APE and integral projection model (IPM) approach by means of an ANOVA type III (Fig. 3.7).

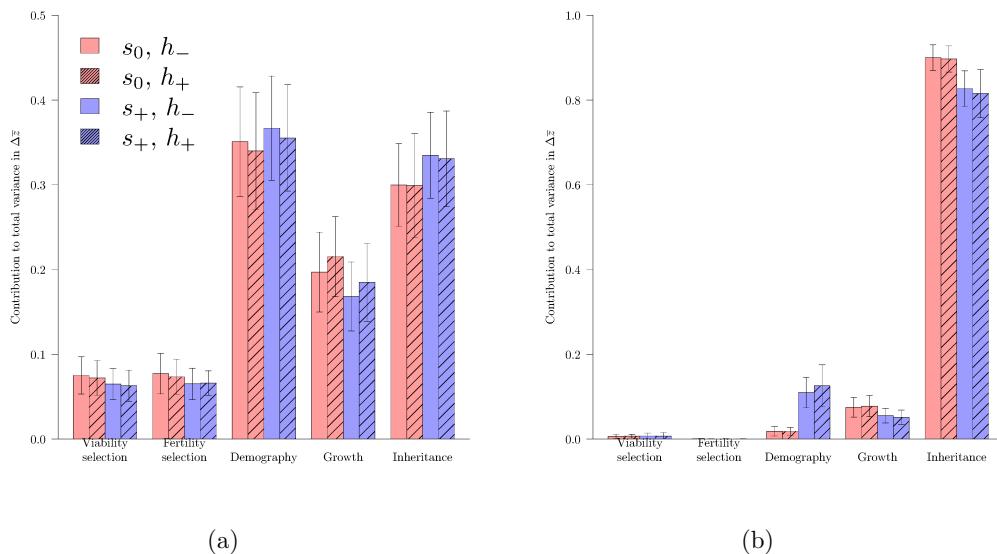


Figure 3.7: Variance partitioning of processes underlying changes in $\Delta\bar{z}$ according to a) the age structured Price equation, and b) integral projection models. Demography is the sum of the two demographic terms, inheritance is the sum of ODC and OMD (Appendix 3.8). Red bars indicate scenarios without selection (s_0), blue bars scenarios with selection (s_+). Solid bars indicate scenarios with low heritability (h_-), shaded bars scenarios with high heritability (h_+). Errors bars show variation across replicates (± 1 SD). Note the different scale on the y-axis in a) and b).

Although the absolute contributions to $\Delta\bar{z}$ are very similar (Fig. 2c and Fig. 2d), the second order contributions to \bar{z} differ greatly between the APE and the IPM (figure 3.7). The main reason for this is that the IPM smoothens the contributions over time, through the statistical models that it is based on. A variance decomposition relies on the variation in the contributions. When the absolute values are smoothed, the outcome of the decomposition may change drastically, as observed in figure 3.7.

Chapter 4

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

The world is always full of the sound of waves. The little fishes, abandoning themselves to the waves, dance and sing, and play, but who knows the heart of the sea, a hundred feet down? Who knows its depth?

— Eiji Yoshikawa, *Musashi* (1935)

If you can't control your peanut butter, you can't expect to control your life.

— Bill Watterson, *Calvin and Hobbes* (1985-1995)

Timothée Bonnet, Peter Wandeler, Glauco Camenisch and Erik Postma. In review for PloS Biology.

4.1 Abstract

Despite being heritable and under selection, trait dynamics often do not appear to match those predicted by evolutionary theory. Indeed, conclusive evidence for contemporary adaptive evolution of quantitative traits remains rare for wild vertebrate populations, and stasis seems to be the norm. This so-called 'stasis paradox' highlights our inability to predict evolutionary change. This is especially concerning within the context of rapid anthropogenic environmental change, and its underlying causes are therefore hotly debated. Applying a quantitative genetic framework to individual-based long term data for a wild rodent population, we show that in this population stasis is an illusion: The population has evolved to become lighter, and this genetic change is an adaptive response to a change in snowfall patterns. Whereas both this evolutionary change and the selective pressures that drive it are not apparent on the phenotypic level, by estimating selection at the genetic level we were able to identify the relevant phenotypic selective pressure, as well as uncover the accompanying evolutionary response. We thereby demonstrate that natural populations can show a rapid and adaptive evolutionary response to novel selective pressures, and that ex-

plicitly (quantitative) genetic models are able to provide us with an understanding of the causes and the consequences of selection that is superior to purely phenotypic estimates of selection and evolutionary change.

4.2 Introduction

Given the rapid anthropogenic environmental changes experienced by organisms around the world, there is an increasing need for an ability to understand and predict the evolutionary dynamics of wild populations (Parmesan 2006; Merilä and Hendry 2014). Despite good empirical examples of the adaptive evolution of traits with a simple genetic architecture (Hof et al. 2011; Karell et al. 2011; Lamichhaney et al. 2016), the picture is very different for quantitative traits, which are a function of many genes of small effect (Wellenreuther and Hansson 2016). Although it are these traits that are of most interest to evolutionary biologists (Roff 2007; Walsh 2014), predictive models of quantitative trait evolution have largely failed when applied to data from wild populations (Merilä, Sheldon, and Kruuk 2001).

Although there is an abundant literature showing that both directional selection (Kingsolver et al. 2001; Kingsolver et al. 2012) and heritable genetic variation (Mousseau and Roff 1987; Postma 2014) are common, these pre-requisites of Darwinian evolution rarely allow us to explain evolutionary trends retrospectively, let alone to make predictions for the future (Merilä, Sheldon, and Kruuk 2001; Morrissey, Parker, et al. 2012). For example, both natural and sexual selection almost universally favour larger and heavier individuals (Blanckenhorn 2000). Furthermore, morphological traits are generally moderately heritable (Mousseau and Roff 1987; Postma 2014), and averaged across the 151 estimates compiled in (Postma 2014), the heritability of body mass is 0.33 ± 0.02 . Nevertheless, while species do tend to get larger over geological time-scales (Cope 1887; Alroy 1998; Heim et al. 2014; Baker et al. 2015), this rate of evolution is orders of magnitude slower than what could be predicted from the strength of selection and heritability observed in contemporary populations (Merilä, Sheldon, and Kruuk 2001; Bell 2010; Gotanda et al. 2015).

On the whole, conclusive evidence for the contemporary adaptive evolution of quantitative traits in wild vertebrate populations is remarkably scarce and elusive (Merilä, Sheldon, and Kruuk 2001; Morrissey, Parker, et al. 2012), and good examples —Trinidadian guppy life-histories (Reznick and Bryga 1996), human reproductive timing (Milot et al. 2011), timing of pink salmon migration (Kovach, Gharrett, and Tallmon 2012) and big-horn sheep horn size (Pigeon et al. 2016)—can be counted on one hand. Furthermore, of these studies, (Pigeon et al. 2016) reported a response to harvesting-induced, artificial rather than natural selection, and despite considerable effort to uncover any evolutionary consequences of climate change (Charmantier and Gienapp 2014; Gienapp and Brommer 2014; Merilä and Hendry 2014; Crozier and Hutchings 2014), only (Kovach, Gharrett, and Tallmon 2012) were able to demonstrate an adaptation to climate.

Our apparent inability to reconcile predictions of evolutionary change based on estimates of selection and genetic variation with the (lack of) of genetic change observed, i.e. the 'stasis paradox' (Merilä, Sheldon, and Kruuk 2001), is a major concern in urgent

need of a resolution. Given how commonly evolutionary predictions fail to capture observed trait dynamics, some have concluded that there are fundamental problems that prohibit the application of quantitative genetic methods to natural populations (Steiner and Tuljapurkar 2012; Coulson et al. 2015). However, there are in fact three theoretically well-developed (quantitative genetic) explanations for this mismatch.

First, although the great majority of studies base their observed rate of evolution solely on phenotypic changes, evolutionary change does not need to be apparent at the phenotypic level. Instead, it may be masked by phenotypically plastic changes, which may be several-fold larger and/or go opposite to the genetic change (Postma, Visser, and Van Noordwijk 2007). For instance, while a change in the environment may generate selection favouring an increase in the frequency of alleles promoting fat accumulation, at the same time this may create a food shortage, leading to a plastic decrease of fat reserves. As a consequence, an evolutionary trend may be masked by a counteracting plastic change, also referred to as “cryptic evolution” (Merilä, Kruuk, and Sheldon 2001; Hadfield, Wilson, and Kruuk 2011).

Second, this potentially flawed phenotypic estimate of the the *observed* rate of evolution is typically compared to a *prediction* derived from the univariate breeder’s equation, i.e. the product of selection and heritability, where selection is quantified as the covariance between the trait of interest and relative fitness. In natural systems, selection however rarely acts on traits in isolation (Lande and Arnold 1983). If these traits are genetically correlated to the focal trait, they may significantly alter the focal trait’s evolutionary trajectory (Schluter, Price, and Rowe 1991; Morrissey, Walling, et al. 2012). While the role of genetic correlations among traits within the same individual (Teplitsky, Robinson, and Merilä 2014) or between the sexes (Poissant, Wilson, and Coltman 2010) has received substantial attention, the potential role of genetic constraints resulting from genetic correlations between traits expressed in different individuals has received far less attention. In particular parent-offspring conflict, i.e. a genetic trade-off between offspring quality and fecundity (Trivers 1974), may however constrain the evolution of size (Kölliker et al. 2015), with positive directional selection on offspring size counterbalancing selection against investment per offspring on the level of parents (Rollinson and Rowe 2015).

Finally, even in the absence of selection on correlated traits, it is challenging to obtain an estimate of the strength of natural selection that is unbiased by the existence of a third, non-genetic variable that influences both the trait and fitness (Rausher 1992). Although the univariate breeder’s equation assumes that the covariance between phenotype and fitness is solely the result of a causal relationship between the two (Morrissey, Kruuk, and Wilson 2010; Morrissey, Parker, et al. 2012), this assumption is likely to be violated, especially in natural populations. For instance, a trait that plastically responds to food availability, such as body mass, is likely to covary with fitness at the phenotypic level, irrespective of the causal effects of the trait on fitness: individuals that have access to more food are heavier and reproduce more (van Noordwijk 1988; Schluter, Price, and Rowe 1991). Because the fitter individuals are not genetically different in terms of body mass, this covariation has no evolutionary consequences, even if body mass is heritable (Rausher 1992).

While these difficulties have been discussed previously, and studies regularly note that the mismatch between the observed and predicted response may be attributable

to any of them, they rarely account for them in an explicit, quantitative manner. Therefore, we here apply a comprehensive analytical framework to long-term individual-based body mass data for a wild rodent population, which shows an apparent mismatch between observed rates of phenotypic change and predicted rates of genetic change. We use information on within-population relatedness and individual-level trait measurements (C. Henderson 1950; Lynch and Walsh 1998) to obtain a statistically robust estimate of the direction and rate of genetic change (Postma 2006; Hadfield et al. 2010; Morrissey, Parker, et al. 2012). Subsequently we disentangle the role of genes and the environment in shaping the covariance between body mass and fitness, and identify the target of selection. This allows us to directly compare the observed genetic change to a range of evolutionary predictions, and to thereby resolve the stasis paradox and provide a deeper understanding of selection and evolution in this biological system.

4.3 Results and discussion

Based on ten years of data on an alpine population of snow voles (García-Navas et al. 2015; Bonnet and Postma 2016) (*Chionomys nivalis*, Martin 1842), we find that relatively heavy individuals both survive better ($p = 0.04$) and produce more offspring per year ($p = 0.003$). Assuming causality, this generates a strong phenotypic estimate of selection favouring heavier individuals (selection differential $S = 0.86$ g, $p < 10^{-5}$). In line with other morphological traits (Mousseau and Roff 1987; Postma 2014), variation in body mass has a significant additive genetic component ($V_A=4.34$ g 2 , 95%CI [2.40;7.36]), which corresponds to a heritability (h^2) of 0.21 (95%CI [0.11;0.29]). Similarly, we find significant additive genetic variance in fitness ($V_A=0.10$; 95%CI [0.06;0.19], $h^2 = 0.06$ 95%CI [0.04;0.12]), measured as relative lifetime reproductive success (rLRS).

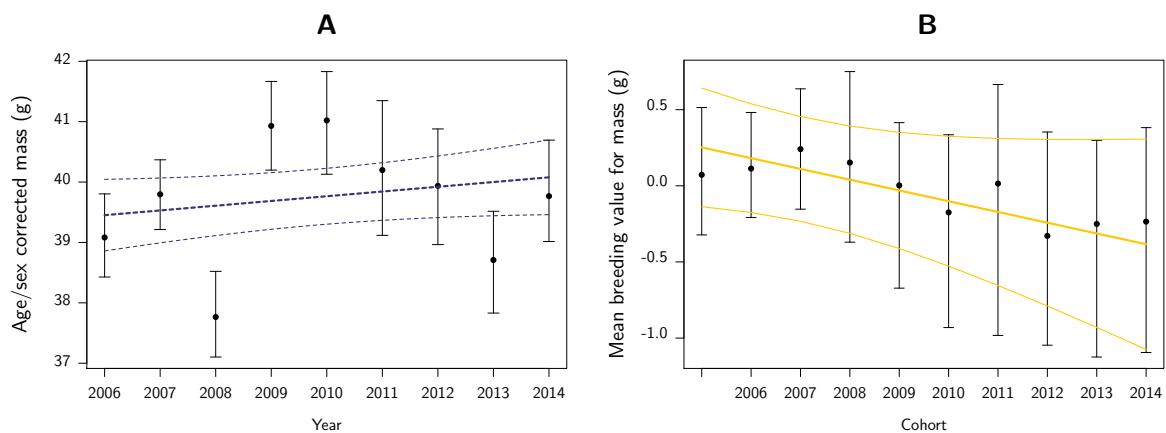


Figure 4.1: Temporal variation in mass and estimated breeding values for mass. (A): Year-specific mean mass corrected for age, sex and date of measurement, with 95%CI. (B): Cohort-specific mean estimated breeding value for mass with their 95%CI and the trend in breeding value with 95%CI. Note the different scale on the y-axes.

Given these estimates of selection (S) and heritability (h^2), the breeder's equation ($R = h^2S$) predicts an adaptive evolutionary response (R) in body mass (Lynch and Walsh 1998; Morrissey, Parker, et al. 2012), i.e. an increase in the mean breeding value for body mass over time, of 0.17 g/year (95%CI [0.07;0.28]; Fig. 4.2A UBE). However, after correcting for changes in demographic structure (i.e. accounting for sex and age effects, see Fig. 4.5), over the past nine years (approximately eight generations), the change in mean body mass is not significant and small at best (0.08 g/y; 95%CI [-0.02;0.18]; $p=0.14$). This apparent mismatch between the predicted evolutionary change based on the breeder's equation and the phenotypic change observed provides yet another example of the stasis paradox (Merilä, Sheldon, and Kruuk 2001).

To test whether the predicted positive genetic trend, i.e. an increase in breeding values, is being masked by an opposing phenotypically plastic response (Merilä, Sheldon, and Kruuk 2001; Hadfield, Wilson, and Kruuk 2011), we directly estimated the additive genetic covariance between mass and fitness. Based on the Robertson-Price's equation, this provides an unbiased estimate of the rate of genetic change per generation (Robertson 1966; G. R. Price 1970; Morrissey, Parker, et al. 2012). Contrary to our prediction, this estimate of genetic change in mass is strongly negative and highly significant ($p_{MCMC} < 0.001$; Fig. 4.2A GCPE). When normalized by a mean generation time of 1.2 years, this provides a rate of evolutionary change of -0.29 g/year (95%CI [-0.55; -0.07]) or approximately 8,600 Darwins, which is in line with other rates of "micro-evolution" (e.g. between 3,700 and 45,000 Darwins in the Trinidadian guppies (Reznick et al. 1997)). Importantly, this rate of evolution is unlikely to have happened solely through genetic drift ($p_{MCMC} < 0.001$; Fig. 4.6 and 4.7) (Hadfield et al. 2010), and therefore most likely reflects a response to selection favouring genetically lighter individuals.

This result was confirmed by an independent estimate using best linear unbiased predictors (BLUPs) of breeding values for mass: Taking into account the non-independence of BLUPs and sampling variance (Postma 2006; Hadfield et al. 2010), we find that predicted breeding values have declined over the past nine years (-0.07 g/year, $p_{MCMC}=0.06$; Fig. 4.1B & Fig. 4.2A TPBV), and this despite the BLUPs approach being potentially biased towards the phenotypic trend (Postma 2006) (i.e. in this case toward zero). This negative trend, combined with the fact that the phenotypic mean has either remained constant or has shown a slight increase (see above), implies that the plastic component of body mass must have increased. Although the cause of this increase remains unknown, population size has declined over the study period (Fig. 4.5), which may have resulted in an increase in the per-capita resource availability (i.e. density dependence). Alternatively, the absolute food availability or quality may have improved. Interestingly, although these environmental changes may be coincidental, they may also be a direct result of a change in the selection regime or the evolutionary change toward smaller size (Cooke et al. 1990; Hadfield, Wilson, and Kruuk 2011).

As the phenotypic selection differential ($\sigma_{P(m,\omega)}$) is equal to the sum of the additive genetic and environmental covariances between mass and rLRS ($\sigma_{A(m,\omega)}$ and $\sigma_{E(m,\omega)}$, respectively) (Robertson 1966; G. R. Price 1970; Morrissey, Parker, et al. 2012), it follows that because $\sigma_{P(m,\omega)}$ is positive and $\sigma_{A(m,\omega)}$ is negative, the environmental covariance must be large and positive (Fig. 4.2B LRS). In other words, while environmental

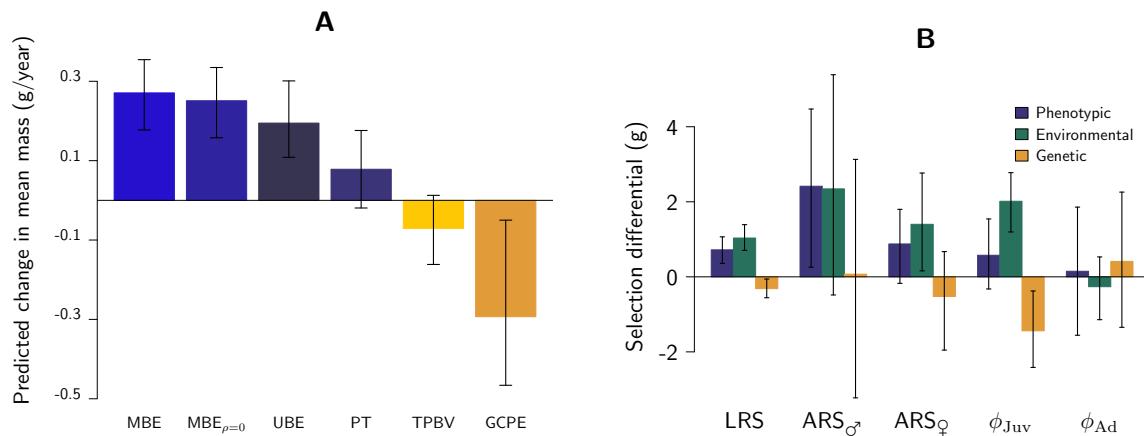


Figure 4.2: Predicted and observed rates of evolutionary change. (A): Rates of evolutionary change predicted by (from left to right) the breeder's equation in its multivariate form (MBE), the multivariate breeder's equation while constraining the genetic correlations to zero ($MBE_{\rho=0}$), and the univariate breeder's equation (UBE), followed by the phenotypic trend (PT), the trend in predicted breeding values (TPBV) and the genetic change estimated by the Price equation (GCPE). (B): Phenotypic, genetic and environmental selection differential for total selection (LRS), fertility selection in males (ARS_{σ^2}) and females (ARS_{ϕ^2}), viability selection in juveniles (ϕ_{Juv}) and in adults (ϕ_{Ad}). Both panels show posterior modes, with vertical lines indicating 95%CI.

conditions that make voles heavy (for instance abundance of food or lack of parasites) also make them successful at reproducing and surviving, there is no causal *positive* relationship between breeding values for mass and fitness (Fig. 4.3). It is this difference in sign between $\sigma_{A(m,\omega)}$ and $\sigma_{E(m,\omega)}$ which represents an extreme violation of the breeder's equation (which assumes $\sigma_{A(m,\omega)} / (\sigma_{A(m,\omega)} + \sigma_{E(m,\omega)}) = h^2$). Hence, our initial prediction of evolution was wrong, demonstrating that phenotypic estimates of selection may provide severely biased predictions of the evolutionary trajectories of wild populations. But *why* is evolution taking place in a direction that is opposite to apparent phenotypic selection?

Indirect selection may be acting on body mass through one or more traits with negative genetic correlations with mass (Lande and Arnold 1983; Morrissey, Walling, et al. 2012). However, the genetic correlations among the three morphological traits for which we have data—body mass (m), body length (b) and tail length (t)—are all positive (estimates and 95%CI: $\rho_{m,b} = 0.79 [0.06; 0.93]$; $\rho_{m,t} = 0.40 [0.01; 0.66]$; $\rho_{t,b} = 0.56 [-0.04; 0.85]$), and the predicted response based on the multivariate breeder's equation (Fig. 4.2A MBE) is very similar to that based on its univariate counterpart (Fig. 4.2A UBE), as well as to that based on a multivariate breeder's equation constraining the correlations to zero (Fig. 4.2A $MBE_{\rho=0}$). Furthermore, for parent-offspring conflict between size and fertility to constrain the evolution of size, the genetic correlation between juvenile size and adult annual reproductive success should be negative (Rollinson and Rowe 2015). In our study population this correlation was 0.21, 95%CI [$-0.24; 0.74$]), arguing against a role for a trade-off between fertility and offspring size in driving the observed evolution toward smaller sizes. Although we cannot exclude that selection on other, unmeasured, traits does indirectly shape body mass evolution,

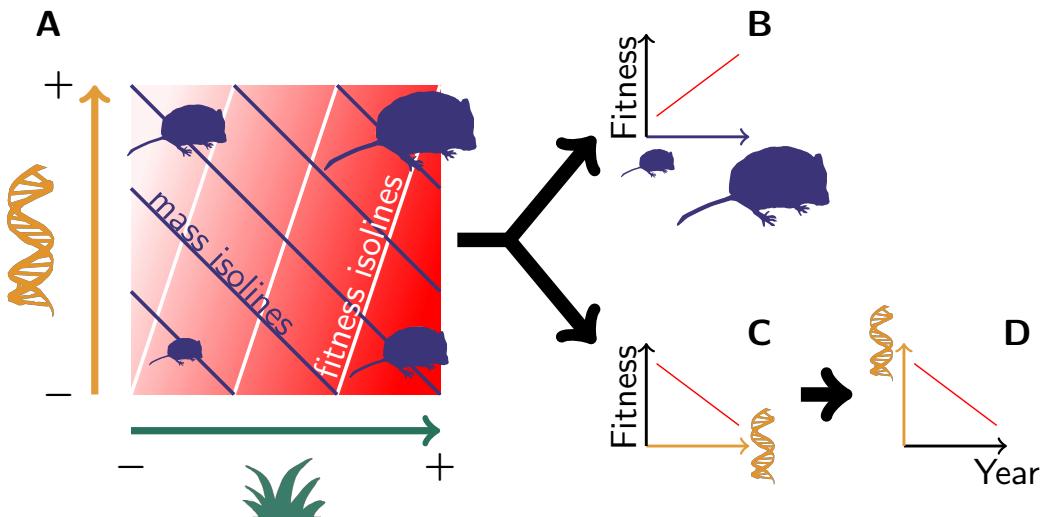


Figure 4.3: Schematic representation of why adaptive evolution goes opposite to apparent selection. (A) Both genetic variation (orange) and environmental variation (green) contribute to phenotypic variation in mass (purple) in an additive and equivalent way: large voles are the result of genes for being large and of a favourable environment. Therefore, mass isolines form an angle of 45° with both axes. However, the effect the fitness effect of the environmental component of mass variation is not the same as the fitness effect of the genetic component of mass variation: Fitness increases (the redder the fitter) with increasing environmental effects on mass, but decreases with increasing genetic effects on mass, as illustrated by fitness isolines (white). This pattern leads to (B) a phenotypic selection for heavier voles, along with (C) a “genetic selection” for lighter genotypes, thus leading to (D) an evolution towards lighter voles. Panel (A) should not be mistaken for a genotype-by-environment interaction: genetic and environmental effects are additive both for mass and for fitness. The additive effects on mass do not map on the additive effects on fitness, however.

genetically correlated traits are not more likely to constrain than to facilitate adaptation (Agrawal and Stinchcombe 2009). This suggests that the counter-intuitive direction of evolution is really due to selective pressures acting on mass, but given that selection acts on phenotypes rather than genotypes, which aspect of an individual’s body mass is the subject of negative selection?

To identify the fitness component that is negatively associated with genes for being heavy, we computed sex- and age-specific genetic covariances between mass and fitness components. Whereas the genetic covariances between mass and both relative annual reproductive success and adult survival are close to zero in both sexes (Fig. 4.2B), the genetic covariance between mass and over-winter survival is negative in juveniles (-0.98 [-2.44;-0.18] on a logit scale, $p_{MCMC}=0.01$). Because the genetic correlation between juvenile and adult mass is positive ($r_A = 0.88$; 95%CI [0.39;1]) and significantly different from 0 ($p=0.004$) but not 1 ($p=0.35$), selection on juvenile mass can shape genetic variance for mass at all ages, and thereby contribute to the observed negative genetic change (Chevin 2015). While this shows that negative viability selection of juvenile mass is responsible for the genetic change toward smaller individuals, how come survival is higher for heavier phenotypes *and* lighter genotypes?

Juvenile mass covaries positively with both within- and between-year survival ($p = 0.009$ and $p = 1.3 \times 10^{-6}$, respectively). However, juveniles can only be captured when they first leave their burrow, at an age of approximately three weeks (Janeau and Aulagnier 1997) and a weight of 12 to 20 g, and they may continue to be captured until the end of the season, when they can reach weights of up to 50 g. Because of growth, mass measurements are therefore not directly comparable among juveniles. Indeed, at least part of the positive estimated selection on juvenile mass is likely to be mediated by the simultaneous increase, with age, of mass and of the probability of survival to the next year (Hadfield 2008). In addition, viability selection introduces non-random missing data, which results in biased estimates of viability selection on mass (Hadfield 2008; Steinsland et al. 2014). This led us to hypothesise that the positive phenotypic association between juvenile body mass and survival was largely the result of ontogeny non-random missing data, whereas the negative genetic association is driven by selection imposed by the necessity to have completed growth before the end of the growing season.

The (co)variance decompositions presented above have the advantage that they do not make causative statements. For example, a genetic covariance describing the rate of evolution has a self-contained, tautological, meaning and does not make any assumptions with respect to its causes (S. A. Frank 2012). However, if we are to identify the target of juvenile viability selection, we must adopt a more traditional hypothesis testing framework. Although, as we emphasised above, inferring a causal relationship between a trait and fitness based on their covariance requires great care, we set out to test the hypothesis that when the period favourable for growth is limited, selection favours lighter juveniles, as they require less time to reach their adult size.

To obtain an estimate of viability selection that is unbiased by growth and non-random missing data due to mass-dependent mortality occurring after the first capture (Hadfield 2008), we used a Bayesian model to simultaneously infer birth dates and growth curves for all juveniles observed at least once, irrespective of when and how often they have been captured. Although we cannot account for viability selection acting before the first capture, this model enabled us to quantify viability selection on age-corrected juvenile mass—i.e. asymptotic or predicted adult mass—, and thereby compare all individuals at the same developmental stage, irrespective of their fate.

Inferred birth dates revealed that snow fallen during the preceding winter is a major ecological factor constraining the onset of reproduction in spring, with reproduction starting on average 40 days after the snow has melted (SE 4.5, $p = 4 \times 10^{-5}$) (Fig. 4.4A). As a consequence, juveniles only have a limited amount of time to grow and reach their adult mass before the return of winter. As growth rate and predicted adult mass are slightly negatively correlated (correlation -0.077; 95%CI [-0.150; -0.002]), juveniles with a smaller adult mass on average require less time to complete development. If individuals that have not completed development before the arrival of winter pay a survival cost, for example due to trade-offs between growth and vital physiological processes (Stearns and Koella 1986; Owens and Wilson 1999), this generates selection for small size, especially for juveniles born toward the end of the season (Fig. 4.4D and Fig. 4.8).

To test this, we quantified the strength of survival selection acting on predicted adult

mass, which was slightly negative when averaged over all years and the complete reproductive season ($p_{MCMC}=0.13$), but interacted strongly and significantly with the number of days between birth and the first snowfall of that year ($p_{MCMC}=0.008$). This implies that individuals born closer to the first snowfall are more strongly selected for a low adult mass, and that the length of the snow-free period in a given year determines the total selection experienced by the population in that year. Interestingly, at our field site, the length of the snow-free period in the years 2008 to 2014 has been significantly shorter than during the preceding six years (Fig. 4.4B). The latter coincides with a period of exceptionally high snowfall, low temperatures and a long duration of snow cover, across the Swiss Alps (Beniston 2012).

Our model estimates that in 2006 and 2007, when the snow-free period was long (Fig. 4.4B), most juveniles reached their adult mass before the first snowfall, and there was hence no selection on asymptotic mass ($\beta = -0.002$, SE= 0.0006, $p_{MCMC}=0.47$, Fig. 4.4C; D; 4.7). However, in all subsequent years, the snow-free period was much shorter, and there was selection for a lower asymptotic mass ($\beta = -0.10$, SE=0.0008, $p_{MCMC}=0.009$). This suggests that the shortening of the snow-free season, and thereby selection for lower asymptotic mass, is a novel phenomenon that the population is currently in the process of adapting to. Although model complexity and data availability prohibit disentangling genetic and environmental sources of variation in asymptotic mass among individuals and over time, and we cannot rule out the possibility that the selective pressure we identified is not causative (Morrissey 2014), the cohort born in 2013 had an estimated adult mass that was 1.02 g smaller than the cohort born in 2006 ($p=0.05$). This shrinkage is predicted to have increased population-level juvenile survival by 2.5%, and to have contributed positively to population recovery (Fig. 4.5).

4.4 Conclusion

We have exploited a case of apparent evolutionary stasis to gain a deeper insight into the evolutionary dynamics of natural populations, and the selective pressures that shape them. Whereas estimates of selection and evolution based on phenotypic data alone can easily mislead our understanding of the selective and evolutionary processes in natural populations, a quantitative genetic framework applied to individual-based long-term data allows us to unravel evolutionary and environmental changes over time, and to obtain unbiased estimates of selection. This has resolved a case of apparent evolutionary stasis, and provided a comprehensive empirical demonstration of contemporary adaptive evolution in response to a climatic fluctuation.

4.5 Methods

Snow vole monitoring. Monitoring of the snow vole population began in 2006 and the present work uses data collected until the fall of 2014. The snow vole monitoring was authorised by the *Amt für Lebensmittelsicherheit und Tiergesundheit*, Chur, Switzerland. The study site is located at around 2030m above sea level, in the central eastern Alps near Churwalden, Switzerland (46°48' N, 9°34' E). It consists of scree, inter-

spersed with patches of alpine meadows and surrounded by habitat unsuitable for snow voles: a spruce forest to the West, a cliff to the East and large meadows to the North and South. In accordance with it being considered a rock-dwelling specialist (Janeau and Aulagnier 1997), at our study site it is almost never captured outside of the rocky area. Given that it is ecologically fairly isolated, we are able to monitor the whole population. Trapping throughout the whole study area takes place during the snow-free period, between late May and mid-October. One trapping session consists of four trapping nights. Analyses presented here are based on a total of two (in one year), three (in three years) or five (in five years) trapping sessions per season. All newly-captured individuals weighing more than 14 g are marked with a subcutaneous passive transponder (PIT, ISO transponder, Tierchip Dasmann, Tecklenburg). Additionally, an ear tissue sample is taken (maximum 2 mm in diameter) using a thumb type punch (Harvard Apparatus) and stored in 100% ethanol at -20°C. DNA extracted from these samples is genotyped for 18 autosomal microsatellites developed for this population (Wandeler, Ravaoli, and Bucher 2008), as well as for the *Sry* locus to confirm the sex of all individuals. Finally, another Y-linked marker as well as a mitochondrial marker is used check for errors in the inferred pedigree (see below). An identity analysis in CERVUS v.3.0 (Marshall et al. 1998) allows us to identify animals sampled multiply, either because they lost their PIT, or because at their first capture as a juvenile they were too small to receive a PIT. All the analyses were carried out in R (R Core Team 2014). Specific packages are referenced below.

Pedigree inference. Parentage was inferred by simultaneously reconstructing paternity, maternity and sibship using a maximum likelihood model in MasterBayes (Hadfield, Richardson, and Burke 2006). Parentage was assigned using a parental pool of all adults present in the examined year and the previous year, assuming polygamy and a uniform genotyping error rate of 0.5% for all 18 loci. As it is known that in rare cases females reach sexual maturity in their year of birth (Janeau and Aulagnier 1997), we matched the genotypes of all individuals against the genotypes that can be produced by all possible pairs of males and females. We retrieved the combinations having two or less mismatches (out of 18 loci) and ensured that parental links were not circular and were temporally consistent. This way, we identified eight young females as mothers of animals born in the same year, with a known father but a mother not yet identified. All of these females were relatively heavy (>33 g) at the end of the season and their home-ranges matched those of their putative offspring. Finally, the pedigree was checked using a polymorphic Y-linked locus developed for this population (Wandeler and Camenisch 2011), as well as a fragment of the mitochondrial DNA control region, amplified using vole specific primers (Haring, Herzig-Straschil, and Spitzemberger 2000). There were no inconsistencies between the transmission of these three markers and the reconstructed pedigree. The final pedigree had a maximum depth of 11 generations and a mean of 3.8 generations. It consisted of 987 individuals with 458 full-sibling pairs, 3010 half-sibling pairs, 764 known maternities and 776 known paternities, so that, excluding the base population, 86% of the total parental links were recovered.

Traits. The recapture probability from one trapping session to the next was estimated to be 0.924 (SE 0.012) for adults and 0.814 (SE 0.030) for juveniles using mark-recapture models. Thus, with three trapping session a year, the probability not to trap an individual present in a given year is below 10^{-3} . Not surprisingly, no animal was captured in year y , not captured in $y + 1$, but captured or found to be a parent of a juvenile in $y + 2$ or later. Therefore, capture data almost perfectly matches over-winter survival. However, as is almost always the case in these type of studies, we are unable to separate death from permanent emigration. Importantly however, as both have the same consequences on the population level, this does not affect our evolutionary predictions.

Annual and lifetime reproductive success (ARS and LRS, respectively) were defined as the number of offspring attributed to an individual in the pedigree, either over a specific year or over its lifetime. 56 individuals born of local parents were not captured in their first year, but only as adult during the next summer, probably because they were born late in the season and we had only few opportunities to catch them. This means that we miss a fraction of the juveniles that are not observed in their first year and die, or emigrate, during the following winter. We acknowledge that this means that our measures of ARS and LRS partly conflate adult reproductive success and the viability of those juveniles that were never observed, but our measures are the most complete measures of reproductive success available in this system.

We used relative LRS (ω) as proxy for fitness (Lande and Arnold 1983), where $\omega_i = \frac{\text{LRS}_i}{\frac{1}{N_{s,t}} \sum_{j=1}^{N_{s,t}} \text{LRS}_{j,t}}$. Here, $N_{s,t}$ is the number of individuals of same sex as the focal individual i , present in the cohort t , so that $\frac{1}{N_{s,t}} \sum_{j=1}^{N_{s,t}} \text{LRS}_{j,t}$ is the sex-specific, cohort-specific mean of LRS. The latter is required as the mean LRS differs between males and females due to imperfect sampling (Morrissey, Parker, et al. 2012). In addition, we used cohort-specific means in order to account for variations in population size.

Generation time was defined as the mean age of parents at birth of their offspring (Charlesworth 1994).

Mass (m) was measured to the nearest gram with a spring scale. Both body length (b), measured from the tip of the nose to the base of the tail, and tail length, measured from the tip to the base of the tail (c), were measured to the closest mm with a calliper while holding the animal by the tail.

Selection. Selection differentials were estimated using bivariate linear mixed models, as the individual-level covariance between fitness and mass (corrected for sex, age and cohort). However, while this provides the best estimate of the within-generation change in trait mean due to selection (Lande and Arnold 1983), because the distribution of fitness is not Gaussian, it cannot be used to estimate confidence intervals. Hence, the statistical significance of selection was tested using a univariate over-dispersed Poisson generalized linear mixed model (GLMM) in which LRS was modelled as a function of individual standardized mass and including sex and age as covariates and cohort as a random effect. Note that the latter estimates the effect of mass on a transformed scale, and therefore cannot be directly used to quantify an effect of selection on the original scale measured in grams (Mitchell-Olds and Shaw 1987).

The significance based on the basis of the GLMM was confirmed by non-parametric bootstrapping. Similarly, we tested for the significance of selection through ARS only, using an over-dispersed Poisson GLMM including sex as a fixed effect, and year and individual as random effects.

The estimation of survival selection is facilitated by the fact that the year-to-year individual recapture probability is effectively 1. Therefore, selection on year-to-year survival was tested for by a binomial GLMM. This model included sex, age and their interaction as fixed effects, and year as a random effect.

In order to integrate the uncertainty in the estimation of selection with the uncertainty in the estimation of heritability when predicting the rate of evolution, selection differentials and gradients were also obtained from the multivariate animal model presented below.

Quantitative genetic analyses. We used uni- and multivariate animal models to estimate additive genetic variances, covariances and breeding values (C. R. Henderson 1984; Lynch and Walsh 1998; Kruuk 2004) with MCMCglmm (Hadfield 2010). All estimations were carried out in a Bayesian framework in order to propagate uncertainty when computing composite statistics such as heritabilities and rates of genetic change (Stinchcombe, Simonsen, and Blows 2014). All estimates provided in the text are posterior modes and credibility intervals are highest probability density intervals at the 95% level. All the animal models were run for 1,300,000 iterations with a burnin of 300,000 and a thinning of 1,000, so that the autocorrelations of each parameter chain was less than 0.1. Convergence was checked graphically and by running each model twice.

Univariate models: We first carried out univariate model selection, fitting models without an additive genetic effect, to determine which fixed and random effects to include. Based on AICc (Burnham and Anderson 2002), and fitting the models by Maximum of Likelihood in lme4 (Bates et al. 2015), we obtained a model that predicts the mass $m_{i,t}$ of individual i at time t by: age, as a factor (juvenile or adult); sex as a factor; the interaction between age and sex; Julian dates and squared Julian dates, which were centered and divided by their standard deviations in order to facilitate convergence; the interaction between age and Julian date; the interaction between sex and Julian date; the three way interaction between age, sex and Julian dates; a random intercept for individual; and a random intercept for year. The inclusion of year accounts for non-independence of observation within years, while individual accounted for the non-independence of repeated measurements made on the same individual (Kruuk and Hadfield 2007). We then fitted an animal model by adding a random intercept modelling variance associated with mother identity (Kruuk 2004), and a random intercept modelling additive genetic variance. Although it was not included in the best models, we kept inbreeding coefficient (estimated from the pedigree) as a covariate, because leaving it out could bias the later estimation of additive genetic variation (Boer and Hoeschele 1993). Nevertheless, animal models fitted without this covariate gave indistinguishable estimates.

Multivariate models: Univariate animal models can be expanded to multivariate models in order to estimate genetic correlations, genetic gradients and genetic differentials.

$$[m, l, t, \omega] \sim bX + D_1a + D_2m + D_3p + D_4y + Ir.$$

Here X , D_1 , D_2 , D_3 and D_4 are design matrices relating observations to the parameters to estimate, b is a matrix of fixed effects, a , m , p and y are random effects accounting for the variance associated with breeding value, mother, permanent environment and year, respectively. The fixed part of the model matches that used for each trait in univariate models.

The most important aspect of this model is that a , the matrix of breeding values, follows a multivariate normal distribution:

$$a \sim MVN(\mathbf{0}, A \otimes G) \quad (4.1)$$

where A is the relatedness matrix describing the relatedness among all individuals, and G is the G-matrix, containing the additive genetic variances and covariances among all traits.

$$G = \begin{pmatrix} \sigma_A^2(m) & \sigma_A(m, l) & \sigma_A(m, t) & \sigma_A(m, \omega) \\ \sigma_A(m, l) & \sigma_A^2(l) & \sigma_A(l, t) & \sigma_A(l, \omega) \\ \sigma_A(m, t) & \sigma_A(l, t) & \sigma_A^2(t) & \sigma_A(t, \omega) \\ \sigma_A(m, \omega) & \sigma_A(l, \omega) & \sigma_A(t, \omega) & \sigma_A^2(\omega) \end{pmatrix}. \quad (4.2)$$

For any trait z , $\sigma_A(z, \omega)$ is the genetic differential, that is, the predicted rate of evolutionary change according to Robertson's secondary theorem of natural selection, or Price equation applied to genetic variation (Robertson 1966; G. R. Price 1970; Morrissey, Parker, et al. 2012). The Price equation is generally presented as a prediction of evolutionary change over the next generation, but it has also been used as a description of change (Heywood 2005; S. A. Frank 2012; Coulson and Tuljapurkar 2008). We use this prediction retrospectively, as an estimation of the mean evolutionary change that has occurred during the study period, which makes the assumption that ω is a good measure of fitness, because when "real fitness" is used, the equation is a mathematical tautology, i.e. it is exact (S. A. Frank 2012). A deviation from this perfect fitness measure could come from random Mendelian segregation or systematic meiosis distortion. Our results were robust to the use of an annualized measure of fitness (annual reproductive success plus twice survival), and to standardizing fitness across all individuals, within years, within cohorts, and within sexes.

For two traits z and y , the genetic correlation is $\frac{\sigma_{A(z,y)}}{\sigma_A(z)\sigma_A(y)}$. The vector of selection differentials on the three traits (S) was estimated as the sum of the vectors of covariances between traits and ω in the variance-covariance matrices for a , p and r ; which was equivalent to the selection differential computed in the paragraph on selection above.

We excluded the among-year level covariance from the selection differential, because (i) covariation between mass and fitness at the level of year does not correspond to selection as it does not occur among individuals (ii) due to the standardization of relative fitness at the level of cohorts, the among year variance and covariances involving ω were effectively zero ($\sigma_Y^2(\omega) < 10^{-8}$). Let G' be a subset of G excluding the column and the row that contain ω . The vector of selection gradients on the three traits (β) was estimated as $(G' + P' + R')^{-1}S$, where P' and R' are the equivalent of G' for permanent environment effects and for residuals, respectively.

The prediction of the multivariate breeders equation is obtained by $\Delta\bar{Z}' = G'\beta$, while the multivariate breeders equation ignoring genetic correlations is obtained by multiplying the G' matrix by the identity matrix (Morrissey, Walling, et al. 2012): $\Delta\bar{Z}' = (G' \times I)\beta$.

To investigate the potential role of parent-offspring conflict, we estimated the genetic correlation between parental ARS and offspring mass using a bivariate animal model. For juvenile mass, we used predicted adult mass (i.e. age-corrected juvenile mass; see below). The model included sex, Julian dates and squared Julian dates as fixed effects for offspring mass, and only sex for ARS.

$$[m_O, ARS_P] \sim bX + D_1a + D_2y + Ir.$$

Test of genetic correlations: We used ASReml-R (Gilmour et al. 2014; Butler et al. 2009) to test the genetic correlation between mass in adults and in juveniles against 1 and 0, by considering them as two separate traits. We first ran an unconstrained model and then reran it with the genetic correlation parameter set to 0.99 (and not exactly to 1 because ASReml cannot invert matrices with perfect correlations), or 0 respectively. The fit of the unconstrained model was then compared to that of the two constrained models using a likelihood ratio test with one degree of freedom (Wilson et al. 2009).

Birth date and growth prediction: Using the Bayesian programming environment JAGS (Plummer 2003), we fitted a multivariate Bayesian model to mass measurements of all 613 juveniles with mass data, and to their overwinter survival. The model simultaneously estimated individual growth curves—that is onset of growth (although this is referred to as “birth date” hereafter, this actually is the projected time when mass was zero, i.e. at conception), individual growth rates and asymptotic masses of all juveniles—and the effect of asymptotic mass on overwinter survival. The model clustered juveniles from the same mother born in the same year into litters (see e.g. (Cornulier et al. 2009) for a similar approach), assuming a maximum of five litters per year and assuming that successive litters are at least 20 days apart (Janeau and Aulagnier 1997). Preliminary model selection, assuming no differences in asymptotic masses among individuals, selected a monomolecular growth model ($\Delta DIC > 80$) over Gompertz and logistic models, as defined in (English, Bateman, and Clutton-Brock 2012). The model accounted for measurement error in mass, assuming that the standard deviation of the errors was that observed in animals measured multiple times on the same day (2.05g). In order to estimate the overall viability selection on

asymptotic mass, we performed within the model a logistic regression of year-to-year survival on sex and asymptotic mass. In order to test for the effect of the length of snow free period on the selection on asymptotic mass, we reran the full model including time until the first snow fall and its interaction with asymptotic mass in the logistic regression. We use the estimates of these two models to predict the survival probability as a function of asymptotic mass for every year, or for groups of years, depending on the distribution of birth dates and on the timing of the first snow fall.

Three MCMC chains were run for 6,300,000 iterations, with a burnin of 300,000 and a thinning of 6,000. Convergence was assessed by visual examination of the traces, and by checking that the $\hat{R} < 1.01$. Convergence was not achieved for the litter affiliations of 25 individuals as well as for one asymptotic mass, thus generating a bit more uncertainty in the estimations. The fit of the model was assessed using posterior predictive checks on the predictions of individual masses ($p=0.46$) and survival probabilities ($p=0.49$). The JAGS code for this model can be found at <https://github.com/timotheenivalis/SelRepSel>.

Acknowledgements

Thanks to Lauren Richardson, Tim Coulson, Alastair Wilson and two anonymous reviewers for constructive comments. Thanks to Wolf U. Blanckenhorn, Jarrod D. Hadfield, Lukas F. Keller, Marc Kéry, Hanna Kokko, Chelsea J. Little, Pirmin Nietlisbach, Barbara Tscharren and Ashley E. Latimer for comments and discussions on earlier versions of this work. Thanks also to the many field helpers. Weather data were provided by MeteoSwiss. The snow vole monitoring was authorised by the *Amt für Lebensmittelsicherheit und Tiergesundheit*, Chur, Switzerland. T.B. is funded by a Swiss National Science Foundation project grant (31003A_141110) awarded to EP.

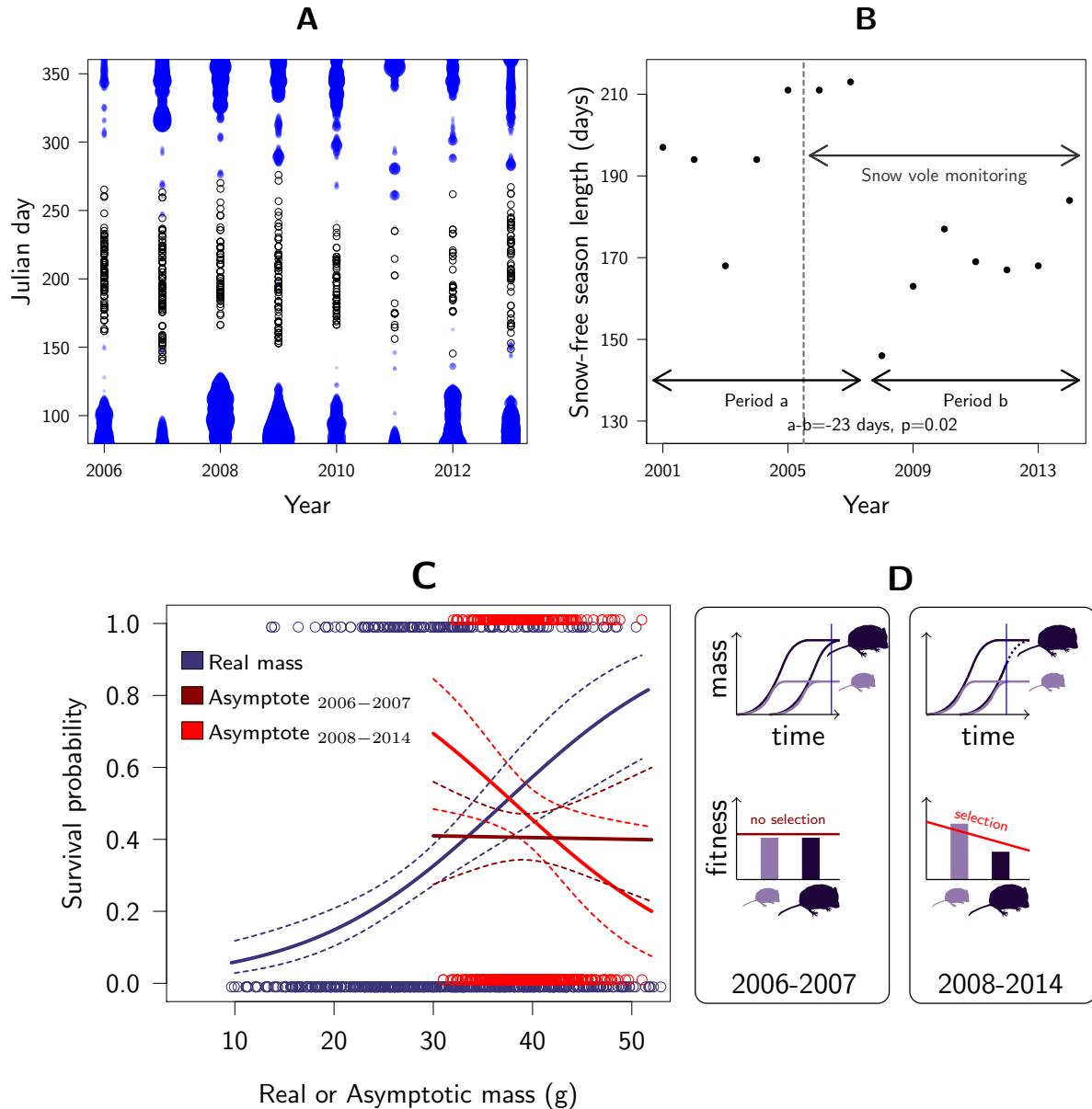


Figure 4.4: Snow-free season, timing of reproduction and selection for asymptotic mass. (A): Births (black dots) only occur during the snow free season (snow depth in blue), (B): which in 2008-2014 has been shorter than in the preceding 8 year. Therefore, (C) despite a positive phenotypic selection on body mass (blue), asymptotic mass was selectively neutral in 2006-2007 (brown), and was negatively selected in 2008-2014 (red), as a result of (D) the selective disappearance of heavy individuals that were born too close to the onset of winter (blue vertical line) during 2008-2014.

4.6 Supplementary information

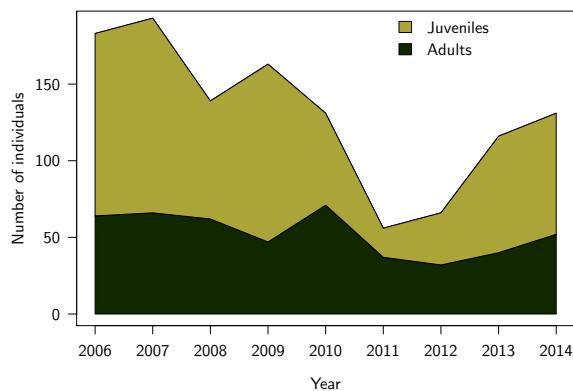


Figure 4.5: Temporal variation in population size and age-structure. Number of individual adults and of juveniles captured in each year.

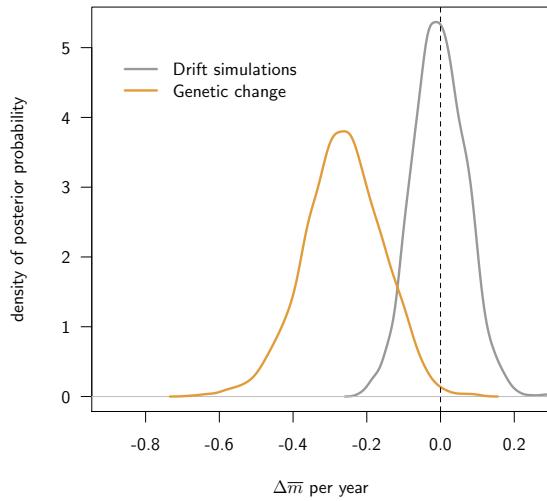


Figure 4.6: Estimation of the rate of genetic change and rate of change expected under drift. The posterior distributions of the realized rate of genetic change, estimated by the Price equation, exceeds that expected under genetic drift $p = 0.009$.

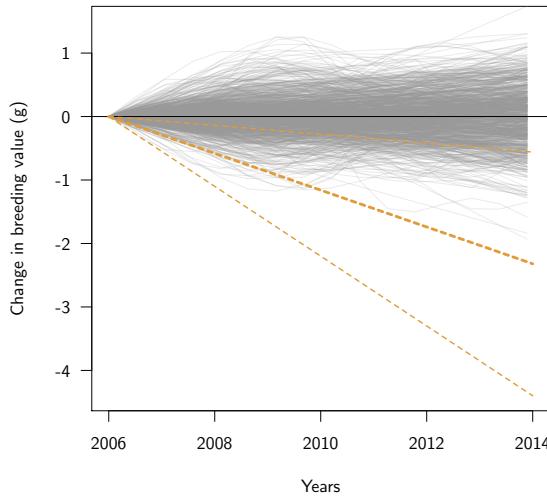


Figure 4.7: Estimation of the time-trajectory of genetic change and evolutionary trajectories simulated with drift only. Evolution of breeding values for mass are shown relative to the year 2006. The yellow lines show the mode and 95% credibility interval of the rate of evolution estimated by the Price equation within an animal model. The gray lines show 1,000 simulations of genetic drift, based on the real population pedigree and on the posterior distribution of genetic variance for mass estimated by the animal model. The probability that the observed rate of evolution happened due to drift is only 0.009, less than could be understood from the overlap between the two distributions. It is, however important to notice that the two distributions are not independent, but that small (/large) values of change due to drift are simulated for small (/large, respectively) posterior samples of estimate rate of evolution.

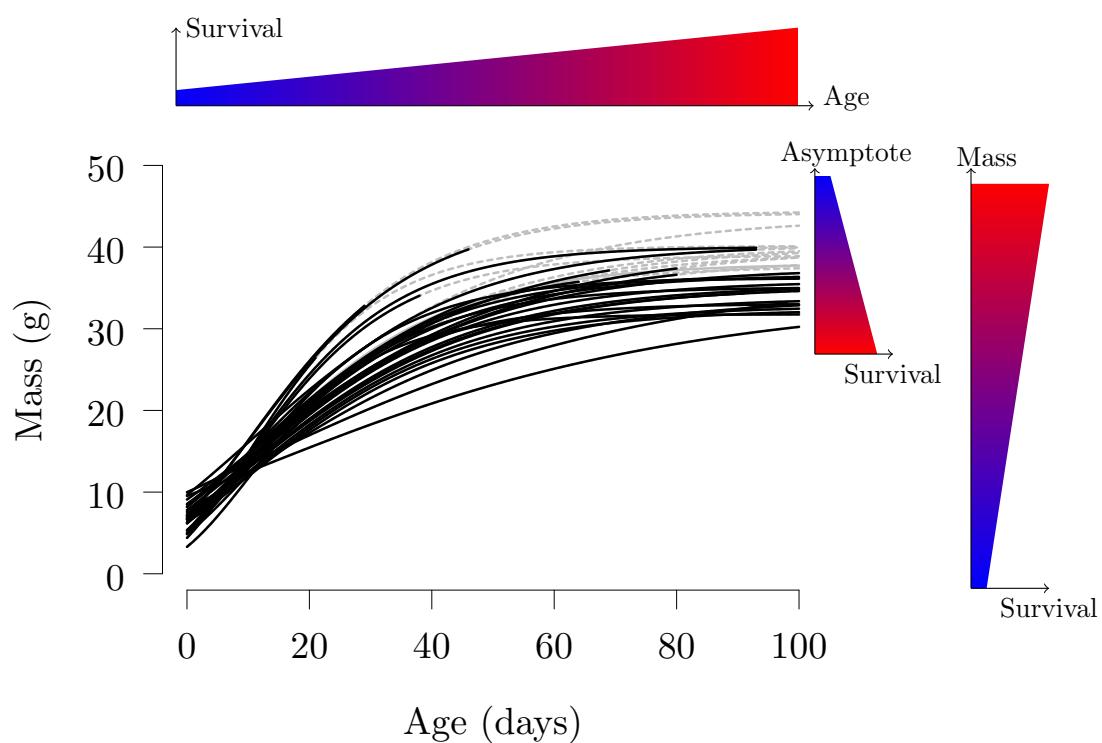


Figure 4.8: Conceptual illustration of the selection for asymptotic mass despite apparent selection for mass. Black lines represent simulated individual growth trajectories, and they are prolonged by grey dashed lines after an individual death. The probability of surviving between the time of measurement and the next year increases with age. Because mass increases with age, there is apparent selection favouring heavier individuals. However, it is still possible for viability selection at a given developmental stage, such as asymptotic mass, to be negative. Because genetic variation is related to asymptotic mass, but not to age, the expected genetic change will be toward lower masses.

Chapter 5

Fluctuating selection but no fluctuating evolution in a wild rodent population

All entities move and nothing remains still.

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

It is impossible to live in the past, difficult to live in the present and a waste to live in the future.

— Frank Herbert, *Dune* (1965)

Timothée Bonnet and Erik Postma. Submitted to Evolution.

5.1 Abstract

Temporal fluctuations in the strength and direction of selection are often suggested to slow down evolution, both over geological and contemporary time-scales. The prevalence of fluctuating selection and its relevance for evolutionary dynamics remains unclear however, especially on contemporary time scales: Unbiased empirical estimates of variation in selection are still scarce, and the question of how much variation in selection translates into variation in genetic change has been largely ignored. Using long-term individual-based data for a wild rodent population, we quantify the amount of fluctuating selection on body mass. Subsequently, we estimate the evolutionary dynamics of mass, and test for a link between fluctuating selection and evolution. We show that, over the past 10 years, phenotypic selection on body mass has fluctuated significantly. While this variation is largely the result of variation in fertility rather than viability selection, viability selection is the main driver of adaptive evolution in this system. Accordingly, we found that the strength and direction of genetic change remained stable over the study period. Thus, the rate of genetic change was similar in years where total selection favoured heavier or lighter individuals. These results demonstrate that, over shorter time-scales, fluctuating selection is not necessarily evolutionary relevant.

Introduction

Selection shapes biodiversity in time and space, explaining the general match between organisms and their environment (Darwin 1859; Endler 1986). Linking the sources of natural and sexual selection to the dynamics of genetic evolution has been a focus of evolutionary biology during the last century (e.g. R. Fisher 1958), but for most of the 20th century this goal has been hampered by the lack of an unified framework to quantify selection (Wade 2006). This changed with the development of regression-based methods to measure the strength and direction of selection (Lande 1979; Lande and Arnold 1983), which have enabled the estimation of selection gradients in a large variety of traits and biological systems (Kingsolver et al. 2001; Stinchcombe et al. 2008). This bonanza of estimates has shown that directional selection is stronger and more common than balancing selection, for both morphological and life-history traits (Kingsolver et al. 2001; Hereford, Hansen, and Houle 2004). At first sight, this pattern is contrary to expectations (Kingsolver and Diamond 2011): As most traits are heritable (Postma 2014), they are predicted to evolve towards their fitness optimum, with directional selection progressively being replaced by balancing selection. However, most traits evolve only very slowly and within a limited phenotypic range (Hendry and Kinnison 1999; Merilä, Sheldon, and Kruuk 2001; Brookfield 2016).

One explanation for this paradox is that fitness landscapes are not constant over time, and populations are evolving towards a continuously changing fitness optimum (Fisher and Ford 1947; Lande 1976). Whereas at any point in time directional selection may be strong, average selection gradients may be weaker, and if selection fluctuates not only in strength but also in direction, average selection may even be zero (Figure 5.1 (A-C)). Given that fluctuating selection may slow down evolutionary adaptation, or even bring it to a halt (Jones, Arnold, and Bürger 2004; Estes and Arnold 2007), it constitutes an appealing explanation for the commonly observed lack of evolutionary change, i.e. evolutionary stasis, as well as for the commonness of directional selection (Merilä, Sheldon, and Kruuk 2001; Robinson et al. 2008; Bell 2010). However, although fluctuating selection as an explanation for “macro-evolutionary” stasis is gaining theoretical and empirical support (Uyeda et al. 2011; Estes and Arnold 2007; Voje et al. 2015), our understanding of the importance of fluctuations in selection in shaping the evolutionary dynamics of natural populations on a much smaller time scale, e.g. from year one year to the next, is hampered by the lack of a clear answer to two questions: (i) Does phenotypic selection really fluctuate, in strength and/or direction, over short time scales? (ii) If it does, do these fluctuations translate into fluctuations, in speed and/or direction, of genetic change?

The first question seemingly received a positive answer with the publication of a synthetic analysis of temporal replicates of selection from 89 studies, which came to the conclusion that phenotypic selection does indeed vary and reverses its direction among years (Siepielski, Dibattista, and Carlson 2009). However, Morrissey and Hadfield 2012 showed that most of these fluctuations can be ascribed to sampling variation, and that when this is accounted for, directional selection is in fact remarkably constant over time, both in magnitude and direction.

Instead of estimating the variance of the distribution of temporal estimates of selection, as in (Siepielski, Dibattista, and Carlson 2009), tests for fluctuating selection

must estimate the variance of the temporal distribution of selection Morrissey and Hadfield 2012. As of yet, Chevin, Visser, and Tufto 2015 are among the few to have done this: Using a random regression approach, they found that phenotypic selection on laying date fluctuated over a short time period in a population of great tits (*Parus major* Linnaeus, 1758). The generality of this finding however needs to be confirmed across across a wider range of species, populations and traits, using the same robust approach.

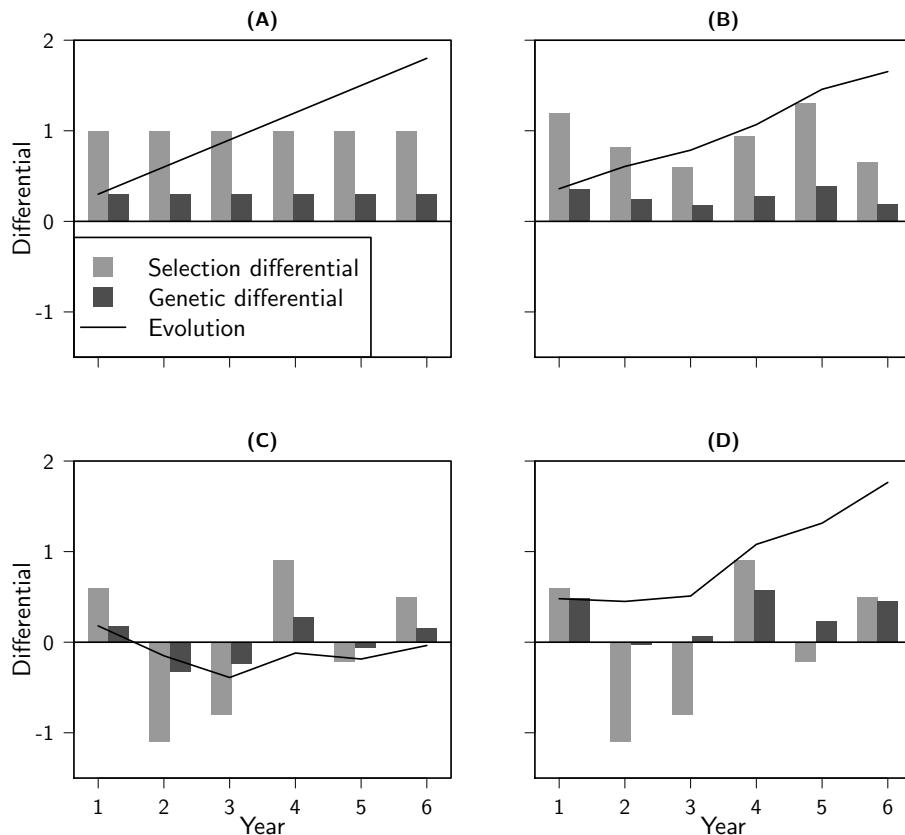


Figure 5.1: Evolutionary change under constant and fluctuating selection regimes. In (A), selection is constant among years. Following the breeder's equation, the genetic differential (i.e. the response to selection) is equal to the product of the selection differential and the heritability, which is here set to 0.3. The resultant cumulative response to selection, i.e. the evolutionary trajectory, is described by a straight line. In (B), selection fluctuates but does not reverse, and mean selection and the rate of evolution are only slightly reduced compared to (A). In (C), selection fluctuates and reverses, resulting in fluctuating and reversing evolution, and thereby evolutionary stasis over the time frame considered. In D, selection fluctuates and reverts as in C, but selection is partly non-causal and mediated by an unobserved environmental factor (i.e. a key assumption of the breeder's equation is violated). As a consequence, selection and evolution are uncoupled and despite fluctuating selection the rate of evolution is similar to (A).

In addition to showing statistically significant variation in selection, two more points must be investigated to assess the evolutionary relevance of fluctuating selection. First, the precise pattern of fluctuation matters. Even in the presence of fluctuating selection, evolution will only come to a halt if the direction of selection changes

regularly and the mean selection differential equals zero (Blanckenhorn 2000; Hunt et al. 2004; Morrissey and Hadfield 2012) (see Figure 5.1 (B)). Second, phenotypic selection, defined as a non-null phenotypic covariance between a trait and relative fitness, does not necessarily lead to an evolutionary response (Rausher 1992) (see Figure 5.1 (D)): Estimates of phenotypic selection provide a poor predictor of genetic change when the assumptions of the breeder's equation are violated, and in particular when selection is disproportionately dominated by an environmental covariance between the trait of interest and fitness (Price and Liou 1989; Rausher 1992; Bonnet and Postma 2016).

While this is one of the general explanations for apparent evolutionary stasis, it is particularly relevant within the context of fluctuating selection: As fluctuating selection is often thought to be driven by environmental fluctuations (Bell 2010; Chevin and Haller 2014), these may disproportionately shape the environmental component of selection. If fluctuating selection is not the result of fluctuations in the causal effects of the focal trait on fitness, it will not result in fluctuations in the additive genetic covariance between the trait and fitness, i.e. in fluctuating evolution (Robertson 1966; G. R. Price 1970; Morrissey, Parker, et al. 2012). Hence, fluctuating selection affects evolutionary dynamics only if the variation in the sign and strength of selection is, at least partly, coupled to variation in the sign and pace of genetic change.

Here we take advantage of the ten-year long monitoring of a population of snow voles (*Chionomys nivalis* Martins, 1842) to i) quantify fluctuating selection on body mass, ii) describe the temporal dynamics of evolution in mass, and iii) quantify the relationship between fluctuating selection and evolution. To this end, we first estimate directional selection on a year-to-year basis to quantify the variation in selection estimates. We then explicitly model these fluctuations of directional selection within a random regression framework in order to account for sampling variance. Based on the sign of annual selection estimates, as well as on the ratio of the variance in selection over the mean strength of selection, we also assess the probability of selection reversal. These analyses are performed for total selection, as well as for fertility and viability selection separately. Subsequently, we use a quantitative genetic framework to describe the general pattern of evolution over the study period, and estimate the rate of evolution of mass on a year-to-year basis. Finally, we combine analyses of selection and estimations of evolution to assess the coupling between variation in the strength and sign of selection and evolution.

Material and methods

Study population

Since 2006, a wild population of snow voles (*Chionomys nivalis* Martins, 1842) has been monitored intensively. This population is located in the Swiss Alps, near Chur (N $46^{\circ}48'$, E $9^{\circ}34'$; 2,030 m.a.s.l.). The study area consists of 5 ha of scree with sparse vegetation, surrounded by meadows, forest and a steep cliff. Because the snow vole shows an overwhelming preference for rocky environments (Janeau and Aulagnier 1997; Luque-larena, López, and Gosálbez 2002), the monitored population is ecologi-

cally fairly isolated. Every year, snow voles have been live-trapped during two to five trapping sessions, between mid-June and early October. To this end, the study area is overlaid with a 10×10 m grid consisting of a total of 559 cells with stable geographic coordinates. A trapping session consists of four trapping nights, necessary to cover all four quarters of the study area.

During each trapping session, a Longworth trap (catch-and-release trap, Penlon Ltd, Oxford, UK) filled with hay and baited with apple, hamster food and peanut butter is placed in every cell. Individuals captured for the first time are ear-clipped (2mm diameter, thumb type punch, Harvard Apparatus, Massachusetts, USA) and individually marked with a subcutaneous PIT tag (ISO transponder, Tierchip Dasmann, Tecklenburg, Germany). Ear-clips are preserved in 95% ethanol + 5% TE. For each capture, we record individual identity, geographic coordinates, body mass, body length, tail length, sex and age.

Ear clips are stored at -20°C until DNA extraction. All individuals are genotyped for 18 autosomal microsatellites, using snow vole-specific primers (Wandeler, Ravaioli, and Bucher 2008; García-Navas et al. 2015). In addition, the sex of all individuals is confirmed by sequencing the *Sry* locus (Gubbay et al. 1990). Finally, we sequence the mitochondrial control region, and all males are genotyped for one Y-linked microsatellite and three Y-linked insertion-deletions (Wandeler and Camenisch 2011). Based on the autosomal microsatellite genotypes, we reconstruct the pedigree of the population using the maximum likelihood based program COLONY (Wang 2004; Jones and Wang 2010) and a Bayesian R package MasterBayes (Hadfield, Richardson, and Burke 2006; R Core Team 2016). The consistency of the pedigree is then checked for consistency using the Y-linked markers and the mitochondrial haplotypes. This procedure allows the identification of most of the parental links (91%) as well as the identification of likely immigrants (individuals first captured as adults and with two unknown parents). This high-quality pedigree is used to define annual reproductive success and lifetime reproductive success, as well as to estimate the relatedness among all pairs of individuals.

A mark-recapture analysis has shown that between-session recapture probabilities are very high (adults: $92.4\% \pm 1.1$; juveniles: $81.1\% \pm 3.0$). Therefore, the year-to-year recapture probability is effectively 1, which means that the non-capture of an individual in a given year can be directly equated with death or permanent immigration without the need for mark-recapture modeling (García-Navas et al. 2015).

Fitness measures

We considered three measures of fitness: (i) survival from one year to the next, $\phi_{i,t}$, based on whether an individual i observed in year t is observed again in year $t+1$ ($\phi_{i,t} = 1$) or not ($\phi_{i,t}=0$); (ii) annual reproductive success, $\rho_{i,t}$, the number of juveniles born from i during the year t according to the inferred pedigree; (iii) an annualized measure of overall fitness $F_{i,t} = 2\phi_{i,t} + \rho_{i,t}$. Because animals present at the beginning but dying early on that year are less likely to reproduce, ρ does not perfectly isolate the contribution of fertility to overall fitness independently of viability. Nevertheless, ρ and ϕ are only little correlated at the individual level (Pearson's product-moment correlation -0.054 ; SE = 0.027), which suggests that these two statistics capture dif-

ferent aspect of fitness. Still, it is best to keep in mind that ρ contains a viability component, and any variation in selection estimated using ρ might partly capture variation in viability selection.

Measures of juvenile and adult body mass

As demonstrated in Bonnet and Postma 2016, the overall positive covariance between viability and mass is a consequence of both mass and survival probability increasing with age, and as variation in age is not heritable and cannot respond to selection, this phenotypic covariance has no evolutionary consequences (also see (van Noordwijk 1988; Rausher 1992)). Indeed, accounting for juvenile growth by projecting juvenile masses to the same age reveals viability selection favouring lighter juveniles (Bonnet and Postma 2016). As in (Bonnet and Postma 2016), we therefore correct juvenile mass for age by fitting individual growth curves based on juveniles mass measurements. Using the Bayesian programming environment JAGS Plummer 2003, we estimated for every individual a birth date, a growth rate and an asymptotic body mass. Preliminary model selection assuming no differences in asymptotic masses among individuals selected a monomolecular growth model ($\Delta\text{DIC} > 80$) over Gompertz and logistic models, as defined in English, Bateman, and Clutton-Brock 2012.

Short-term fluctuations and measurement error in mass were accounted for by assuming that the standard deviation of the deviations between "real" and observed mass was the standard deviation observed in animals measured multiple times on the same day (2.05g). Convergence was assessed by visual examination of the MCMC traces, and by checking that the $\hat{R} < 1.01$. Only for one individual asymptotic mass convergence was not achieved.

This approach provided a single, age-corrected, measure of mass per juvenile that can easily be correlated to the measures of annual fitness (for which we also have a single measure per individual). The correlation between the estimated asymptotic size and the observed adult size of the juveniles surviving to become adult was 19.9%. This correlation is relatively low and partly illustrates the inaccuracy of the estimation, but is also lowered by the fact that adult mass continues to vary throughout an individual's life (i.e. the repeatability of adult mass is not 1).

In adults, within-year variation in mass is much smaller, but in both sexes mass tends to increase in early summer and decrease in late summer, with a mean predicted amplitude of about 1 g. To account for this, we modelled adult mass measurements as a function of a simple and a quadratic effect of Julian date. We used the mean, per individual and per year, of the residuals, as adult mass for a given year. Not doing this correction, and using the averaged non-corrected mass instead, did not change the result in any noticeable way.

In order to obtain standardized selection gradients, we standardized our corrected mass measurements across all years by subtracting their mean and dividing by their standard deviation.

Selection analysis

Selection was estimated with a series of generalized linear models (GLMs) and generalized linear mixed models (GLMMs), regressing fitness measures on mass. Mixed models were fitted in `lme4` and confidence intervals computed by likelihood profiling (Bates et al. 2015).

Using the annualized measure of overall fitness, $F_{i,t}$, we first estimated selection on a year-by-year basis using a quasi-Poisson GLM with a log link, where the expected fitness of individual i at time t is predicted from :

$$\log(F_{i,t}) = \mu_{F,t} + \beta_{F,a,t}a_{i,t} + \beta_{F,s,t}s_i + (\beta_{F,z,t})z_{i,t}, \quad (5.1)$$

where $a_{i,t}$ is the age (adult or juvenile) of individual i at year t , s_i is the sex of i , $z_{i,t}$ is the mean mass of i at t , $\mu_{F,t}$ is the intercept of the regression, $\beta_{F,a,t}$ is the effect of age, $\beta_{F,s,t}$ is the effect of sex and $\beta_{F,z,t}$ is the strength of selection on mass. Because we used a log link, $\beta_{F,z,t}$ is a selection gradient *sensu* Lande and Arnold 1983 (Smouse, Meagher, and Lobak 1999; Firth et al. 2015).

The variation in the yearly estimates of selection ($V(\hat{\beta}_{F,z,t})$) gives a first idea about the temporal dynamic of selection, but as it is the sum of real variation in selection and of sampling variance, it will always overestimate the real variation in selection (Morrissey and Hadfield 2012).

Second, we estimated selection across all years, by fitting a quasi-Poisson GLM to pooled data from all the years, without taking into account temporal variation:

$$\log(F_{i,t}) = \mu_F + \beta_{F,a}a_{i,t} + \beta_{F,s}s_i + \beta_{F,z}z_{i,t}. \quad (5.2)$$

Third, we directly estimated variation in selection by fitting a random regression to the full dataset. Thus, we modified model 5.2 to a quasi-Poisson GLMM by including a random intercept and a random slope of mass:

$$\log(F_{i,t}) = \mu_{F,t} + \beta_{F,a}a_{i,t} + \beta_{F,s}s_i + (\beta'_{F,z} + \zeta_{F,t})z_{i,t}, \quad (5.3)$$

where $\beta'_{F,z}$ is the median selection estimate, μ_t is the random deviation of the intercept in year t and $\zeta_{F,t}$ is the deviation of selection (i.e. the slope) in year t . Both μ_t and ζ_t are assumed to be normally distributed, but their covariation is not estimated.

$$\mu_{F,t} \sim \mathcal{N}(0, \sigma_{F,\mu}^2) \quad (5.4)$$

$$\zeta_{F,t} \sim \mathcal{N}(0, \sigma_{F,\zeta}^2). \quad (5.5)$$

The main parameter of interest in this equation is $\sigma_{F,\zeta}^2$, the temporal variation in selection excluding sampling variance (Chevin, Visser, and Tufto 2015). We tested for the statistical significance of $\sigma_{F,\zeta}^2$ using a likelihood ratio test (LRT) (see e.g. Pinheiro and Bates 2000; Crainiceanu and Ruppert 2004) comparing model 5.2 and 5.3. Because variance components cannot be negative, we assumed that the LRT statistic follows an even mixture of χ_1^2 and χ_0^2 (Self and Liang 1987), which in practice means that p -values from a χ_1^2 have to be divided by 2. The median selection estimate ($\beta'_{F,z}$) from model 5.3 differs from the selection estimate across all years ($\beta_{F,z}$) in model 5.2 if the estimate

of $\sigma_{F,\zeta}^2$ is different from 0 and data are not perfectly balanced among years. The latter, $\beta_{F,z}$, can be seen as the best estimation of the overall selection, while the former, $\beta'_{F,z}$, can be seen as the selection occurring in a "normal" year. The ratio $\beta'_{F,z}/\sigma_{F,\zeta}$ gives an idea of the likelihood of a reversal in the direction of selection. Indeed, assuming that the annual selection gradients follow a Gaussian distribution (as the random regression does), this ratio is similar to a Z-value. Values between -1 and 1 indicate frequent reversals (more than 32% of the time), and values above 2 or below -2 indicate that reversals are unlikely (less than 2.5% of the time).

We repeated these analyses for annual reproductive success (ρ), again using a quasi-Poisson GLMM, and for over winter survival (ϕ), using a logistic regression. Because only a few juveniles (9 out of 764) have been found to reproduce in their first year, for ρ we restricted our analyses to adults. For ϕ , we excluded the last year (2015) because we do not yet know who has survived the subsequent winter. For ρ , as for F , the estimates of strength of selection are directly selection gradients sensu (Lande and Arnold 1983) because we use a log link (Smouse, Meagher, and Lobak 1999; Firth et al. 2015). This is not the case for ϕ , but because there are no interactions involving z , the sign and strength of estimates of selection are still interpretable qualitatively. The main parameters of interest, the variances in the slope of selection, are written $\sigma_{\phi,\zeta}^2$ and $\sigma_{\rho,\zeta}^2$, for viability and fertility, respectively.

Inference of evolution and the contribution of fluctuating selection

We estimated all quantitative genetic parameters by fitting animal models (C. Henderson 1950; C. R. Henderson 1975; C. Henderson 1976) using the R package `MCMCglmm` (Hadfield 2010). This Bayesian package allows to extract and combine full posterior distributions of parameters, and unless stated otherwise, all calculations were done on the posterior distributions (rather than on point estimates) in order to propagate uncertainty and account for covariation between parameters. For all models, we run a MCMC chain long enough to obtain 1,000 posterior samples, with a thinning interval large enough so that the autocorrelation of any parameter was below 10%, and added a burnin of about 20% of the total iterations.

Because additive genetic variation in fitness is a prerequisite for a response to selection, we first estimated the heritability of our fitness proxy F , using a univariate animal model assuming a Poisson distribution with a log link. The model included an intercept, age, sex and their interaction as fixed effects, and additive genetic effects, individual identity (i.e. permanent environment effects), mother identity and year as random effects. Heritability was estimated after transformation from the latent scale to the data scale, by integrating over all the random effects and fixed effects (Morrissey 2015; de Villemereuil et al. 2016), using the R package `QGglmm` (de Villemereuil et al. 2016).

We then used two approaches to infer the yearly rates of evolution in body mass: 1) a univariate approach based on BLUPs regression (C. Henderson 1950; Hadfield 2012) and 2) a multivariate approach based on the Robertson-Price identity (G. R. Price 1970; Morrissey, Parker, et al. 2012; Bonnet and Postma 2016).

For the first approach, we fitted a univariate animal model on body mass data, in-

cluding age and sex as fixed effects, and a random additive genetic, permanent environment (i.e. individual identity), maternal (maternal identity) and year effect. For every two successive years, we computed the genetic change in mass between the two sets of living individuals using best linear unbiased predictors (BLUPs) for breeding values (following Hadfield 2012). We simulated genetic drift down the pedigree of the snow vole population (following Hadfield et al. 2010, and using the function `rbv()` in `MCMCglmm`), and computed the range of genetic change between every two successive years that genetic drift can produce. We also visualized the temporal dynamics of genetic evolution of mass, by fitting a time spline (i.e. a smoother) to the breeding values of all individuals alive in each year. The spline was fitted using a generalized additive model in the R package `mgcv` (Wood 2011). We estimated a time spline for each posterior sample of the distributions of individual breeding values, in order to obtain the posterior distribution of evolution. To quantify the coupling of variation in selection and variation in evolution, we computed the correlations between the yearly estimates of selection gradients and the change in breeding value to the next year. We used the posterior distribution of changes in breeding values, but only the point estimate of annual selection gradients, to obtain a posterior distribution of correlations.

For the second approach, we would have ideally estimated the genetic and environmental selection gradients for every year by fitting a multivariate animal model with mass in each year considered as a different trait. However, although we did initially fit such a model, because of data limitations it did not reach convergence and the priors dominated the posterior distribution. Instead we therefore split the data in two groups of years: those where our estimate of selection gradient (as estimated above) was positive, and those where it was negative. We considered mass in these two groups of years as two different traits (M_+ and M_- , respectively). We fitted a trivariate animal model to both body mass traits and our annualized measure of fitness (F). This model allows the estimation of an additive genetic covariance between mass and fitness for the two year classes. Based on the Robertson-Price equation, these covariances provide a direct and unbiased expectation of the rate of evolution during the two groups of years (Robertson 1966; G. R. Price 1970; G. Price 1972; S. A. Frank 2012; Morrissey, Parker, et al. 2012). By measuring fitness on a yearly basis we remove the assumption of non-overlapping generations. We compare and explain the advantages and drawbacks of the two approaches in the discussion.

The trivariate animal model can be written as

$$[M_+, M_-, F] \sim bX + Z_1a + Z_2m + Z_3p + Z_4y + Ir,$$

where X , Z_1 , Z_2 , Z_3 and Z_4 are design matrices relating mass and fitness observations to the parameters to estimate, b is a matrix of fixed effects, a , m , p and y are random effects accounting for the variance associated with additive genetic, maternal, permanent environment and year effects, respectively. Residuals r are assumed to be normally distributed and independent, and are therefore associated to observations by an identity matrix I . The fixed part of the model matches that used for each trait in univariate models (see above).

The matrix of breeding values \mathbf{a} follows a multivariate normal distribution

$$\mathbf{a} \sim MVN(\mathbf{0}, \mathbf{R} \otimes \mathbf{A})$$

where \mathbf{R} is the relatedness matrix between all individuals, and \mathbf{A} is the additive genetic variance covariance matrix between the three traits.

$$\mathbf{A} = \begin{pmatrix} \sigma_A^2(M_+) & \sigma_A(M_+M_-) & \sigma_A(M_+F) \\ \sigma_A(M_+M_-) & \sigma_A^2(M_-) & \sigma_A(M_-F) \\ \sigma_A(M_+F) & \sigma_A(M_-F) & \sigma_A^2(F) \end{pmatrix},$$

where $\sigma_A^2(M_+)$ and $\sigma_A^2(M_-)$ is the additive genetic variation for mass in years with positive selection and negative selection respectively, $\sigma_A(M_+M_-)$ is the covariance additive genetic in mass between the two group of years, $\sigma_A^2(F)$ is the additive genetic variation in fitness across years, which is the genetic differential of fitness itself (R. Fisher 1958), and finally, $\sigma_A(M_+F)$ and $\sigma_A(M_-F)$ is the additive genetic covariation between fitness and mass in years with high selection, and low selection, respectively. We computed the genetic gradients for both groups of years as $\beta_{A+} = \sigma_A(M_+F)/\sigma_A^2(M_+)$ and $\beta_{A-} = \sigma_A(M_-F)/\sigma_A^2(M_-)$. The additive genetic correlation between mass on the two groups of years was computed as $\sigma_A(M_+M_-)/\sigma_A(M_+)\sigma_A(M_-)$.

Environmental selection differentials $\sigma_E(M_+F)$ and $\sigma_E(M_-F)$ were obtained from the sum of the covariances between mass and fitness in the random effect variance-covariance matrices for permanent environment, maternal identity and residuals. The environmental variances $\sigma_E^2(M_+F)$ and $\sigma_E^2(M_-F)$ were obtained by summing the variance components of the same random effects. The environmental selection gradients were then obtained as $\beta_{E+} = \sigma_E(M_+F)/\sigma_E^2(M_+)$ and $\beta_{E-} = \sigma_E(M_-F)/\sigma_E^2(M_-)$.

Finally, the phenotypic selection gradients were recovered as $(\sigma_A(M_+F) + \sigma_E(M_+F))/(\sigma_A^2(M_+) + \sigma_E^2(M_+))$ and $(\sigma_A(M_-F) + \sigma_E(M_-F))/(\sigma_A^2(M_-) + \sigma_E^2(M_-))$.

Results

Yearly estimates of selection

Yearly estimates of selection gradients showed considerable variation (standard deviation=0.167) around the mean selection estimate for all years pooled together ($0.082 \pm SE 0.028$; Figure 5.2 (A)). Estimates of total selection were mostly positive, but appeared to have reversed in three years. Although the standard deviation of the yearly estimates was greater than the overall selection gradient, a large portion of this variation must be attributable to sampling error. Indeed, yearly selection was estimated with much less precision than overall selection, and the mean standard error of the yearly estimates was 0.097. Fertility and viability selection gradients showed similar patterns and were either positive or close to zero, except for two years (Figure 5.2 (B-C)). The standard deviations of the estimates of viability and fertility selection were high, but so were the mean standard errors of these estimates (table 5.1)

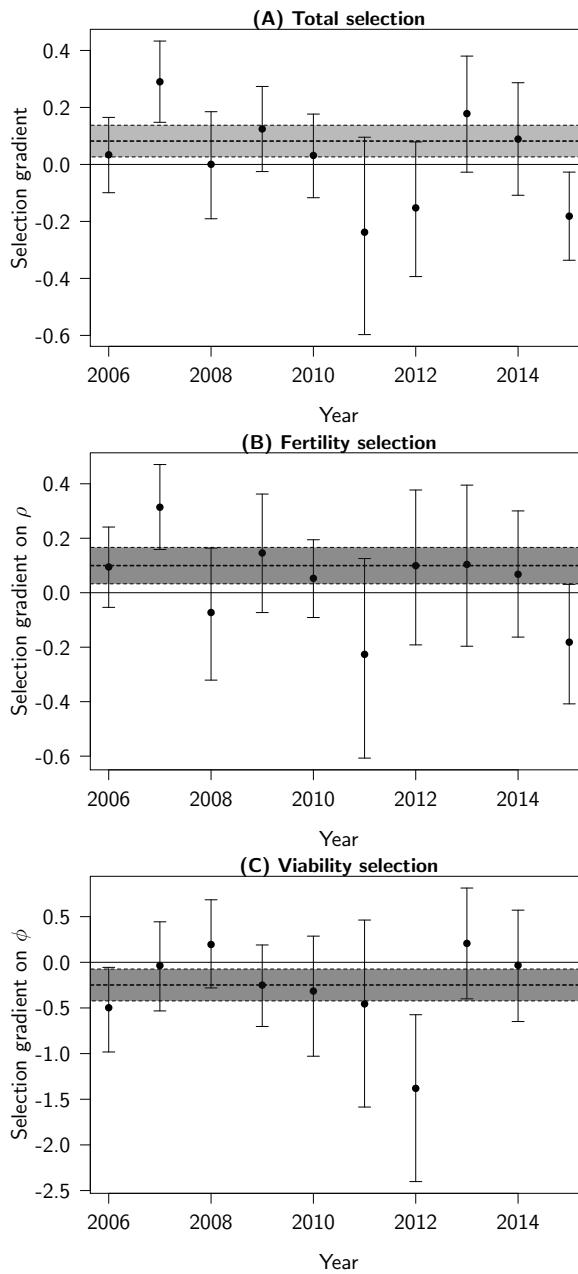


Figure 5.2: Estimates of overall, viability and fertility selection gradients, year-by-year and across all years. (A) Selection was estimated as the slope of absolute annual fitness (ARS + twice survival) on body mass, on the transformed scale of a Poisson GLM. (B) Selection was estimated as the slope of survival on body mass, on the transformed scale of a binomial GLM. (C) Selection was estimated as the slope of annual reproductive success on body mass, on the transformed scale of a Poisson GLM. Year-by-year estimates (black dots with 95%CI error bars) were obtained by fitting separate GLMs for each year. The overall estimate (dashed line with 95%CI as a grey polygon) was produced by pooling all years together.

Table 5.1: Selection and temporal variation in selection for total selection (F), fertility selection (ρ) and viability selection (ϕ).

Selection	β_z (SE)	SD _{year}	SE _{year}	β'_z (SE)	σ_ζ 95%CI	p($\sigma_\zeta > 0$)	$ \beta'_z / \sigma_\zeta$
Total	0.082 (0.028)	0.167	0.097	0.036 (0.044)	0.117 [0.063;0.218]	$8 \cdot 10^{-6}$	0.309
Fertility	0.1 (0.034)	0.160	0.117	0.052 (0.044)	0.111 [0.053;0.212]	$3 \cdot 10^{-4}$	0.466
Viability	-0.248 (0.089)	0.484	0.319	-0.217 (0.098)	0.109 [0;0.425]	0.36	1.998

Notes: β_z is the selection gradient across all years, as estimated from a Generalized Linear Model (GLM), given with its standard error (SE); SD_{year} is the standard deviation of selection gradients estimated from year-specific GLMs; SE_{year} is the mean standard error on those year-by-year estimates; β'_z is the selection gradient on an average year, estimated from a random regression Generalized Linear Mixed Model (GLMM), given with its standard error; σ_ζ is the standard deviation in selection, estimated from this random regression GLMM, given with 95% confidence interval computed by likelihood profiling; p($\sigma_\zeta > 0$) is the p -value from a likelihood ratio test for the significance of σ_ζ ; $|\beta'_z| / \sigma_\zeta$ is the ratio of the absolute median year selection over the standard deviation in selection, and indicates the likelihood of absence of reversal in the direction of selection.

Statistical significance of variation in selection

Fitting equation 5.3, we estimated $\sigma_{F,\zeta} = 0.117$ (95%CI [0.063;0.218]). Allowing for annual variation in the selection gradient significantly improved the fit of the model ($\Delta\text{log-likelihood} = 9.3$, one-sided $\chi^2 = 18.59$, df=1, $p=8 \cdot 10^{-6}$). Fitting a non-zero covariation between the random intercept and the random slope did not change the likelihood of the model ($\Delta\text{log-likelihood} = 1.3$, two-sided $\chi^2 = 2.69$, df=1, $p=0.10$). Given $|\beta'_z| / \sigma_\zeta = 0.309$, the reversal of selection is very likely (table 5.1).

Variation in fertility selection was estimated as $\sigma_{\rho,\zeta} = 0.111$ (95%CI [0.053;0.212]), which is larger than the median selection gradient ($\beta_{\rho,z} = 0.052$ SE= 0.044). Allowing for fluctuating fertility selection improved significantly the fit of the model to the data ($\Delta\text{log-likelihood} = 6.1$, one-sided $\chi^2 = 12.13$, df=1, $p=3 \cdot 10^{-4}$). There however was little support for fluctuation in viability selection (table 5.1): Variance in viability selection was not significantly different from zero ($\sigma_{\phi,\zeta} = 0.109$; 95%CI [0;0.425]), accounting for variation in viability selection did not significantly improve the fit of the model ($\Delta\text{log-likelihood} = 0.07$, one-sided $\chi^2 = 0.13$, df=1, $p=0.36$), and the reversal of viability selection was unlikely.

Fluctuation of evolution

There was a small but significant amount of additive genetic variation in our proxy of annual fitness: On the latent scale of the Poisson model, the additive genetic variation was estimated to be 0.028 [0.001;0.082]. On the scale of the data, this translates into an additive genetic variation of 0.052 [0.001;0.105] and a heritability of fitness of 1.18% [0.03%;3.17%]. This is comparable to the heritability of life-time fitness in Bonnet and Postma 2016, which used a lifetime rather than annual measure of fitness.

We found significant additive genetic variation in age-corrected mass (1.99 g^2 [0.91;2.68]; heritability = 17% [10%;25%]). As already shown in Bonnet and Postma 2016, the evolutionary trend from 2006 to 2014 was toward smaller breeding values for mass (Fig. 5.3). There is however some visual indication of a stabilization and

possibly a reversal of evolution in the last two years.

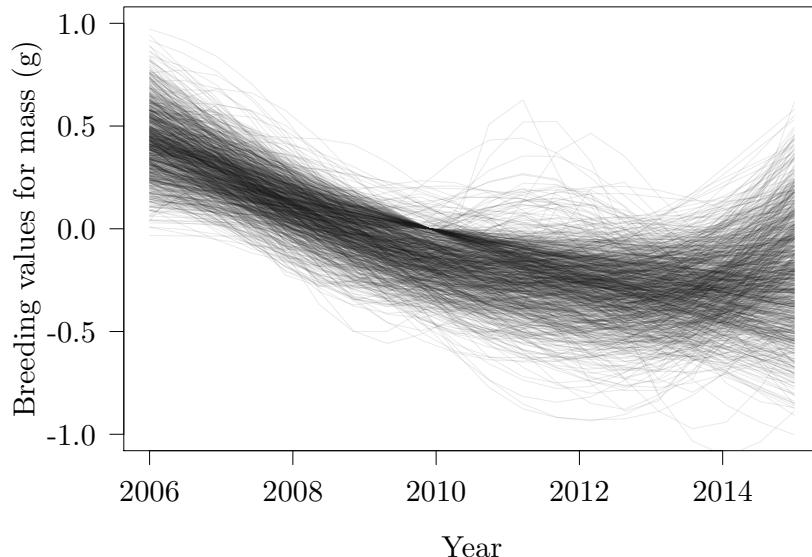


Figure 5.3: Temporal dynamics of mean breeding value for mass. Each line was obtained from a different MCMC posterior sample, by fitting a time-spline to the mean of estimated breeding values among individuals alive in any given year.

The same pattern emerges when looking at the posterior distribution of change in breeding values between any two successive years (Fig. 5.4). None of the year-to-year changes are statistically different from zero (Fig. 5.4), but because they are largely negative or null, they sum up to a strongly negative and statistically significant trend from 2006 to 2014 (Bonnet and Postma 2016). Similarly, although none of the observed changes are stronger than what could be expected due to drift alone (Fig. 5.4), drift cannot explain the cumulative change (Bonnet and Postma 2016).

From selection to evolution

As discussed above, the correlation between selection gradients and change in breeding values from one year to the next is estimated with a lot of uncertainty and is not statistically significantly different from zero. Nevertheless, the most likely value was positive (mode 0.36, 95%CI [-0.39; 0.64]).

As expected, in years with positive selection (based on selection gradients from year-by-year GLMs, see above), the selection gradient reconstructed from our trivariate animal model was positive, while it was negative for years with negative selection gradients (fig. 5.5). Importantly however, the genetic gradients were negative in both groups of years (fig. 5.5) and did not differ from each other ($\beta_{A+} - \beta_{A-} = -0.004$, 95%CI[-0.080;0.076], $p_{MCMC} = 0.91$). On the other hand, the environmental gradients differed from each other ($\beta_{E+} - \beta_{E-} = 0.075$, 95%CI[0.038;0.137], $p_{MCMC} < 0.001$), with β_{E+} being significantly positive, and β_{E-} slightly negative. Moreover, during years of positive selection, the genetic and environmental gradients were of opposite sign (fig. 5.5), and significantly different ($\beta_{A+} - \beta_{E+} = -0.123$, 95%CI[-0.218;-0.028], $p_{MCMC} = 0.006$). On the other hand, during years of negative selection,

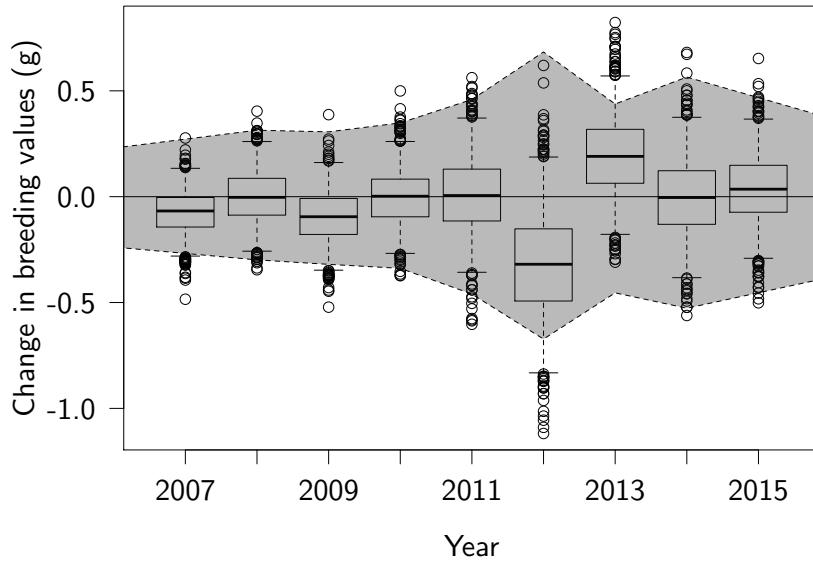


Figure 5.4: Posterior distribution of the change in breeding values for mass relative to the mean in the previous year, and the range of change possible due to genetic drift. For each posterior sample of the estimated breeding values, we computed the difference between the mean breeding value of individuals alive in year t and individuals alive in year $t - 1$. Box plots show the median, the first and third quartiles, quantiles 2.5% and 97.5% and outliers. The grey envelope shows the 95% interval of year-to-year evolution simulated with drift only.

the genetic and environmental gradients were both negative (fig. 5.5), and not significantly different ($\beta_{A-} - \beta_{E-} = -0.010$, 95%CI[-0.138;-0.058], $p_{MCMC} = 0.424$). Finally, the genetic correlation between mass in positive selection years and mass in negative selection years was strong and positive (0.82, 95%CI [0.46; 0.93]).

Discussion

Here we have shown that selection on body mass fluctuates in a natural population of snow voles. These fluctuations originate from changes in fertility rather than viability selection. In addition, we have shown that body mass has been genetically evolving in this population, but that both the rate and direction of evolution were relatively constant. As a consequence, changes in the direction of phenotypic selection did not result in concordant changes in the direction of evolution.

Below, we discuss the methodological challenges posed by the quantification of the variation in selection and evolution, and highlight our contributions to their resolution. We then clarify why the signs of selection and evolution can be different, even though the temporal variability in selection and in evolution are likely to be correlated. We also discuss what our analyzes can tell about the mechanisms of fluctuating selection, and what is needed to answer the questions that are outside the reach of our analyses. Finally, we discuss the importance of the timescale when studying variation in selection and evolution.

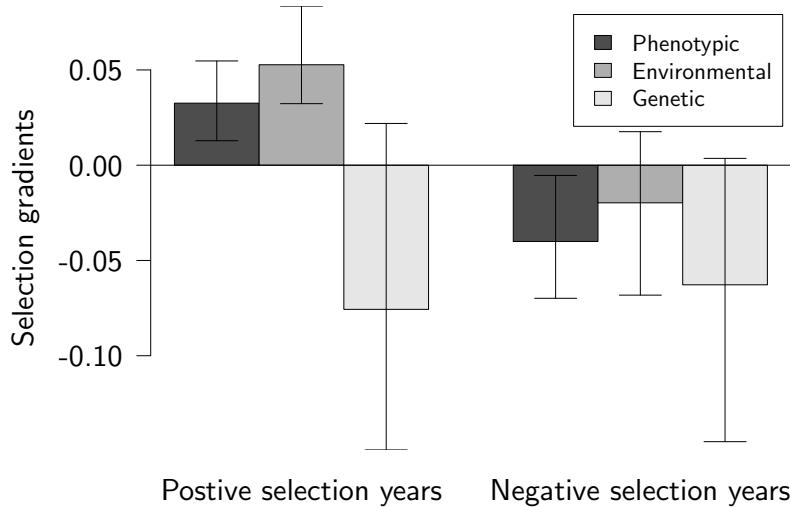


Figure 5.5: Phenotypic selection gradients and their decomposition into environmental and genetic gradients for years with positive selection on mass and for years with negative selection on mass. Error bars show 95% confidence intervals.

The modeling of evolution and selection

The random regression method, proposed by Morrissey and Hadfield 2012 and developed further by Chevin, Visser, and Tufto 2015, provides a statistically rigorous way to quantify and test for the significance of variation in selection (Chevin, Visser, and Tufto 2015). On its own, however, a random regression does not address the evolutionary relevance of fluctuating selection. To establish the latter, two additional issues need to be investigated: (i) Variation in the strength of selection will reverse the direction of evolution only if it fluctuates not only in strength, but also in direction (see Figure 5.1B and C); (ii) Selection does not always lead to a genetic response to selection (Rausher 1992; Morrissey, Kruuk, and Wilson 2010; Merilä, Sheldon, and Kruuk 2001), and fluctuating selection might have no influence on evolutionary dynamics (see Figure 5.1D).

To address the first issue, we considered where the distribution of selection gradients, estimated by a random regression, is located relative to zero. If this distribution is centered around zero, selection reversal is frequent, while if the distribution does not overlap much with zero, selection reversal is rare. We evaluated the likelihood of selection reversal by calculating the ratio of the absolute median selection gradient over the standard deviation of selection gradients ($|\beta_z'| / \sigma_{\zeta}$). Reversal becomes less likely as this ratio increases. As only 10 years are included in our analysis, this ratio may not exactly comply with a Z-distribution and we are therefore reluctant to translate it into a probability. Nevertheless, it gives a qualitative assessments of the likelihood of reversal and could be developed further into a more quantitatively rigorous measure.

To address the second issue, we estimated the coupling between variation in selection and variation in genetic change. While we were able to satisfactorily show that selection and evolution are uncoupled, the exercise proved to be challenging. In a first

approach, we computed the correlation between selection and year-to-year changes in breeding values by relating the full distribution of the change in BLUPs for breeding values to point estimates of selection gradients. Therefore, the uncertainty accompanying the selection estimates was not propagated to this correlation. This in contrast to the trivariate animal model, which estimates selection and evolution within the same model. Consequently, the latter approach allows the integration of uncertainty in both selection and evolution when comparing genetic and environmental gradients, and to take into account the non-independence of their posterior distributions. Unfortunately however, this multivariate approach is particularly data-hungry, and the snow vole population is too small to estimate year-specific genetic parameters. As a consequence, we were forced to compare groups of years with negative and positive selection. Nevertheless, whenever the population size allows for it, we advocate the use of year-specific multivariate animal models for the investigation of the question at hand.

Uncoupling of selection and evolution

We found that both evolution and viability selection did not fluctuate and were always negative, whereas fertility selection fluctuated significantly between positive and negative values (table 5.1, figure 5.3). This is in line with our previous finding that directional evolution towards lower body mass is driven by viability rather than fertility selection (Bonnet and Postma 2016). Together this implies that evolution and total selection are partly uncoupled in this system, and in particular that their signs do not match.

Nevertheless, evolution is not completely independent of selection: While we do not observe a significant correlation between selection and evolution among years—probably because of strong genetic drift—the most likely value is positive (see Results). Simple algebra shows that a positive correlation is indeed expected. For a trait z , a selection gradient is the ratio of the phenotypic covariance between trait and relative fitness, over the phenotypic variance in the trait:

$$\beta_P = \frac{\sigma_P(z, F)}{\sigma_P^2(z)}.$$

Assuming a standard quantitative genetic model in which there is no correlation or interaction between the genetic effects and the environmental effects, z can be decomposed into additive genetic effects and environmental effects $z = a + e$. Because all covariances between additive genetic effects and environmental effects are null, from this it follows that the phenotypic covariance ($\sigma_P(z, F)$, i.e. the selection differential) can be decomposed into an additive genetic ($\sigma_A(z, F)$) and an environmental covariance ($\sigma_E(z, F)$). Therefore, the phenotypic selection gradient (β_P) can be written as:

$$\beta_P = \frac{\sigma_A(z, F) + \sigma_E(z, F)}{\sigma_P^2(z)}.$$

According to the Robertson-Price identity (Robertson 1966; G. R. Price 1970), $\sigma_A(z, F)$

is the expected rate of genetic change. From the above it follows that the phenotypic selection gradient is likely to be positively correlated with evolution (provided the latter is non-zero). Therefore, although their signs are opposite in all years, years with more positive selection gradients go with less negative genetic change, and vice versa.

What drives fluctuations in selection?

Although our random regression and quantitative genetic models give a thorough description of the dynamics of selection and evolution in this population, they do not provide direct insight into the mechanisms underlying these dynamics. We have shown that selection fluctuates, and thus that the relationship between mass and fitness changes at the population level, but why does it change? Different processes may lead to the same distribution of directional selection gradients, and based on the analysis of selection gradients alone it is difficult to distinguish fluctuations due to a moving fitness optimum from those due to a change in the distribution of phenotypes among years (Chevin and Haller 2014). The latter may play a role here as we find substantial variation between years in both the mean phenotype (ranging between 38.6 g and 40.6 g) and its standard deviation (ranging between 3.1 g and 4.4 g).

Indirectly we can nevertheless gain some insights into the ecological drivers of variation in selection. We have found that it is fertility rather than viability selection that is fluctuating (table 5.1), and that fertility selection is independent of genetic variation for mass (Bonnet and Postma 2016). If the environmental variance in fitness is interpreted as the environmental variance in individual quality (see Wilson and Nussey 2010, for a discussion of individual quality in an evolutionary context), the environmental covariation between fertility and body mass can be considered to capture variation in 'body condition'. In agreement with this interpretation, in the snow voles the year-specific environmental covariation is either positive or null (see Fig. 5.5), but never significantly negative. Although this is simply a reformulation, it reveals that fluctuating selection in this system is driven by variation in the condition-dependence of body mass. For example, a favorable territory—for instance with high food availability and low parasitic prevalence—simultaneously increases mass and reproductive success, and thus body condition, but the strength of this covariation will depend on, for example, the mean and variance in habitat quality within a given year.

If we are to gain a deeper understanding of the dynamics of the fitness landscape and the ecological drivers of selection, we ultimately need to move beyond the estimation of variance parameters, toward a more mechanistic understanding of the genetic and ecological sources of phenotypic variation and their covariance with fitness (Morrissey and Hadfield 2012). However, good examples where we know the ecological driver of variation in selection are scarce. Some notable exceptions are beak size in Darwin finches (Grant and Grant 2002) and reproductive timing in great tits (Husby, Visser, and Kruuk 2011). Both of these, as well as the present study, rely on individual-based long-term monitoring, difficult and costly to upkeep, but necessary to disentangle the causes and consequences of selection in natural populations (Clutton-brock and Sheldon 2010).

Timescale

Despite fluctuations in the strength and direction of phenotypic selection, the rate and direction of evolution was constant over the course of the study period. Thereby our findings are at odds with the idea that fluctuating selection causes short-term evolutionary stasis. Nevertheless, fluctuating selection may be a driver of short-term evolutionary dynamics in other natural populations, where the selection measured by regression-based methods is causal and not dominated by an environmental covariation between traits and fitness. Moreover, it is unlikely that fluctuating selection will not be evolutionary relevant on some longer time scales, in the snow vole population and in other populations. Indeed, over geological time scales, bounded fluctuations of phenotypic evolution are increasingly attributed to responses to fluctuating selection, rather than to sampling variation and evolutionary stasis (Uyeda et al. 2011; Voje et al. 2015). Unless the environment is perfectly constant, causal selective pressures are likely to change over longer time periods, either because the fitness landscape changes, or because the phenotypic distribution changes through evolutionary adaptation or phenotypic plasticity.

Fluctuating selection and evolution might go undetected because the time frame is too short. For instance in the snow vole population, adaptive evolution and the causal selective pressure causing it are probably related to a short-term climatic anomaly which is likely to be reversed by global climate change (Bonnet and Postma 2016). On the other hand, fluctuating selection and evolution might also be missed because the temporal resolution at which selection is estimated is too low, smoothing out very short-term changes in selection and the rate of genetic change. This would also be a possibility in the snow vole population where the causal selective pressure varies within a year: selection is null early in the reproductive season and increases throughout summer (Bonnet and Postma 2016).

Conclusion

While our results do not argue against the evolutionary relevance of fluctuating selection in general, they warn against interpreting any phenotypic fluctuating selection in terms of fluctuating evolution: as the dynamics of selection and evolution can be uncoupled on certain time scales, fluctuating selection does not necessarily provide a general explanation for evolutionary stasis. Thereby we have highlighted the danger of relying on temporally replicated phenotypic estimates of selection to understand and predict the evolutionary dynamics of natural populations. Instead, quantifying the evolutionary relevance of fluctuating selection requires a joined analysis of selection and evolution.

Acknowledgments

Thanks to Jarrod D. Hadfield, Lukas F. Keller, Marc Kéry and Pirmin Nietlisbach for comments on earlier versions of this work. Thanks to the many field helpers over 10 years. Weather data were provided by MeteoSwiss. The snow vole monitoring was

Abstract

authorised by the *Amt für Lebensmittelsicherheit und Tiergesundheit*, Chur, Switzerland. T.B. is funded by a Swiss National Science Foundation project grant (31003A_141110) awarded to EP.

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)

Le mal qui est dans le monde vient presque toujours de l'ignorance, et les bonnes intentions peuvent faire du mal autant que la malveillance si elles n'ont pas la compréhension. / The evil that is in the world always comes of ignorance, and good intentions may do as much harm as malevolence, if they lack understanding.

— Albert Camus, *La Peste* (1947)

6.1 Overview

In this thesis, I investigated the causes and consequences of variation in fitness in a wild snow vole population. I showed that in this population, variation in proxies of individual fitness is not purely stochastic, but is underlain by variation in latent fitness (chapter 2). Furthermore, this variation in latent fitness has an additive genetic component, revealing the existence of natural selection and adaptive evolution (chapter 4 and 5). I also explored ways to decompose the causes of phenotypic changes and identified the quantitative genetic animal model as a appropriate tool to estimate evolution (chapter 3). Using this tool in various ways, I showed that body mass is 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 genetic variation for fitness (chapters 4 and 5), and body mass therefore evolved over the study period.

Below, I will further discuss both the insights brought by this thesis and the remaining challenges when it comes to understanding the causes of phenotypic variation and how wild populations respond to environmental change.

6.2 The causes of phenotypic variation

This thesis has brought some new insights into the causes of phenotypic variation in general, and fitness variation in particular, but there still is much more to learn. Below, I comment on two promising directions that were only touched upon in this thesis, but that have the potential to improve the predictive understanding of the causes and

consequences of fitness variation. These are the effects an individual's genes may have on other individuals, and the study of the molecular basis of genetic variation using genomic methods.

6.2.1 The effect of the genes of others

Using quantitative genetic methods, I decomposed the phenotypic variation of morphological and life-history traits into components related to additive genetic effects, maternal effects and permanent environments. This decomposition was sufficient to measure the rate of evolution of the direct genetic effects of an individual's genes (as summarised by its breeding value) (chapter 4), that is, the direct action of an individual's genes on itself. However, an individual's genes may have effects reaching out beyond itself, to the environment, and thus other individuals (Dawkins 1982). This may happen through interactions between individuals (indirect genetic effects, e.g. maternal effects, McAdam, Garant, and Wilson 2014), or through the pleiotropic action of genes expressed in kin at different life-stages (e.g. genetic conflict, Trivers 1974).

Indirect genetic effects could play an important role in shaping selection and evolution in the snow vole population. Indeed, in the snow voles, genes within an individual are likely to affect the phenotype of another individual during at least two types of situations. First, related females tend to form clusters of territories, and the presence of kin could suppress reproduction in subordinate females (García-Navas, Bonnet, Waldvogel, Camenisch, et al. 2016). Moreover, as in all placental mammals, maternal effects on offspring phenotypes are prevalent from pregnancy to weaning. Maternal effects have been studied extensively in natural populations (Wolf and Wade 2009), but estimates of the genetic component of maternal effects remain scarce (McAdam, Garant, and Wilson 2014). Nevertheless, genetic maternal effects may add to any direct genetic variation and thereby increase evolutionary potential (McGlothlin and Galloway 2014; McAdam, Garant, and Wilson 2014; Mcfarlane et al. 2015). In the snow vole population, preliminary analyses have shown the presence of additive genetic maternal effects for body mass (results not shown), and these could be subject to selection and evolve adaptively. Note that in chapter 4, maternal genetic effects were not explicitly modelled, and their evolution was instead assigned to phenotypic plasticity. A full account of body mass evolution should explicitly capture this evolution in addition to that of direct additive genetic effects.

In addition to indirect genetic effects, the effect of others' genes matters for evolution in the presence of genetic conflicts, that is, genetic trade-offs between traits expressed in different individuals. For four decades, genetic conflicts between parents and offspring have been thought to be a major constraint on the evolution of size (since Trivers 1974), but the idea resisted empirical tests despite behavioural studies showing patterns consistent with it (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 earwigs (*Forficula auricularia*, Linnaeus 1758). Moreover, Rollinson and Rowe 2015 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 ge-

netic conflict constrains the evolution of body mass, and found qualitative evidence that this 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 we 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, whereas the genetic size-number trade-off is an explanation for evolutionary stasis of body size or mass, mass is evolving in the snow vole population (chapter 4). This makes it more difficult to formulate an expectation for the genetic covariance between mass and litter size in the presence of such a trade-off. Finally, although in theory it is possible to measure the genetic trade-off using quantitative genetics, the exact model that would need to be fitted has as of yet not been described and the modelling tools that are required are currently unavailable (Hadfield 2012; Rollinson and Rowe 2015). This means that at the moment 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

On several occasions during this PhD, we considered using high-throughput sequencing (van Dijk et al. 2014) to retrospectively obtain high-density marker data for the snow vole population (tissue samples are kept in -80° freezers for most of the individuals trapped in the last ten years). As of yet we did not obtain the funding necessary for this, and time-limitations prevented further lab work and the development of a bioinformatic pipeline. As I discussed in chapter 1, molecular approaches to measuring selection and evolution are in general inferior to quantitative genetic approaches. Nevertheless, individual-based genomic data could bring complementary insights to my empirical chapters.

To start, individual-based genomic data could (marginally) improve quantitative genetic parameter estimates (Bérénos et al. 2014) by: (i) allowing for the use of *realized* relatedness in quantitative genetic animal models, instead of the relatedness *expected* from the pedigree; and (ii) providing some relatedness information about individuals with unknown parents (for which there is no information at all in a pedigree). More importantly, individual-based genomic data would allow the identification of some of the genetic loci underlying phenotypic variation and quantitative evolution. This task is generally a challenging one in small populations (Wellenreuther and Hansson 2016), but the snow vole population presents three rare advantages that would ease it considerably.

First, at least one trait, body mass, has been evolving during the last decade, and some adaptive molecular evolution must have happened. The search for the molecular basis of evolution would therefore start with the knowledge that there is something to find, and with indications on what functional types of genes are likely to be involved. Second, because in natural populations with complex demographics and mating patterns, there is in general no null-expectation for the effect of drift. This makes it difficult to show that evolution at a genetic locus is due to selection and not only due to drift. A pedigree provides such a null expectation, however. Simulating

the random dropping of alleles down our pedigree would result in a null distribution of changes in allele frequencies against which to test for the effect of selection on each genetic locus. This method was successfully employed to show contemporary adaptive evolution at 67 genetic loci in a wild population of Florida scrub-jays (Nancy Chen, Evolution conference, 2016, Austin, USA). Third, thanks to the availability of life-history data, it would be possible to correlate the allelic variation of the evolving loci to success and failure in various life-stages. Therefore, the combination of genomic and life-history data can pinpoint when selection occurs in life, and what kind of molecular mechanism selection acts on. Altogether, individual-based genomic data could therefore refine not only our molecular understanding of phenotypic variation, but also provide clues regarding the ecological nature of selection.

6.3 Predicting responses to environmental change

Anthropogenic environmental change has stimulated research aiming at understanding and predicting the response of natural populations to environmental change (Parmesan 2006; Chevin, Collins, and Lefèvre 2012; Smallegange and Coulson 2013; Charmantier and Gienapp 2014), but massive challenges continue to hinder this research agenda. At the moment, it is already the retrospective study of phenotypic and demographic responses that remains inconclusive (Merilä, Sheldon, and Kruuk 2001; Mc Carty 2001; Charmantier and Gienapp 2014; Brookfield 2016), and in most cases prospective prediction therefore seems to be well out of reach. During my Ph.D., I confronted three challenges that must be tackled to improve the predictive abilities of evolutionary ecology. Below, I discuss the problems with measuring selection, predicting the response to selection, and integrating evolutionary and demographic responses.

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 consequently biologists have attempted to understand its causes and mechanisms. More recently, the study of selection assumed a more applied goal as researchers hope to predict the response of natural populations to the selective pressures imposed by environmental change (Chevin, Lande, and Mace 2010; Coulson, Tuljapurkar, and Childs 2010; Merilä and Hendry 2014). The principle of natural selection appears so simple: In a given environment, individuals with a phenotype that favours survival and fertility contribute more to the next generation. Given the level of research attention devoted to such a simple process, it may at first sight seem surprising that our understanding of natural selection has developed so slowly, and that its study remains so difficult. 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 covariance-based methods (Robertson 1966; G. R. 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 measure the proportional effect of selection per unit of phenotypic variation, and allows decomposing 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). This has revealed several general patterns. For instance, directional selection is stronger and more common than suggested by early evolutionists, whereas stabilizing selection appears to be rare. Also, 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-based methods faces at least three difficulties that might severely hamper their significance and explain the general absence of genetic responses 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. Since, fitness is rarely observable directly, fitness proxies must be used instead. Hence, 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: a certain phenotype might increase survival but decrease fertility, so that the net selection on the trait is null, despite covariation with both fitness components (Thompson et al. 2011; Kingsolver et al. 2012; Brookfield 2016). Fortunately, this bias in general appears to be minor, 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 (Hadfield 2008). Still, estimates of evolution using the Price equation (that is, selection on the genotype) or using the trend in BLUPs for breeding values (that is, not using any information about selection nor fitness) agree qualitatively with my corrected estimates of selection (chapter 4 and 5), suggesting that my proxies for fitness are 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). 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 indirect selection. 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, using covariance- and regression-based methods to estimate phenotypic selection essentially measure the statistical association between traits and relative fitness. Selection is however a causal association, be it direct or indirect. If the statistical association is entirely mediated by an environmental covariance between traits and fitness, there is no selection and no possibility of a 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 parasite-free—is likely to lead to large mass, high survival, and high fertility. Accordingly, phenotypic estimates of natural selection on mass and size are overwhelmingly positive, but mass does not evolve as predicted from its heritability (Blanckenhorn 2000; Kingsolver et al. 2012). In the snow voles, an excess of environmental covariance does underlie the apparent selection on mass (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 not without its own limitations. First, manipulations are work-intensive, time-consuming and must be designed carefully in order to manipulate the trait of interest without affecting any other traits. Moreover, manipulations cannot easily be applied to all traits. While an experimental approach has been widely used to study selection on brood size, it is not clear to me how one could manipulate body mass in a controlled way (that is, without accidentally affecting other traits). Instead of an experiment, 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. This made understanding the mechanism of this selection and measuring its strength a matter of hypothesis testing. This approach could be used in other systems, but it requires sufficient phenotypic and relatedness data to fit bivariate animal models. In addition, although in the snow vole a single fitness component drove evolution, multiple fitness components could be involved, which would complicate 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: there is always a risk of false-positives and a correlation does not prove causation.

6.3.2 Evolutionary response

Once a measure of phenotypic selection and a heritability estimate is obtained, it is straightforward to formulate a prediction of genetic response based on the breeder's

equation (Lush 1937; Falconer and Mackay 1996). We have already seen (chapter 4 and 5) that such a prediction is often unreliable in natural populations, however. Estimates of selection might not correspond to causal selection, and unmeasured selection acting on genetically correlated traits might constrain evolution. I have shown that estimating the genetic component of selection, or the rate of evolution, can test whether selection has been measured appropriately to be predictive (chapter 4).

Nevertheless, most attempts to understand the evolutionary response to environmental change do not measure genetic parameters. Thus, the alarming lack of evidence for evolutionary responses to climate change probably originates primarily from a lack of tests for genetic change (Charmantier and Gienapp 2014; Gienapp and Brommer 2014; Merilä and Hendry 2014; Crozier and Hutchings 2014). Ignoring the genetic properties (e.g. the heritability) of the trait of interest (e.g. Forcada and Hoffmann 2014; Coulson and Clegg 2014; Traill, Schindler, and Coulson 2014) easily leads to underestimating, or incorrectly dismissing, the potential to respond to selection and the actual evolutionary response (Nietlisbach and Hadfield 2015; Chevin 2015; Pigeon et al. 2016). Similarly, the evolutionary potential of small populations was dismissed by population matrix simulations while ignoring genetic-based arguments (see chapter 2). Moreover, methods based on phenotypic covariances do not distinguish between the presence and the absence of heritable variation, and cannot be used alone to predict an evolutionary response (chapter 3).

Therefore, all the chapters of this thesis illustrate that a genetic approach, be it based on quantitative genetics or population genetics, is necessary to measure evolution, and can more reliably identify the selective causes and constraints shaping adaptation. Attempts to understand the evolutionary dynamics of natural populations based on phenotypic observations only (e.g. Smallegange and Coulson 2013) are a gamble that might pay off on special occasions, but is unlikely to be reliable in general.

6.3.3 Demographic response to environmental change

This thesis is almost exclusively concerned with traits and their evolutionary dynamics. In the context of understanding the response of natural population to environmental change, such an investigation is legitimate: Whether a trait distribution changes through demographic, plastic, or genetic mechanisms has different consequences on the fate of the population (Chevin and Lande 2010). Nevertheless, for most applications, and to the eyes of the society, it is unimportant whether animal and plant populations respond to climate change primarily through migration, through plastic changes, or through evolution. Instead, the primary motivation of the research on the response to environmental change is to ascertain whether populations will persist or go extinct, and how managers can affect the outcome.

The question of the persistence of a population is primarily a demographic one. The evolutionary approach that was mine during this PhD is not sufficient to ascertain the fate of the snow vole population, but it might be a useful first step. Indeed, it is now widely acknowledged that evolutionary processes can act on the same time scale as ecological ones, and that they can significantly affect demographics (Hairston et al. 2005; Ellner, Geber, and Hairston 2011; Chevin, Lande, and Mace 2010; Turcotte, Reznick, and Hare 2011). For instance, theory and laboratory experiments support the

existence of *evolutionary rescue*, that is, adaptive genetic change within a population that prevent the extinction of the population (Gonzalez et al. 2013; Schiffers et al. 2013). Still, empirical evidence for evolutionary rescue in the wild remains extremely limited (Vander Wal et al. 2013).

In chapter 4, I inferred that the genetic response to selection tended to increase mean juvenile survival over the study period. All other things being equal, evolution therefore had a positive demographic effect and contributed to the recovery of population size. As of yet, however, it is unclear how to quantify the demographic effect of evolution. To the best of my knowledge, an appropriate methodological framework is still lacking. Indeed, traditional demographic models used to predict population resilience ignore individual heterogeneity and genetic change (Kendall et al. 2011; Vindenes and Langangen 2015; Plard et al. 2016). On the other hand, quantitative genetic studies focus on estimating rates of evolutionary change, but mostly ignore their possible consequences for the dynamics of populations (Coulson, Tuljapurkar, and Childs 2010; Chevin, Collins, and Lefèvre 2012). It is now acknowledged that the integration of evolutionary and demographic aspects is crucial for predicting trait dynamics, population resilience and viability (Schoener 2011; Pelletier et al. 2012; Chevin, Collins, and Lefèvre 2012; Merilä and Hendry 2014). However, only in the last year have methods that could start to address this question in the wild been proposed (Vindenes and Langangen 2015; Coulson et al. 2015; Childs, Sheldon, and Rees 2016), and these should certainly be followed up.

6.4 General conclusion

Natural selection is a potent force that shapes the evolution of natural populations, but its causes and consequences can be blurred by the complexity of natural populations. Understanding the process of adaptation requires isolating selection from the stochasticity in fitness components, to disentangle evolution from other drivers of phenotypic change, and to mechanistically link genetic change to selective pressures. Being able to do so, thanks to an individual-based monitoring including genetic relatedness, I provided a rare example of contemporary adaptive evolution. More examples of evolution in action certainly await to be described, and will make our understanding of the response to environmental change more general and more predictive. This thesis shows in several ways that crucial insight is more likely to come from studies that explicitly study the genetic aspects of selection and phenotypic changes.

Bibliography

- 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.
- Agrawal, A. F. and Stinchcombe, J. R. 2009. 'How much do genetic covariances alter the rate of adaptation?' *Proceedings of the Royal Society of London B: Biological Sciences* **276** (1659): 1183–91.
- 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.
- Albert, A. and Anderson, J. 1984. 'On the existence of maximum likelihood estimates in logistic regression models.' *Biometrika* **71** (1): 1–10.
- Alroy, J. 1998. 'Cope's Rule and the Dynamics of Body Mass Evolution in North American Fossil Mammals.' *Science* **280** (5364): 731–734.
- 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.
- 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.
- Baker, J., Meade, A., Pagel, M., and Venditti, C. 2015. 'Adaptive evolution toward larger size in mammals.' *Proceedings of the National Academy of Sciences, USA* **112** (16): 5093–5098.
- Barfield, M., Holt, R. D., and Gomulkiewicz, R. 2011. 'Evolution in stage-structured populations.' *The American Naturalist* **177** (4): 397–409.
- 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.
- 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.

Bibliography

- Bell, G. 2010. 'Fluctuating selection: the perpetual renewal of adaptation in variable environments.' *Philosophical Transactions of the Royal Society B: Biological Sciences* **365** (1537): 87–97.
- Beniston, M. 2012. 'Is snow in the Alps receding or disappearing?' *Wiley Interdiscip Rev Clim Change* **3** (4): 349–358.
- Bérénos, C., Ellis, P. a., Pilkington, J. G., and Pemberton, J. M. 2014. 'Estimating quantitative genetic parameters in wild populations: a comparison of pedigree and genomic approaches.' *Molecular Ecology* **23**:3434–3451.
- Blanckenhorn, W. 2000. 'The evolution of body size: what keeps organisms small?' *Quarterly Review of Biology* **75** (4): 385–407.
- 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.
- Boer, I. de and Hoeschele, I. 1993. 'Genetic evaluation methods for populations with dominance and inbreeding.' *Theoretical and Applied Genetics* **86**:245–258.
- Bonduriansky, R. 2012. 'Rethinking heredity, again.' *Trends in Ecology & Evolution* **27** (6): 330–6.
- Bonnet, T. and Postma, E. 2016. 'Successful by chance? the power of mixed models and neutral simulations for the detection of individual fixed heterogeneity in fitness components.' *The American Naturalist* **187** (1): 60–74.
- Brandon, R. and Beatty, J. 1984. 'The propensity interpretation of 'fitness'. No interpretation is no substitute.' *Philosophy of Science* **51** (2): 342–347.
- 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.
- 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.
- Butler, D. G., Cullis, B. R., Gilmour, A. R., and Gogel, B. J. 2009. *ASReml-R reference manual*.
- 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.
- 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**.
- 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. 1994. 'Models of age-structured populations.' In *Evolution in Age-Structured Populations*, 2nd, 28–30. Cambridge: University of Cambridge Press.
- . 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.
- Charmantier, A. and Gienapp, P. 2014. 'Climate change and timing of avian breeding and migration: evolutionary versus plastic changes.' *Evolutionary Applications* **7** (1): 15–28.
- Charmentier, A., Garant, D., and Kruuk, L. E. B. 2014. *Quantitative Genetics in the Wild*. Oxford: Oxford University Press.
- Chevin, L. M. and Haller, B. C. 2014. 'The temporal distribution of directional gradients under selection for an optimum.' *Evolution* **68** (12): 3381–3394.
- 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.
- Chevin, L.-M., Collins, S., and Lefèvre, F. 2012. 'Phenotypic plasticity and evolutionary demographic responses to climate change: taking theory out to the field.' Edited by A. Hoffmann. *Functional Ecology* **27** (4): 967–979.
- Chevin, L.-M. and Lande, R. 2010. 'When do adaptive plasticity and genetic evolution prevent extinction of a density-regulated population?' *Evolution* **64** (4): 1143–1150.

Bibliography

- 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.
- Chevin, L.-M., Visser, M. E., and Tufto, J. 2015. 'Estimating the variation, autocorrelation, and environmental sensitivity of phenotypic selection.' *Evolution* **69** (9): 2319–2332.
- 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. 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.
- Cooke, F., Taylor, P. D., M., F. C., and Rockwell, R. F. 1990. 'Directional selection and clutch size in birds.' *The American Naturalist* **136**:261–267.
- Cope, E. D. 1887. *The origin of the fittest : essays on evolution*. 532. [Http://www.biodiversitylibrary.org/bibliography/24160](http://www.biodiversitylibrary.org/bibliography/24160). New York :D. Appleton, <http://www.biodiversitylibrary.org/item/65400>.
- Cornulier, T., Elston, D. a., Arcese, P., Benton, T. G., Douglas, D. J. T., Lambin, X., Reid, J., Robinson, R. a., and Sutherland, W. J. 2009. 'Estimating the annual number of breeding attempts from breeding dates using mixture models.' *Ecology Letters* **12** (11): 1184–93.
- Coulson, T. and Clegg, S. 2014. 'Population biology: Fur seals signal their own decline.' *Nature* **511**:414–415.
- 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.

- 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.
- Crozier, L. G. and Hutchings, J. A. 2014. 'Plastic and evolutionary responses to climate change in fish.' *Evolutionary Applications* **7** (1): 68–87.
- 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.
- Dawkins, R. 1982. *The extended phenotype*. Oxford University Press, Oxford, U.K.
- de Villemereuil, P., Schielzeth, H., Nakagawa, S., and Morrissey, M. 2016. 'General methods for evolutionary quantitative genetic inference from generalised mixed models.' *bioRxiv*. eprint: <http://dx.doi.org/10.1101/026377>.
- 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.
- 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.
- Ellegren, H. and Sheldon, B. C. 2008. 'Genetic basis of fitness differences in natural populations.' *Nature* **452**:169–175.
- 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.
- Endler, J. A. 1986. *Natural selection in the wild*. Princeton, NJ: Princeton University Press.
- 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.
- English, S., Bateman, A. W., and Clutton-Brock, T. H. 2012. 'Lifetime growth in wild meerkats: incorporating life history and environmental factors into a standard growth model.' *Oecologia* **169** (1): 143–53.
- Estes, S. and Arnold, S. J. 2007. 'Resolving the paradox of stasis: models with stabilizing selection explain evolutionary divergence on all timescales.' *The American Naturalist* **169** (2): 227–244.
- 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.

Bibliography

- 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.
- Falconer, D. S. and Mackay, T. F. C. 1996. *Introduction to quantitative genetics*. 4th ed. Edited by Longman. Essex, U.K.
- Firth, J. A., Hadfield, J. D., Santure, A. W., Slate, J., and Sheldon, B. C. 2015. 'The influence of nonrandom extra-pair paternity on heritability estimates derived from wild pedigrees.' *Evolution* **69** (5): 1336–1344.
- Fisher, R. A. 1941. 'Average excess and average effect of a gene substitution.' *Annals of Eugenics* **11** (1): 53–63.
- Fisher, R. a. and Ford, E. B. 1947. 'The spread of a gene in natural conditions in a colony of the moth Panaxia dominula L.' *Heredity* **1** (2): 143–174.
- Fisher, R. 1930. *The genetical theory of natural selection*. First. Oxford university press, Oxford, Great Britain.
- . 1958. *The genetical theory of natural selection*. Second. Dover Publications, New York.
- 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.
- Forcada, J. and Hoffman, J. 2014. 'Climate change selects for heterozygosity in a declining fur seal population.' *Nature* **511**:462–465.
- Fox, G. A. and Kendall, B. E. 2002. 'Demographic stochasticity and the variance reduction effect.' *Ecology* **83** (7): 1928–1934.
- Frank, S. A. 2012. 'Natural selection. IV. The Price equation.' *Journal of Evolutionary Biology* **25** (6): 1002–1019.
- 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.

- García-Navas, V., Bonnet, T., Waldvogel, D., Wandeler, P., Camenisch, G., and Postma, E. 2015. 'Gene flow counteracts the effect of drift in a Swiss population of snow voles fluctuating in size.' *Biological Conservation* **191**:168–177.
- 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.
- Gienapp, P. and Brommer, J. E. 2014. 'Evolutionary dynamics in response to climate change.' In *Quantitative genetics in the wild*, edited by A. Charmentier, D. Garant, and L. E. B. Kruuk, 254–273. Oxford: Oxford University Press, Oxford.
- Gilmour, A. R., Gogel, B. J., Cullis, B. R., Welham, S. J., and Thompson, R. 2014. *ASReml User Guide Release 3.0*. VSN International Ltd, Hemel Hempstead, UK.
- Gonzalez, A., Ronce, O., Ferriere, R., and Hochberg, M. E. 2013. 'Evolutionary rescue: an emerging focus at the intersection between ecology and evolution.' *Philosophical Transactions of the Royal Society B: Biological Sciences* **368** (1610): 20120404.
- 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.
- Gotanda, K. M., Correa, C., Turcotte, M. M., Rolshausen, G., and Hendry, A. P. 2015. 'Linking macro-trends and micro-rates: re-evaluating micro-evolutionary support for Cope's rule.' *Evolution* **69** (5): 1345–1354.
- Grant, P. R. and Grant, B. R. 2002. 'Unpredictable evolution in a 30-year study of Darwin's finches.' *Science* **296** (5568): 707–711.
- Gubbay, J., Collignon, J., Koopman, P., Capel, B., Economou, A., Munsterberg, A., Vivian, N., Goodfellow, P., and Lovell-Badge, R. 1990. 'A gene mapping to the sex-determining region of the mouse Y chromosome is a member of a novel family of embryonically expressed genes.' *Nature* **346** (6281): 245–250.
- 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. 2008. 'Estimating evolutionary parameters when viability selection is operating.' *Proceedings of the Royal Society of London B: Biological Sciences* **275** (1635): 723–734.
- . 2010. 'MCMC Methods for Multi-Response Generalized Linear Mixed Models: The MCMCglmm R Package.' *Journal of Statistical Software* **33** (2): 1–22.
- . 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., 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.

Bibliography

- 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.' *Molecular Ecology* **15** (12): 3715–3730.
- 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.
- Hansen, T. F., Pélabon, C., and Houle, D. 2011. 'Heritability is not evolvability.' *Evolutionary Biology* **38**:258–277.
- Haring, E., Herzig-Straschil, B., and Spitsenberger, F. 2000. 'Phylogenetic analysis of Alpine voles of the *Microtus multiplex* complex using the mitochondrial control region.' *Journal of Zoological Systematics and Evolutionary Research* **38**:231–238.
- Heim, N. A., Knope, M. L., Schaal, E. K., Wang, S. C., and Payne, J. L. 2014. 'Cope's rule in the evolution of marine animals.' *Science* **347** (6224): 867–870.
- Henderson, C. R. 1975. 'Best linear unbiased estimation and prediction under a selection model.' *Biometrics* **31**:423–447.
- Henderson, C. R. 1984. *Applications of Linear Models in Animal Breeding*. Guelph: University of Guelph Press.
- 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.
- Hendry, A. P. and Kinnison, M. T. 1999. 'Perspective: the pace of modern life: measuring rates of contemporary microevolution.' *Evolution* **53** (6): 1637–1653.
- Hereford, J., Hansen, T., and Houle, D. 2004. 'Comparing strengths of directional selection: how strong is strong?' *Evolution* **58** (10): 2133–2143.
- 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.
- Hof, A. E. van't, Edmonds, N., Dalíková, M., Marec, F., and Saccheri, I. J. 2011. 'Industrial melanism in British peppered moths Has a singular and recent mutational origin.' *Science* **332** (6032): 958–960.
- Hosmer, D. W., Lemeshow, S., and Sturdivant, R. X. 2013. *Applied logistic regression, third edition*. Hoboken, New Jersey: Wiley.
- 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>.
- Hunt, J., Bussière, L. F., Jennions, M. D., and Brooks, R. 2004. 'What is genetic quality?' *Trends in Ecology and Evolution* **19** (6): 329–333.
- Husby, A., Visser, M. E., and Kruuk, L. E. B. 2011. 'Speeding up microevolution: The effects of increasing temperature on selection and genetic variance in a wild bird population.' *PLoS Biology* **9** (2): e1000585.
- 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. Charmentier, D. Garant, and L. E. B. Kruuk. Oxford: Oxford University Press.
- Jones, A. G., Arnold, S. J., and Bürger, R. 2004. 'Evolution and stability of the g-matrix on a landscape with a moving optimum.' *Evolution* **58** (8): 1639–1654.
- Jones, J. S. 1987. 'The heritability of fitness: bad news for good genes?' *Trends in Ecology & Evolution* **2** (2): 35–36.
- Jones, O. R. and Wang, J. 2010. 'COLONY : a program for parentage and sibship inference from multilocus genotype data.' *Molecular Ecology Resources* **10**:551–555.
- 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.
- Karell, P., Ahola, K., Karstinen, T., Valkama, J., and Brommer, J. E. 2011. 'Climate change drives microevolution in a wild bird.' *Nature Communications* **2**:208.
- 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 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., Gibert, 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.' *Evolutionary Ecology* **26** (5): 1101–1118.

Bibliography

- Kirschner, M. W. and Gerhart, J. C. 2010. 'Facilitated variation.' In *Evoution. The extended synthesis*, edited by M. Pigliucci and G. B. Müller. The MIT press, Cambridge, Massachusetts, USA.
- 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.
- Kovach, R. P., Gharrett, A. J., and Tallmon, D. A. 2012. 'Genetic change for earlier migration timing in a pink salmon population.' *Proceedings of the Royal Society of London B: Biological Sciences* **279** (1743): 3870–3878.
- 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.
- Kruuk, L. E. B. and Hadfield, J. D. 2007. 'How to separate genetic and environmental causes of similarity between relatives.' *Journal of Evolutionary Biology* **20** (5): 1890–903.
- Lamichhaney, S., Han, F., Berglund, J., Wang, C., Almen, M. S., Webster, M. T., Grant, B. R., Grant, P. R., and Andersson, L. 2016. 'A beak size locus in Darwins finches facilitated character displacement during a drought.' *Science* **352** (6284): 470–474.
- Lande, R. and Arnold, S. J. 1983. 'The measurement of selection on correlated characters.' *Evolution* **37** (6): 1210–1226.
- Lande, R. 1976. 'Natural selection and random genetic drift in phenotypic evolution.' *Evolution* **30** (2): 314–334.
- . 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.
- 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.
- 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.

- 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.
- . 2014. *Evolution and Selection of Quantitative Traits*. Online version 30 December 2014. Sunderland, Massachusetts, USA: Sinauer.
- Mackay, T. 2001. 'The genetic architecture of quantitative traits.' *Annual Review of Genetics* **35**:303–309.
- Marshall, T., Slate, J., Kruuk, L. E., and Pemberton, J. M. 1998. 'Statistical confidence for likelihood-based paternity inference in natural populations.' *Molecular Ecology* **7**:639–655.
- 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.
- 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.
- Mc Carty, J. P. 2001. 'Ecological Consequences of Recent Climate Change.' *Conservation Biology* **15** (2): 320–331.
- McAdam, A. G., Garant, D., and Wilson, A. 2014. 'The effects of others' genes: maternal and other indirect genetic effects.' In *Quantitative Genetics in the Wild*, edited by A. Charmentier, D. Garant, and L. E. B. Kruuk. Oxford: Oxford University Press.
- 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., Kruuk, L. E., and Sheldon, B. C. 2001. 'Cryptic evolution in a wild bird population.' *Nature* **412**:76–79.

Bibliography

- Merilä, J., Sheldon, B. C., and Kruuk, L. E. B. 2001. 'Explaining stasis: microevolutionary studies in natural populations.' *Genetica* **112**:199–222.
- Merilä, J. and Hendry, A. P. 2014. 'Climate change, adaptation, and phenotypic plasticity: The problem and the evidence.' *Evolutionary Applications* **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.
- Milot, E., Mayer, F. M., Nussey, D. H., Boisvert, M., Pelletier, F., and Réale, D. 2011. 'Evidence for evolution in response to natural selection in a contemporary human population.' *Proceedings of the National Academy of Sciences, USA* **108** (41): 17040–17045.
- Mitchell-Olds, T. and Shaw, R. G. 1987. 'Regression analysis of natural selection: statistical inference and biological interpretation.' *Evolution* **41** (6): 1149–1161.
- Morris, D. 1998. 'State-dependent optimization of litter size.' *Oikos* **83**:518–528.
- 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.
- Morrissey, M. B. 2014. 'Selection and evolution of causally covarying traits.' *Evolution* **68** (6): 1748–61.
- Morrissey, M. B. 2015. 'Evolutionary quantitative genetics of nonlinear developmental systems.' *Evolution* **69** (8): 2050–2066.
- Morrissey, M. B. and Hadfield, J. D. 2012. 'Directional selection in temporally replicated studies is remarkably consistent.' *Evolution* **66** (2): 435–442.
- 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.
- Morrissey, M. B., Walling, C. a., Wilson, A. J., Pemberton, J. M., Clutton-Brock, T. H., and Kruuk, L. E. B. 2012. 'Genetic analysis of life-history constraint and evolution in a wild ungulate population.' *The American Naturalist* **179** (4): E97–E114.
- Mousseau, T. A. and Roff, D. A. 1987. 'Natural selection and the heritability of fitness components.' *Heredity* **59** (2): 181–197.
- 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.
- 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.
- 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).
- Owens, I. P. F. and Wilson, K. 1999. 'Immunocompetence: A neglected life history trait or conspicuous red herring?' *Trends in Ecology & Evolution* **14** (5): 170–172.
- 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.
- 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.
- 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.
- 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.
- Pelletier, F., Moyes, K., Clutton-Brock, T. H., and Coulson, T. 2012. 'Decomposing variation in population growth into contributions from environment and phenotypes in an age-structured population.' *Proceedings of the Royal Society of London B: Biological Sciences* **279** (1727): 394–401.
- Pigeon, G., Festa-Bianchet, M., Coltman, D. W., and Pelletier, F. 2016. 'Intense selective hunting leads to artificial evolution in horn size.' *Evolutionary Applications*: in press.
- Pigliucci, M. and Müller, G. B. 2010. *Evoution. The extended synthesis*. The MIT press, Cambridge, Massachusetts, USA.
- 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.

Bibliography

- 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.
- 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.
- Plummer, M. 2003. 'JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling.' In *Proceedings of the 3rd International Workshop on Distributed Statistical Computing*. Vienna, Austria.
- Poissant, J., Wilson, A. J., and Coltman, D. W. 2010. 'Sex-specific genetic variance and the evolution of sexual dimorphism : a systematic review of cross-sex genetic correlations.' *Evolution* **64** (1): 97–107.
- 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.
- 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., Visser, J., and Van Noordwijk, A. J. 2007. 'Strong artificial selection in the wild results in predicted small evolutionary change.' *Journal of Evolutionary Biology* **20**:1823–1832.
- 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.
- Price, G. 1972. 'Extension of covariance selection mathematics.' *Annals of Human Genetics* **35**:485–490.
- Price, P. W. 1996. *Biological evolution*. Saunders College Publishing, Philadelphia, PA.
- Price, T. and Liou, L. 1989. 'Selection on clutch size in birds.' *The American Naturalist* **134** (6): 950–959.
- R Core Team. 2014. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. <http://www.R-project.org/>.
- _____. 2016. *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.

- 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.
- Rees, M. and Ellner, S. P. 2009. 'Integral projection models for populations in temporally varying environments.' *Ecological Monographs* **79** (4): 575–594.
- Reznick, D. N. and Bryga, H. A. 1996. 'Life-history evolution in guppies (*Poecilia reticulata*: Poeciliidae).5. Genetic basis of parallelism in life histories.' *The American Naturalist* **147** (3): 339–359.
- Reznick, D. N., Shaw, F. H., Rodd, F. H., and Shaw, R. G. 1997. 'Evaluation of the Rate of Evolution in Natural Populations of Guppies (*Poecilia reticulata*).' *Science* **275**:1934–1937.
- Robertson, A. 1966. 'A mathematical model of the culling process in dairy cattle.' *Animal Production* **8**:95–108.
- Robinson, M. R., Pilkington, J. G., Clutton-Brock, T. H., Pemberton, J. M., and Kruuk, L. E. B. 2008. 'Environmental Heterogeneity Generates Fluctuating Selection on a Secondary Sexual Trait.' *Current Biology* **18** (10): 751–757.
- Roff, D. a. 2007. 'A centennial celebration for quantitative genetics.' *Evolution* **61** (5): 1017–1032.
- 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.
- 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.
- Rousset, F. 2004. *Genetic structure and selection in subdivided populations*. Princeton, New Jersey, USA.
- 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.

Bibliography

- 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.
- Schlüter, D., Price, T. D. P., and Rowe, L. 1991. 'Conflicting selection pressures and life history trade-offs.' *Proceedings of the Royal Society of London B: Biological Sciences* **246** (1315): 11–17.
- Schoener, T. W. 2011. 'The newest synthesis: understanding the interplay of evolutionary and ecological dynamics.' *Science* **331** (6016): 426–429.
- 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.
- Shaw, R. G. and Shaw, F. H. 2014. 'Quantitative genetic study of the adaptive process.' *Heredity* **112** (1): 13–20.
- Siepielski, A. M., Dibattista, J. D., and Carlson, S. M. 2009. 'It's about time : the temporal dynamics of phenotypic selection in the wild.' *Ecology Letters* **12**:1261–1276.
- 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.
- Smouse, P. E., Meagher, T. R., and Lobak, C. J. 1999. 'Parentage analysis in *Chamaelirium luteum* (L.) Gray (Liliaceae): why do some males have higher contributions?' *Journal of Evolutionary Biology* **12** (12): 1069–1077.
- Speakman, J. R. 2008. 'The physiological costs of reproduction in small mammals.' *Philosophical Transactions of the Royal Society B: Biological Sciences* **363** (1490): 375–398.
- Stearns, S. C. and Koella, J. 1986. 'The evolution of phenotypic plasticity in life-history traits: Predictions of reaction norms for age and size at maturity.' *Evolution* **40** (5): 893–913.
- 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.
- Steinsland, I., Larsen, C. T., Roulin, A., and Jensen, H. 2014. 'Quantitative genetic modeling and inference in the presence of nonignorable missing data.' *Evolution* **68** (6): 1735–1747.

- 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* **62** (9): 2435–40.
- Stinchcombe, J. R., Simonsen, A. K., and Blows, M. W. 2014. 'Estimating uncertainty in multivariate responses to selection.' *Evolution* **68** (4): 1188–1196.
- 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.
- Teplitsky, C., Robinson, M. R., and Merilä, J. 2014. 'Evolutionary potential and constraints in wild populations.' In *Quantitative genetics in the wild*, edited by A. Charnier, D. Garant, and L. E. B. Kruuk. Oxford: Oxford University Press.
- 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.
- Tienderen, P. H. van. 2000. 'Elasticities and the link between demographic and evolutionary dynamics.' *Ecology* **81** (3): 666–679.
- 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.
- 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.
- Trivers, R. L. 1974. 'Parent-offspring conflict.' *American Zoologist* **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.
- Tuljapurkar, S., Steiner, U. K., and Orzack, S. H. 2009. 'Dynamic heterogeneity in life histories.' *Ecology Letters* **12** (1): 93–106.
- 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.
- 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.
- Uyeda, J. C., Hansen, T. F., Arnold, S. J. ., and Pienaar, J. 2011. 'The million-year wait for macroevolutionary bursts.' *Proceedings of the National Academy of Sciences* **108** (38): 15908–15913.
- van Dijk, E. L., Auger, H., Jaszczyzyn, Y., and Thermes, C. 2014. 'Ten years of next-generation sequencing technology.' *Trends in Genetics* **30** (9): 418–426.

Bibliography

- van Noordwijk, A. J. 1988. 'Two-stage selection in which the first stage only reduces the environmental variation in body size in the great tit.' In *Acta XIX Congressus Internationalis Ornithologici, Vol. II*, University of Ottawa Press, Ottawa,
- Vander Wal, E., Garant, D., Festa-Bianchet, M., and Pelletier, F. 2013. 'Evolutionary rescue in vertebrates: evidence, applications and uncertainty.' *Philosophical transactions of the Royal Society of London B: Biological sciences* **368** (1610): 20120090.
- 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.
- 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.
- Voje, K. L., Holen, Ø. H., Liow, L. H., and Stenseth, N. C. 2015. 'The role of biotic forces in driving macroevolution: beyond the Red Queen.' *Proceedings of the Royal Society B* **282** (April 2016): 1–9.
- 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.
- Walsh, B. 2014. 'Special issues on advances in quantitative genetics: introduction.' *Heredity* **112** (1): 1–3.
- Wandeler, P. and Camenisch, G. 2011. 'Identifying Y-chromosomal diversity by long-template PCR.' *Molecular Ecology Resources* **11**:835–841.
- Wandeler, P., Ravaoli, R., and Bucher, T. B. 2008. 'Microsatellite DNA markers for the snow vole (*Chionomys nivalis*).'*Molecular Ecology Resources* **8**:637–639.
- Wang, J. 2004. 'Sibship reconstruction from genetic data with typing errors.' *Genetics* **166**:1963–1979.
- 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.
- Wellenreuther, M. and Hansson, B. 2016. 'Detecting Polygenic Evolution: Problems, Pitfalls, and Promises.' *Trends in Genetics* **32** (3): 1–10.

- 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.
- Wilkins, J. S. 2009. *Species. A history of the idea*. Berkeley, California, USA: University of California Press.
- Wilson, A. J. and Nussey, D. H. 2010. 'What is individual quality? An evolutionary perspective.' *Trends in Ecology & Evolution* **25** (4): 207–14.
- 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.
- 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.
- Wood, S. 2011. 'Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models.' *Journal of the Royal Statistical Society (B)* **73** (1): 3–36.
- 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.' *Nature Genetics* **42** (7): 565–569.

Chapter 7

Acknowledgements

I don't know half of you half as well as I should like; and I like less than half of you half as well as you deserve.

— J.R.R. Tolkien, *The Fellowship of the Ring* (1954)

Je feins l'adulte, mais, secrètement, je guette toujours le scarabée d'or, et j'attends qu'un oiseau se pose sur mon épaule, pour me parler d'une voix humaine et me révéler enfin le pourquoi du comment. / I pretend to be adult, but secretly I still look out for the golden beetle, and I wait for a bird to land on my shoulder, talk to me with a human voice, and at last reveal me the how and the why.

— Roman Kacew, dit Romain Gary, *La Promesse de l'aube* (1960)

Looking back, it is amazing to notice how much life is influenced by a myriad of people. These acknowledgements are for a Ph.D. thesis, but they are bound to look back in time well before the Ph.D. onset.

First and foremost, many thanks to Erik Postma for being such a great supervisor. Erik set the track for a fascinating and successful Ph.D. project when he realized the potential of the snow vole monitoring and wrote a thoughtful SNF proposal full of surprisingly correct guesses about the biology of the population. Erik taught me statistical techniques, ideas from quantitative genetics and presentation techniques. More importantly though, he trained me to consider logical arguments critically, read scientific publications in depth (in particular during epic journal clubs “rejecting” about half of the papers, based on technical or logical flaws) and thereby to learn from the mistakes of others and a bit less from mine. For four years, Erik’s door was literally always open to answer my “little questions” and review my work. Finally, it impossible not to mention that Erik introduced me to running, pushed me to run more often, longer and faster, and thus kept me fit, healthy and entertained.

Not being one of his students, and given his very tight schedule, I am especially grateful to Lukas Keller for long conversations, old publication mining and spontaneous suggestions. His seemingly encyclopaedic knowledge, and his almost as rich library, not only solved a few crucial issues relevant to my Ph.D., but also kept acute my appetite for the wonders of population genetics, organismic biology and statistics.

Thanks to Arpat Ozgul for his constant enthusiasm about the project and for supporting its continuation during the postdoc to come. Also, thanks for being a bridge to the ideas and slang of population ecology and demography, thus clarifying “trans-

Chapter 7 Acknowledgements

lation” issues.

Thanks to Barbara Tschirren for her precious suggestions when trying to understand the voles in a more mechanistic way, molecular and physiological. Thanks for letting me take your brand new respirometer in the humidity, low pressures and shakiness of the field.

The most important analyses of my Ph.D. were run with the package `MCMCglmm`, and I was extraordinarily lucky to count his creator, Jarrod Hadfield, among my committee members. Besides technical support, Jarrod helped clarifying some confusing concepts and results from quantitative genetics. Also, thanks to him for giving me the opportunity to review for the journal *Evolution*.

Thanks to Marc Kéry for his help and encouragements with complicated JAGS models, but also for some sound life-planing advice.

Thanks to Peter Wandeler for initiating the snow vole monitoring in Churwalden, while I was still a high-school student, unaware of the scree little happy folks.

I could never thank enough Glauco Camenisch for the help he provided in the lab, on the field, in front of the computer and in the kitchen. Glauco saved me many weeks of work and made constant efforts to improve the quality of the data set. With Glauco, the other backbone of the research group is Ursina Tobler. Ursina always keeps the administrative duties minimal on our side, but makes sure everything works fast and smoothly in the background. Thanks to her for making the live of students so easy.

During the last three years, the majority of my scientific discussions were with Koen van Benthem. I am grateful for his many ridiculous questions and discoveries, most of them turned out to be puzzling or insightful. A particular thank for pushing me towards \LaTeX and Zurich Tonhalle. I wish him to crack the pattern behind prime numbers, and to bring more mathematical rigour into ecology and evolution. Thanks also to him for taking the lead (i.e. first first authorship or sixth last authorship) on our “Price’s equation review” (from which Price’s equation was eventually extirpated), a long and tedious, but incredibly educational, collaboration, for which I also have to thank Marjolein Bruijning and Eelke Jongejans.

Vicente García-Navas can do research faster than a snow vole can run into a trap, and publishes it in good journals before the apple is all eaten. It was very stimulating and good for my CV to work with him.

Andres Hagemayer was an exemplar Master student, very serious, autonomous and efficient. Thanks to him for almost two years of a fruitful collaboration that taught me a lot about teaching.

Thanks to Cindy Canale for helping with respirometry, for kind encouragements and nice parties. Many thanks to Dominique Waldvogel and Martina Schenkel for precious and cheerful help on the scree.

For four years, the majority of my non-sleeping time was spent in the company of the “permanent residents” of the office Y13-J-34. I will remember this office as home (although I kept my word and never spent a night there) thanks to all our passionate discussions, our ire at wrong papers, our debugging struggles and our coffee breaks. Special thanks to Pirmin Nietlisbach for hosting me on Mandarte island; to Philipp Becker for teaching me how to cross-country ski and catch dippers; and to Judith Bachmann for nice frog expeditions and for translating the summary of this thesis into German (and not letting Koen include a mention of whales and cathedrals). This

office was such a great place to work, live and laugh with the essential contribution of its intermittent occupants, Erica Ponzi, Vanja Michel, Rien van Wijk and Johann Hegelbach.

Thanks to the Kokonuts—Hanna Kokko, Isobel Booksmythe, Anaïs Tilquin, Nina Gerber, Susanne Schindler and Xiang-Yi Li—for inspiring journal clubs, where you can sit on an orange primitive salamander (definitely not a pony). Christine Grossen and Daniel Croll are among the kindest humans I have ever met. In particular, thanks for welcoming me during my first days in Zürich and for lending me a piano for two years. Merci to Chelsea J. Little for being a great adventure buddy, a cheering friend and an example of scientific brightness.

Many thanks (sadly impersonal so that the acknowledgments do not become a thesis chapter on their own right) to the Zürich folks for four very fun and enriching years. In a jumble, thanks to Alexandra Jansen van Rensburg , Alexander Nater , Irene Weinberger , Simon Evans , Jasmin Winkler , Joel Pick , Mélissa Lemoine , Dennis Hansen , Stefanie Muff , Rassim Khelifa , Debbie Leigh , Josh van Buskirk , Kasia Sluzek , Frédéric Guillaume , Béatrice Nussberger , Benedikt Gehr , Nina Vasiljevic , Ulrich Reyer , Hedwig Ens , Wolf Blanckenhorn , Tina Cornioley , Sam Cruickshank , Mollie Brooks , Juan Pablo Busso , Franziska Lörcher , Jobran Chebib , Cini Gabriella Gall , Andreas Sutter , Aurélie Garnier , Nino Magg , Inge Juszak , Gianalberto Losapio

...

Special thanks to Ashley E. Latimer for forcing me to cook and go birdwatching during stressful periods, for teaching me some English and for inspiring conversation bridging gaps between population genetics and evolution over geological time-scales.

I never had the opportunity to thank all the great people I have met since the beginning of my scientific life (because I didn't write a report for the two first internship, and acknowledgements were *prohibited* (sic) for my master thesis). Still, I owe them a lot, and in particular many people were instrumental in setting the path toward a Ph.D.. In 2011, while I was trying to escape from leave engineering for fundamental research, without having much clue of what it was about, Jean-François Martin gave me the opportunity to try it out by tutoring a gap year entirely dedicated to research. Pierre-André Crochet provided determinant help to find my way in the cloud of biology and reach a first research experiment, in the form of an internship at University of Oslo CEES. There, Glenn-Peter Sætre was an amazing first supervisor, extremely patient with my childish English, my misadventures in the lab and my perfect ignorance in evolutionary genetics, and was very supportive to find housing and survive one semester in Norway without any income. Thanks to him and all his team, especially Tore Oldeide Elgvin and Anna Fijarczyk, I had a great time that convinced me that fundamental research in evolutionary biology would be a good way to spend my time. This might be where I learned the most about what the life of an evolutionary biologist was about: from sequencing DNA, to searching and understanding scientific literature; from the quasi-universal fascination of scientists for coffee and Friday beer, to the poetic insight of Kimura's neutral theory, and many more things. Eventually, thanks to them for making me publish my first research papers, although I did not do a lot and was probably not really understanding the little I did. I then went to Chizé CEBC where I worked with a crew of amazing people including Adrien Pinot, Vincent Bretagnolle, David Pinaud, Vincent Lecoustre, Edoardo Tedesco, thanks to

Chapter 7 Acknowledgements

them and all the other Chizéens for all they taught me. A particular thank to Mathieu Authier for his (partially-successful) attempt to convert me to Bayesianism, and to Laurent Crespin for forcing me to go deeper into Mark-Recapture modelling and maximum of likelihood. Last but not least, Bertrand Gauffre was crucial in my scientific development, thanks to him for our many conversations, for taking me seriously, for introducing me to Richard Dawkins' writings and for pushing me to do complete my Masters in Montpellier. The B2E Master in Montpellier was a time of hard studying, and great fun, during which some teachers opened my mind to new scientific horizons, thanks in particular to Patrice David, Isabelle Olivieri, Olivier Gimenez and Michel Raymond for that. Thanks to Raphaël Leblois for introducing me to the coalescent theory and its powerful thought experiments, as well as to C++ programming. Coding ForwardBackward and its improved sequels was a crazy adventure, full of traps and wonders. Thanks to François Rousset for giving me a small sight at what it means to understand population genetics, statistics and logic. Thanks again to PAC for his patient help explaining me again and again evolutionary concepts and writing tricks. Also, thanks to Nicolas Bierne for his PhD offer and our fascinating discussion on oceanic streams, gene flow and the elusive eel, I still think about this alternative scientific and life path with curiosity and envy.

Thanks to all the Montpelliérains, in particular Émeric Figuet, Léo Grasset, Joane Elleouet, Paul Sanders, Pascal Milesi, Julie Landes... our conversations account for a quarter of what I know in evolutionary biology. Out of the academic world, I could always count on the support of Gaëlle Jeanne Duranthon during the last 12 years. Thanks also to my two adoptive families who kept me cheerful with meals, games, music, remote places, short nights and craziness during the last decade. These are the Goret family with an otter, a weasel, a canary, a baboon, a snail, a goat, a bee, a goose, a panda, a wolf, a bear, an owl, a lion and a chameleon; and the Margicon (*©une marque déposée par terre*) family: Nounou, Chonchon, Caca, Doudou, Mirabelle and their mates/cats/penguins.

My early interest in ecology and evolution was much reinforced by wandering through forests, meadows and moorlands of Lacaune mountains, contemplating glittering and diverse carabids, excitement of long days monitoring bird migration, and astonishing conversations on conservation and nature management. There I learned that while we still know so little about the world, it takes only some sweat and patience to discover new things. This early encounter with the natural world would not have been sustainable without the naturalists who taught me so much and kept me amazed. Thanks in particular to Amaury Calvet and other members of the "Ligue pour la Protection des Oiseaux du Tarn" as well as to the members of BeaOsea: Adrien Chaigne, Camille Denozière, Aurélien Salesse, Denis Guillaumin and Manon Ghislain.

Some high school teachers were particularly influential not so much by what they taught, but rather by how they taught to learn, to see things in another way or by opening unexpected opportunities. Many would deserve acknowledgements for doing that at diverse degrees, but here are a few whose classes remain very alive: Benoît Leviandier, Mdm Rolland, Jacky Cariou, Jean Leblanc.

Being born in such a family made life rather easy. Thanks especially to my parents Francine and Claude, to my siblings Élodie-Kiyomi and Cyrille, to my grand-parents Juliette, Louis, André and Simone and to my uncle Laurent. They kept supporting,

sustaining and trusting me over long years of studies they could less and less understand.

Chapter 8

CV Timothée Bonnet

Personal Data

BIRTH DATE: November 1st 1988
NATIONALITY: French
WORK ADDRESS: Department of Evolutionary Biology and Environmental Studies
University of Zürich
Winterthurerstrasse 190
CH-8057 Zürich
Switzerland
PHONE: +41 (0)44 635 47 66
EMAIL: timothee.bonnet@ieu.uzh.ch
PERSONAL PAGE <http://www.ieu.uzh.ch/staff/phd/bonnet.html>
LANGUAGES: French (mother tongue), English (fluent), Spanish (basic)

Education

OCT 2012 - Current	<p>PhD student in Zürich evolutionary biology PhD program. <i>Individual-level causes and population-level consequences of variation in fitness in an alpine rodent.</i> Under the supervision of Dr Erik Postma.</p>
<i>Courses:</i>	
MAY 27-28th 2013	NGS for Model and Non-Model Species, by K. Shimizu & al.
JUN 22-29th 2013	Evolutionary Biology Workshop in Guarda, by D. Ebert & S. Bonhoeffer
OCT 10-11th 2013	Workshop on Integral Projection Models, C. Merow & al.
OCT 14-18th 2013	Evolutionary Demography, by D. Levitis & al.
JAN 13-17th 2014	Bayesian Population Analysis using WinBUGS, by M. Kéry & M. Schaub
NOV 6-7th 2014	Advanced Software Carpentry, by M. D. Robinson & al.
MAY 18-19th 2015	Advanced NGS, by S. Wider & H. Lischer
SEP 2011 - JUN 2012	<p>M.Sc. in evolutionary biology and ecology. University Montpellier II, France. <i>Neutral processes and biased mitochondrial introgression</i> at the center for population biology and management (CBGP). Supervised by Drs Raphaël Leblois, Pierre-André Crochet and François Rousset.</p>
<i>Research project:</i>	
SEP 2008 - SEP 2011	<p>B.Sc Biology: National engineering school in biology and agronomy, Montpellier Supagro, France. Specializations in biodiversity conservation, ecology, phylogeny, population genetics, GIS.</p>
<i>Research projects:</i>	
JAN-AUG 2011	<p><i>Population dynamics of rodents, and agricultural practices</i> at Chizé Centre for Biological Studies (CEBC), France. Supervised by Drs Bertrand Gauffre and Vincent Bretagnolle.</p>
SEP-DEC 2010	Centre for Ecological and Evolutionary Synthesis in Oslo, Norway: lab work, genetic data analysis and article redaction. Genetic identification, speciation, hybridization and role of gonosomes in flycatchers and sparrows. Supervised by Prof. Glenn-Peter Sætre.
MAY-AUG 2010	<i>Meadow birds phenology, conservation and agriculture</i> at LPO (Birdlife international), in Grenoble, France.

Seminars

Invited seminars

- Bern, Switzerland, March 9th 2016.
- *Body mass selection in an alpine rodent: does it fluctuate? does it matter?* Radboud Universiteit Nijmegen, the Netherlands, December 4th 2014.
- *Variation in fitness: proximal and ultimate causes.* CNRS Brunoy, France. October 28th 2014.
- *Individual-level causes of variation in fitness in an alpine rodent.* IEU, Zurich, September 29th 2014.

Contributed seminars

- *The stasis that wasn't: Adaptive evolution goes against phenotypic selection in a wild rodent population* Evolution 2016, Austin, Texas, USA, June 17th-21st 2016.
- *The stasis that wasn't: Adaptive evolution goes against phenotypic selection in a wild rodent population* Biology16, the Swiss conference on organismic biology, Lausanne, Switzerland, February 11th-12th 2016.
- *The stasis that wasn't: Adaptive evolution goes against phenotypic selection in a wild rodent population* 3rd Young Natural History scientists Meeting, Paris, France, February 2nd-6th 2016.
- *Rapid adaptive evolution opposite to phenotypic selection. Or why snow voles get smaller despite selection for larger individuals* European Society for Evolutionary Biology (ESEB) 15th, Lausanne, Switzerland. August 10th-14th 2015.
- *Evolution outreach through dirtiness* Poster at the ESEB Workshop on Teaching Evolution, Lausanne, Switzerland. August 9th 2015.
- *Successful by chance? The power of mixed models and neutral simulations for the detection of individual fixed heterogeneity in fitness components* GDR Ecological Statistics meeting, Lyon, France, March 12th-13th 2015.
- *Why voles do not become beavers: indirect relationships between traits and fitness counteract selection for larger individuals in a snow vole population* Poster at Biology15, the Swiss conference on organismic biology, Dübendorf, Switzerland, February 12th-13th 2015.
- *Fluctuating selection and genetic gradients on snow vole mass* Wild Animal Models Biennial Meeting, University of St Andrews, U.K. July 21st-25th 2014
- *Lord of the scree by chance or by merit? Dynamic vs. fixed heterogeneity in an alpine rodent population.* Evolutionary Demography Society (EvoDemoS) first meeting, in Odense, University of South Denmark. October 5th-10th 2013.

- *Climatic variability, viability selection and demography in an alpine rodent.* European Meeting of PhD Students in Evolutionary Biology (EMPSEB) 19th, at university of Exeter, U.K. September 3rd-7th 2013.
- *Neutral processes and cyto-nuclear discordant introgression.* Colloquium Petit Pois Déridé, Avignon, France. August 29th 2012.

Skills

Scientific

Biology | Evolutionary biology, population and quantitative genetics, demography.

Statistics | Generalized Linear Mixed Models, Bayesian methods, Mark-Recapture analysis.

Mathematics | Linear algebra, analysis, probabilities.

IT

O.S | Microsoft Windows, Linux (Ubuntu), MacOS X

Scripting and programming | R/S4, C/C++, Bash shell, BUGS/JAGS, Matlab, L^AT_EX
Some projects visible on GitHub:
<https://github.com/timotheenivalis>

Software | Arlequin, Genemapper, Mega, Sequencher, CLC Workbench, BioEdit, Genepop, Mark, Ucare, ArcGIS,...

Miscellaneous

February 2013 | Far from help first aid course (2 days)

Teaching

AUG 2016 | Field course

MAR 2016 | 10 afternoons of practicals in Population Ecology

MAR 2015 | One hour practical in quantitative genetics

FEB 2015 | One day introductory course to L^AT_EX(self-organized)

JUL 2014 - *current* | One Master student

DEC 2013 | 3 weeks supervision of a Bachelor student project

Reviewing activity

Journals

Publons merit: 21 (<https://publons.com/a/822275/>)

2015-2016 | Evolution

2015-2016 | Molecular Ecology

2016 | Heredity

2015 | Proceedings of the Royal Society B: Biological Sciences

2015 | Ecology and Evolution

Conferences

2016 | Jury member for biodiversity and conservation at YNHM
16, Paris

External activities

Scientific outreach

Contributions to:

- Scientific "speed-dating" with the public at ESEB 15 and Biology 16 conferences

- Dans les testicules de Darwin. (7 articles, 2013 - 2015)
<http://danslestesticulesdedarwin.blogspot.ch>
- Un pied dans le plat. (1 article, 2012)
www.unpieddansleplat.fr/menu_gauche/alimentation_sante/laitage_et_cancer.php

Ornithology and naturalism

2012 - *current* | Member of the regional rare bird comity Tarn-Aveyron
http://www.faune-tarn-aveyron.org/index.php?m_id=20025

Competitive funding

2016 | Travel grant to attend the 3rd Young Natural History Scientists Meeting, Paris, France

2012 | PhD fellowship at IEU UZH

Peer-reviewed publications

ISI citations: 30

- van Benthem, K.*, Bruijning, M.* , **Bonnet, T.***, Jongejans, E., Postma, E. & Ozgul, A.. Disentangling evolutionary, plastic and demographic processes underlying trait dynamics: A review of four frameworks. *Accepted in Methods in Ecology and Evolution.*(*co-first authors)
- **Bonnet, T.** & Postma, E. **2016.** Successful by chance? The power of mixed models and neutral simulations for the detection of individual fixed heterogeneity in fitness components. *The American Naturalist* 187(1). Recommended by Faculty of 1000

- García-Navas, V., **Bonnet, T.**, Waldvogel, D., Camenisch, G. & Postma, E. **2016.** Consequences of female philopatry for reproductive success and mate choice in an Alpine rodent. *Behavioral Ecology*.
- García-Navas, V., **Bonnet, T.**, Bonal, R. & 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., Wandeler, P., Camenisch, G. & Postma, E. **2015.** Gene flow counteracts the effect of drift in a Swiss population of snow voles fluctuating in size. *Biological Conservation* 191: 168–177.

- Bonnet, T., Crespin, L., Pinot, A., Bruneteau, L., Bretagnolle, V. & Gauffre, B. 2013. How the common vole copes with modern farming: Insights from a capture-mark-recapture experiment. *Agriculture, Ecosystems & Environment* 177: 21-?27.
- Elgvin, T.O., Hermansen, J.S., Fijarczyk, A., Bonnet, T., Borge, T., Sæther, S. a, Voje, K.L. & Sætre, G.P. 2011. Hybrid speciation in sparrows II: a role for sex chromosomes? *Molecular Ecology* 20: 3823-?3837.
- Bonnet, T., Slagsvold, P.K. & Sætre, G.P. 2011. Genetic species identification of a Collared Pied Flycatcher from Norway. *Journal of Ornithology* 152: 1069-?1073.

Completed manuscripts

- Bonnet, T., Wandeler, P., Camenisch, G. & Postma, E.. Adaptive evolution goes against phenotypic selection in a wild rodent population. *In review in PLoS Biology*.
- Bonnet, T. & Postma, E.. Fluctuating selection but no fluctuating evolution in a wild rodent population. *Submitted to Evolution*.
- Bonnet, T., Leblois, R., Rousset, F. & Crochet, P.A.. A reassessment of explanations for discordant introgressions of mitochondrial and nuclear genomes. *To be submitted to Evolution*.

Addresses of co-authors

Koen J. van Benthem, Arpat Ozgul, Glauco Camenisch and Erik Postma

Department of Evolutionary Biology and Environmental Studies
University of Zurich-Irchel
Winterthurerstrasse 190
CH-8057 Zurich
Switzerland

Marjolein Bruijning and Eelke Jongejans

Animal Ecology and Physiology Department
Radboud University
Heyendaalseweg 135
6525 AJ Nijmegen
The Netherlands

Peter Wandeler

Musée d'histoire naturelle de Fribourg
Chemin du Musée 6
CH-1700 Freiburg
Switzerland