ECOLOGICAL GENETICS OF POPULATIONS EXPERIENCING CHANGING ENVIRONMENTAL CONDITIONS

ARILD HUSBY

SUBMITTED FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

UNIVERSITY OF EDINBURGH 2009

ABSTRACT

A major goal in evolutionary biology is to understand how ecological factors shape the phenotypic and genetic variation that we observe in natural populations and in this thesis I examine how rapid changes in temperature have influenced phenotypic and genetic variation in morphological and life history traits in long-term studies of great tits. In **Chapter 1** I review what is known about the effects of environmental change on natural populations, and outline the quantitative genetic framework that is available to study genetic variation in natural populations. Much focus on the effects of climate change has concerned species' phenology, far less attention has been given to other traits. In **Chapter 2** I examine the effects changing environmental conditions have had on the proportion of females that produce second broods. Temperature operates mainly through indirect effects (such as food abundance) but may also have more direct effects. In **Chapter 3** I show that over a 36 year period body size have declined in line with predictions from Bergmann's rule and I explore the genetic basis of this decline and the environmental factors involved. Although we can learn much from population level responses, there is a great deal of additional information to be gained by studying between-individual responses. In **Chapter 4** I therefore compare the multivariate pattern of between-individual variation in phenotypic plasticity and its genetic basis for laying date and clutch size, in two great tit populations. Environmental changes may also directly affect the expression of genetic variance as well as the strength of selection acting on a trait, and in **Chapter 5** I show that, for laying date, the environment induces a positive covariance between strength of selection and the expression of additive genetic variance, something that may enhance the rate of adaptation. Finally, in **Chapter 6** I discuss and summarise the wider implications of the findings from this thesis.

DECLARATION

The work presented in this thesis would not have been possible were it not for the many researchers and field assistant that have collected data. As they are too numerous to mention I have only indicated if the data used was collected by the Netherlands Institute of Ecology (NIOO) or by the Edward Grey Institute (EGI), Oxford University. More details about contribution to each chapter are provided in the table below (MEV = Marcel E. Visser; LK = Loeske Kruuk; AH = Arild Husby).

	Chapter 2	Chapter 3	Chapter 4	Chapter 5
Initial idea	MEV/LK	MEV/AH	LK	AH/LK
Materials	NIOO	NIOO/AH	NIOO/EGI/AH	NIOO/AH
Analyses	AH	AH (J. Hadfield helped with MCMC analysis)	AH (MEV did 'sliding window' analysis)	AH
Wrote manuscript	AH in collaboration with co-authors	AH in collaboration with co-authors	AH in collaboration with co-authors	AH in collaboration with co-authors

I hereby confirm that all analysis and words contained in this thesis are my own.

Arild Husby, Edinburgh 28.10.2009

ACKNOWLEDGMENTS

I would like to thank all the people who have helped me to complete this thesis. First of all my supervisors, Loeske and Marcel, who have skilfully guided me through these three years: I have thoroughly enjoyed working with you both and learnt much! Thank you, Loeske, for dissecting all my manuscripts and helping me put them together again. I have learned a great deal from the many revisions and discussions we have had about the chapters presented in this thesis. I hope I also someday will develop such keen eyes for spotting logical flaws. Thank you, Marcel, for making sure I kept a firm sight on the biology behind the questions, for many enlightening discussions and for encouragement throughout. And, not least, for providing much needed beer both in the field and at the ESEB poster session!

There are many people in Edinburgh to thank, especially the members of the WEG (Wild Evolution Group), and, in particular, Alastair and Dan. Alastair answered my questions about random regression models and provided insightful comments and suggestions for Chapter 4 and 5, for which I am very grateful. Dan provided crucial help in getting the NIOO database up and running during the first few months, and by doing so made sure I did not waste my entire first autumn on non-sense Access queries. He also gave very helpful comments on Chapter 3 and 4 in this thesis. Thanks guys!

I would also like to thank two previous WEG members, Matt and Tom, who both took care of me and invited me along for cinema, dinner parties and football when I first arrived in Edinburgh, something I greatly appreciated. I should also like to thank Matt personally for the many whisky nights that started well and ended less well, as well as for his thoughtful comments on Chapter 5 and for many random (regression) discussions. Tom also gave very useful comments on Chapter 5. Josephine deserves special thanks for providing useful feedback on Chapter 6 on a very short notice. Thanks to Jarrod for helping to sort out the MCMC animal models and for useful discussions. I would also like to thank all other members of WEG for helpful, critical and inspiring discussions about quantitative genetics (and other things) throughout my time in Edinburgh.

A few ‘outsiders’ also deserve special mentioning: in particular, Ben Sheldon (Oxford University) who I had the pleasure to collaborate with on Chapter 4: thank you for the useful discussions and comments on those analyses. Many thanks also to Sabine Hille (University of Bodenkulture Vienna) for collaborating on Chapter 3 and for hosting me during my stay in Vienna. There are also many people at NIOO to thank for making my stays there enjoyable, in particular Henri, Stein Are, Andrey (who made my first Dutch field season a very enjoyable experience), Francisco, Kate, Luc and Sonja as well as all the people I no longer remember the names of that have stayed in the guest house during my stays. I enjoyed all the dinner parties and especially the Tuesday night football match during my stays.

I would also like to thank Stein Are (now at the University of Oslo) for collaboration related to the cost of reproduction in great tits, which, although not included in this thesis, I look forward to continue (and finish off) very soon.

On the social side of things I would like to thank everyone that have made sure I have had good times in Edinburgh: Alex L. for many great badminton games, running trips and late Sunday guitar nights, Matt & Tabi for lovely dinner evenings and hiking trips (the peak district was great), Konrad & Marie for many hiking trips, occasional birding trips and many nice evenings. All the people I have shared office with throughout the times (Matt, Tom, Michelle, JC, Alex H, Olivia, Adam, Anjie) for many good chats. To whom I may have forgotten to thank, this is for you: thanks!

Finally I would like to thank my family.
To my parents who have always been supportive and kept a keen interest in my doings. I doubt I would have embarked on this project were it not for the many times we have had out in the mountains, by the lakes, summer, winter, skiing, hiking and at all times keeping a look out for birds. You sparked the interest!

To my dearest Elina, thank you for all the discussions about science but even more for the discussions without. You have been patient, supportive and loving, always. Also many thanks for providing comments on all but one of the chapters in here.

TABLE OF CONTENT

CHAPTER 1

General introduction

All photos in thesis: © Arild Husby

1. 1 General introduction

A major goal in evolutionary biology is to understand how ecological factors influence phenotypic and genetic variation in natural populations and how individuals adapt to the environmental conditions they experience. The goal of this thesis is to obtain a better insight into how changes in environmental conditions influence phenotypic and evolutionary dynamics using long-term studies of great tits (*Parus major*). One such important environmental factor is temperature and recent decades has seen an increase in global temperatures by on average 0.6 °C since 1900, and a rate of increase in temperatures since 1976 greater than at any other time during the last 1000 years (Houghton et al. 2001). The increase in temperatures is furthermore expected to continue, something which offers an excellent opportunity to study the consequences rapid environmental changes may have on natural populations. Such consequences have been, and continue to be, well documented on the phenotypic level (reviewed in Parmesan 2006) and include advances in the phenology of many plant and animal taxa (Crick et al. 1997; Parmesan 2006; Franks et al. 2007; Parmesan 2007; Charmantier et al. 2008), changes in breeding patterns (Husby et al. 2009), shifts in migration distance (Visser et al. 2009b) and pattern (Spear and Ainley 1999) in many bird species, changes in population size (Both et al. 2006), body size (Yom-Tov 2001; Teplitsky et al. 2008) and in the distribution of plants and animals (reviewed in Parmesan and Yohe 2003), to name but a few examples.

Hence, it is clear that climatic changes, and associated alterations in habitat and biotic interactions, may have a profound influence on a large number of traits from many different taxa and will impose natural selection on ecologically important

traits. Such changes are likely to be one of the main driving forces in evolution (Endler 1986) and may, ultimately, lead to adaptation and speciation (Schluter 2000).

This thesis explores some of the phenotypic and genetic consequences that increased temperature, and its associated effects, have had, and continue to have, on several long-term study populations of great tits (*Parus major*) in their natural environment. It is only by studying populations in their natural habitat that we will be able to reach a thorough understanding of, and insight into, the consequences that changing environmental conditions have on selection pressures and the phenotypic and genetic variation which, ultimately, will allow us to understand how individuals adapt to the environment they inhabit.

Darwin (1859) was the first to notice that variation between individuals in their appearance (phenotype) and in their ability to reproduce or survive (fitness) creates an opportunity for some individuals to become more successful than others (selection). This is a fundamental principle in all of biology and in this introductory chapter I will explore these principles and their dependency on the environment in more detail. I will firstly discuss the ubiquity of variation in natural populations and look at how natural selection operates and how to measure it before I turn to how one can measure the genetic basis of variation in natural populations. Finally, I will introduce the study species and systems used in this thesis and outline the goals of this thesis.

1.2 Variation in traits in natural population

A general observation of individuals in natural populations is that they vary for almost any character measured. This variation is either continuous or discrete and this distinction is an important one indeed: it separates the fields of quantitative

genetics and population genetics respectively. For example, body mass is a trait where almost every individual differs from others; thus when we summarise variation in body size this generally gives rise to a normal distribution (see Fig 1.1) and such traits are analysed within the field of quantitative genetics (see below). In contrast, for discrete traits, individuals only fall into a few categories, as is for example the case of melanism in the peppered moth (*Biston betularia*) where individuals either belong to a dark or a light morph (Kettlewell 1973, see below) and this type of traits are analysed in the field of population genetics. Such discrete variation will tend to approach a continuous distribution however when the number of categories increases.

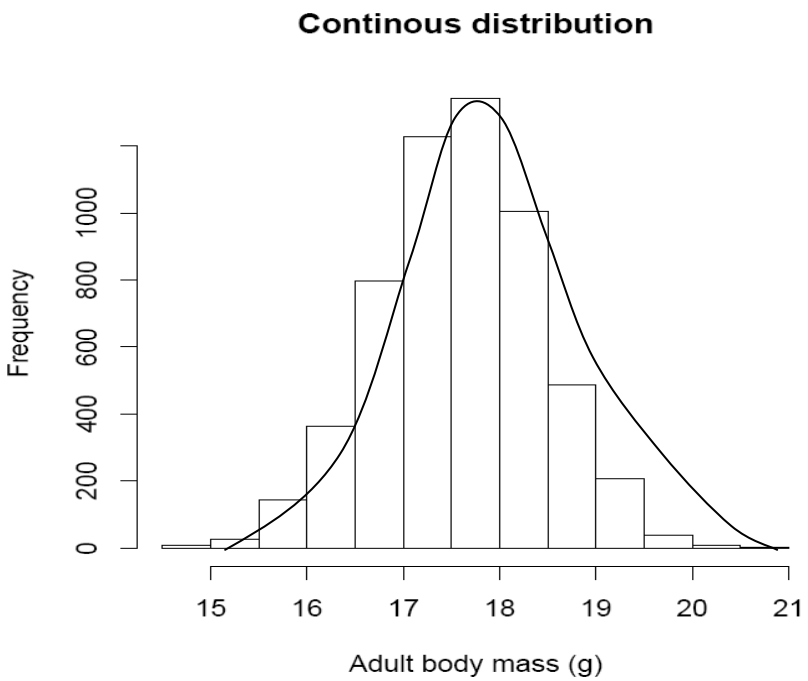


Fig. 1.1 Distribution of adult body mass in great tits from the Hoge Veluwe population (n = 5659) form a normal (or Gaussian) distribution. I will return to this example in more detail in Chapter 3.

Importantly, variation between individuals is not only apparent for morphological traits, but is also present on the cellular, biochemical and, ultimately, DNA level. It is

this variation that is fundamental to natural selection and evolution: without variation there will be no selection and no evolutionary change.

1.3 Fitness variation in natural populations

Variation between individuals is also found in the number of offspring they produce and in how long they live. These two components, fecundity and viability, are the fundamental properties that define an individual's reproductive success or its so-called fitness. Without differences in fitness, natural selection can not act and adaptation can not occur. Fitness itself however is a difficult concept (see for instance Brommer et al. 2004) and it is often difficult to obtain accurate measures of fitness in natural populations thus it is more common that studies use components of fitness (like, say, survival to a particular age) than total fitness. Perhaps the 'easiest' way is to follow individually marked animals or plants over their lifespan and record the number of offspring they produce. This is a time-consuming, but very rewarding, exercise and a number of systems exist where such longitudinal studies are possible, predominantly in ungulates (Clutton-Brock et al. 1997; Coltman et al. 2005; Gratten et al. 2008) and birds (Lack 1968; van Balen 1973; Gustafsson et al. 1995; Charmantier et al. 2004; Husby et al. 2006) where individuals are relatively easy to catch and mark. I will not discuss the concept of fitness any further here and refer the reader to a excellent recent review on the role of fitness in evolutionary genetics by H. Allen Orr (Orr 2009).

1.4 Selection in natural populations

The fact that individuals differ both in their fitness (w) and in their trait value means that we can measure the association between the two to find out if individuals with a particular trait value have a larger or smaller fitness compared to individuals with a different trait value. If there is a non-random association between fitness and the trait value then the trait is said to be under selection. More formally, the strength of selection acting on a trait is defined as the covariance between relative fitness (ω) and relative trait value (*z*):

$$S = \mathrm{cov}[\omega, z] \qquad (1.1),$$

where *S* is referred to as the selection differential and is a measure of the strength of the association (Lande 1979; Lande and Arnold 1983). Large (positive or negative) values of *S* mean that the trait is under strong selection whereas if *S* is small the trait is under only weak selection. Note that because *z* is standardized (zero mean, unit variance) the covariance is equal to the linear regression coefficient in a least squares regression (i.e. $S = \beta$). If *z* is not standardized then the selection differential, *S*, is not identical to the regression coefficient, β, and is termed the selection gradient. Of course the above model only deals with linear selection (β coefficient), or so-called directional selection, but in a linear regression framework it is easy to extend the model to deal also with non-linear selection by including the second order term of *z* (γ coefficient). If $\beta = 0$ and $\gamma > 0$ this will lead to disruptive selection and if $\beta = 0$ and $\gamma < 0$ we will have stabilizing selection (Arnold and Wade 1984a, b). The development of this method thus allowed researchers to assess the strength of selection in natural populations with the use of such a simple method as a linear regression. This stimulated a large amount of research on selection in natural populations: only a few years after Lande and Arnold's landmark papers, Endler

(1986) reviewed studies of selection in natural populations, and concluded that selection in the wild was commonplace. A later review with much larger sample size estimated a mean standardized selection gradient of |0.16| (Kingsolver et al. 2001). Kingsolver and colleagues also found that morphological traits were generally under stronger selection than life history traits and that quadratic selection was generally weak and symmetrically distributed around 0 (i.e. stabilising selection was not stronger or more common than disruptive selection).

A great benefit of this regression approach to measuring selection is that it can very easily be expanded to include multiple traits, which allows one to control for the effects of indirect selection (i.e. a trait may appear under selection due to a correlation with another trait under selection), and that it is directly relevant to quantitative genetic models for the evolution of multiple traits (see Lande and Arnold 1983 for more details). It is important to realise here that even if a trait is under selection this will not lead to any evolutionary change (i.e. allele frequency change) unless the trait is heritable. This is discussed in the following paragraphs.

1.5 The genetics of trait variation in natural populations

A major challenge in biology is to understand the genetic basis of variation in quantitative traits (Mackay et al. 2009). Indeed, the observation that natural populations harbour a large amount of genetic variance (e.g. Lewontin 1974) is among the most important in evolutionary biology. This is also surprising as selection is commonplace and will generally tend to reduce genetic variance (Barton and Keightley 2002). Although the observation that natural populations are often genetically diverse came largely in the 1960s with the discovery of molecular markers, interest in the genetic basis of trait variation and its relationship to the

environment goes further back. Pioneering work on habitat-correlated genetic variation in plants was carried out as early as in the 1920s (Turesson 1922) and is an good example of a ecological genetic study. The term ecological genetics was first coined by E.B. Ford in his book called *Ecological genetics* (Ford 1964) where he referred to it as the study of genetic variation and selection in natural populations. Perhaps the most famous example of ecological genetics however are the studies carried out by H.B.D Kettlewell in the 1950s (Kettlewell 1973) on frequency changes in melanism in the peppered moth (*Biston betularia*). Peppered moths occur in a dark (melanic) morph and a light morph, where the white is normally much more abundant than the black. The genetics of this colour polymorphism is thought to mainly (but not only) be controlled by a single locus with two alleles (C and c) where the C allele is dominant. Light morphs were homozygous cc and dark morphs homozygous CC (heterozygous Cc were phenotypically similar to the dark morph). Museum collections in Britain made in the eighteenth century contained just the light morph, and it was not until 1848 that the dark morph was registered near Manchester. The interesting thing that Kettlewell noticed was that the dark morph increased in frequency in polluted industrialised areas whereas it did not in unpolluted areas and he suggested that this was due to selective bird predation. In polluted areas the dark morph offered more camouflage when the moths rested on the dark tree trunks than did the light morph and so were less predated. Further support for this came from the observation that dark morphs decreased in frequency when air pollution decreased. This research has however been the subject of much debate recently, see for instance Cook (2003) for an overview on the current status.

The peppered moth example provides a nice illustration that knowledge of the genetic basis of traits in natural populations can help us understand both the ultimate and proximate reasons for the observed variation and changes in traits. For example, without knowing that colour polymorphism in this species was genetically determined it would not have been possible to relate the observed dramatic changes in colour frequency to air pollution and predation risk. However, armed with this knowledge the peppered moth has become a prime example of Darwinian evolution.

In the vast majority of cases however we do not know the exact genetic mechanism underlying the observed phenotype in natural populations. And only very rarely are traits influenced by just one or a few loci (the field of population genetics); more often, variation in traits is influenced by tens or even hundreds of loci each with small effect. If a trait is influenced by many loci of small effect then the variation in trait values will approximate a normal distribution (see Fig 1.1) and as such be analysed using techniques developed in the field of quantitative genetics (Falconer and Mackay 1996).

1.5.1 Quantitative genetics

Quantitative genetics deals with continuous traits (as in Fig 1.1). Inferences about the genetic basis of phenotypic variation come from comparing the phenotypes of individuals of known genetic relationships, instead of knowledge of each individual's underlying genotype. In general, individuals that are more closely related resemble each other more, although the similarity varies with respect to the trait under study. The properties we can observe in connection with a continuous trait are the means, variances and covariances and these are the crucial parameters upon which quantitative genetics rely (Falconer and Mackay 1996).

An individual's phenotype is made up by its genotypic and environmental effects, that is:

$$P = G + E \quad (1.2),$$

where P is the phenotypic value, G the genotypic value and E the environmental deviation. As noted above, we do not know the genotypic value of an individual directly but it is possible to obtain a measure of this by studying the transmission of genes from parents to offspring. We refer to the effect that the genes from a parent have on the trait value of its offspring as that individual's breeding value. Breeding values therefore, unlike genotypic values, can be measured: if an individual is mated randomly to individuals from the population, its breeding value is twice the mean deviation of the resulting progeny from the population mean (Falconer and Mackay 1996). The breeding value is a fundamental property in quantitative genetics and the variance in breeding values is the additive genetic variance. I shall return to this below.

As I have already emphasized it is the *variation* between individuals that is of particular interest in evolutionary genetics. As such we can write the phenotypic variance in the population as a straightforward extension of (1.2) as:

$$V_P = V_G + V_E \quad (1.3),$$

where V_P refers to the phenotypic variance, V_G the genetic variance and V_E the environmental variance. The genetic variance can be further partitioned into its additive (V_A), dominance (V_D) and epistatic (V_I) gene interactions to give:

$$V_P = V_A + V_D + V_I + V_E \quad (1.4).$$

This variance partitioning allows us to separate and estimate the relative importance of each component (expressed as ratios) and thus provide an insight into the relative role of genes and environment in determining the phenotype. The two

non-additive components of variance (V_D and V_I) are notoriously difficult to estimate, but it is thought that they are relatively unimportant in contributing to genetic variance in complex traits (Hill et al. 2008) and they are furthermore not transmitted from parents to offspring (Falconer and Mackay 1996). They are therefore largely ignored and considered part of the residual variance and I shall not consider them further here.

Individual variance components can be expressed as ratios to determine their relative value. By far the most important ratio is the additive to total phenotypic variance, V_A/V_P, which expresses the extent to which the phenotype is determined by parentally transmitted genes. This ratio is referred to as the narrow sense heritability, or more commonly, *heritability* (h^2) and determines the degree of resemblance between relatives (Falconer and Mackay 1996; Lynch and Walsh 1998).

There are a number of ways in which the additive genetic variance can be estimated but all are based on comparing the covariance in phenotype across individuals that are related to differing degrees. The most common methods are parent-offspring regression and different full and half-sib designs. These methods are relatively simple and have been widely used in evolutionary genetic studies of natural populations (see for example Merilä and Sheldon 2001). However it is also well known that they are relatively imprecise and will generally return upwardly biased estimates of V_A, mainly due to the inability to separate out environmental effects. More advanced methods have therefore been developed that can handle this, and other problems, and among them is the 'animal model'.

1.5.2 The Animal Model

Like much of the framework of quantitative genetics, the animal model was originally developed to analyse data arising from breeding programmes on domesticated animals, and particularly to provide an estimate of an individual's genetic merit or so-called breeding value. Since the general principles were first published by Henderson (1950) it has evolved greatly in terms of general usage, something that has been particularly aided by computational advances (Lynch and Walsh 1998). Although the animal model has a long history in animal breeding context, its application in evolutionary studies of natural populations is much more recent (Reale et al. 1999; Kruuk et al. 2000).

In its simplest form the animal model aims to partition the observed phenotypic variance of a polygenic trait (i.e. a trait influenced by many genes of small effect) into its additive genetic basis and its environmental basis such that observation of individual *i* is expressed as

$$y_i = \mu_i + a_i + e_i \quad (1.5),$$

where y_i is the phenotypic trait value of individual *i*, μ_i is the fixed environmental effect such as age, sex etc of individual *i*, a_i is the additive genetic values of individual *i* and e_i is the environmental effects affecting individual *i*. The additive genetic value in the term a_i above represents the average additive effects of genes an individual receives from both parents and is its breeding value. Thus, $a_i = ½\ a_s + ½\ a_d + m_i$, where a_s and a_d is the breeding values of the sire (father) and dam (mother) respectively, and m_i is the Mendelian sampling variance (deviation of the individual's breeding value from the mean of the sire and dams breeding values). We can also express equation (1.5) in matrix notation:

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Zu} + \mathbf{e} \quad (1.6),$$

where **y** is the vector of observations, **b** is the vector of fixed effects, **u** the vector of additive genetic effects and **e** the vector of environmental effects. **X** and **Z** are the design matrices relating the phenotypic observations to the fixed and random effects respectively. This equation is a particular form of a linear mixed model and as such can be solved using REML methods (see Kruuk 2004 for more details).

It is not immediately obvious where the information about relatedness comes into equation 1.6, however there are two matrices that are not 'visible' in this equation but that are very important. Both of these matrices define the covariances of the random effects: we define the matrix **G** as describing the covariances among the breeding values (**u** vector) and, similarly, the matrix **R** describes the covariance among the residual errors (**e** vector). The matrix G is thus defined as

$$\mathbf{G} = V_A\mathbf{A} \quad (1.7),$$

where **A** is the additive genetic relationship matrix with each element being twice the coefficient of coancestry (the probability that the two alleles at a locus in a individual are identical by decent). It is thus this **A** matrix that we can obtain from the pedigree information.

The covariance matrix for the residual errors is defined as:

$$\mathbf{R} = V_E\mathbf{I} \quad (1.8),$$

where V_E is the environmental (residual) variance and **I** the identity matrix (1 on diagonal and zero elements on off-diagonal) which assumes that the errors are uncorrelated something that is not always the case.

Although these two covariance matrices do not appear in the equation for the mixed model (1.6), they are heavily involved when solving equation 1.6.

For example, the covariance matrix (**V**) for the vector of observations (**y**) is defined as:

$$\mathbf{V} = \mathbf{ZGZ}^{\mathbf{T}} + \mathbf{R} \quad (1.9),$$

and, for example, the estimates of the breeding values ($\hat{\mathbf{u}}$) or the so-called BLUPs (Best Linear Unbiased Predictors) can be obtained as:

$$\hat{\mathbf{u}} = \mathbf{GZ}^{\mathbf{T}}\mathbf{V} - 1(\mathbf{y} - \mathbf{Xb}) \quad (1.10).$$

More details about the animal model can be found in Lynch and Walsh (1998), Kruuk (2004) and Postma & Charmantier (2007).

There are some features that make the animal model particularly attractive: firstly, it, unlike the traditional approaches (see above), utilises all genetic relationships within the pedigree rather than, say, only that between sibs or parent-offspring relationships. Secondly, it can handle large and unbalanced datasets, something that is a frequent feature of studies on natural populations and it is easy to control statistically for various environmental effects (like age, sex, year of birth etc). The animal model also allows us to predict the breeding values of individuals which can be used to assess microevolutionary trends as a change in breeding values over time is equivalent to evolutionary change (Kruuk 2004). Finally, as we shall see in some of the chapters in this thesis, the animal model is also easily expanded to include more random effects, such as, for instance, repeated measurements (which can take into account the fact that error are not uncorrelated, see above) and maternal effects (a particular kind of environmental effect), interactions between random effects, higher order functions of the random effects and multiple traits.

1.6 How do populations respond to changes in the environment?

Two main mechanisms allow populations to respond to environmental changes: one is evolutionary change (change in allele frequency between generations) and the

other phenotypic plasticity (phenotypic adjustment). Traditionally, they have been viewed as two separate processes (Via et al. 1995) and, for convenience, I will also treat the two separately here.

1.6.1 Microevolution

I have already discussed how to measure selection and how we can estimate the heritability in natural populations. Together these two parameters can help us predict the expected response to selection through what is known as the breeder's equation (Falconer and Mackay 1996):

$$R = h^2 S \qquad (1.11),$$

where R is the response to selection in one generation, and h^2 and S the heritability and selection differential respectively. Therefore if a trait is heritable and under selection we expect the trait to evolve in the direction predicted by the selection differential.

The prediction of selection responses has been very successful in artificial breeding experiments (Falconer and Mackay 1996), but much less so when applied to natural populations (Merilä et al. 2001c). There are many reasons for this, including: biased heritability estimates, spatiotemporal variation in selection, biased selection estimates, selection on correlated traits, changing environmental conditions and lack of statistical power. These have all been extensively covered in Merilä et al. (2001c) and so will not be repeated here.

There are many different approaches to demonstrating evolutionary change, but in the context of this thesis the use of predicted breeding values is particularly interesting. As the breeding value is a measure of each individual's genetic composition, a change in mean breeding values over time can be taken to represent a

genetic change over time and hence as a demonstration that microevolution has taken place (Kruuk 2004). This approach has been used extensively during the last 8 years or so to examine microevolutionary change in a many different traits and species (e.g. Kruuk et al. 2001; Merilä et al. 2001a, b; Coltman et al. 2003; Charmantier et al. 2004; Garant et al. 2004; Wilson et al. 2007).

Recently some criticisms have been raised against this approach however, the main arguments focussing on the ability to reliably separate environmental and genetic effects in the prediction of breeding values and the problem of ignoring the uncertainty associated with the estimates.

Firstly, Postma (2006) pointed out that because the estimated breeding values (see equation 1.10 above) will always be more similar to the individual's phenotype than its true breeding value they will to some extent also include a environmental component. Therefore, regressing mean breeding values against time without explicitly incorporating the environmental component in the regression analysis will lead to biased results. In other words, he cautioned that information in the pedigree may not be sufficient to separate environmental and genetic year effects and thus interpreting the year trend in breeding values is difficult and requires careful formulation of a valid null hypothesis.

More recently, Hadfield and collaborators have also criticised the breeding value approach, both for using them to estimate selection on the genetic component of a trait (Hadfield 2008) and for assessing microevolutionary trends (Hadfield et al. in press). Here however I focus only on the criticism put forward for using breeding values to assess microevolutionary trends. The main argument Hadfield et al. put forward is that the often large uncertainty in the predicted breeding values is often ignored in analyses of microevolutionary trends, something that will lead to anti-

conservatism. That is, it will tend to give a *P*-value that is too low and thus lead researchers to be overly confident that evolutionary change has taken place. As a remedy they advocate the use of a Bayesian animal model and use of the posterior mode of the slope to estimate the genetic trend over time (Hadfield et al. in press). This should give the same estimate of the annual change (slope) but will properly take into account uncertainty in the predicted breeding values and thus provide more reliable *P*-values. In accordance with this Hadfield et al. (in press) re-analysed two previous studies (Garant et al. 2004; Wilson et al. 2007) that had demonstrated a microevolutionary trend and found that with the new Bayesian approach the trends were no longer significant.

1.6.2 Phenotypic plasticity

Phenotypic plasticity is a genotype's change in phenotype across an environmental gradient (Scheiner 1993) and as such is a within-individual adjustment to environmental heterogeneity. Although phenotypic plasticity was first recognised intuitively as early as around 1900 (Baldwin 1896) it took another 40 years until researchers became particularly interested in plasticity, and it has recently received renewed interest (Schlichting and Pigliucci 1998; Lande 2009). The presence of plasticity within a population is thought to result in more rapid adjustment toward the optimal phenotype and thus to be an important mechanism for dealing with environmental heterogeneity. A central question is whether plasticity itself can be considered as a trait which would thus be able to evolve in response to selection. This, of course, requires that plasticity has a genetic basis (or, equivalent, that there is GxE) and there are two main approaches to measuring the genetic basis of phenotypic plasticity (Via et al. 1995). The 'character state approach' treats the same

trait as different traits in each environment, an idea originally proposed by Falconer (1952) whereas with the 'reaction norm approach' plasticity is measured using continuous covariance functions (Kirkpatrick et al. 1990). It is important to realise however that the two methods are interchangeable when the environmental variation is discrete (Via et al. 1995; Kruuk et al. 2008).

Scheiner and Lyman (1989) claimed that it is only possible to estimate the genetic basis of phenotypic plasticity with an experimental approach, but much has happened since. Kruuk (2004) first suggested that genotype-environment interactions could be examined within a animal model framework and in a recent review Nussey and colleagues (2007) summarised the framework available to analyse the genetic basis of phenotypic plasticity using data from longitudinal studies of natural populations.

The key point for analysing the genetic basis of plasticity is that genotypes must be replicated and distributed across the environmental gradient. Most longitudinal studies fulfil this criteria to varying degree. Secondly, we need information about the phenotype of related individuals and their genetic relationship. This information is often available from long term studies where animals (or plants) are individually marked (Pemberton 2008), something which make it possible to use the animal model to examine the genetic basis of variation in plasticity (Nussey et al. 2007).

A central point here is to examine if there is individual variation in the reaction norm slopes, something that can be achieved by using a so-called 'random regression animal model'. These models utilize covariance functions to estimate covariances between the regression coefficients (Meyer 1998) in an animal model framework (Lynch and Walsh 1998; Kruuk 2004). The individual breeding values

can thus be modelled as linear (or higher order) functions along some continuous environmental measure and we can thus estimate the genetic variance for slopes in the population. If there is genetic variation in plasticity, i.e. a GxE interaction, this can lead to changes in the amounts of genetic variance (and covariance) expressed over the environment (Scheiner 1993) and as such play an important role in shaping phenotypic evolution in natural populations (Kruuk et al. 2008).

How do GxE interactions come about? There are two main molecular genetic mechanisms that has been proposed to be responsible for GxE interaction (Via et al. 1995; Mackay and Anholt 2007). The 'allelic sensitivity' hypothesis propose that alleles have varying effect on the phenotype in different environments and the 'gene regulation' hypothesis suggests that environmental change leads to differences in the expression of genes that influence the trait. Unfortunately, we are unable to separate between the two with traditional quantitative genetics, but the increasing use of genomics should alleviate this to some respect. Indeed, it has been proposed that 'understanding the genetic basis of GxE interactions is the next frontier in the analysis of complex traits' (Mackay and Anholt 2007).

1.7 Study species and populations

In order to understand how environmental variation influences phenotypic and evolutionary patterns it is important to have data collected in years when environmental conditions have been different. In this thesis I use long-term studies of great tits (*Parus major*) to address these issues. The great tit is a small (14 - 22 g), mainly resident, passerine bird species abundantly distributed throughout Europe, North Africa and some parts of Asia (Gosler 1993). It normally breeds in natural

holes, but readily accepts nest boxes for breeding which make it ideally suited for long-term studies. It forms mainly monogamous pairs with egg laying normally commencing in early April in most of Western Palaearctic, although it can be up to a month later in the most northern parts of its geographic distribution. Laying dates in the great tit are very variable between years, something that is largely determined by annual spring temperatures (van Balen 1973; Perrins and McCleery 1989). There is currently great interest in understanding the proximate reasons behind this variation (Visser et al. 2009a) and the phenotypic and genetic basis of between individual variation in how birds adjust their laying date (Brommer et al. 2005; Nussey et al. 2005; Charmantier et al. 2008, Chapter 4 in this thesis). Clutch size is on average 9 eggs (range 6 - 15) with only the female incubating, but both sexes providing parental care. The main diet consists mainly of insects (and in particular Lepidoptera) in summer, but they normally feed on seeds and nuts in winter. For chick feeding Lepidoptera are particularly important and often comprise > 80 % of the diet.

The Netherlands Institute of Ecology (NIOO) has been monitoring a number of great tit nest box populations for several decades, some of which (Hoge Veluwe, Oosterhout, Liesbos and Vlieland) have been continuously monitored since 1955 (van Balen 1973). In each population, annual information on individuals' laying date, clutch size, number of fledglings and number of clutches has been collected. In addition, parents are caught and ringed during chick feeding and the juveniles are also ringed. As a result, we can reconstruct individual pedigrees in great detail and obtain relatively accurate fitness measurements for all individuals. This pedigree information forms the basis of all quantitative genetic estimates in this thesis and was therefore treated in greater detail in the section on animal models above and will be discussed more in the discussion.

Another interesting long-term time series that deserves special mention is that on caterpillar frass (droppings) that have taken place, mainly in Hoge Veluwe from 1985 onwards (Visser and Holleman 2001), but also to a lesser extent in some of the other populations. By collecting and measuring the frass it is possible to estimate the maximum caterpillar biomass and thus the date when caterpillars where most abundant. This information can subsequently be used to calculate the difference in days between when the great tit chicks require most food and the peak in caterpillar abundance, something that is referred to as 'mismatch' (Visser et al. 1998). I will return to the use of this measure in Chapter 2.

1. 8 Goals and outline of this thesis

The goal of this thesis is to obtain a better insight into how ecological processes influence phenotypic and evolutionary dynamics in natural populations. In particular I focus on the effects that changes in one particular ecological variable, temperature, has had on populations of great tits in their natural habitats. Global temperatures have risen by around 0.6 °C since 1900 (Houghton et al. 2001) and this temperature increase has had widespread and well documented effects on both great tit populations (e.g. Visser et al. 1998; Husby et al. 2009; Charmantier et al. 2008) and other systems (reviewed in Parmesan (2006); see also beginning of this chapter). The increase in temperature is furthermore expected to continue and it is therefore important to understand how natural populations will be able to adjust to these changes. Importantly different populations may respond differently. For instance, although many bird populations have advanced their timing of breeding (e.g. Crick et al. 1997) as a direct consequence of the increase in temperatures, other populations have not changed their timing of breeding appreciably (Visser et al. 1998).

Understanding such discrepancies is a major goal for biologists and can only be achieved by a detailed and integrative approach. In this thesis I therefore examine how changes in environmental conditions have influenced phenotypic and genetic variation in both morphological and life history traits using different long-term studies of great tits.

In particular, I examine the consequences that changing environmental conditions, and specifically increasing temperatures, has had on: the proportion of females producing a second clutch (Chapter 2), body size (Chapter 3), the multivariate patterns of phenotypic plasticity in lay date and clutch size and their genetic basis (Chapter 4), and the covariance between expression of genetic variance and selection on laying date (Chapter 5). Below I outline in some more detail the questions I address in the individual chapters of this thesis.

It is well documented that timing of breeding/flowering has advanced in many species of birds and plants (e.g. Parmesan and Yohe 2003). Much less attention has been paid to other traits, such as the number of reproductive events in a breeding season. In many short-lived species of birds the number of clutches a female produces is closely related to the reproductive success and as such an important trait that, ultimately, will determine the population viability. In **Chapter 2** I explore the causes behind a decrease in the proportion of females producing a second clutch in four long term (1955 – 2004) study populations of great tits in the Netherlands.

Increasing spring temperatures can also have more direct influences on traits than through their effect on food dynamics. For example, comparisons between populations have shown that, in many homoeothermic vertebrates, body size correlates negatively with external temperatures, something that is known as 'Bergmann's rule' (Bergmann 1847; Mayr 1956). In **Chapter 3** I use long term data

(1979 – 2008) on three populations of great tits to examine if increasing temperatures have led to a within-population decline in body size as predicted by Bergmann's rule. Long-term and more rapid changes in body size as a result of temperature fluctuations have been shown in many bird (Yom-Tov 2001) and mammal species (Smith et al. 1995). However, these changes are often inferred to be a result of evolutionary change despite little or no evidence to support these claims. We use the animal model to estimate the genetic basis of two body size traits, mass and tarsus length, in these three populations and assess the rate of microevolutionary change using both the breeding value approach and the newly proposed Bayesian approach (see above).

As discussed earlier, another important mechanism by which populations may adapt to changes in the environment besides evolutionary change is phenotypic plasticity, a genotype's change in phenotype across an environmental gradient (Scheiner 1993). There is currently considerable interest in individual variation in phenotypic plasticity and its genetic basis (Brommer et al. 2005; Nussey et al. 2007; Brommer et al. 2008; Charmantier et al. 2008; Husby et al. provisionaly accepted) as it is becoming increasingly clear that environmental changes are rapidly causing a change in the phenology in a wide variety of taxa (Parmesan 2006). Longitudinal studies offer a unique opportunity to study individual variation in phenotypic plasticity, as well as, in systems where pedigree information exists, the genetic basis to such plasticity (Nussey et al. 2007). Although several recent studies have examined the intra-individual variation in plasticity and its genetic basis, comparison between the different studies is hampered by the use of different environmental variables, statistical approaches and data structure. In **Chapter 4** I address these issues by directly comparing the phenotypic and genetic variation in phenotypic

plasticity for two important life-history traits, clutch size and laying date, using two long-term study populations of great tits from Wytham Woods, UK and Hoge Veluwe, Netherlands.

Genotype by environment interactions (GxE's) will lead to changes in additive genetic variance across the environment (Kruuk et al. 2008). Similarly, the annual strength of selection may also vary with prevailing environmental conditions (Endler 1986). As the response to selection is determined by the heritability and the strength of selection acting on the trait, an interaction between annual measures of h^2 and selection may act to either 'fuel' or hamper an evolutionary response. How the two parameters covary in natural populations is however still very poorly understood (but see Wilson et al. 2006). In **Chapter 5** I examine the dynamics between environmentally induced changes in heritability of laying date and strength of selection and in particular how their interaction depends on changes in spring temperature.

Finally, in **Chapter 6**, I summarise and discuss the insights gained as a result of the findings presented in this thesis and focus particularly on new questions that this work has raised. I will also briefly present what I believe is the future of quantitative genetic studies in the wild and their co-existence along molecular genetics.

CHAPTER 2

Decline in the frequency and benefits of multiple brooding in great tits as a consequence of a changing environment.

Husby, A., Kruuk, L. E. B. & Visser, M. E. 2009. *Proceedings of the Royal Society B-Biological Sciences* **276**: 1845-1854.

2.1 Summary

For multiple brooded species, the number of reproductive events per year is a major determinant of an individual's fitness. Where multiple brooding is facultative, its occurrence is likely to change with environmental conditions, and as a consequence, the current rates of environmental change could have substantial impacts on breeding patterns. Here we examine temporal population-level trends in the proportion of female great tits (*Parus major*) producing two clutches per year ("double brooding") in four long term study populations in The Netherlands, and show that the proportion of females that double brood has declined in all populations, with the strongest decline taking place in the last thirty years of the study. For one of the populations, for which we have data on caterpillar abundance, we show that the probability that a female produces a second clutch was related to the timing of her first clutch relative to the peak in caterpillar abundance, and that the probability of double brooding declined over the study period. We further show that number of recruits from the second clutch decreased significantly over the period 1973 - 2004 in all populations. Our results indicate that adjustment to changing climatic conditions may involve shifts in life history traits other than simply timing of breeding.

2.2 Introduction

Many species differ not only in the number of offspring they produce, but also in the number of breeding attempts per season, such that multiple breeding (more than one reproductive event in a season) is a common reproductive strategy in a variety of taxa (Verhulst et al. 1997 and references therein). Whenever reproductive costs exist, life-history theory predicts that parents face a compromise between current reproduction and future reproduction in order to maximize their own fitness

(Williams 1966; Stearns 1992). Long-lived species are expected to favour their own survival at the expense of their current brood of offspring, whereas short lived species should invest more in the current breeding attempt (Drent and Daan 1980). This trade-off is often invoked to explain costs of reproduction across seasons, but also holds for within season reproductive decisions such as how many breeding attempts to make within a year. Furthermore, it is likely that prevailing environmental conditions and hence resource availability shape these costs and hence the likelihood of double brooding. In this paper we use long-term, individual-based data on four populations of great tits (*Parus major)* in the Netherlands, to test whether current changes in environmental conditions are affecting breeding patterns in relation to double brooding.

Several studies of birds have investigated the intra-seasonal costs of multiple breeding, most of them experimentally (Lindèn 1988; Verboven and Verhulst 1996; Verhulst et al. 1997; Verhulst 1998; Brinkhof et al. 2002; Parejo and Danchin 2006), but some also using either longitudinal studies (Tinbergen et al. 1985) or a combination of the two (Verboven and Verhulst 1996; Verboven et al. 2001). Experimental studies of clutch size (e.g. Parejo and Danchin 2006) and brood size (e.g. Lindèn 1988) manipulations are normally used to investigate the determinants of multiple breeding, but experimental delay/advance of hatching date (e.g. Brinkhof et al. 2002) is also frequently used. These studies show that delaying hatching date, as well as increasing clutch size and/or brood size, commonly leads to a lower probability of initiating a second clutch. These findings suggest that differences between populations in the occurrence and extent of multiple breeding may be causally related to differences in mean laying date and/or number of fledglings in the nest, which again might be linked to differences in habitat between populations.

In short lived species, differences in annual fecundity play a major role in determining population growth (Sæther and Bakke 2000). Furthermore, many such species often have two or more breeding attempts per season and the variance in individual fecundity can often be better explained by the number of breeding attempts than by the number of young produced from each breeding attempt (Klomp 1970; Nagy and Holmes 2005 and references therein; Weggler 2006). Understanding what factors determine the decision to initiate multiple breeding attempts per season is therefore interesting not only from a life-history point of view, but also from a conservation perspective as it determines population growth rate and thus the future viability of a population.

In many taxa the main variable determining reproductive success is the abundance of prey items (for a review see White 2008): for example, long-term individual-level studies of great tit populations have shown that the synchrony of breeding with the peak in caterpillar abundance is the primary determinant of reproductive success (Perrins 1970; van Balen 1973; Verboven and Visser 1998; Visser et al. 2006). Because of the impact increasing spring temperatures have on the timing of the peak in caterpillar abundance, and the close link between temperature and timing of reproduction in birds, there has recently been great interest in such systems as they provide an ideal way in which to test the impact of climate change on natural populations (e.g. Visser et al. 1998; Visser et al. 2003; Gienapp et al. 2005; Both et al. 2006; Visser et al. 2006; Charmantier et al. 2008). As a consequence of the warming spring temperatures many bird species have advanced their laying date (Crick et al. 1997; McCleery and Perrins 1998), but see also Barbraud & Weimerskirch (2006) and Visser et al. (1998; 2003). For instance, great tits breeding in Wytham Woods near Oxford, UK have advanced their laying date by

around 14 days over the past 47 years (1961-2007) (Charmantier et al. 2008). Furthermore, the observed advance in laying dates in long term studies often show a 'broken stick' pattern where there is little or no change in laying dates in the period from the 1950s to 1970s, but in the later period from the 1970s onward there is a strong advancement. This pattern furthermore coincides with a similar pattern of increase in spring temperatures (e.g. McCleery and Perrins 1998), again emphasizing the importance spring temperature has on the timing of reproduction in birds.

Based on the negative relationship between an individual's laying date and the probability of producing a second clutch (e.g. Verboven and Verhulst 1996; Brinkhof et al. 2002), we might therefore expect that the proportion of females producing two clutches per season should increase as laying dates become increasingly earlier. This prediction is, however, in marked contrast to the observed patterns in populations of great tits in The Netherlands (Visser et al. 2003), which show a decline in the proportion of females producing a second clutch.

The aim of this paper is to understand the reasons behind this decline in double brooding in four geographically separated populations of great tits in the Netherlands, using data from over a 50 year period (Table 2.1). Establishing the causes of the decline is important for understanding the effects a changing environment can have on natural populations. Timing of breeding relative to the peak in caterpillar abundance is an important predictor of the likelihood of initiating a second clutch in this species, and here we examine if changing climatic factors has caused a shift in the relationship between the likelihood of double brooding and relative timing of breeding over the course of the study. Because spring temperature changes have been particularly pronounced in the last three decades, we analysed the time series 1955 – 2004 and 1973 – 2004 separately. Where we find a significant

decline in proportion of females double brooding, we also tested the hypothesis that the decline reflects changing selection patterns, i.e. that the benefits of double brooding relative to the costs have decreased over the study period.

2.3 Materials and methods

2.3.1 Study area, field procedures and data

Data were collected at four different localities in The Netherlands, Vlieland (VL), Hoge Veluwe (HV), Oosterhout (OH) and Liesbos (LB) (Table 2.1). Because a storm damaged the pine plantation in the Hoge Veluwe population in the winter of 1972/73, and nest boxes were subsequently re-located, we treated HV1 (1955-1972) and HV2 (1973 - 2004) as two (temporally, not spatially) separate populations. For more details about the study populations see van Balen (1973).

In all areas nest boxes were visited at least once every week during the breeding season (April – June). Population size in an area in a given year was defined as the number of first clutches. The laying date of the first egg of the clutch was calculated from the number of eggs found during the weekly checks, assuming that one egg was laid per day. Number of eggs and/or young in the nests was counted, and when the young were 7 – 10 days old the parents were caught on the nest using a spring trap. Parents already ringed were identified and unringed birds were given a metal ring with a unique number. Young were ringed at day 7 - 10 (HV1, HV2, OH, LB) or at day 10-15 (VL). Females that were unknown (i.e. females not captured) were excluded from the analysis, and we also excluded nestlings from such nests because of the missing maternity. In total 224 recruits were produced from nests in which the female was unknown (for the entire study period

and all populations). Because of the small number (224 out of 9510, 2.3 %), it is highly unlikely that excluding such nests will bias the results in any way. Laying dates are presented as the number of days after 31st March (day 1 = 1st April, day 31 = 1st May).

We used the mean of daily average temperatures from the period 1st of March until 20th April from the De Bilt meteorological station of the Royal Netherlands Meteorological Institute (KNMI), for consistency with other studies on the same study populations (see e.g. van Balen 1973; Gienapp et al. 2006).

In some population/ year combinations only a small proportion (< 50 %) of the adults were caught and ringed and this makes it difficult to estimate any reasonable survival and/or recruitment rates for these population/year combinations, hence they were excluded from the analysis. We further excluded years if large-scale experiments (>70 % of the population manipulated) were carried out that affected parental survival or recruitment probability. Excluded population/year combinations are given in Table 2.1. In all years, we also excluded individual nests in which manipulations took place (except for the viability selection analysis, see below).

Several studies have emphasized the importance of timing of breeding relative to food abundance for the probability of producing a second clutch (e.g. Verboven et al. 2001; Brinkhof et al. 2002; Nagy and Holmes 2005), hence where we had the necessary data we used the timing of an individual's first clutch relative to the peak in caterpillar abundance (or "mismatch") to predict the individual probability of producing a second clutch (see below); see also Verboven et al. (2001) for a similar approach. Caterpillar peak dates have been collected during the period 1985 – 2004 for the Hoge Veluwe population (Visser et al. 2006) and we used these data to estimate the temperature period (i.e. average temperatures during a given time

interval) that gave the best prediction of caterpillar peak date (see also Visser et al. (1998)). The period with the highest r^2 value was 8th March – 17th May with a predicted caterpillar peak of 105.471 – 5.968* temperature ($r^2 = 0.78$, Visser et al. 2006). We subsequently used this relationship to estimate the caterpillar peak in Hoge Veluwe for years in which we had no information (see Visser et al. 2006 for more details). The difference in days between when the chicks from the first clutch are 12 days old, and demand most food, relative to the estimated peak date in caterpillar abundance, was used as an approximation to the "mismatch" experienced by the birds (see Verboven et al. 2001).

2.3.2 Spatiotemporal variation in proportion of females producing second clutches

We tested for differences among populations and changes with time in the proportion of females producing two clutches by defining a second clutch as a clutch produced following a successful first clutch by the same female, i.e. only nests where at least one chick fledged from the first clutch were included in the analyses. Thus all replacement clutches (where the first clutch failed and the pair produced a new clutch) were excluded (see Verboven and Verhulst 1996; Verboven et al. 2001; Visser et al. 2003). First and second clutches were matched on the basis of female identity as in Verboven & Verhulst (1996), thus we can be absolutely certain that it was the same female who produced a second clutch.

We first analysed population-level trends in multiple breeding for the two periods 1955 – 2004 and 1973-2004. We defined the proportion of females producing a second clutch in a given population in a given year as the ratio of number of ringed females producing a second clutch divided by the total number of

ringed females (note that females that had a replacement clutch are excluded). We included the proportion of females double brooding as a response variable in a generalized linear model, GLM (quasibinomial family argument to correct for overdispersion, see Table 2.2) with the following terms: population as a factor and year, temperature (1^{st} March – 20^{th} April, see above), annual population density (mean centred), the population annual mean laying date of the first clutch and its quadratic (to test for a non-linear relationship), the population annual mean clutch size of the first clutch and its quadratic and the variance in the laying date of the first clutch as continuous covariates. We also included the two-way interaction between population and year in the model to test for spatiotemporal trends.

2.3.3 Individual level analyses

As it is only for the HV2 population we have substantial data on caterpillar peak dates, we used this population to investigate in detail the variables determining the probability of an individual starting a second clutch, and whether these had changed over time. We defined whether an individual produced a second clutch or not as a binomial trait in a logistic regression mixed model (GLMM) with year and mismatch of the first clutch (see above) as continuous covariates and the two-way interaction in the model (thus testing for a temporal change in the relationship between the probability of double brooding and the amount of mismatch experienced). Female identity and year (as a factor) were included as random effects to account for repeated measures of females and repeated measures within years. It has been demonstrated previously that mismatch is a better predictor of the proportion of females starting a second clutch than absolute timing (Verboven et al. 2001) and thus we only used mismatch for this analysis.

2.3.4 Selection analyses

As we found no indication that the rate of decline in the proportion of females double brooding differed between populations for the period 1973-2004, we quantified selection on double brooding by testing for associations with female fecundity and survival in all four populations jointly. An offspring was classified as a recruit to the breeding population if it was seen again in the population in subsequent years after its year of birth. For the fecundity selection analysis, we estimated the fitness of an individual female from her annual reproductive success, defined as the number of offspring recruiting to the breeding population from each breeding season. We then tested for differences between single versus double brooders in their annual reproductive success. For the viability selection analysis, we used the survival of a female to subsequent years to test whether double brooding affected adult female survival rates. Estimates of fecundity and viability selection were based on recapture data under the assumption that nestlings and adults not returning to the study area in subsequent years had died. The use of recruitment as a measure of fecundity selection represents a reasonable fitness measure relative to other broods in the same year as it is only those individuals who recruit that will contribute to any response to selection.

When analyzing selection on adult female survival we included females that had been manipulated (note however that females who were removed or had their partner removed were excluded from the dataset altogether (n = 2, 2 and 7 individuals in the HV, LB and VL population respectively (years as in Table 2.1)) as there is no indication that experimental manipulations, such as clutch size manipulations, influence adult survival in this species (Tinbergen and Both 1999).

Consequently, sample sizes for the viability analysis (n = 5468) are higher than for the fecundity analysis (n = 4475). Because of repeated measurements of the same female over time, we included female identity and year as random effects (as factors) in a GLMM model with female survival as response variable (0/1). Breeding category was included as a two level factor (single brooded vs. double brooded) in the analysis in order to determine if fitness varied between single versus multiple breeding individuals.

All selection models reported here are without laying date fitted as covariate, as we wanted to consider selection on double brooding without removing any of the associated variation in laying date. However, we also repeated the selection models including laying date to control for the laying date associated variation in reproductive success and survival. Although laying date had a significant effect on fecundity (b = -0.017, se = 0.003, χ^2 = 30.19, d.f. = 1, P < 0.0001), as well as on viability (b= -0.009, se= 0.004, χ^2 = 4.18, d.f. = 1, P = 0.041), its inclusion in the models did not qualitatively change the conclusions, and thus we do not report the results from the laying date corrected selection models here. We did, however, correct for population density (mean centred), defined as the population specific number of breeding females, in all selection analysis reported here.

2.3.5 Recruitment from first and second clutch

Differences in total annual reproductive success between single and multiple breeding females could be due to additional recruitment from the second clutch, or due to differences in recruitment from the first clutch, implying systematic differences in the type of birds that produce two clutches rather than a benefit of the second clutch *per se*. To test for this, and to explore possible explanations for the

decline in second brooding over time, we firstly compared recruitment from the first clutch between females who proceeded to have a second clutch and those that did not, and secondly, amongst those that did produce a second clutch, we considered the number of recruits produced from the second clutch only.

For the analysis of recruitment from the first clutch we fitted a GLMM with number of recruits from the first clutch as response variable and included breeding category (single vs. double brooder) as a factor and year as a covariate as well as the interaction between breeding category and year. Female identity and year were included as random factors.

In order to determine the contribution of the second clutch to parental reproductive success in a given year, and whether this changed over the course of the study, we also analysed how the number of recruits from the second clutch had changed over time including year as continuous covariate. Again, female identity and year were included as random factors (see above).

2.3.6 Statistical analysis

Statistical significance was estimated from the GLMMs with the appropriate error structure. Thus we used binomial error structure for both the analysis of the probability of double brooding and for the viability selection analysis. Similarly, Poisson error structure was applied to the fecundity selection analysis and for the analysis of number of recruits from the first and second clutch. All models were fitted using Schall's technique (Schall 1991) and significance levels of variables were assessed from their Wald test statistics, distributed as χ^2 on the appropriate degrees of freedom (Sokal and Rohlf 1995; Gilmoure et al. 2006). We used ASREML version 2.0 (Gilmoure et al. 2006) as a plug-in to R version 2.7.0

(RDevelopmentCoreTeam 2007) for all analysis except the GLM models which were run directly in R. For the GLM models significance of terms was assessed using the change in deviance between the reduced and complete model and tested against the *F* - distribution (because of over-dispersion, see Crawley (2002)) with degrees of freedom equal to the difference in number of parameters between the two models (Crawley 2002).

In general we fitted a global model containing all explanatory variables of interest as well as their interactions. A final model was then determined by step-wise exclusion of the least significant terms, starting with the non-significant highest order interactions and then non-significant main effects. Significance of main effects was tested separately without any interaction effects fitted.

TABLE 2.1. General information of the study populations including study period, sample size (number of breeding birds for each population), number of second clutches, the total number of recruits from first clutch, the total number of recruits from second clutch and years excluded (due to large scale manipulations or low proportion adults ringed).

Study site	**Full name**	**Study period**	**Nr of breeding birds**	**Nr of second clutches**	**Nr recruits first clutch**	**Nr recruits from second clutch**	**Year(s) excluded**	**Coordinates (long./lat.)**
HV1	Hoge Veluwe	1955-1972	1520	431	1159	82	1955-58	52° 05' N, 05° 50' E
HV2	Hoge Veluwe	1973-2004	3831	723	2636	234	1973 (re-organising of study area, see Methods)	52° 05' N, 05° 50' E
LB	Liesbos	1955-2004	1570	176	536	20	1955-57, 1967-75, 1982, 1991, 1996, 1998-99	51° 35' N, 04° 40' E
OH	Oosterhout	1955-2004	1205	96	799	16	1956-66, 1971, 1988, 1990-92, 1996-97	51° 55' N, 05° 50' E
VL	Vlieland	1955-2004	4380	1287	3540	488	1955, 1960-65, 1967-69, 1971-75, 1990 - 2004	53° 15' N, 05° 00' E

2.4 RESULTS

2.4.1 Spatiotemporal trends in proportion of females producing second broods

For the period 1955 – 2004 the proportion of females producing second clutches decreased in all the four study populations, but the rate varied significantly between populations, resulting in a significant interaction term between population and year (Table 2.2A, Fig. 2.1). There was a quadratic relationship with the mean lay date of the population and a positive relationship with the mean clutch size in the first clutch (see Table 2.2A). Furthermore, the proportion of females producing second clutches declined with both increasing population density and increasing temperature. We found no indication that the proportion of females producing a second clutch was related to the variance in laying date ($F_{1,113} = 2.29$, $P = 0.13$) nor of any quadratic relationship with the mean clutch size ($F_{1,112} = 0.06$, $P = 0.80$).

For the period 1973 – 2004 we did not find any significant interaction term between population and year ($F_{3, 84} = 1.514$, $P = 0.22$), suggesting that for this period all populations showed a significant negative decline (see Table 2.2A/B). Again, we did not find any indication that the proportion of females producing a second clutch was related to the variance in laying date ($F_{1, 85} = 1.65$, $P = 0.20$) nor of any quadratic relationship with clutch size ($F_{1, 78} = 0.039$, $P = 0.85$). Furthermore there was no quadratic relationship with mean laying date for this period ($F_{1, 86} = 3.32$, $P = 0.07$).

To investigate the significant interaction term between population and year, i.e. to study the spatiotemporal double brooding patterns in more detail, we used population specific models (see Table 2.2B), correcting for population density, for each of the two periods. Although temporal trends were negative for all populations for both time periods, only the HV1/2 population showed a significant temporal decline over the period 1955 – 2004 (Table 2.2B). For the period 1973 – 2004

however, there was a significant decline in proportion of females double brooding for the HV2, LB and OH population (Table 2.2B).

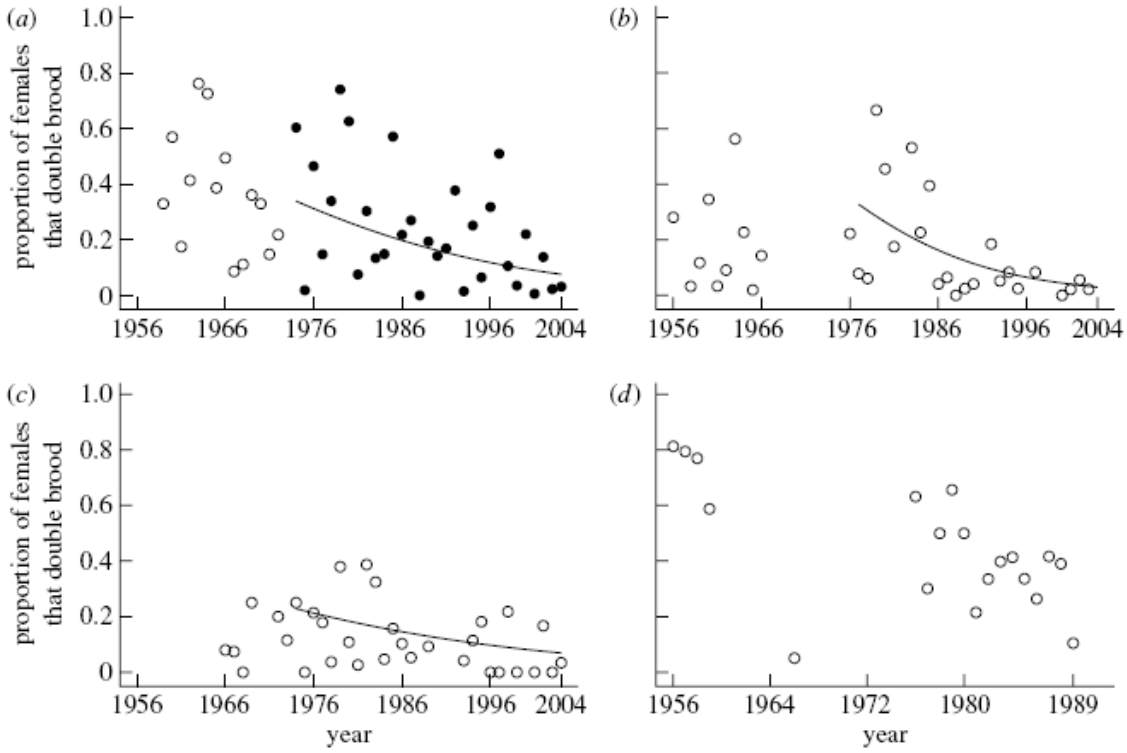


Figure 2.1. Temporal trends in proportion of females double brooding in the four study populations. The curves were fitted using the logistic equation from generalized linear models separately for each population (see Table 2.2B for equations on logit scale).

Table 2.2A. Analysis of variance table ('type 3') for the minimal adequate model from a generalized linear model (GLM) for the population-level analysis of the proportion of females second brooding produced each year, all populations combined, for the period 1955 -2004 and 1973 -2004. The analysis was corrected for over-dispersion ($\varphi = 6.19$, $\varphi = 5.85$ for the period 1955- 2004 and 1973 - 2004 respectively), and has a total of 129 (1955 – 2004) and 95 (1973 – 2004) population year combinations; see text for further details. Note that ff the same model is fitted excluding temperature, significance of year increases (χ^2=23.355 and χ^2= 48.425, for the '55 – '04 and '73 – '04 period respectively).

Parameter	**Deviance**		**d.f.**	***P*- value**		**β ± se**	
	'55- '04	*'73- '04*		*'55 – '04*	*'73 – '04*	*'55 –'04*	*'73 –'04*
Population	135.501	86.909	4(3)	< 0.001	< 0.001		
Year	23.200	37.891	1	< 0.001	< 0.001		-0.066 (0.011)
Av. laying date	3.368	36.418	1	0.066	< 0.001	0.220 (± 0.123)	-0.132 (0.023)
Av. laying date ^{2}	8.564		1	0.003		- 0.007 (± 0.003)	
Av. clutch size	20.734	17.630	1	< 0.001	< 0.001	0.546 (± 0.122)	0.529 (0.127)
Population density	18.862	8.155	1	< 0.001	0.004	- 0.014 (± 0.003)	-0.009 (0.003)
Temperature	18.300	18.581	1	< 0.001	< 0.001	- 0.35 (± 0.084)	-0.395 (0.095)
Population x Year	16.152		4(3)	0.003		See Table 2B	See Table 2B

Table 2.2B. Population specific GLMs for temporal change in proportion of females double brooding in each year corrected for population density (mean centred, estimates not given) for each of the two time periods. Note that the coefficient estimates of year given here are on logit scale and are the ones used to draw the lines in Fig 2.1. Also note that some populations have some years within the given time periods excluded (see Table 2.1 for full details).

Population	**Deviance**		**β ± se**		***P* -value**	
	'55 –'04	*'73 –'04*	*'55 –'04*	*'73 –'04*	*'55 –'04*	*'73 –'04*
Hoge Veluwe	5.118	9.101	-0.024 (0.011)	-0.061 (0.021)	0.024	0.002
Liesbos	2.266	9.905	-0.024 (0.016)	-0.102 (0.036)	0.132	0.002
Oosterhout	1.352	4.277	-0.022 (0.019)	-0.046 (0.024)	0.245	0.039
Vlieland	0.092	0.787	-0.012 (0.040)	-0.051 (0.058)	0.762	0.375

2.4.2 Temporal change in the individual probability of producing a second clutch in the HV2 population

For the HV population, where we have data on caterpillar peak dates, the probability of an individual female starting a second clutch was negatively related to the difference in timing of her first clutch and the caterpillar food peak (Table 2.3, Fig. 2.2A). The probability of producing a second clutch was also negatively related to year (see Table 2.3), indicating that the probability of double brooding had decreased over the course of the study, even after controlling for the effects of individual mismatch (Fig 2.2B). There was, however, no indication that the slope between the probability of producing a second clutch and the degree of mismatch had changed significantly over time, as the interaction between year and mismatch was not significant (Table 2.3). Thus, the intercepts had declined over time whereas the slopes had not (Fig. 2.2B; note that the lines are back-transformed and if plotted on logit scale the lines would be parallel with a decline in intercepts over time). There was also no indication of any non-linear relationship with mismatch (quadratic term for mismatch: $\chi^2 = 2.05$, d.f. = 1, $P = 0.15$).

Table 2.3. Individual probability of double brooding for the HV2 population in relation to the amount of mismatch experienced and temporal trend from a GLMM with binomial error structure (see Methods). Note that these are the Wald statistics from a full model, all main effects were significant when tested separately and main effect estimates are from these models (see main text, $n = 2153$).

Term	χ^2	**d.f.**	***P***	**Estimate ± s.e.**
Year	6.92	1	0.009	-0.073 ± 0.029
Mismatch	148.27	1	< 0.001	-0.188 ± 0.015
Year x Mismatch	0.01	1	0.92	-0.001 ± 0.002

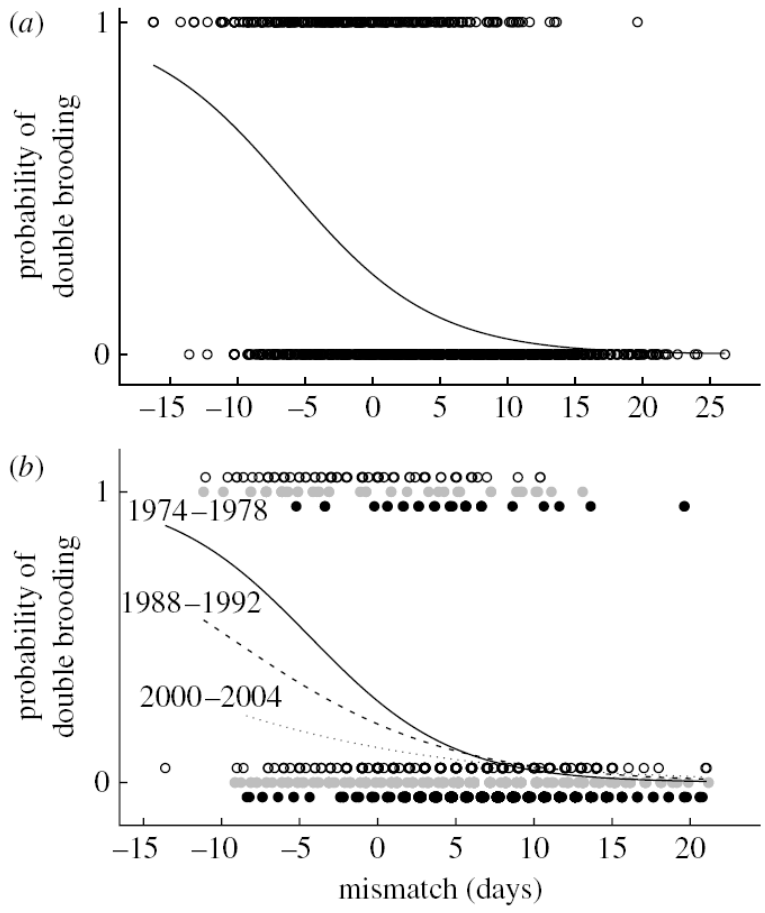


Figure 2.2a) The probability of producing a second clutch in relation to how mismatched an individual was to the food peak for the HV2 population [Equation on logit scale: -1.15 – 0.188*mismatch] fitted in a binomial GLMM. Negative values of

mismatch indicate that the period in which the brood requires large amount of food occurs before the seasonal peak in caterpillar abundance. Positive values indicate that the period was after the caterpillar peak. **Figure 2.2b)** The probability of producing a second clutch in relation to the amount of mismatch experienced for the first 5 years of the study (open circles, solid line, equation on logit scale: -0.944 – 0.219*mismatch), the mid 5 years (grey circles, dashed line, equation on logit scale: -1.379 – 0.145*mismatch) and the last 5 years (closed circles, dotted line, equation on logit scale: -1.992 – 0.993*mismatch) for the HV2 population. Each prediction line is restricted to the data range for the respective period. Equations are from a binomial GLMM for 5 year periods; note that predictions are back-transformed and if plotted on logit scale the lines would be parallel.

2.4.3 Viability selection analysis

Combing data from all four populations for 1973-2004, we did not find any significant difference in survival between females who had been single brooded versus double brooded in a given year ($\chi^2 = 0.69$, d.f. = 1, $P = 0.41$). There was weak indication of a temporal decline in survival ($\chi^2 = 3.36$, d.f. = 1, $P = 0.07$), but no significant interaction between breeding category and year ($\chi^2 = 1.95$, d.f. = 1, $P = 0.16$), suggesting that survival between single and multiple breeding females had not changed differently over time. Nor did survival between single and double brooded females differ in different populations (population-breeding category interaction: $\chi^2 = 6.87$, d.f. = 3, $P = 0.08$). Furthermore, we did not find any differences in temporal patterns between the different populations (population-year interaction: $\chi^2 = 1.99$, d.f. = 3, $P = 0.57$), nor any indication that survival differed between single and double brooded individuals over time between the different populations (three-way interaction: $\chi^2 = 2.65$, d.f. = 3, $P = 0.45$). Population density had a negative effect on survival (b = -0.005, se= 0.002, $\chi^2 = 10.95$, d.f. = 1, $P < 0.001$).

2.4.4 Fecundity selection analysis

Multiple breeding females had, on average across the 1973-2004 period and the four study populations, significantly more recruits than females who only produced a single brood in a given year (1.32 ± 0.05 SE vs 0.75 ± 0.02 SE respectively; χ^2 = 50.771, d.f. = 1, $P < 0.001$). There was a marginally significant negative temporal change in fecundity (b = -0.016, se = 0.008, χ^2 = 3.93, d.f. = 1, P = 0.047), suggesting that total number of recruits has decreased. We found no significant interaction between breeding category and year (χ^2 = 0.84, d.f. = 1, P = 0.36), suggesting that differences in fecundity between single and multi brooded females had not changed differently over time. Although there was no significant interaction, both single brooded and double brooded females showed a negative trend in the number of recruits produced over time (b = -0.017, se = 0.008, χ^2 = 4.06, d.f. = 1, P = 0.044 and b = -0.005, se = 0.01, χ^2 = 0.41, d.f. = 1, P = 0.52 for single and double brooded females respectively). We did not find any indication that the number of recruits from single and double brooded females had changed differently over time in the different populations (χ^2 = 4.98, d.f. = 3, P = 0.17) or that number of recruits produced by single and double brooded females differed significantly between populations (χ^2 = 2.25, d.f. = 3, P = 0.52). Not surprisingly we found large spatial variation in number of recruits produced, as evident from a highly significant population x year interaction (χ^2 = 30.64, d.f. = 3, $P < 0.001$). Population density had furthermore a strong negative effect on the number of recruits produced (b = -0.009, se = 0.001, χ^2 = 79.91, d.f. = 1, $P < 0.001$).

2.4.5 Recruitment from first clutch

As there was no indication of different temporal change in total fecundity between double brooded and single brooded individuals, we tested whether there had been any temporal change between single and double brooded individuals in the number of recruits produced from the first clutch during the 1973-2004 period. Because there was a significant interaction between population, year and breeding category (χ^2 = 13.93, d.f. = 3, P = 0.003), suggesting that patterns differed between single and double brooded females over time in the different populations, we analyzed each population separately for ease of interpretation.

We did not find any indication that number of recruits from the first clutch changed differently over time for any of the populations except in HV2 (χ^2 = 9.96, d.f. = 1, P = 0.002), where single brooded individuals showed a stronger decline (b= -0.021, se = 0.01) than did double brooded individuals (b= 0.003, se= 0.01). However, overall there was no significant difference in the number of recruits from the first clutch produced by single or double brooded females for this population (χ^2 = 1.53, d.f = 1, P = 0.22, main effect tested separately). Also for the LB and OH population (P = 0.15 and P = 0.79 respectively) there was no difference in number of recruits produced from the first clutch between single and double brooded females, suggesting that in these populations the observed difference in total number of recruits is due to a fitness benefit of having a second clutch. For the VL population however, double brooded females had significantly less recruits from the first clutch than did single brooded females (χ^2 = 4.52, d.f. = 1, P = 0.03). Apart from the different temporal pattern in number of recruits from the first clutch between single and double brooded individuals found in the HV2 population, only LB showed a weak temporal trend (b= -0.027, se= 0.014, χ^2 = 4.36, d.f. = 1, P = 0.04) in number

of recruits produced from the first clutch. For both HV, LB and VL there was a strong negative effect of increased population density on the number of recruits produced from the first clutch, but there was no such trend for OH ($\chi^2 = 0.07$, d.f. = 1, $P = 0.79$).

2.4.6 Recruitment from second clutch

The decline in proportion of females producing a second clutch suggests a decline in the fitness benefits of producing a second clutch, but these were not apparent in any of our analyses of female survival, fecundity or recruitment from first clutch as described above. However we did find a significant decline in number of recruits produced from the second clutch during the 1973-2004 period (b = -0.053, se = 0.023, $\chi^2 = 5.88$, d.f. = 1, $P = 0.01$, Fig. 2.1.3). Not surprisingly, the number of second-clutch recruits produced from the different populations also varied considerably ($\chi^2 = 17.45$, d.f. = 3, $P < 0.01$). However there was no interaction between population and year ($\chi^2 = 5.11$, d.f. = 3, $P = 0.17$), suggesting that the fitness benefits of producing a second clutch have declined at a similar rate in all four populations during this period. Number of recruits produced from the second clutch was furthermore negatively related to the population density (b = - 0.007, se = 0.003, $\chi^2 = 4.78$, d.f. = 1, $P = 0.03$).

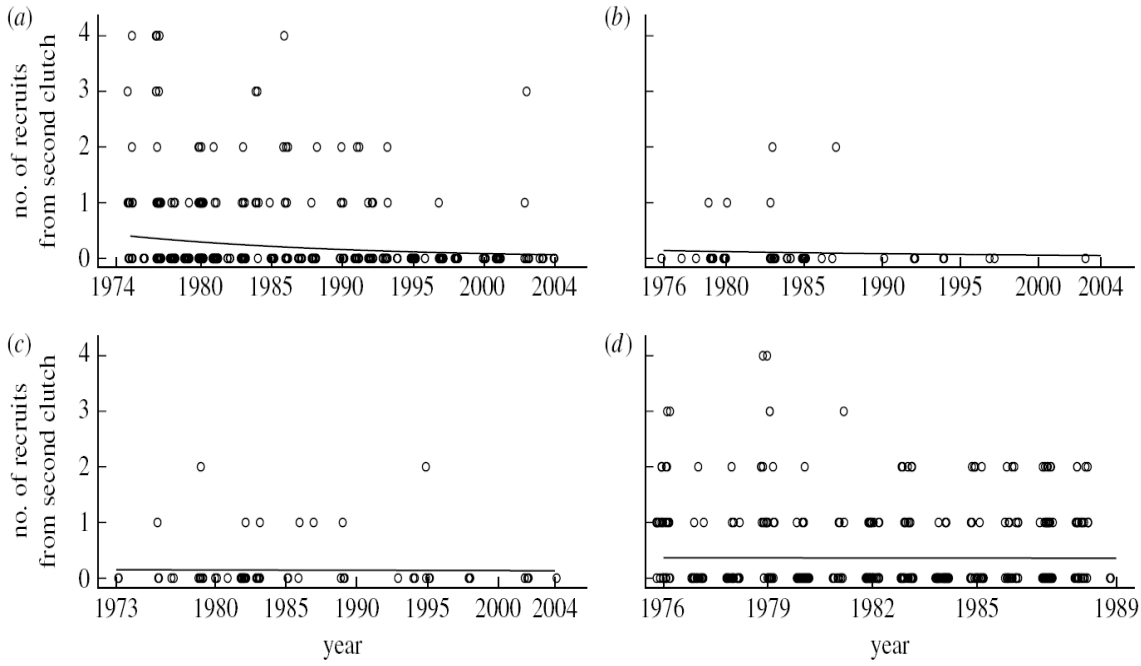


Figure 2.3. Temporal trend in number of recruits from the second clutch for each population, fitted from the equations from population specific GLMM's with Poisson error structure (see text for further details). The equations fitted were (all on log scale): –0.858 -0.062*year, -1.865 – 0.037*year, -1.891 – 0.004*year, -0.999 – 0.002*year for the HV2, LB, OH and VL population respectively. Note that some population-year combinations are missing (see Table 2.1 for details).

2.5 DISCUSSION

We have shown here that the proportion of females double brooding has declined over a 50 year period in all four main study populations of great tits in The Netherlands (Fig. 2.1, Table 2.2B), although a significant decline was found only for the HV2 population for the whole period. However, for the period 1973 – 2004 both the HV2, LB and OH populations showed a decline in the proportion of females producing a second clutch indicating that the decline has been strongest in the later part of the study. This decline coincides with increasing spring temperatures during the same period, and with analyses of changes in lay date in great tits, which have reported more rapid changes in the last three decades (e.g. McCleery and Perrins 1998).

For the HV2 population the probability of producing a second clutch was closely related to the degree of mismatch (difference in days between when chicks from first clutch are 12 days old, and the peak in caterpillar abundance) experienced by the female (Fig. 2.2A). As predicted from the population-level analyses, this probability had changed over the study period, with individuals breeding in the later years of the study having a significantly lower probability of starting a second clutch than individuals in the earlier part of the study (see Fig. 2.2B), but there was no evidence of a temporal change in how mismatch influenced the probability of double brooding (Table 2.3). Further, we found no indication of viability selection operating differentially on single versus double brooded individuals. Although double brooded individuals did, on average across the whole study period in the four populations, produce more recruits than single brooded individuals, the number of recruits from the second clutch alone declined over time for all four populations during the period 1973-2004 (Fig. 2.3) – suggesting that the benefits of double brooding *per se* have changed over time.

Surprisingly, given the importance of multiple breeding on reproductive success, this is, to our knowledge, the only study to have considered temporal trends in the occurrence of multiple brooding and associated patterns of selection. Hence, we have very limited information about whether other populations of facultative multiple brooded bird species are experiencing a similar decline in number of clutches produced during the breeding season. We have only been able to find two studies mentioning temporal trends in proportion of females double brooding. Visser *et al*. (2003) compared large scale responses to climate change on laying dates across a Europe wide range of great tit and blue tit (*Cyanistes caeruleus*) populations and showed a decrease in the proportion of second clutches produced in both species; our

results confirm and extend these analyses. Additionally, Møller (2007) studied the interclutch interval in a population of barn swallows (*Hirundo rustica*) in Denmark in relation to climate change, and briefly noted that the proportion of birds producing a second clutch had not changed over the study period (1975 – 2005).

In many bird species, it is commonly found that early breeding individuals have a higher probability of producing a second clutch than late breeding individuals (e.g. Brinkhof et al. 2002), and it is thought that this is because of a decline in food abundance throughout the season rather than a seasonal change in the quality of breeders, at least in the great tit (Verboven and Verhulst 1996). The importance of timing relative to the food peak for determining the probability of producing a second clutch has been demonstrated previously in great tits (Verboven et al. 2001) as well as other species (Simons and Martin 1990; Nagy and Holmes 2005), and our findings support this (see Fig. 2.2A). For instance, Nagy & Holmes (2005) provided food supplements to females after the time of laying of the clutch in the Neotropical Black-throated Blue warbler (*Dendroica caerulescens*) and demonstrated that food supplemented females produced significantly more second clutches compared to the control females who did not receive food supplementation. It is clear from this experiment, as well as similar studies (Simons and Martin 1990), that food limitation during the breeding season can be a constraint for the production of multiple clutches.

The observed relationship between the probability of double brooding and the mistiming also suggests that if the mistiming increases (i.e. a shift towards larger, more positive, mismatch values) birds will be less likely to initiate a second clutch. Based on previous work in the Hoge Veluwe population we know that there has been an increase in mistiming over the course of the study (Visser et al. 1998). While

hatching date of caterpillars has advanced by 0.74 days per year in the period 1985 – 2004(Visser et al. 2006), great tits in the same population have only advanced their laying date by 0.18 days per year over a thirty year period 1973 – 2003 (Gienapp et al. 2006). The close relationship between the probability of producing a second clutch and the amount of mismatch experienced (Fig. 2.2A), together with an increase in mismatch because of climate change, will lead to a decline in the probability of producing a second clutch and subsequently to a decline in the proportion of females double brooding (Fig. 2.1, Table 2.2B).

In support of this we found that there had been a temporal decline in the probability to produce a second clutch (Fig. 2.2B). We did not find that the relationship between the probability to produce a second clutch and mismatch had changed over the study period however. The temporal decline in overall probability of double brooding suggest that individuals from the early part of the study experiencing a given mismatch were more likely to produce a second clutch than individuals experiencing the same amount of mismatch in the later part of the study (Fig. 2.2B).

This change suggests the possibility of changing selection patterns for birds producing a second clutch. We found, however, no indication of any temporal change in survival over the course of the study. Moreover, there was also no detectable survival difference between single and double brooded females. Differences in survival between single and multiple breeding individuals have been investigated in several different bird species before with little consensus (e.g. Bryant 1979; Verhulst 1998; Brinkhof et al. 2002; Nagy and Holmes 2005). Although we did not find any difference in survival, reproductive costs are often obscured in empirical studies by confounding effects such as territory quality (Reznick 1985).

Changes in selection pressure can also be brought about through changes in fecundity selection. In common with other studies (e.g.Weggler 2006) we found that individuals who produced two clutches per season did have higher reproductive success (measured as total number of recruits produced) than individuals who produced only a single clutch, although this is the mean across the years 1973 -2004 for the four study populations. The higher reproductive success of multiple brooded females was the same for all the four populations suggesting that a second clutch has been an important component of reproductive output. We also found an indication that the number of recruits produced, by both single and double brooded females, had declined, although this was only significant for the single brooded females. More importantly however, we found a strong decline in the number of recruits produced from the second clutch (Fig. 2.3) for all four populations during the years 1973-2004. Although the reason for this decline is not clear, it is likely that it is related to an increase in mistiming (Visser et al. 1998). The increasing mistiming with the main food source has large implications for the reproductive success of the great tits as has been previously demonstrated in this species (Nussey et al. 2005; Visser et al. 2006), and is also suggested by the negative trend in the total number of recruits produced, as well as the decline in number of recruits from the second clutch, found in this study. Unfortunately, data on caterpillar peak dates have been collected to a much lesser extend in the VL and OH population than in HV, and is unavailable for the LB population, so it is difficult to say if the increase in mistiming is also happening here, but given that the pattern of temperature increase is similar in all populations this is quite likely. Nevertheless, the observed decline in number of recruits produced from the second clutch clearly illustrates the decreasing fitness benefit of double brooding (Fig. 2.3).

To summarise, we show here that the proportion of females producing a second clutch in four populations of great tits in The Netherlands declined over a 50 year period, and that this decline was particularly strong during the later part (1973-2004) of the study (Table 2.2B). The reasons for this decline were twofold: firstly, it is likely that due to the strong negative relationship between the probability of producing a second clutch and the amount of mismatch experienced (Fig. 2.2A), observed in the HV population, and an simultaneous increase in mismatch over the study period for this population (Visser et al. 1998) birds are less likely to initiate a second clutch. The observed temporal decline in the probability of double brooding over the study period (Fig. 2.2B) supports this view. Secondly, there was a temporal decline in number of recruits produced from the second clutch (Fig. 2.3), which can be one of the reasons behind the temporal decline in the probability of double brooding (Fig. 2.2B). Taken together, this suggests that changing environmental conditions are important in determining the number of clutches a female produces and that drastic environmental changes have the potential to change an important life history trait in this species. The observed decline in proportion of females producing a second clutch can have important consequences for population dynamics as the number of clutches produced during the breeding season is a major component of reproductive success in this as well as in other multi-brooded species. Other studies investigating temporal trends in multiple breeding and associated patterns of selection would be very valuable as it is becoming increasingly clear that adjustment to changing climatic conditions might involve more than simply a change of timing of breeding.

CHAPTER 3

Phenotypic but no genetic decline in body size in three passerine bird populations under a warming climate.

Husby, A, Hille, S.H and Visser, M.E, *unpublished manuscript*.

3.1 Abstract

Several studies have recently noted that body size in birds and mammals have declined as ambient temperatures have increased and interpreted this as a consequence of Bergmann's rule, which states that body size should decrease with increasing temperature as an adaptive adjustment in endothermic vertebrates. However, this assumes that, firstly, the decline is due to increasing temperatures and, secondly, that the phenotypic change is due to a genetic change, something that is rarely considered. Here we use data from three long-term (1979-2008) study populations of great tits (*Parus major*) to look for temporal trends in two measures of body size (adult body mass and tarsus length), the patterns of selection on these traits, their genetic basis and whether there has been a micro-evolutionary change. Our results show that adult body mass decreased in all our three study populations, and that tarsus length decreased in two of the three populations. These declines were inconsistent with the patterns of selection as there was positive directional selection on adult body mass in two populations (and no selection on adult body mass in one) and tarsus length was not under selection in any of our three populations. Although adult body mass and tarsus length were heritable in all populations, the observed phenotypic change was not due to microevolutionary change. The environmental effects that caused this decline were the annual peak date in the caterpillar biomass and the synchrony of breeding of the birds with this peak, rather than temperature. Thus, in this system environmental deterioration, rather than increased temperatures, has led to the decline in body mass. Our results thus caution against interpreting recent phenotypic body size declines simply as a genetic response to Bergmann's rule.

3.2 Introduction

The effects of increase in the global temperature are widespread and diverse, influencing numerous traits in many different taxa (reviewed in Parmesan 2006). Consequently it is important to understand the potential consequences such changes may have on natural systems (Visser 2008), and the use of long-term studies of natural populations have proved invaluable in providing insight into this complicated question (Stenseth et al. 2002). For instance, a long-running population study of great tits in the Netherlands has been able to demonstrate that increasing spring temperatures have caused the birds to become mistimed with their food peak (Visser et al. 1998) leading to a decline in the proportion of females producing a second clutch (Husby et al. 2009) and in reproductive success (Nussey et al. 2005; Husby et al. 2009).

Similar uses of long-term data have been able to demonstrate that as global temperatures have increased, body size in many bird (e.g. Yom-Tov et al. 2006) and mammal species (Millien 2004) have declined. These declines have been suggested to be an adaptation to climate change and as adhering to Bergmann's rule (Bergmann 1847; Mayr 1956), which originally stated that within a genus of endothermic vertebrates (i.e. birds and mammals), species occupying warmer geographic regions are smaller than species occupying colder regions (Mayr 1956; Freckleton et al. 2003). The reason for the expected decrease in body size in warm temperatures is that a small body size increases the surface-to-volume ratio and thus the loss of energy increases due to conduction, which in turn maximise heat loss and aids thermoregulation (Mayr 1956). Conversely, a large body size will generally decrease the surface-to-volume ratio and help minimise heat loss.

This 'rule' has later been extended to also concern temporal changes within a population in the same geographical region (e.g. Mayr 1956) and has been documented in many endothermic (Millien and Damuth 2004; Yom-Tov and Yom-Tov 2006; Teplitsky et al. 2008) and ectothermic vertebrates (Partridge and Coyne 1997) as well as in some invertebrate (Blanckenhorn and Llaurens 2005) species. It has furthermore also been confirmed in recent meta-analyses of endothermic vertebrates (Ashton et al. 2000; Freckleton et al. 2003). This covariation between body size and temperature has been shown over a wide range of temporal and spatial scales. For example, Smith et al. (1995) used fossilised fecal pellets of the bushy-tailed woodrat (*Neotoma cinerea*) to estimate how body size had changed with temperature fluctuations during the last 25000 years in the United States. They found that in periods with warm climate fecal pellets (and thus the closely correlated body size) were smaller than in cooler periods and thus this species seems to follow Bergmann's rule. Other examples include declines in body size in several passerine bird species in Israel over a fifty year period (Yom-Tov 2001) and an increase in body size in house sparrows (*Passer domesticus*) in North America with increasing latitude (and thus decreasing temperature) (Johnston and Selander 1964). Consequently, Bergmann's observation has established itself as one of the best-known 'rules' in ecogeography (Yom-Tov 2001; Meiri et al. 2007; Guillaumet et al. 2008), although there are also many examples where no temperature-size clines have been found (e.g. Adams and Church 2008).

It is important to point out however, that there are many different selection pressures influencing body size in addition to that of adaptation to external temperatures (Mayr 1956) and thus it is important to examine patterns of selection when exploring causes of changes in body size.

The temperature- body size clines proposed by Bergmann (1847) are commonly interpreted as an evolutionary adaptation to conserve heat loss in cold climates, or, conversely, dissipate heat in warm climates (Adams and Church 2008). Whether body size clines are due to an evolutionary adaptation, which assumes that genetic change has taken place, or to a phenotypic response to temperatures (as for instance an increase in cell size and numbers, see Blanckenhorn and Llaurens 2005) is not known (Partridge and Coyne 1997). It certainly is true that many studies have demonstrated a decline in body size with increasing temperature, but very few studies have actually explicitly considered if the body size decline is due to a genetic change or not, although in many cases the observed phenotypic change is, unfortunately, claimed to be due to a genetic change (see for instance Smith et al. 1995). The importance of genetic versus environmental effects of body size clines was addressed by Laugen and colleagues (Laugen et al. 2005b) who collected common frogs (*Rana temporaria*) across a 1600 km latitudinal gradient in Sweden and carried out a common garden experiment. Their findings only gave partial support to Bergmann's rule as size increased from the South to central Sweden , but then declined further North, generating a non-linear size – latitude gradient (although there were few sampled locations from the mid latitude range which may have obscured a more detailed examination of this pattern). They were also able to show that the size of lab-reared tadpoles were significantly correlated with the size of their wild caught parents, suggesting that geographic size variation may be genetically determined in this species.

Another study examining whether changes in body size are due to genetic or environmental factors took advantage of the animal model approach (see below); Teplitsky et al. (2008) examined the relationship between temperature and body size

in red-billed gulls (*Larus novaehollandiae scopulinus*) and compared the phenotypic trend with that of the temporal change in breeding values, but could find no indication that the size trend was due to evolutionary change.

The fact that so few studies have tried to examine whether the size clines are genetic or environmental in origin is unfortunate for several reasons. Firstly predictions about expected change in body size will be inaccurate, or even wrong, in cases where phenotypic and genetic patterns contrast. Secondly, and equally importantly, such inferences will lead to a general misconception about how populations are expected to adapt to climate change and severely overestimate the rate of micro-evolutionary changes.

The importance of explicitly considering the genetic change is further highlighted by some recent empirical studies that have shown that a phenotypic change need not be mirrored at the genetic level (see also review by Gienapp et al. 2008; e.g. Teplitsky et al. 2008), and that a genetic change may even be in opposite direction to the observed phenotypic change (e.g. Merilä et al. 2001a). Support for evolutionary change, therefore, must come from demonstrating that a genetic change has taken place.

Convincingly demonstrating a genetic change is no easy task and has, in the past, often required an experimental approach (as in the study by Laugen et al. (2005) above), but recently the increased application of the animal model (Henderson 1950; Kruuk 2004) in natural populations has opened the possibility for researchers to address this question even without experimental work. The animal model can estimate each individual's genetic merit, or their breeding value, and a change in mean breeding values over time can be taken to represent a genetic change (Kruuk 2004). The vast majority of recent longitudinal studies on natural populations

so far have used this method to detect evolutionary trends (e.g. Merilä et al. 2001a; Charmantier et al. 2004; Garant et al. 2004; Teplitsky et al. 2008). However, some recent work suggest that the test used to assess significance of the genetic trend may often be anti-conservative, leading researchers to conclude that evolutionary change has taken place even when it has not (Hadfield et al. in press).

Changes in body size may also be due to other environmental factors than changes in temperature, which is generally ignored in studies on temporal changes in body size. For example, factors such as food availability (McAdam and Boutin 2003) and population density (Damuth 1981) are known to lead to changes in body size related traits (e.g. body mass or tarsus length) in many species. Consequently, by considering a broad range of environmental variables when examining changes in body size we can increase our understanding of the proximate causes of observed (or lack of) change in body size traits.

Our goals in this study were to examine the causes of phenotypic temporal patterns in adult body mass and tarsus length in three long term populations (1979-2008) of great tit (*Parus major*) in the Netherlands. According to Bergmann's rule we predicted that great tits would have become smaller over time due to the recent increase in temperatures (Visser et al. 1998; Husby et al. provisionaly accepted). Thus, we firstly examined if adult body mass and tarsus length had changed significantly over the study period and, secondly, if the phenotypic changes were due to selection or changes in selection pressure over time. However, as there must also be a heritable basis of traits for there to be a response to selection we also examined quantitative genetic basis of adult body mass and tarsus length in each population. We then tested whether the observed phenotypic changes over time were due to micro-evolutionary change, using two methods: firstly, by examining the change in

breeding values using a linear mixed model (Kruuk 2004), and secondly, using a newly proposed Bayesian approach (Hadfield et al. in press). Finally, we used information on a range of different environmental measures in addition to temperature to try to disentangle what factors are most important in driving the observed phenotypic changes.

3.3 Materials and methods

3.3.1 Study species, study area, field procedures and data

Great tits are small (14-22 g) insectivorous passerines distributed throughout most of Europe and some parts of Asia (Gosler 1993). The data used in this study were collected at three different populations in The Netherlands, Hoge Veluwe (HV), Oosterhout (OH) and Vlieland (VL) as part of an ongoing long-term study first started in 1955 (van Balen 1973). Systematic collection of adult body mass, however, only started in the late 1970's and thus we restricted our analyses to individuals that were caught between 1979 and 2008 (but note that this includes individuals born before 1979). For more details about the study populations see van Balen (1973).

In all areas nest boxes were visited at least once every week during the breeding season (April – June). The number of eggs and/or young in the nests was counted, and when the young were 7 – 10 days old (10-15 on VL) all chicks were ringed and the parents were caught on the nest using a spring trap. At day 15 chicks were weighed and measured. All adult body mass measurements were taken with a Pesola spring balance to the nearest 1/10 g and tarsus length measured with a sliding calliper to the nearest 1/10 mm. Generally, between one and three measurements of adult body mass and tarsus length were available per bird (for body mass 797 out of 8319 individuals were measured more than three times, note that there was a total of

12709 records from these 8319 individuals, see Table 3.1). Parents already ringed were identified and unringed (immigrant) birds were given a metal ring with a unique number.

As variation in adult body mass is influenced by a large number of different factors (time of day when bird was caught, breeding status, age etc.) that could potentially bias our results, we only used measurements taken of adult breeding birds (i.e. adults caught during chick feeding) between April and July caught between 0650 am and 2100 pm and that had not been subject to manipulations. Relatively few birds (n = 1360/7019, 464/2205 and 1623/6932 records for HV, OH and VL respectively) fall outside these selection criteria and restricting the dataset in this way makes the measurements more precise.

For tarsus length, we calculated mean tarsus length for each individual across its lifetime, using adult measurements, or when chicks were 12 days or older. Thus we used one individual measurement of tarsus length in all analysis. There are two reasons for this, firstly, the within-individual variation in tarsus length was large in some populations (notably HV and OH) presumably due to the high number of fieldworkers that have measured tarsus in presumably slightly different ways something that resulted in low repeatability (HV: 0.28, OH: 0.43, VL: 0.67). Secondly, there is no reason to expect tarsus measurements to change within an individual's lifetime as tarsi are fully grown at the age of 12 days in great tits (Björklund 1996), something that was also evident in our dataset from the highly non-significant test when comparing the within-individual tarsus measurements between the nestling stage and adult stage ($F_{1,6726} = 0.043$, $P = 0.84$). Hence, all within-individual variation in tarsus length is simply due to measurement errors, something that may lead to inflated residual variance and underestimation of

heritability (see Table 1 in Åkesson et al. 2008 for example). Additionally, it has been shown that measurement error can yield biased estimates of the selection differentials (Mitchell-Olds and Shaw 1987) and so using the average of all measurements will also reduce this error source. Note that for the tarsus length analysis we retained breeding birds which had been manipulated as such manipulations are highly unlikely to cause changes in tarsus length. For more information about the traits and populations used in this study see Table 3.1.

Table 3.1. Trait means, standard deviations and sample size for the different traits in each population (HV= Hoge Veluwe, VL = Vlieland, OH = Oosterhout) for great tits. Adult body mass measurements were restricted to the criteria's outlined in the methods (see above).

	Adult body mass			Tarsus length		
Population	HV	OH	VL	HV	OH	VL
Trait mean (SD)	17.64 (0.83)	17.73 (0.86)	17.47 (0.87)	19.79 (0.61)	19.68 (0.66)	19.47 (0.65)
Number of records (= Number of individuals for tarsus)	5659	1741	5309	3219	877	3331

3.3.2 Population level trends

As both adult body mass and tarsus length were normally distributed we used linear mixed models (LMM) to test for temporal changes in adult body mass and tarsus length. Because body mass depend on when during the day the bird is caught, the time of year, breeding status and also varies between sexes we included sex, catching

date, age of the individuals, age of their chicks, time of day when captured and the identity of the measurer as fixed effects. Year was included as a covariate to test for temporal changes in body mass and we also included the interaction between sex and year to test for sex specific temporal change. Individual identity and year of birth were included as random effects in the adult body mass analysis.

For the tarsus length analysis we only included year as a covariate and sex as factor as well as their interaction as fixed effects. Note also that only year of birth was used as random effect in the analysis of tarsus length as we used the average tarsus length of an individual in the analysis (see above) so each individual was only represented once.

Our goal was to study within-population temporal changes in adult body mass and tarsus length and therefore we performed separate analysis for each population. In addition, we also included all three populations in one global model and tested for between-population differences in temporal trends.

All linear mixed models were fitted using ASREML-R v 2.0 (Gilmour et al. 2006) using Schall's technique (Schall 1991). Statistical significance of fixed effects was assessed from their conditional Wald *F* test statistics to respect principles of marginality and we used a backward deletion procedure thus starting with a global model and deleting the least significant terms by a step wise procedure until only significant effects were included.

3.3.3 Selection analysis

We estimated selection on each trait in each population using the yearly number of offspring that an individual recruited to the breeding population (a recruit means that an offspring was recorded as a breeding bird in the population in subsequent years

after its year of birth) as an estimate of fitness, as it is only offspring that recruit to the breeding population that will contribute to any selection response. Note that the use of number of recruits as fitness measure do not incorporate direct viability selection and so neglect a potentially important component for selection on body size related traits, although number of recruits will partly (indirectly) incorporate this information as longer lived individuals generally have more recruits than short lived individuals.

For each population we conducted year-specific analyses (excluding 2008 as fitness measurements were not yet available at the time of analysis) by standardizing each trait (adult body mass and tarsus length) to have zero mean and unit variance (creating *z*-scores) within each year and population. Yearly fitness values (number of recruits) were divided by the mean number of recruits produced in the given year and population to give relative fitness scores (ω) for each individual. Standardized selection differentials (*S*) for adult body mass and tarsus length were measured using least squares regression technique as the covariance between relative fitness and standardized trait values, i.e. $S = \mathrm{Cov}(\omega, z)$ (Lande and Arnold 1983). Although selection differentials themselves (and their standard error) are often unbiased even when assumptions of the least-squares regression technique is violated (Lande and Arnold 1983), estimating statistical significance of the selection differentials is more problematic. For instance, relative fitness (ω) rarely, if ever, follows a Gaussian distribution and so inferences from a least-squares regression will be unreliable. Statistical significance of the standardized selection differentials were therefore assessed with a Generalized Linear Mixed Model (GLMM, log link function) including individual identities as random effects to account for repeated measurements on individuals. We provide the yearly standardised selection

differentials, their standard error and significance (as estimated from the GLMM analysis) in Table S3.1 for adult body mass and Table S3.2 for tarsus length.
We also tested for a temporal change in selection pressure by regressing annual standardized selection differentials against year with a least-squares regression. The least-squares regression analyses were done using R 2.8.0 (RDevelopmentCoreTeam 2007) and the GLMM analysis in ASREML-R (Gilmour et al. 2006).

3.3.4 Pedigree

Quantitative genetic studies require knowledge about the relationship between individuals within a population. We reconstructed a pedigree based on social information (i.e. from field observations) and so the pedigree can contain errors through the paternal line (due to extra-pair paternity, EPP). Although the rate of EPP is unknown in the HV and OH population, it has been estimated to be as low as 3.5% extra-pair young in the VL population (excluding one nest in which all offspring were sired by an extra-pair male, Verboven and Mateman 1997). Furthermore, there is no reason to expect the EPP levels to be higher in the two other populations, as it is generally low also in other populations of great tits (e.g. Lubjuhn et al. 1999). Such small levels of EPP has only a negligible effect on the estimated additive genetic variance component, and thus h^2 estimates, when samples are large as in this study (Charmantier and Reale 2005). In some cases chicks were cross-fostered, in which case we used the genetic parents rather than the social parents in the pedigree. To preserve sibship information and maximise pedigree information, we dummy-coded parents whenever information about either the male or female was missing.

3.3.5 Quantitative genetic analyses

For all populations we used a residual maximum likelihood (REML) mixed model approach in the context of an 'animal model' to estimate variance components of adult body mass and tarsus length. The animal model uses information about the relatedness between all individuals in the population in a mixed model framework to partition the phenotypic trait variance under study into its additive genetic variance component and environmental (and other non-genetic) variance components (Henderson 1950; Lynch and Walsh 1998; Kruuk 2004).

Because we have repeated measures on the same individuals in different environments for adult body mass we also estimated the permanent environment effect, i.e. the within-individual variance associated with environmental effects (or non-additive effects such as dominance or epistasis, Kruuk 2004). We also included year of birth as a random effect to account for temporal heterogeneity in environmental effects on the phenotypes. Note that year of birth was also included as a covariate in all animal model analyses to avoid bias in the breeding values (EBV) towards the phenotypic trend (Postma 2006). We included the trait specific fixed effects that were significant in the phenotypic analysis also in the animal model analysis (see Table 3.2).

Thus, after correcting for the fixed effects mentioned above, our animal models partitioned the phenotypic variance in adult body mass into the following components:

$$V_P = V_A + V_{PE} + V_{YOB} + V_R \quad (1),$$

where V_A is the additive genetic variance, V_{PE} is the permanent environmental variance, V_{YOB} is the variance associated with the year of birth (common environment variance) and V_R is the residual variance. For tarsus length we used the same model as above, but because we did not use repeated measures on tarsus length

(see above) the V_{PE} component could not be fitted. Heritability of the two traits were subsequently calculated as $h^2 = V_A / V_P$ after accounting for fixed effects (Falconer and Mackay 1996). We also report the coefficient of additive genetic variance ($CV_A = 100\sqrt{(V_A)}$/trait mean) as a measure for comparison between populations and with other studies (Houle 1992). In addition, we report here the phenotypic standard deviation in the raw data (see Table 3.1), i.e. the variance before conditioning on the fixed effects, so that a 'standardized' heritability can be calculated (as suggested by Wilson 2008) and to allow our estimates to be used more easily in future comparative studies of heritability.

To examine if there had been a genetic change in population composition over time we calculated the best linear unbiased predictor (BLUP) of each individual's estimated breeding value (EBV), which is the sum of the contributions of the individuals own phenotype and that of related individuals phenotype scaled by the relatedness to the focal individual. The EBVs were subsequently used to test for microevolutionary change, as changes in EBVs reflect changes in additive genetic effects resulting from selection and/or drift, using a linear mixed effects model with year as a random effect, following (e.g. Merilä et al. 2001a; Garant et al. 2004; Teplitsky et al. 2008).

However, because it has recently been suggested that assessing the probability that evolutionary change has taken place using the above method can lead to inflated *P*-values (Hadfield et al. in press), we also assessed the significance of the temporal change in EBVs from the posterior distribution (of change in breeding values over time) using a Bayesian MCMC animal model approach (Hadfield et al. in press). We included the same fixed effects and random effects as described under the animal model analysis, and we used weakly informative priors set to ¼ of the

phenotypic variance. Mixing and convergence of the chains was assessed by visual inspection of the time series of the MCMC iterations. As this method is very recent we present the results from both the BLUP and Bayesian analyses for comparison of the two methods within this study and also for comparison with previous studies that have only used the BLUP approach.

Although some studies infer microevolutionary trends by comparing the predicted response to selection (as assessed from the breeders equation) and the observed response we do not consider this approach here. This is because it does not exclude the possibility that some of the change is due to plasticity and it has been shown numerous times that the breeders equation does not work well for studies on natural populations (Merilä et al. 2001c).

All animal models were run in the software ASREML version 2.0 (Gilmour et al. 2006) and the Bayesian MCMC animal model was run in R using the MCMCglmm package (Hadfield in revision).

3.3.6 Environmental variables

Although the observed decline in adult body mass in the three populations is consistent with the expectation from Bergmann's rule, other environmental factors may also contribute to this decline. Thus, to try to disentangle what environmental factors could be the cause of the decline we used information about a range of environmental variables: the beech crop index (Perdeck et al. 2000), the date of the caterpillar peak (Visser et al. 2006), the synchrony between the peak in caterpillars and mean laying date (Visser et al. 2006), the temperature period during February – March as a proxy for winter severity (Perdeck et al. 2000), the annual mean temperature, and, finally, population density (measured as number of breeding pairs). Unfortunately, information on beech crop index, date of caterpillar peak and synchrony (difference between caterpillar peak and yearly mean laying date) was only available for the HV population from 1985 onwards (see Visser et al. 2006), and so we had to restrict our analysis to the HV population and the years 1985-2008. We thus included adult body mass as response variable in a linear mixed model with the above environmental factors included as explanatory variables and individual identity and year as a random effects to account for repeated measures on the same individuals across years. For tarsus length we used a linear model, as within-individual variation in tarsus length was averaged away (see above), with the above environmental variables included as explanatory variables. All environmental variables were mean centred before analysis.

3.4 Results

3.4.1 Population level trends

Adult body mass declined significantly in all three (HV, OH and VL) populations (b_{hv} = -0.013 ± 0.004 g/yr, b_{oh} = -0.012 ± 0.006 g/yr, b_{vl} = -0.015 ± 0.005 g/yr, see Table 3.2a, Fig. 3.1) over the study period. There were strong effects of the time of day when the bird was measured, the day of capture (measured as April days, see Methods), age of the chicks at the time of capture and observer identity (Table 3.2a). There was also a strong effect of sex, suggesting that sexes differ in their mean adult body mass, but no suggestion that there had been any sex specific decline in adult body mass as indicated by a non-significant sex by year interaction in all populations (Table 3.2a). Because of the very similar negative trend in all three populations there was no significant population by year interaction when all populations were tested simultaneously in the same model ($F_{2,11345.8}$ = 0.013, P = 0.98).

Tarsus length declined significantly over time in the HV and OH population (b_{hv} = -0.009 ± 0.002 mm/yr, b_{oh} = -0.010 ± 0.004 mm/yr, Table 3.2b, Fig. 3.2), but showed a significant increase in the VL population (b_{vl} = 0.007 ± 0.002 mm/yr, Table 3.2b, Fig. 3.2). There was no indication of any sex by year interaction, but tarsus length differed significantly between the sexes in all populations (Table 3.2b). The between population difference in temporal patterns was also clearly indicated by a highly significant population by year interaction when all populations were tested in the same model ($F_{2,7113.6}$ = 51.70, P < 0.001).

Table 3.2. ANOVA table from the linear mixed model analysis of adult body mass (Table 3.2a) and tarsus length (Table 3.2b) for the three study populations. Note that in Table 3.2b we used the within-individual average tarsus measurements and so we did not control for observer identity etc. as such variation was averaged away. In Table 3.2a year of birth and individual was fitted as random effects and only year of birth was fitted in Table 3.2b.

Table 3.2a – Bodymass	HV		OH		VL	
Variable	df	F	df	F	df	F
Year	1/133.8	10.58**	1/1706.7	3.90*	1/396.2	9.49**
Sex	1/3253.7	199.90***	1/943.5	92.68***	1/2988.2	422.4***
Catching date (April date)	1/3068.2	81.16***	1/1109.6	11.56***	1/3146.0	200.7***
Age of chicks	1/3566.3	36.69***	1/1279.8	5.059*	1/3546.9	25.92***
Age of individual	1/3401.5	68.07***	1/1235.0	25.84***	1/3486.8	72.92***
Time of day when captured	1/3925.2	247.70***	1/1320.6	80.84***	1/3702.5	308.3***
Observer identity	36/3285.4	3.387***	24/1423.7	4.02***	40/2600.0	3.41***
Year x Sex	1/3325.1	2.00	1/935.6	4.51	1/2988.6	2.65

Table 3.2b – Tarsus length	HV		OH		VL	
Variable	df	F	df	F	df	F
Year of birth	1/36.9	14.46***	1/34.1	8.11**	1/36.2	7.93**
Sex	1/3196.5	824.50***	1/863.4	317.60***	1/3315.6	935.10***
Year X Sex	1/3207.2	0.008	1/867.1	0.36	1/3326.4	0.13

*$P < 0.05$, **$P < 0.01$, ***$P < 0.001$

Fig. 3.1. Temporal patterns in yearly average adult body mass (corrected for time of measurement) in the three study populations as a function of year (1979 – 2008). All trends were statistically significant; see Table 3.2a for details.

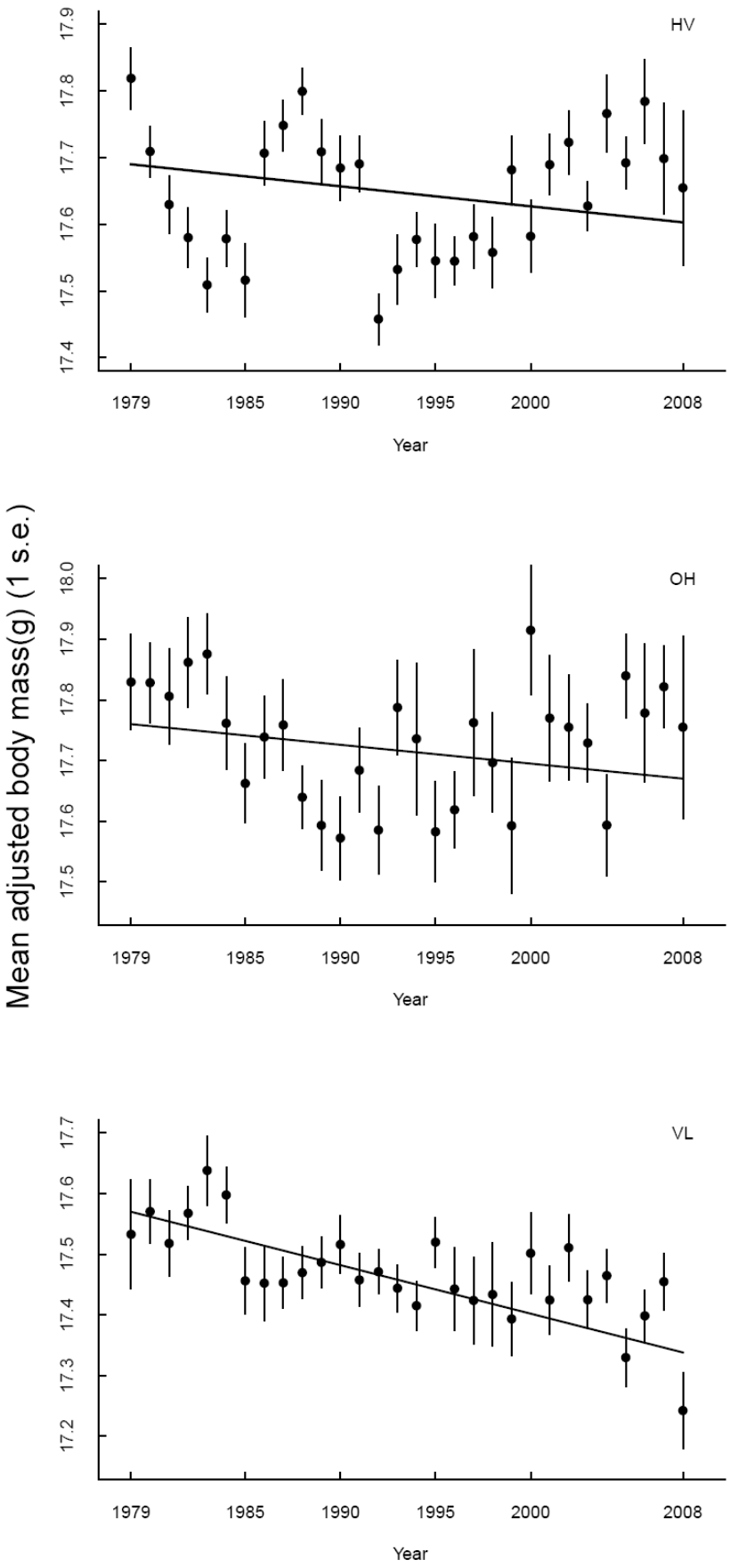

HV
OH
VL
Mean adjusted body mass(g) (1 s.e.)
Year
1979
1985
1990
1995
2000
2008
17.4
17.5
17.6
17.7
17.8
17.9
18.0
17.2
17.3

Fig 1

3.4.2 Selection analyses – Adult body mass

Analyses across the study period showed that there was overall positive directional selection on adult body mass in the HV (S = 0.11, se = 0.029, χ^2_1= 23.65, P < 0.001) and OH population (S = 0.042, se = 0.037, χ^2_1= 6.47, P = 0.011), but not in the VL population (χ^2_1= 0.05, P = 0.832). The year-specific analyses further demonstrated that there was overall little indication of selection on adult body mass with only seven selection differentials being significant, although all significant differentials were positive across all three populations (Table S3.1). Selection differentials also differed considerably in sign between-years within the same population and within years between the three populations (see Table S3.1). Also, the strength of the selection differentials changed over the study period in the OH population (b = 0.012, se = 0.005, P = 0.02), but not in the HV (b = -0.007, se = 0.004, P = 0.069) or VL population (b = 0.004, se = 0.002, P = 0.089). Consequently, in the OH populations the strength of selection on adult body mass increased over the study period.

Fig 3.2. Temporal patterns in yearly average tarsus length in the three study populations as a function of year of birth (1979 – 2007). All trends were statistically significant; see Table 3.2b for details.

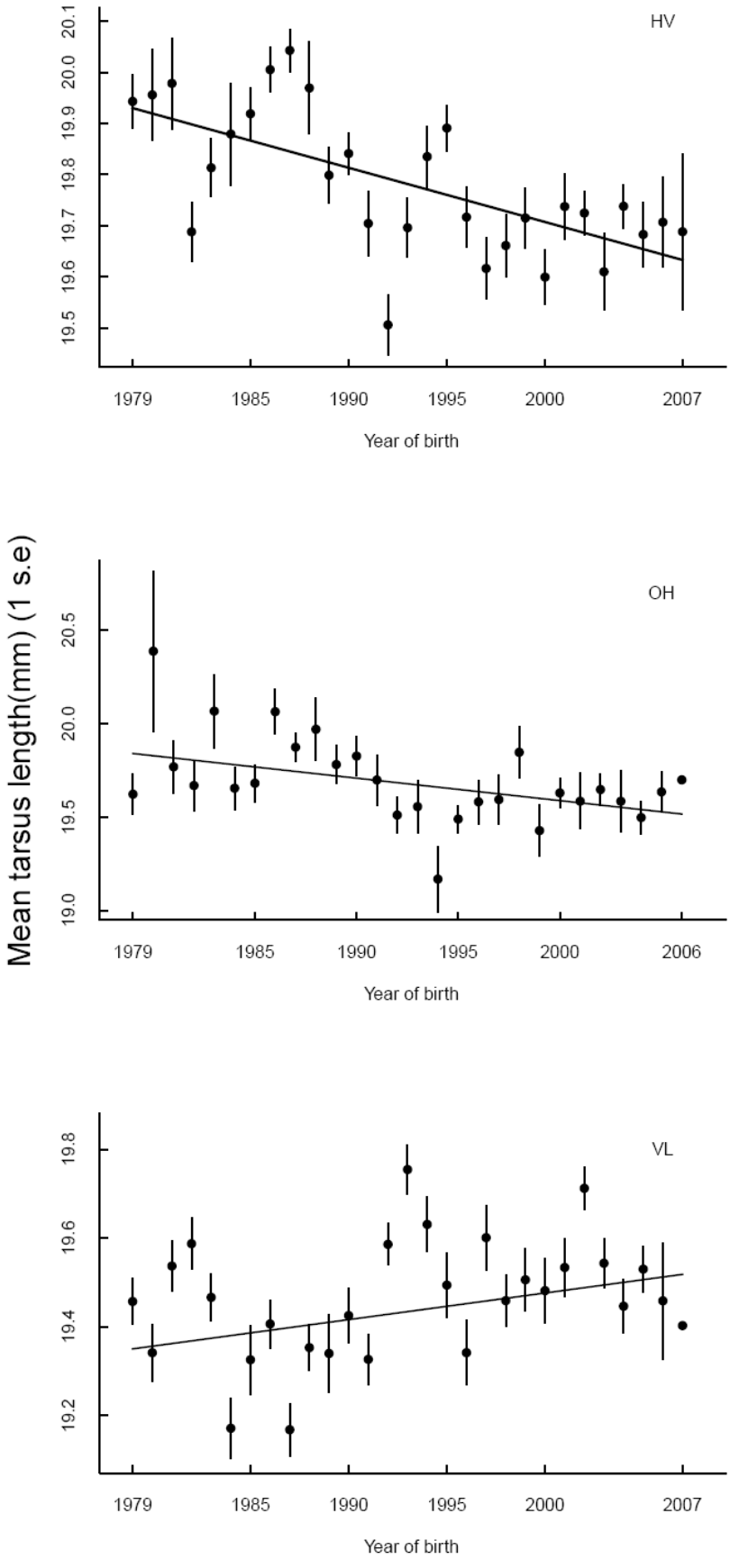

HV
OH
VL
Mean tarsus length(mm) (1 s.e)
Year of birth
1979
1985
1990
1995
2000
2006
2007
19.0
19.2
19.4
19.5
19.6
19.7
19.8
19.9
20.0
20.1
20.5

Fig 2

3.4.3 Selection analyses – Tarsus length

There was no indication that tarsus length was under directional selection when we analysed the data across all years in the HV (χ^2_1= 3.24, P = 0.072), OH (χ^2_1= 0.15, P = 0.697) or VL (χ^2_1= 1.16, P = 0.282) populations. Again, the year specific analyses showed that there were very few years in which selection was significant and the selection differentials differed in sign both within years between populations and within populations across years (Table S3.2). We did not find any indication that selection on tarsus length had changed over the study period in any of the three populations (P = 0.55, P = 0.63 and P = 0.70 for the HV, OH and VL population respectively), reflecting, to some degree, the considerable year-to-year variation in both the strength and direction of the standardized selection differentials.

3.4.4 Quantitative genetic analyses – estimation of heritability

The heritability of both adult body mass and tarsus length was significantly greater than zero in all three populations (Table 3.3, adult body mass; HV: t_{5658} = 14.06, OH: t_{1740} = 7.61, VL: t_{5308} = 14.97, all $P < 0.0001$; tarsus length; HV: t_{3218} = 13.49, OH: t_{876} = 5.90, VL: t_{3330} = 15.38, all $P < 0.0001$) and did not differ between populations (adult body mass; HV/OH: t_{7398} = 1.07, P = 0.28 , HV/VL: t_{10966} = 1.21, P = 0.22, OH/VL: t_{7048} = -0.23, P = 0.82; tarsus length; HV/OH: t_{4094} = 1.82, P = 0.07, HV/VL: t_{6548} = 1.17, P = 0.24, OH/VL: t_{4206} = 1.12, P = 0.26). Also, both V_A estimates as well as C_{VA} estimates indicated that there was little between population variation in the two traits at the genetic (Table 3.3) as well as at the phenotypic level (Table 3.1).

Table 3.3. Components of phenotypic variance (V_P) in adult body mass and tarsus length and their standard error (SE) as estimated from an animal model (V_A: additive genetic variance, V_{PE}: permanent environment variance, V_{YOB} : year of birth variance, V_R: residual variance), CV_A is the coefficient of additive genetic variance and h^2 the heritability. Note that V_{PE} was not fitted in the tarsus model as we used the average tarsus measurements in this analysis.

	Adult body mass			Tarus length		
	HV	OH	VL	HV	OH	VL
V_P	0.616 (0.015)	0.648 (0.029)	0.638 (0.017)	0.292 (0.008)	0.319 (0.016)	0.339 (0.010)
V_A	0.312 (0.026)	0.282 (0.044)	0.287 (0.025)	0.162 (0.014)	0.130 (0.025)	0.167 (0.014)
V_{PE}	0.109 (0.021)	0.108 (0.034)	0.141 (0.017)	-	-	-
V_{YOB}	0.006 (0.003)	0.003 (0.005)	0.004 (0.002)	0.009 (0.003)	0.009 (0.006)	0.012 (0.004)
V_R	0.189 (0.005)	0.259 (0.013)	0.207 (0.006)	0.122 (0.011)	0.179 (0.021)	0.160 (0.009)
CV_A	3.17	2.99	3.07	2.03	1.83	2.10
h^2	0.506 (0.036)	0.434 (0.057)	0.449 (0.030)	0.553 (0.041)	0.407 (0.069)	0.492 (0.032)

3.4.5 Quantitative genetic analyses –testing for microevolution

The linear mixed models on breeding values over time showed a contrasting pattern to the observed phenotypic pattern. Whereas there was no significant temporal trend in the breeding values for adult body mass in the HV population (b_{hv} = -0.0008 ± 0.001, Table 3.4), there was a significant temporal increase in the OH population (b_{oh} = 0.004 ± 0.001, Table 3.4) and in the VL population (b_{vl} = 0.0019 ± 0.0008, Table 3.4).

The breeding values for tarsus length increased over time in the VL population (b_{vl} = 0.002 ± 0.0006, Table 3.4), but for the HV and OH population there was no significant temporal trend (Table 3.4).

When we repeated the above analysis using Bayesian MCMC animal models and assessed the significance of the trend in EBV from the posterior distribution, there was no indication that the breeding values for adult body mass or tarsus length had changed significantly over the study period in any of the three study populations (see Table 3.4).

Hence, the MCMC results do not support the significant results from the linear mixed model analysis, probably because linear mixed model may often return inflated p-values (Hadfield et al. in press). Consequently, we favour the results from the MCMC models and conclude that there had been no genetic change for adult body mass or tarsus length in any of our study populations.

Table 3.4. LMM and MCMC analyses assessing significance of temporal trends in estimated breeding values for adult body mass and tarsus length. For the LMM analysis the mean (β),the 95% confidence interval and the test statistic for the LMM model (χ^2) is reported and for the MCMC analysis the estimate of the mean (β) of the posterior distribution of estimated genetic change and its 95% confidence interval. Estimates in bold are significant at $P < 0.05$.

	Adult body mass						Tarsus length					
	HV		OH		VL		HV		OB		VL	
LMM Analysis	χ^2_1	β (95% CI)	χ^2_1	β (95% CI)	χ^2_1	β (95% CI)	χ^2_1	β (95% CI)	χ^2_1	β (95% CI)	χ^2_1	β (95% CI)
Year	0.80	-0.0008 (-0.0027, 0.0010)	7.25	**0.0042** (0.0011, 0.0073)	4.77	**0.0019** (0.0001, 0.0037)	0.27	0.0003 (-0.0009, 0.0015	2.30	0.0018 (-0.0006, 0. 0042)	10.57	**0.0021** (0.0007, 0.0035)
MCMC Analysis	β (95% CI)		β (95% CI)		β (95% CI)		β (95% CI)		β (95% CI)		β (95% CI)	
Year	0.0006 (-0.0024, 0.0038)		0.0042 (-0.0014, 0.0113)		0.0019 (-0.0023, 0.0065)		0.0003 (-0.0038, 0.0034)		0.0021 (-0.0039, 0.0086)		0.0037 (-0.0002, 0.0081)	

3.4.6 Exploring environmental variables causing the decline

When exploring which environmental variables influence variation in adult body mass we found that density had a negative effect on adult body mass (b = -0.071, se = 0.001, $F_{1,1229.4} = 25.64$, $P < 0.001$), whereas synchrony, caterpillar peak date and beech crop index were all positively related to adult body mass (b = 0.077, se = 0.002, , $F_{1,732.5} = 12.69$, $P < 0.001$; b = 0.121, se = 0.003, $F_{1,870.2} = 21.90$, $P < 0.001$ and b = 0.072, se = 0.001, $F_{1,1087.7} = 24.65$, $P < 0.001$ for synchrony, caterpillar peak date and beech crop index respectively). Annual temperatures or annual mean temperatures during February-March was, however, not significant ($F_{1,23.5} = 0.97$, $P = 0.34$ for the annual temperature and $F_{1,22.3} = 0.15$, $P = 0.69$ for February-March temperature period respectively).

As only caterpillar peak date ($F_{1,22} = 22.62$, b = -0.87, se = 0.18, P < 0.001) and synchrony ($F_{1,22} = 9.32$, b = 0.59, se = 0.19, P = 0.006) has changed significantly over the examined period (1985-2008), the decline in adult body mass is most likely due to a change in food conditions and the associated increase in mistiming (see Discussion). In contrast to adult body mass, we found that none of the environmental variables examined had an effect on tarsus length (all $P > 0.14$).

3.5 Discussion

In this article we have provided a detailed analysis of the genetic and environmental causes of phenotypic changes in adult body mass and tarsus length in three long term study populations of great tits in relation to a changing environment. In common with many other recent studies examining the effect of increased temperatures on body size (Millien 2004; Yom-Tov et al. 2006; Teplitsky et al. 2008) we found that adult body mass and tarsus length declined (except tarsus in VL) over the study

period (1979-2008) where at the same time temperatures have increased (Visser et al. 1998). When we examined patterns of selection on adult body mass and tarsus length we found overall weak evidence for selection acting on these traits with large inter-annual variation in strength of selection (Table S3.1, S3.2). As is generally found when examining the genetic basis of avian morphological traits (Merilä and Sheldon 2001) adult body mass and tarsus length were highly heritable (Table 3.3). There was however no indication of microevolutionary change in any of the populations for the two traits. Interestingly, we did not find that temperature influenced body mass, but rather the mistiming between the caterpillar food peak date and the chick rearing period was the more important factor driving the decline in adult body mass.

Our study provides an interesting example of how a detailed examination of phenotypic, genetic and environmental patterns can dissect the causes behind the temperature-size clines and provide deeper insight into the generality of Bergmann's rule, especially in relation to climate change.

3.5.1 Temporal changes in body size traits

There has been a decline in adult body mass (all populations) and tarsus length (HV and OH, but not VL) with a simultaneous increase in temperature. Hence, our results are in line with other studies that have found that there has been a decline in adult body mass when temperatures have increased (e.g. Smith et al. 1995; Teplitsky et al. 2008). Other studies on bird populations have also noted a decline in adult body mass with increasing temperatures, both within (Teplitsky et al. 2008) and across species (Yom-Tov 2001). For instance, Teplitsky et al. (2008) investigated change in adult body mass in red-billed gulls (*Larus novaehollandiae scopulinus*) between 1958 and 2004 and showed that it was decreasing at a rate of 0.28 g/year. In comparison, the

decrease in adult body mass found here was between 0.013 g/year and 0.015 g/year. This is somewhat lower than in two similar studies on temporal change in adult body mass in great tits, which found a decrease of 0.036 g/year over the period 1968 – 2002 (note that the authors used the residuals from a model that controlled for many of the factors that we included as fixed effects, see Methods) (Yom-Tov et al. 2006) and 0.023 residual g/year (Cresswell et al. 2009). Nevertheless the decrease found here is similar to what has been observed in closely related species and other species of birds with similar size as the great tit. Both bullfinches (*Pyhrulla pyhrulla*), blue tits (*Cyanistes caeruleus*) and dunnocks (*Prunella modularis*) in various British populations all showed a decline of around 0.014 residual g/year between 1968 – 2003 (Yom-Tov et al. 2006).

An interesting possible consequence of the declines in adult body mass is that it may cause the trade-off pattern between starvation and predation to change, something that may subsequently lead to changes in adult body mass regulations in relation to predation pressure (Cresswell et al. 2009).

It should also be noted that observations of body mass declines are not confined to adults: also fledgling body mass in great tits has declined over a 36 year study period with 0.09 g/year (Garant et al. 2004). This is something that provides further support for the proximate reason for observed declines perhaps rather being due to other factors than temperature (see Discussion of environmental variables). There is no obvious reason to expect that tarsus length should change in relation to warming temperatures, yet we observed a decline in two populations (HV and OH) and an increase in one (VL). Other studies have also noted a decline in tarsus length (Yom-Tov 2001; Teplitsky et al. 2008). One possible reason for this may be due to a correlated response to the changes in body mass, as body mass and tarsus length, and

morphological traits in general, are often genetically correlated (e.g. Jensen et al. 2003; Charmantier et al. 2004). In the presence of two genetically correlated traits, selection on one may lead to a correlated response in the other trait (Falconer and Mackay 1996).

3.5.2 Selection analysis

Body size related traits have often been found to be under strong directional selection in many studies of birds (Kruuk et al. 2001), mammals (Milner et al. 1999) and other taxa as is evident from a large scale review on the strength of selection (Kingsolver et al. 2001). However, there are also abundant examples where morphological traits are under no apparent selection. For example, Schluter & Smith (1986) found that adult body mass was not related to reproductive success in a population of song sparrows (*Melospiza melodia*), and in house sparrows (*Passer domesticus*) total lifetime reproductive success was also not related to any morphological traits in males nor in females (Jensen et al. 2004). In this study we found that there was, across all years, positive selection on adult body mass in two populations (HV and OH), but no selection on tarsus length (Table S3.2). However, for both traits and in all populations there was large inter-annual variation in the strength and direction of the standardized selection differentials, suggesting that selection was not very consistent (Table S3.1 & S3.2).

In contrast, Charmantier et al. (2004) found significant directional selection on fledgling body mass and tarsus length in two out of three blue tit populations, but, interestingly, when correcting for the effects of indirect selection on the two traits (i.e. looking at the selection gradients) there was no indication of any selection.

This difference in conclusions when looking at direct and indirect selection also illustrates a limitation of this study, as we did not take selection on other traits into account. Although it is possible that, as found by Charmantier et al. (2004), selection on body mass and tarsus length may be biased because of indirect selection, in general, selection resulting from indirect selection tend to be small and hence selection differentials and selection gradients often coincide (Kingsolver et al. 2001).

3.5.3 Heritability of body size traits

The genetic basis of body size related traits in birds are very well studied and hundreds of heritability estimates, normally obtained from either parent-offspring regressions or sib analysis, are available (reviewed in Merilä and Sheldon 2001). In general, avian morphological traits show moderate to high heritability (0.4 - 0.6, Merilä and Sheldon 2001) and our results also fall within this range: heritability of adult body mass varied from 0.43 – 0.51 (Table 3.3) in our three study populations, and similarly the heritability of tarsus length varied from 0.41 – 0.55 (Table 3.3). Heritability estimates are often found to differ between populations of the same species, due to both genetic and environmental differences, but in our three populations the heritability estimates were similar and there were thus no significant population-differences in heritability of adult body mass or tarsus length. Comparison of variance components between populations remain scarce (but see Charmantier et al. 2004), and are mainly concerned with how different environmental conditions affect variance components. For example, Charmantier (2004) and colleagues found that heritability of adult body mass in blue tits (*Cyanistes caeruleus*) was higher in the population with lowest habitat quality, something that was caused by a decrease in V_A as well as an increase in

environmental variance. Unlike the blue tit system studied by Charmantier and colleagues, we do not have any indication that our study populations differ in environmental quality, which may be one reason why there were only small between-population difference in V_A, h^2 or CV_A (Table 3.3).

3.5.4 Microevolutionary change?

Following the recent debate on how significance testing of a trend in breeding values should be performed (Hadfield et al. in press) we used the newly proposed Bayesian method by Hadfield and colleagues (Hadfield et al. in press) as well as the more widely used linear mixed method (e.g. Merilä et al. 2001a) to assess the significance of a genetic change and test for microevolutionary change. In line with what has been suggested by Hadfield et al. (in press), we found that the MCMC results gave non-significant results for the change in breeding values over time, even when the GLMM reported a significant change (Table 3.4). The GLMM results suggested that there had been a temporal *increase* in breeding values for adult body mass in the OH and VL populations (but not HV) and for tarsus length in the VL population (in contrast to the phenotypic *decrease* in body size), but the Bayesian analysis did not support this. Consequently, although the GLMM analysis could have been interpreted as support for 'cryptic microevolution' (Merilä et al. 2001a), the Bayesian analysis leave little evidence for microevolutionary change in these three populations for adult body mass and tarsus length.

The lack of a genetic change in these populations is perhaps not unexpected as there was little indication of consistent selection acting on adult body mass and tarsus length in our populations, and the standardized selection differentials fluctuated both in strength and direction between years (Table S3.1 & S3.2). The

comparison of the estimated change in breeding values over time (i.e. the slope of yearly mean EBVs against time) between the GLMM results, and those obtained from a Bayesian analysis, showed that these were not always identical, although the 95% confidence interval from the MCMC analysis always incorporated the estimate from the GLMM analysis and thus the two estimates were never significantly different. It is unclear why this would be the case, but it may be due to a difference in the individual EBV estimates from the two methods, as the animal model and the Bayesian animal model will only return identical EBV's if the variance components have no sampling variance. Nevertheless, the difference does highlight the fact that testing for microevolutionary change in natural populations is a challenging task.

3.5.5 Exploring other environmental variables

Temperature is frequently taken to be the selective agent responsible for the body size clines underlying Bergmann's rule (Stillwell et al. 2007), perhaps because temperature is often the most obvious environmental variable that changes with latitude. In addition, the use of temperature is supported by the observation that Bergmann's rule is also found along altitudinal gradients (Chown and Klok 2003). However, temperature is unlikely to be the only explanation for these 'Bergmann clines' as they have also been demonstrated in ectotherms, such as insects (Stillwell et al. 2007), where the heat conservation argument obviously does not hold. Exploring other environmental variables may therefore provide additional insight into the generality of Bergmann's rule and/or help us understand the causes behind the observation, especially in the light of responses to climate change.

For example, a recent study on pipefish (*Syngnathus*) and seahorses (*Hippocampus*) of the eastern coast of North America found that only the

polygamous pipefish followed Bergmann's rule but not the monogamous seahorses, and the authors suggested that this could be influenced by differences in their mating system, as ecological and demographical factors do not differ between the two species (Wilson 2009). Similarly, in the seed-feeding beetle (*Stator limbatus*) body size clines were more related to variation in host plant seed size, moisture and seasonality than to temperature variation (Stillwell et al. 2007).

In this study we did not find that annual mean temperature or the mean temperature during the most severe winter months was related to the change in body size. Density, time of the caterpillar peak and the timing between laying date and caterpillar peak were instead stronger predictors of variation in adult body mass than were the two temperature periods. However, as density had not changed over the study period this is unlikely to be the cause of the decline in adult body mass. Further support that decline in adult body mass is due to disruption in the food availability (peak and mistiming) comes from studies that have found that also fledgling body mass in great tits and other passerines have declined over time (Merilä et al. 2001a; Garant et al. 2004). The observed decline in body size is thus not a direct response to increased temperatures but due to decline in environmental quality.

Our study thus highlights the importance of considering multiple environmental factors when considering temperature-size clines and their interpretation in the context of Bergmann's rule and show that size declines may be due to phenotypic plasticity rather than a genetic adjustment to warming climate.

Table S3.1. Yearly sample sizes (*n*), standardized selection differentials (*S*) and standard errors (se) from the selection analysis of adult body mass in the HV, OH and VL population. Significant selection differentials ($P < 0.05$) are marked in bold as assessed from a GLMM (see Methods). The standardized selection differential across all years (not the average of the yearly selection differentials) is given at the end along with the total sample size for each population.

	HV			*OH*			**VL**		
Year	*n*	*S*	se	*n*	*S*	se	*n*	*S*	se
1979	149	**0.301**	0.085	54	0.059	0.137	67	-0.064	0.099
1980	238	-0.160	0.103	72	-0.375	0.235	176	-0.069	0.109
1981	186	**0.408**	0.144	68	-0.229	0.213	229	0.036	0.067
1982	198	0.099	0.116	72	0.368	0.152	272	0.000	0.075
1983	259	-0.030	0.128	75	0.299	0.174	204	-0.030	0.070
1984	208	0.402	0.327	70	-0.084	0.196	252	-0.036	0.104
1985	134	0.136	0.105	83	-0.143	0.150	192	-0.068	0.080
1986	184	0.238	0.131	67	-0.174	0.146	167	-0.099	0.061
1987	277	**0.318**	0.095	71	-0.098	0.128	267	-0.066	0.068
1988	285	0.186	0.252	103	-0.192	0.323	261	-0.092	0.052
1989	195	0.057	0.138	81	-0.118	0.171	209	-0.118	0.102
1990	205	0.081	0.106	84	-0.091	0.166	177	-0.146	0.068
1991	275	-0.048	0.139	83	0.180	0.177	236	0.066	0.068
1992	289	0.078	0.099	61	0.107	0.115	252	0.082	0.089
1993	172	-0.293	0.197	71	0.063	0.175	210	0.076	0.084

1994	202	0.282	0.151	30	-0.183	0.317	213	0.078	0.095
1995	165	0.081	0.106	46	-0.097	0.162	220	-0.093	0.077
1996	308	-0.036	0.152	79	-0.078	0.248	100	0.199	0.102
1997	197	-0.058	0.151	23	0.442	0.477	104	0.229	0.148
1998	170	**0.397**	0.185	53	-0.086	0.201	76	-0.099	0.131
1999	189	0.026	0.202	43	**0.605**	0.232	125	-0.050	0.113
2000	150	0.114	0.082	36	-0.100	0.306	90	0.051	0.100
2001	233	-0.101	0.142	49	0.324	0.330	137	-0.095	0.096
2002	167	-0.049	0.084	70	**0.370**	0.159	138	-0.130	0.071
2003	292	0.007	0.168	85	0.141	0.184	165	0.046	0.115
2004	143	0.040	0.075	70	0.032	0.140	203	0.012	0.077
2005	278	-0.127	0.162	103	0.033	0.177	199	-0.004	0.093
2006	149	0.059	0.134	60	0.494	0.302	239	**0.138**	0.066
2007	120	0.116	0.156	79	0.169	0.264	171	0.098	0.085
Total	6017	**0.115**	0.025	1941	**0.042**	0.037	5351	-0.007	0.016

Table S3.2. Yearly sample sizes (*n*), standardized selection differentials (*S*) and standard errors (se) from the selection analysis of tarsus length in the HV, OH and VL population. Significant selection differentials ($P < 0.05$) are marked in bold as assessed from a GLMM (see Methods). The standardized selection differential across all years is given at the end along with the total sample size for each population. Note that sample size is higher than for the adult body mass selection as we did not use the selection criteria's we did for adult body mass (see Methods).

	HV			*OH*			**VL**		
Year	*n*	*S*	se	*n*	*S*	se	*n*	*S*	se
1979	238	0.094	0.063	29	-0.198	0.180	105	0.008	0.009
1980	211	0.028	0.101	23	-0.437	0.516	230	0.020	0.020
1981	108	0.152	0.209	22	-0.436	0.393	269	-0.007	0.013
1982	137	0.025	0.151	34	-0.276	0.278	337	0.013	0.013
1983	362	**0.255**	0.106	59	**0.494**	0.179	283	-0.045	0.021
1984	322	-0.478	0.243	34	0.201	0.248	310	0.028	0.027
1985	204	0.137	0.080	94	-0.248	0.142	269	-0.017	0.017
1986	287	0.128	0.100	75	0.222	0.147	225	0.000	0.009
1987	392	0.104	0.080	75	-0.061	0.126	306	0.014	0.016
1988	319	-0.054	0.230	106	-0.155	0.296	332	-0.052	0.027
1989	200	-0.105	0.135	80	-0.069	0.176	244	0.020	0.029
1990	249	0.062	0.096	86	-0.160	0.166	234	**-0.139**	0.050
1991	299	0.058	0.135	78	-0.113	0.185	296	0.028	0.019
1992	297	0.043	0.099	66	0.052	0.119	316	0.041	0.026

1993	188	-0.080	0.170	84	0.030	0.154	280	0.002	0.016
1994	204	0.130	0.151	64	0.044	0.234	278	0.007	0.020
1995	174	-0.035	0.101	60	-0.123	0.161	321	-0.114	0.058
1996	309	0.024	0.152	96	0.241	0.228	125	0.127	0.076
1997	197	0.082	0.151	23	-0.383	0.479	133	0.068	0.055
1998	190	0.027	0.162	58	-0.073	0.205	130	0.014	0.036
1999	215	-0.107	0.200	45	0.031	0.252	157	0.025	0.027
2000	155	-0.024	0.084	37	-0.443	0.298	105	0.035	0.029
2001	240	0.034	0.141	61	0.426	0.311	161	-0.083	0.078
2002	176	-0.043	0.080	74	**0.347**	0.163	156	-0.099	0.052
2003	301	0.161	0.168	91	-0.059	0.187	195	0.003	0.032
2004	149	0.024	0.074	74	-0.068	0.135	228	-0.027	0.018
2005	287	-0.330	0.161	107	-0.195	0.173	232	**0.044**	0.019
2006	159	0.113	0.129	81	0.253	0.275	268	-0.009	0.011
2007	156	0.085	0.145	82	**-0.591**	0.248	185	0.087	0.062
Total	6725	0.059	0.023	1898	-0.030	0.038	6710	-0.024	0.015

CHAPTER 4

Contrasting patterns of phenotypic plasticity in reproductive traits in two great tit populations.

Husby, A., Nussey, D.H., Visser, M.E., Wilson, A.J., Sheldon, B.C. & Kruuk, L.E.B, provisionally accepted in *Evolution*

4.1 Abstract

Phenotypic plasticity is an important mechanism via which populations can respond to changing environmental conditions, but we know very little about how natural populations vary with respect to plasticity. Here we use random-regression animal models to understand the multivariate phenotypic and genetic patterns of plasticity variation in two key life history traits, laying date and clutch size, using data from two long-term studies of great tits (*Parus major*) in the Netherlands (HV) and UK (WW). We show that, whilst population-level plasticity for laying date and clutch size in response to temperature was similar in these populations, between-individual variation differed markedly. Both populations showed individual-by-environment interaction (IxE) for laying date, and no genotype-by-environment interaction (GxE). However, there was significantly more variation in plasticity in HV than in WW. For clutch size, we only found significant IxE and GxE in WW; yet, the population comparison indicated no significant difference. From a multivariate perspective, plasticity in lay date was not correlated with plasticity in clutch size in either population, thus individuals may not be plastic for multiple traits. Our results suggest that generalisations about the form and cause of any response to changing environmental conditions will be difficult.

4.2 Introduction

An important series of questions in evolutionary biology is how well populations are adapted to the environment they experience, whether they can adapt if the environment changes and, if so, how rapidly. Recent global change in climate and the associated impact on the phenology and behaviour of a wide variety of species (reviewed in Parmesan 2006) has caused an increased interest in these fundamental

questions (Stenseth et al. 2002; e.g. Orr and Unckless 2008; Visser 2008). Although numerous empirical studies have revealed changes in the average phenotype across changing environmental conditions for a wide variety of characters (clutch size adjustment in relation to population density see Both et al. 2000; laying date in relation to increasing temperatures reviewed in Dunn 2004), the causes behind such population responses are rarely explored. One important mechanism by which individuals can adjust to changing environmental conditions is through phenotypic plasticity, which simply refers to a (genotype's) change in phenotypic expression across an environmental gradient (Scheiner 1993). Although phenotypic plasticity can be fundamental in allowing populations to deal with environmental change (Price et al. 2003; reviewed in Ghalambor et al. 2007) there is relatively little knowledge about the extent to which plasticity varies between individuals and a possible genetic basis to such variation in natural populations (Nussey et al. 2007). Yet, such knowledge is important in order to determine how quickly natural populations may respond to environmental changes they might experience. In particular it is of interest to know if variation in plasticity itself is heritable and so might have the potential to evolve under selection.

Most recent studies on variation in plasticity in natural populations have focused on how breeding time (laying date) of birds responds to changes in spring temperature and a genetic basis to this variation in plasticity (e.g. Nussey et al. 2005; Brommer et al. 2008; Charmantier et al. 2008) . This interest is not surprising as it is particularly important to understand how life history traits closely related to fitness, like seasonal timing of reproduction and/or number of young produced, will change with the environment. Separating the average population-level pattern into individual-level patterns can be achieved using longitudinal studies, where repeated

measures of the same individuals under a range of environmental conditions are available, and a linear mixed model framework (Nussey et al. 2007). Longitudinal studies also frequently offer the advantage that a pedigree can be constructed (Pemberton 2008) and this pedigree information can be used in the context of an animal model (Henderson 1950; Lynch and Walsh 1998; Kruuk 2004) to separate the genetic and non-genetic components of variance in the traits under study, including the plasticity component. Similarly, we can partition variation in plasticity into its genetic and non-genetic component, for instance, a genetic basis to variation in plasticity (GxE interaction) would mean that plasticity itself is heritable.

Importantly, the animal model is not restricted to the study of single traits, but can readily be extended to incorporate multiple traits (Kruuk 2004). For instance, a bivariate animal model allows us to partition variance in two traits as well as the covariance between them and we can thus estimate the genetic correlation between the two (or more) traits. This is important as adaptation is an inherently multivariate process (Lande and Arnold 1983; Blows 2007) and so knowledge about genetic constraints is crucial if we want to understand evolutionary change. Although the use of bivariate animal models to estimate genetic correlation between traits (both within and across sexes) in natural populations has increased (e.g. Jensen et al. 2003; Wilson et al. 2007; Garant et al. 2008; Robinson et al. 2009), a limitation of all studies of phenotypic plasticity in the wild to date, is that they have only concentrated on a single trait (e.g. Brommer et al. 2005; Nussey et al. 2005; Wilson et al. 2006; Brommer et al. 2008; Charmantier et al. 2008; but see Robinson et al. 2009). This is somewhat surprising as it is well known that genetic correlation between traits can constrain or speed up the rate of adaptation (Lande and Arnold 1983) and, similarly, between-trait correlations in plasticity also have the potential to

speed up or delay adjustment to environmental changes. For instance, we know nothing about the extent to which individuals in natural populations plastic for one trait are also plastic for other traits, i.e. if there is such a thing as a 'generally plastic' genotype, or what has been termed phenotypic integration (Schlichting 1986). Yet, if this was the case, it could allow rapid adjustment to changes in the environment.

Despite the recent methodological advances in the field and the increasing number of studies demonstrating that phenotypic plasticity can be an important mechanism for adapting to changing environmental conditions (e.g. Reale et al. 2003; Charmantier et al. 2008), we know little about the generality of these patterns. For example, two recent studies of phenotypic plasticity in two populations of great tits from the Hoge Veluwe in The Netherlands and Wytham Woods in England both demonstrated population-level plasticity in average laying date in relation to spring temperature (Nussey et al. 2005; Charmantier et al. 2008). However, the individual-level patterns were strikingly different. While the Dutch population showed large between-individual variation in the response to temperature, and also a genetic basis to this plasticity (Nussey et al. 2005), the English population showed no significant between-individual variation in the response, and also no genetic basis to the variation in plasticity (Charmantier et al. 2008). Direct comparison of the results from these two studies is, however, not straightforward as both the definition of the environment (mean temperature used in the Dutch study and the sum of daily maximum temperatures (warmth sum) in the English study) and the data structure (females who bred twice or more in the Dutch study and females who bred three times or more in the English study) differ.

Our aim in this study was to increase our understanding of phenotypic and genetic between-population variation in plasticity patterns by comparing two long

term study populations of great tits at the Hoge Veluwe, the Netherlands (van Balen 1973) and at Wytham Woods, England (Perrins 1965). Our goals were: firstly to eliminate some of the problems related to methodological issues when comparing studies on plasticity by directly comparing two populations of the same species using the same time series and same methodology. Secondly, to explore the multivariate patterns of plasticity for two key life history traits closely linked to fitness, laying date and clutch size, and also to directly compare reaction norm patterns between the two populations.

4.3 Materials and methods

4.3.1 Study species, populations and data collection

Great tits are small (14-22 g) monogamous hole-breeding passerines occupying most of Europe as well as parts of Asia and North Africa (Gosler 1993). Data have been collected in the Hoge Veluwe (HV) National Park, the Netherlands and in Wytham Woods (WW), Oxford, UK continuously since the early 1950s, but because a storm damaged the study area in HV during winter 1972/73 and nestboxes were subsequently relocated, and as one of our goals was a comparison of populations, we only used the years from 1973 to 2006 for both populations. There is evidence to suggest that climate-change related impacts are apparent only in recent decades (e.g. McCleery and Perrins 1998) further justifying the use of the later period. Table 4.1 summarises information about the two populations.

In both areas nest boxes were visited at least once every week during the breeding season (April – June). The laying date of the first egg of a female's clutch (laying date, LD) was calculated from the number of eggs found during the weekly checks, assuming that one egg was laid per day. Number of eggs in the nests was

counted (clutch size, CS) and when the young were 7 – 10 days old, the parents were caught on the nest using a spring trap. We excluded individual records of females who had their clutch size manipulated during egg-laying (i.e. before clutch completion) in the HV population (n = 138). For the WW population there were no such manipulations.

Laying dates are presented as the number of days after 31st March (day 1 = 1st April, day 31 = 1st May). We only used information on the first clutch for both populations, defined as any clutch started within 30 days of the first laid egg in the respective population in any given year. Replacement and second clutches (which comprise less than 3% of breeding attempts in Wytham (Charmantier et al. 2008) and are currently also rare (< 5% of breeding attempts) in the Hoge Veluwe population (Husby et al. 2009) were excluded from the analysis. More details about the HV study population can be found in van Balen (1973) and about the WW population in Perrins (1965) and in Perrins and McCleery (1989).

4.3.2 Environmental variation

To test for a plastic response in clutch size and laying date we used the population-specific local temperature records as a description of environmental conditions. We used a 'sliding window' approach to decide on the climatic time window that best predicted the onset of mean laying date for the two populations. We thus correlated the average temperature within periods of varying start date (beginning with 1st January), end date (30th April) and length (10 day intervals, ranging from a minimum of 10 days to a maximum of 120 days) to the mean laying date in the population each year. The population-specific period with the highest R-squared value was then used for testing for plastic responses. Temperature data for the Hoge Veluwe population

were obtained from the De Bilt weather station of the Royal Dutch Meteorological Institute (KNMI, http://www.knmi.nl/klimatologie/daggegevens) and for the Wytham Woods population from the Radcliffe Observatory (Charmantier et al. 2008). For both populations we used the daily average temperature ((minimum + maximum)/2). For the HV population the period 13th March – 20th April was the best predictor for the onset of laying (R^2=0.656), whereas the equivalent period for the WW population was 15th February – 25th April (R^2=0.669).

Table 4.1. Summary information about the two populations. Laying date (LD, day 1 = 1st April) and clutch size (CS): data are from 1973 – 2006 inclusive. Note that the sample sizes for the two traits are slightly different in the Wytham Woods population due to missing data.

Population	Number of records		Number of individuals		Mean		Variance	
	LD	CS	LD	CS	LD	CS	LD	CS
Hoge Veluwe	3589	3589	2243	2243	24.199	9.016	48.976	3.803
Wytham Woods	7213	7391	4698	4753	25.804	8.671	67.183	2.926

4.3.3 Population-level response

Following the framework outlined by Nussey et al. (2007) we first quantified the average population-level association between laying date and clutch size for both populations in relation to the most informative temperature period for each population (see above).

4.3.4 Individual-level variation in plasticity (testing for IxE)

By adopting a linear mixed effects model framework we can partition any population-level association into individual-specific changes to test whether individuals differ in their response to spring temperature. We thus used phenotypic information to, firstly, estimate the between-individual variation in average laying date (reaction norm elevations) and clutch size, and secondly to estimate the between-individual variation in the response (reaction norm slopes) as well as the covariance between elevation and slope.

4.3.5 Genetic basis to variation in plasticity (testing for GxE)

In order to estimate the genetic basis of IxE variation a pedigree was constructed for the two populations where all ringed females known to have bred were assigned to their social mother and father if they were known. In cases where brood manipulation experiments had been carried out and chicks had been moved between nests, we assigned the genetic parent rather than the social parent. If only one parent was known, we 'dummy coded' the missing parent to preserve sibship information (note that we did not assign a phenotype to this parent). EPP rate has been estimated to be 14% in the WW population using two allozyme loci (Blakey 1994), but the rate is unknown in the HV population. The EPP rate is, however, generally found to be low (3 - 9 %) in other populations of great tits (Verboven and Mateman 1997; Lubjuhn et al. 1999). However, extra pair paternity rates of less than 20% have been shown to have a negligible impact on heritability estimates (Charmantier and Reale 2005). The pedigree for the Hoge Veluwe population included 6907 individuals with 1271 dams and 1295 sires, whereas the pedigree for the Wytham Woods population included 11117 individuals, with 3161 and 3298 dams and sires respectively.

4.3.5.1 Univariate random regression models

Variation in laying date and clutch size was partitioned using an 'animal model' (Henderson 1950; Lynch and Walsh 1998; Kruuk 2004) to give between-individual phenotypic variation (V_I); this variation was subsequently decomposed into its additive genetic (V_A) and, based on repeated measures on individuals across multiple years, a permanent environmental component (V_{PE}). In order to explore patterns of variation in plasticity for laying date and clutch size we first analyzed each trait separately using a univariate 'random regression animal model' (RRAM). These models utilize covariance functions to estimate covariances between the regression coefficients (Meyer 1998) in an animal model framework (Lynch and Walsh 1998; Kruuk 2004). The individual breeding values can thus be modelled as linear (or higher order) functions along some continuous scale (the environmental variable, i.e. spring temperature in this case). Thus laying dates and clutch size records of individual *i* in each standardized annual temperature measurement were analyzed using Legendre polynomials (Kirkpatrick et al. 1990; Gilmour et al. 2006). Temperature measurements were standardized to be within the range -1 to +1, as Legendre polynomials are only defined within this range (e.g. Huisman et al. 2002), using the following equation: -1 + 2(temperature value – minimum temperature value)/(maximum temperature value – minimum temperature value). We only fitted polynomial functions (φ) of a zero and first order (n = 0 or n = 1) due to problems with model convergence, and thus considering linear reaction norms only; however, population-level responses to temperature are linear (Fig. 4.1 c, d). A first order function, $\varphi(ind_i, T)$, applies a linear reaction norm model for individual-specific values across the environment such that variances in elevation and slope of reaction

norms are estimated, as well as the covariance between them (resulting in a 2x2 variance covariance matrix for each random effect).

Thus our model was:

$$\mathbf{y}_i = \mathbf{X}\mathbf{b}_i + \mathbf{Z_1}\varphi(\mathbf{a}_i,n_1,T) + \mathbf{Z_2}\varphi(\mathbf{pe}_i,n_2,T) + \mathbf{Z_3}\mathbf{yr}_i + \mathbf{e}_i \quad (4.1),$$

where $\mathbf{y}_i$ is the vector of the individual trait values (clutch size or laying date) and $\mathbf{X}$, $\mathbf{Z_1}$, $\mathbf{Z_2}$ and $\mathbf{Z_3}$ are the design and incidence matrices relating to the fixed effects and random effects of the additive genetic ($\mathbf{a}_i$), permanent environment ($\mathbf{pe}_i$) and year ($\mathbf{yr}_i$) observations respectively. Fixed effects ($\mathbf{b}_i$ vector) included age as a two level factor (first year breeder or older) to correct for the fact that laying date generally advances with increasing age in great tits (e.g. Wilkin et al. 2006) and that clutch size is often larger in older females (Kluijver 1951; Perrins 1965). In analyses of clutch size (but not laying date) we also fitted terms for population density as it has been shown previously in great tits (e.g. Perrins 1965; Both et al. 2000; Wilkin et al. 2006) that population density often has a negative effect on clutch size (but not laying date). Population density was defined in both populations as the within population sector-specific density in breeding pairs ha^{-1}. The use of this more local measure of density is justified by its correcting for sector-specific differences in the density of nest boxes. Population-specific (see above) standardized spring temperature (T, on the range -1 to +1) was included as a fixed covariate to account for the population-level response in mean trait value. Year ($\mathbf{yr}$ vector) was included as a random effect in order to model variation over years not explained by spring temperature. $\varphi(\mathbf{a}_i,n_1,T)$ is the random regression function of order n_1 of the additive genetic effect of individual *i* and similarly, $\varphi(\mathbf{pe}_i,n_2,T)$, is the random regression function of order n_2

of the permanent environment effect. We included a permanent environment effect (**pe**$_i$ vector) because of the repeated sampling of the same individuals (Kruuk 2004) and this also reduces inflation of estimates of the additive genetic variance due to environmental factors (Kruuk and Hadfield 2007). The error term (**e** vector) was partitioned into three decade specific (1973-1983, 1984-1994, 1995- 2006) groups, thus allowing residual errors to vary between decades. In general, using a heterogeneous error variance structure gave a substantially better fit compared to a model with homogenous error variance (HV; laying date: $\chi^2_2 = 11.56$, $P = 0.003$, clutch size: $\chi^2_2 = 19.78$, $P < 0.001$, WW; laying date: $\chi^2_2 = 26.34$, $P < 0.001$, clutch size: $\chi^2_2 = 1.34$, $P = 0.512$). We also tried modelling the error variance with year-specific estimates but, due to the large number of parameters involved, some of the models failed to converge and thus we do not present the results here.

4.3.5.2 Bivariate random regression animal model

A bivariate random regression animal model is an extension into two dimensional space of the univariate model described above and allows the calculation of covariances between the different sources of variance of the two traits. Hence for each individual (*i*) our model was:

$$\mathbf{CS}_i\ \mathbf{LD}_i = X\mathbf{b}_i + \mathbf{Z_1}\varphi(\mathbf{a}_i, n_1, T) + \mathbf{Z_2}\varphi\ (\mathbf{pe}_i, n_2, T) + \mathbf{Z_3}\mathbf{yr}_i + \mathbf{e}_i \qquad (4.2),$$

where all parameters are as defined for the univariate random regression model. This model estimates the variation in reaction norm components in each trait as well as the between-trait covariances. For instance, a first order function ($n_1 = 1$) for the

additive genetic effect ($\mathbf{a}_i$) would estimate the additive genetic variance-covariance matrix:

$$\begin{bmatrix} \sigma^2_{CSe} & & & \\ \sigma_{CSes} & \sigma^2_{CSs} & & \\ \sigma_{CSe,LDe} & \sigma_{CSs,LDe} & \sigma^2_{LDe} & \\ \sigma_{CSe,LDs} & \sigma_{CSs,LDs} & \sigma_{LDes} & \sigma^2_{LDs} \end{bmatrix} \quad (4.3),$$

where σ^2_{CSe} refers to the variance in reaction norm elevation, *e*, for clutch size, σ^2_{CSs} refers to the variance in reaction norm slopes, *s*, for clutch size and σ_{CSes} refers to the covariance between the two. Similarly, σ^2_{LDe} refers to the variance in laying date elevation, σ^2_{LDs} to the variance in laying date slope and σ_{LDes} to the covariance between the two. These parameters are all as fitted in the trait-specific univariate models (see above). However, in addition to the within-trait variances, we also estimated the between-trait covariances, where $\sigma_{CSe, LDe}$ is the covariance between clutch size elevation and laying date elevation, $\sigma_{CSe, LDs}$ the covariance between clutch size elevation and laying date slope, $\sigma_{CSs, LDe}$ the covariance between clutch size slope and laying date elevation, and finally $\sigma_{CSs, LDs}$ is the covariance between the slopes for clutch size and laying date (and so estimates the covariance in plasticity between the two traits).

Residual variance was defined as for the univariate models, thus using a heterogeneous structure. This provided a substantially better fit compared to a model with homogenous error variance for both populations ($\chi^2_6 = 38.00$, $p < 0.001$ and $\chi^2_6 = 44.04$, $p < 0.001$ for the HV and WW population respectively). Note that we also fitted the covariance between the two traits for the residual and year variance, as not

fitting these will cause the estimate of the additive genetic covariance to be inflated (in a similar way that not fitting a permanent environment effect will inflate the additive genetic variance in a univariate model).

4.3.5.3 Between-population comparison

In order to explicitly compare if the size of the variance components and the plasticity patterns in the two populations were significantly different from each other, we combined the datasets and pedigree information from both populations. Each trait in the two populations was then used in two separate bivariate random regression models (i.e. CS-HV and CS-WW in one bivariate model, and, similarly, LD-HV and LD-WW in a different bivariate model) extended to incorporate the combined dataset and pedigree from each population. Because gene flow between the two populations is negligible, we constrained all covariances between the population-specific traits to be zero. The residual variance was modelled as three (decade-specific for each population) 2x2 unstructured matrices (with covariances constrained to zero). Hence this will model the same residual variance as described above under the univariate analysis. We also modelled population-specific fixed effects (see detailed description under the univariate model).

The population-specific comparison was done by constraining the respective variance components in the two populations to be equal and then optimizing the likelihood under this model. We then used a LRT (see below) to compare the likelihood of this model to that of a model in which they were unconstrained. For more details concerning the use of LRT's to compare matrices, see Shaw (1991).

4.3.6 Statistical analysis

All models were fitted using REML in ASReml v 2.0 (Gilmour et al. 2006). For all models we first fitted a homogenous residual structure, and then a heterogeneous residual structure (see above). To partition the average population-level plasticity into individual-level variation in plasticity, we used a linear mixed effect model framework with increasingly complex variance structure; this logic was also followed for the multi-trait model. In the few cases where the estimated variance components were negative we constrained the matrix to be positive definite. When trying to decompose the genetic basis of plasticity for clutch size in the HV population the model did not converge (due to very small slope variance) and so we fixed the permanent environment matrix to the estimates obtained under the IxE model to obtain convergence; this result should thus be treated with some caution.

In the multi-trait models we additionally fitted the covariance terms between the random regression coefficients, i.e. testing associations between plasticity components and elevation components between the two traits. We first estimated the full 4x4 matrix (see matrix (3) above) and compared this to a model in which all four between-trait covariances were constrained to be zero, thus giving a single test for the significance of sources of between-trait genetic or environmental covariances. Secondly, we constrained all between-trait covariances except that between the elevations of the two traits (i.e. $\sigma_{CSe,\ LDe}$) to zero in order to assess the significance of the between-trait phenotypic and genetic correlation.

All significance testing of (co)variance component(s) was done by calculating the log likelihood ratio and testing against a chi-squared distribution with degrees of freedom equal to the difference in degrees of freedom between the two models tested (Likelihood Ratio Test, LRT) (Pinheiro and Bates 2000). Thus LRT = $-2(L_2 - L_1)$,

where L_1 is the log likelihood of the initial model and L_2 the log likelihood of the model with (co)variance component(s) added.

Although we are here using the 'reaction norm approach' to assess the genetic basis of variation in plasticity, GxE can also be thought of from a 'character state' view as providing environment specific trait values of V_A for the underlying trait (clutch size or laying date). In the presence of GxE, V_A is expected to change across the environmental axis and thus the character state approach provides a useful way to visualize the change in V_A across the environmental conditions. To this end we calculated the environment specific additive genetic covariance matrix, **G**, which was obtained as $\mathbf{G} = \mathbf{zQz}^T$, where **z** is the vector of orthogonal polynomials evaluated at values of standardized temperature measures in the two populations and **Q** is the additive genetic variance-covariance matrix of the random regression parameters for elevation and slope (obtained under model 6 in Table 4.2). Approximate standard errors for the (co)variance components of **G** as a function of the environmental values were calculated according to Fischer et al. (2004), with confidence intervals defined as twice the standard errors. We also calculated the coefficient of variation ($CV = var^{0.5}/mean$) for the additive genetic variance component according to Sokal and Rohlf (1995).

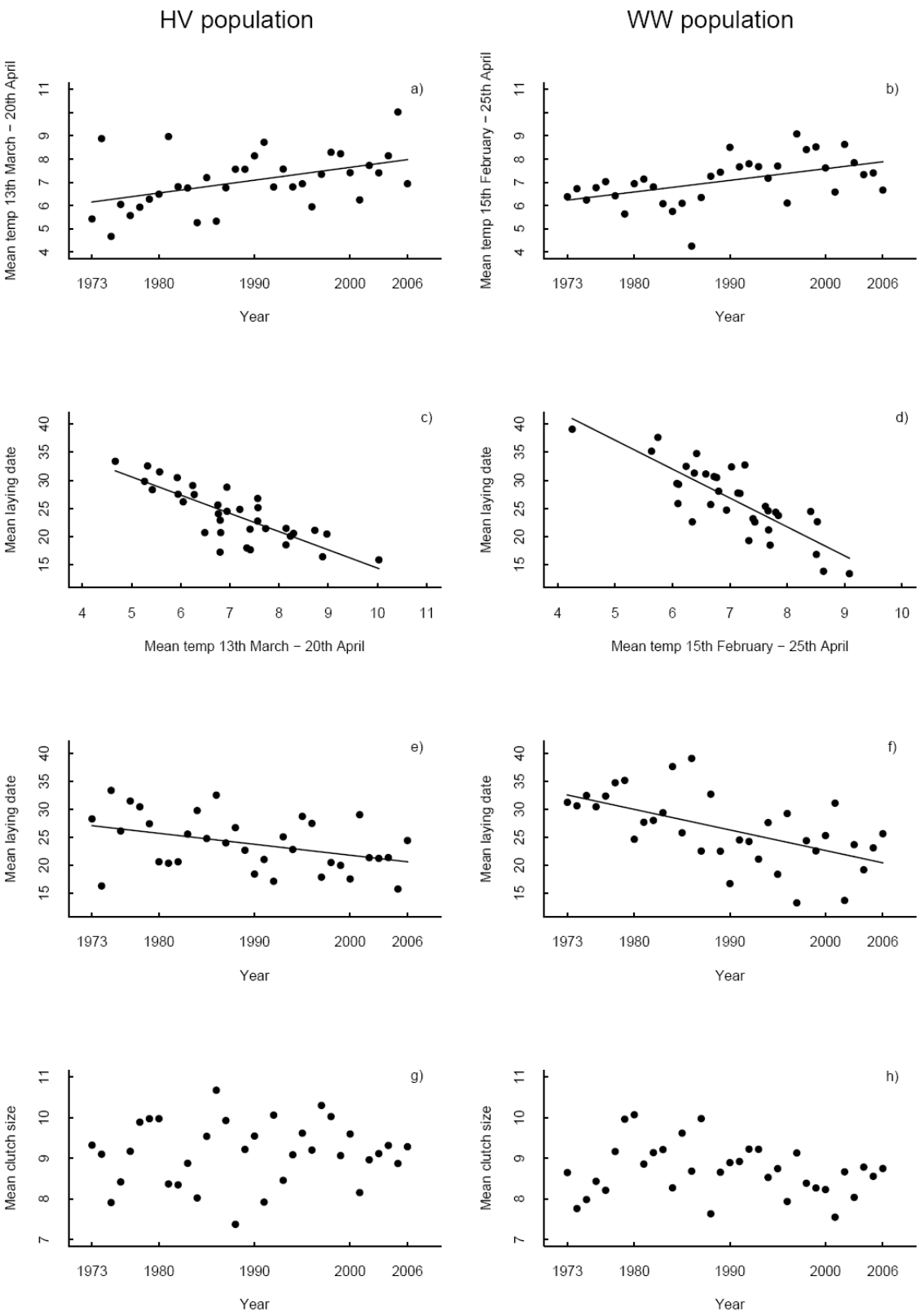
4.4 Results

4.4.1 Population-level patterns

In both populations, spring temperatures increased over the study period at similar rates (b = 0.055, se = 0.019, $F_{1,32} = 8.475$, $P = 0.007$, Fig. 4.1a; b = 0.050, se = 0.015, $F_{1,32} = 10.66$, $P = 0.003$, Fig. 4.1b; for HV and WW respectively). The onset of laying was also closely related to temperature in both populations, although the

response was weaker in HV than in WW (b = -3.256 days $°C^{-1}$, se = 0.421, $F_{1,32}$ = 59.85, $P < 0.001$, Fig. 4.1c; b = -5.158 days $°C^{-1}$, se = 0.635, $F_{1,32}$ = 66.05, $P < 0.001$, Fig. 4.1d; for HV and WW respectively). The close relationship between laying date and spring temperature and the increase in spring temperatures over the study period led to an advancement in laying dates for both populations (b = -0.196 days yr^{-1}, se = 0.079, $F_{1,32}$ = 6.128, $P = 0.007$, Fig. 4.1e; b = -0.367 days yr^{-1}, se = 0.090, $F_{1,32}$ = 16.48, $P < 0.001$, Fig. 4.1f; for HV and WW respectively); as expected given the stronger response to temperature, the increase for the WW population was about twice the rate of that in the HV. We therefore have clear evidence of population-level phenotypic plasticity of lay date in response to variation in spring temperature. In contrast, clutch size did not show a population-level plasticity in relation to spring temperature ($F_{1, 32}$ = 0.191, $P = 0.66$; $F_{1, 32}$ = 0.267, $P = 0.267$) and there was also no temporal change ($F_{1, 32}$ = 0.236, $P = 0.63$, Fig 4.1g; $F_{1, 32}$ = 1.252, $P = 0.27$, Fig 4.1h for HV and WW respectively).

Fig. 4.1 There is a very similar rate of increase in spring temperatures in both (a) HV and (b) WW and a close relationship between the onset of laying and spring temperatures in both the (b) HV and (c) WW population. This has lead to an advancement of the mean laying date in both populations, but this response is weaker in (e) HV than in (f) WW. For clutch size there has been no temporal change in (g) HV or (h) WW. Note that we have used identical y-axis in both populations to aid a visual comparison. See main text for further details.


HV population
WW population
a)
b)
c)
d)
e)
f)
g)
h)
Mean temp 13th March – 20th April
Mean temp 15th February – 25th April
Mean laying date
Mean clutch size
Year
1973
1980
1990
2000
2006

4.4.2 Univariate random regression animal model analysis: laying date

The results from the univariate models of laying date in the two populations provided strong support for between-individual variation in average laying date (Table 4.2a, model 3) and that this variation had a genetic component, i.e. laying date was heritable (Table 4.2a, model 4). Individuals also differed in their response to changing environmental conditions, i.e. in their reaction norm slope, thus there was significant IxE in both populations (Table 4.2a, model 5). However, we did not find statistical support for a heritable basis of the variation in plasticity (no GxE; Table 4.2a, model 6) in either population, although there was more improvement in the model when fitting GxE in HV than in WW (compare model 6 in Table 4.2a for HV and WW). Although we are thus unable to statistically exclude the possibility that the observed IxE is entirely environmental driven, we nevertheless visualised the predictions from the GxE model, using the character state approach, as shown in Table 4.3 and Fig. 4.2. This shows an increase in V_A with increasing temperature in HV (although SEs are large) and no increase in the WW population. The estimated size of the additive genetic variance components for slope in the HV was 3.315 compared to 0.837 in the WW population (see Table 4.S1 for more details).

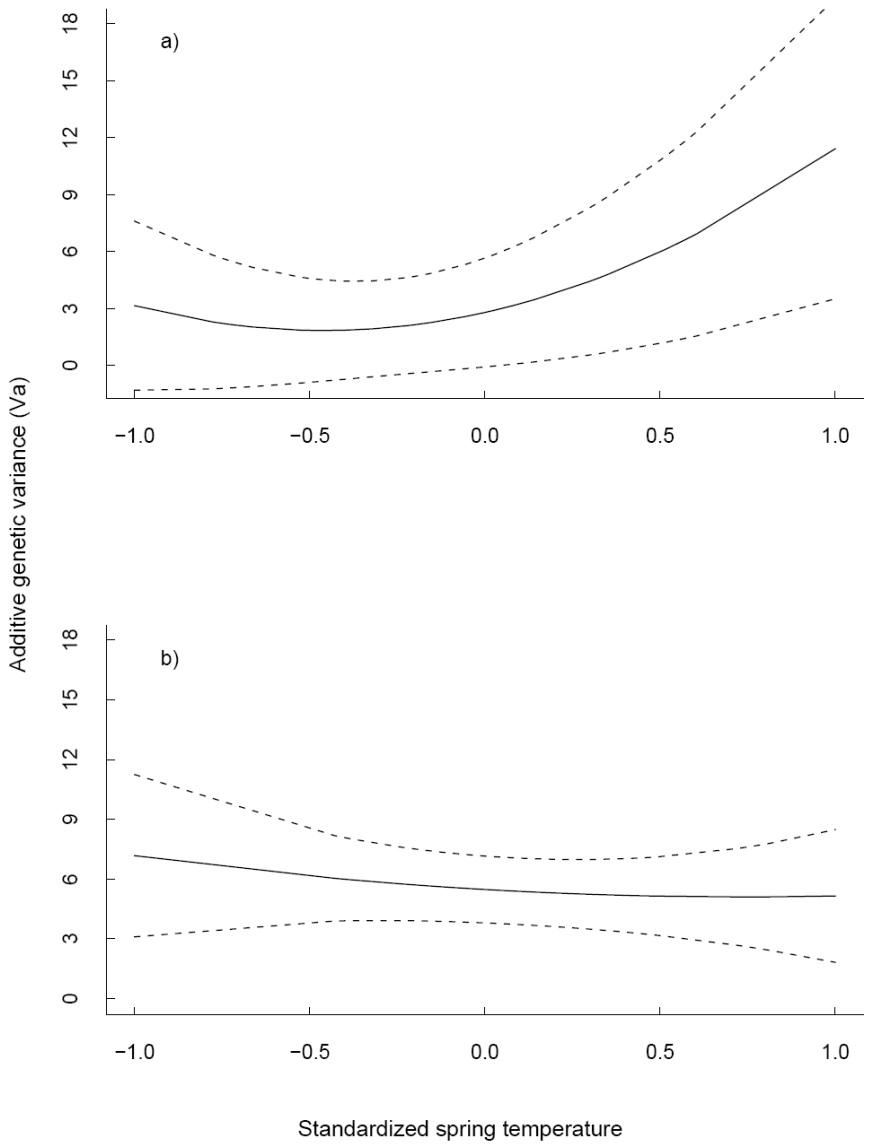


Fig. 4.2 Changes in additive genetic variance for laying date in relation to standardized spring temperature under model 6 in Table 4.2 for the HV population (a) and WW population (b). Dotted lines indicate the approximate 95% confidence interval. Standardized spring temperature of -1, 0 and +1 corresponds to annual mean temperature (°C) of 4.67, 7.34 and 10.03 in the HV population and 4.25, 6.67 and 9.08 in the WW population respectively.

Table 4.2. Results from the univariate random regression analysis of laying date and clutch size in the HV and WW populations.

Table 4.2a: Laying date

			HV			WW		
Model	Variance components	d.f.	LogL	χ^2	*P*- value	LogL	χ^2	*P* - value
1	-		-8009.66			-17178.72		
2	year	1	-7577.87	863.58	< 0.001	-15679.66	2998.12	< 0.001
3	year + V_I	1	-7468.53	218.68	< 0.001	-15382.96	593.4	< 0.001
4	year+ V_{PE} + V_A	1	-7464.81	7.44	0.0064	-15349.77	66.38	< 0.001
5	year + V_{PE} + V_A + IxE	2	-7440.99	47.64	< 0.001	-15344.91	9.72	0.00775
6	year + V_{PE} + V_A + PExE + GxE	2	-7439.11	3.76	0.1526	-15344.52	0.78	0.677

Table 4.2b: Clutch size

			HV			WW		
Model	Variance components	d.f.	LogL	χ^2	*P* - value	LogL	χ^2	*P* - value
1	-		-4151.26			-7627.59		
2	year	1	-3915.57	471.38	< 0.001	-7186.06	883.06	< 0.001
3	year + V_I	1	-3754.35	322.44	< 0.001	-6745.93	880.26	< 0.001
4	year+ V_{PE} + V_A	1	-3749.51	9.68	0.0019	-6709.36	73.14	< 0.001
5	year + V_{PE} + V_A + IxE	2	-3749.51	0	1	-6704.03	10.66	0.0048
6	year + V_{PE} + V_A + PExE + GxE	2	-3749.50	0.02	0.99	-6700.79	6.48	0.0392

Reported χ^2 –values and d.f. are for comparison with the previous model and the *P*- value for the associated LRT test. For each population we used the population specific temperature period (see Methods). All models were fitted with decade specific error variance (see Methods). V_I is the between-individual variance which is split into V_{PE} (permanent environment variance) and V_A (additive genetic variance). IxE is the phenotypic variance-covariance plasticity matrix when no additive genetic variation in plasticity is fitted, PExE is the permanent environment variance-covariance plasticity matrix and GxE refers to the additive genetic variance-covariance plasticity matrix.

Table 4.3. Variance components of laying date evaluated at different standardized spring temperatures for the HV and WW population under model 6 in Table 4.2a.

Population	standardized temperature	*n*	μ ± SD	V_P (SE)	V_A (SE)	V_{PE} (SE)	V_{YR} (SE)	V_R (SE)	CV_A	h^2
HV	-1	143	33.427 ± 5.641	30.144	3.144 (2.227)	2.699 (-)	8.254 (2.116)	16.047 (1.004)	5.305	0.104
	0	83	17.916 ± 5.751	31.047	2.788 (1.435)	6.078 (-)	8.254 (2.116)	13.927 (0.866)	9.320	0.090
	+1	139	15.698 ± 4.938	43.055	11.416 (3.954)	9.458 (-)	8.254 (2.116)	13.927 (0.866)	21.523	0.265
WW	-1	189	39.114 ± 4.987	44.042	7.176 (2.037)	4.077 (2.108)	15.954 (4.021)	16.835 (0.746)	6.849	0.163
	0	319	25.668 ± 4.319	41.252	5.478 (0.839)	5.828 (0.859)	15.954 (4.021)	13.992 (0.611)	9.119	0.133
	+1	291	13.326 ± 4.777	44.247	5.153 (1.666)	9.148 (1.834)	15.954 (4.021)	13.992 (0.611)	17.034	0.116

Note that there are no records in the dataset at exactly 0 standardized temperature and thus the sample size and mean laying date value is given for the nearest temperature record (HV: -0.0019, WW: -0.0012). No standard errors are available for V_{PE} in the HV populations as the associated variance components were fixed at the edge of the parameter space (thus V_P is also without standard errors). V_P is the sum of the variance components (phenotypic variance); V_A the additive genetic variance, V_{PE} the permanent environment variance component, V_{YR} the year variance and V_R the residual variance. Note that V_R is the same for some temp values if they happen to fall within the same decade. CV_A is the coefficient of variance for the additive genetic variance.

4.4.3 Univariate random regression analysis: clutch size

Although neither of the populations showed any population-level plasticity for clutch size in relation to temperature, individuals can still differ in their response if the reaction norm slopes are crossing (see for instance Fig 2g in Nussey et al. 2007). Hence, we also explored individual-level variation in clutch size in relation to temperature. For the HV population, individuals differed in their average clutch size (Table 4.2b, model 3) and part of this variation was due to genetic differences between individuals, i.e. clutch size is heritable (Table 4.2b, model 4). However, we found no indication that individuals differed in how they responded to spring temperature, i.e. there was not any significant variation in plasticity (Table 4.2b, model 5), and there was no improvement in the model when trying to fit a GxE model (Table 4.2b, model 6).

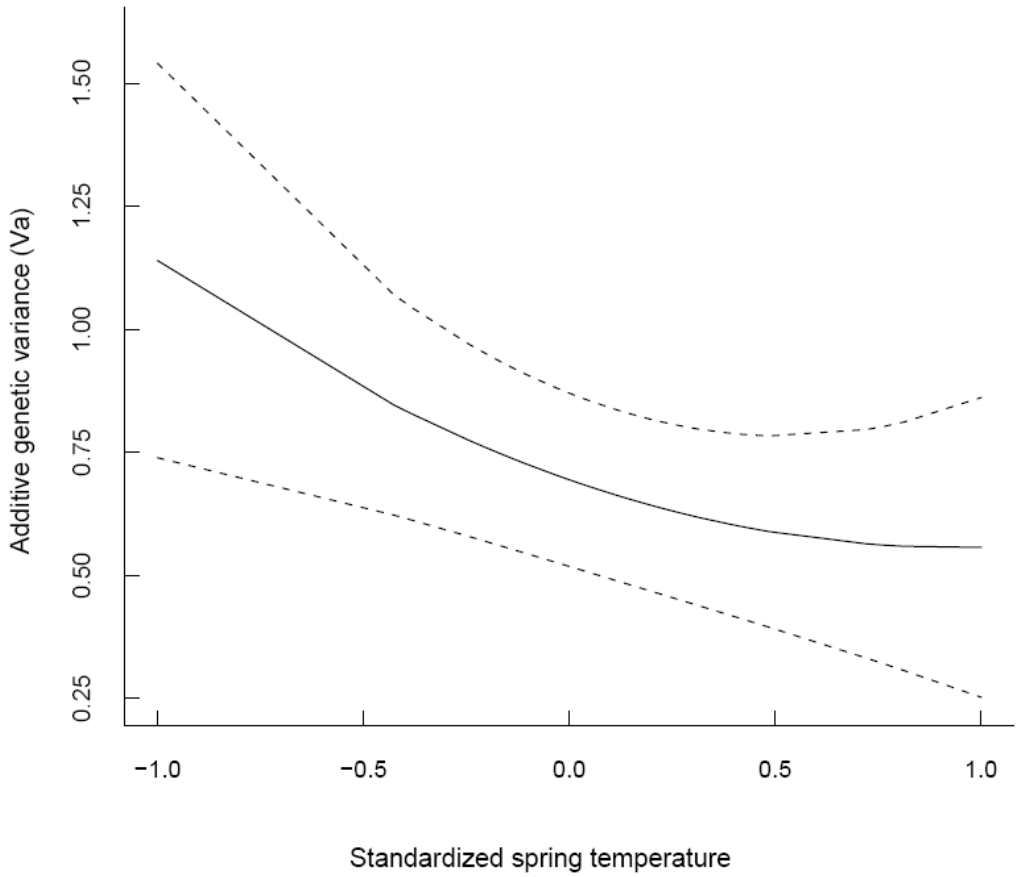
For the WW population we also found a difference between individuals in their average clutch size (Table 4.2b, model 3) and that this difference was partly genetically determined (Table 4.2b, model 4). In contrast to the HV population however, individuals also differed in how they adjusted their clutch size in relation to the temperature (IxE, Table 4.2b, model 5). Furthermore, some of this variation was associated with a genetic basis, i.e. there was significant GxE for clutch size (Table 4.2b, model 6). See Table 4.3 and Fig. 4.3 for an illustration of how the additive genetic variance for clutch size changes with temperature.

Table 4.3. Variance components for clutch size evaluated at the different standardized temperatures for the HV and WW population under model 6 in Table 4.2b.

Population	standardized temperature	*n*	μ ± SD	V_P (SE)	V_A (SE)	V_{PE} (SE)	V_{YR} (SE)	V_R (SE)	CV_A	h^2
HV	NA	3589	9.016 ± 1.950	3.840 (0.184)	0.566 (0.194)	0.851 (0.199)	0.628 (0.164)	1.795 (0.068)	8.344	0.148
WW	-1	189	8.6825 ± 1.606	2.977	1.139 (0.200)	0.279 (0.188)	0.421 (0.108)	1.138 (0.053)	12.291	0.383
	0	319	8.7476 ± 1.599	2.925	0.694 (0.088)	0.598 (0.083)	0.421 (0.108)	1.212 (0.054)	9.523	0.237
	+1	291	9.1306 ± 1.652	3.164	0.557 (0.152)	0.974 (0.164)	0.421 (0.108)	1.212 (0.054)	6.100	0.176

Note that the variance components were evaluated at different standardized spring temperatures for the WW population due to the significant GxE interaction. As there was no IxE and no GxE for clutch size in the HV population the total number of records, overall mean and variance components are reported. As there were no records in the dataset at exactly 0 standardized temperature, sample size and mean laying date value is given for the nearest temperature record (-0.0012). CV_A is the coefficient of additive genetic variance.

Fig. 4.3 Changes in additive genetic variance for clutch size with standardized spring temperature for the WW population under model 6 in Table 4.2 with approximate 95% confidence interval.



4.4.4 Bivariate random regression animal model in HV population

Comparing the full 4x4 phenotypic matrix to a model in which all between-trait covariances were constrained to be zero, indicated that the two traits showed some phenotypic covariance(s) ($\chi^2_4 = 18.60$, $P < 0.001$). We subsequently tested, firstly, the significance of the between-trait covariance in reaction norm elevations. This showed a strong negative phenotypic correlation between clutch size and laying date ($\chi^2_1 = 18.60$, $P < 0.001$, $r_p = -0.264$, SE = 0.047), i.e. individuals that lay on average early have larger average clutch size. Secondly, we tested all three other covariances, but there was no indication of any other covariance term being significant ($\chi^2_3 = 0$, P

= 1), indicating that the phenotypic covariance did not show a significant change with the environmental conditions.

At the genetic level, the model that included covariance between all four genetic reaction norm components was not a significant improvement over the phenotypic model and thus none of the four between-trait genetic covariances were significant ($\chi^2_4 = 6.20$, $P = 0.18$). When explicitly testing the genetic correlation between laying date and clutch size, we found it was positive, although non-significant ($\chi^2_1 = 3.00$, $P = 0.08$, $r_G = 0.560$, SE = 0.387, note that this is slightly different estimate to that given in Table 4 S1 something which is due to problems with obtaining SE on the estimate in that model and so a simpler model was run to obtain the SE of the r_G estimate provided here), and so of opposite sign to the phenotypic correlation between laying date and clutch size found above. Furthermore, a separate test for the three other genetic covariances also did not show any significant effect ($\chi^2_3 = 3.20$, $P = 0.36$) suggesting that there was no significant change in the genetic correlation with environmental conditions. The lack of significant genetic covariances when tested separately is thus consistent with the result from the model in which all genetic covariances were tested simultaneously. For a full breakdown of the bivariate phenotypic and genetic (co)variance plasticity matrices see Table 4.S1.

4.4.5 Bivariate random regression model in WW population

Similarly to the HV population there was a strong indication that some of the phenotypic between-trait covariances were significant ($\chi^2_4 = 16.18$, $P = 0.003$). As above we subsequently tested, firstly, the between-trait covariance in clutch size elevation and laying date elevation which showed a highly significant negative

correlation ($\chi^2_1 = 13.40$, $P < 0.001$, $r_P = -0.341$, SE=0.029). Thus, as for the HV population, early breeding birds have on average larger clutch size than late breeding individuals. When we tested the other three phenotypic between-trait covariances, to see if the phenotypic correlation would change over the range of environmental conditions, they were not significant ($\chi^2_3 = 2.78$, $P = 0.427$) indicating that the phenotypic correlation remained more or less constant.

Testing all additive genetic between-trait covariances there was, in contrast to the HV population, a significant effect ($\chi^2_4 = 12.32$, $P = 0.015$), indicating that one or more of the covariances were significant. As above, we explored, firstly, the additive genetic covariance between clutch size elevation and laying date elevation and this was, in contrast to the HV population, significant ($\chi^2_1 = 9.02$, $P = 0.003$, $r_G = -0.310$, SE = 0.090). Furthermore, when we tested the three other between-trait covariances (i.e. the covariance between clutch size elevation and laying date slope, between clutch size slope and laying date elevation and between clutch size slope and laying date slope), there was no indication that these were significant ($\chi^2_3 = 3.30$, $P = 0.348$).

Our results thus suggests that plasticity for laying date and plasticity for clutch size are not statistically associated and that the phenotypic and genetic covariances between laying date elevation and clutch size elevation did not change with the environment. There is a full decomposition of the phenotypic variance-covariance plasticity matrix in Table 4.S1.

4.4.6 Between-population comparison

To compare the reaction norm patterns in the two populations we fitted laying date and clutch size in the two populations in two separate random regression models.

Thus, one bivariate model where laying dates in HV and WW was included as two separate traits (with no covariance between them) and a second bivariate model where clutch size in HV and WW was included as two separate traits (again with no covariance between them).

There was significantly more between-individual variation in the average laying dates (V_I) in the WW population than in the HV population on the phenotypic level ($\chi^2_1 = 15.14$, $P < 0.001$), but no suggestion that this was the case on the genetic level ($\chi^2_1 = 1.90$, $P = 0.168$), although the estimated additive genetic variance was higher in the WW population than in HV. This implies no significant difference in additive genetic variance for laying date (in the average environment) between the two populations.

As a direct test for differences in IxE pattern for laying date between the two populations, we constrained the population specific variance-covariance matrices to be equal and compared this with a model in which they were unconstrained (see Methods). This test was highly significant ($\chi^2_3 = 19.02$, $P = 0.0003$), confirming that the between-individual reaction norm patterns in these two populations are very different. When we repeated this test on the genetic level, i.e. when comparing the GxE patterns in the two populations, the test was marginally non-significant ($\chi^2_3 = 7.50$, $P = 0.058$). This may reflect a lack of power, but in any case we are thus unable to categorically rule out that the between-population difference is genetic rather than environmental. Nevertheless, this does lend some support to the observations from the univariate random regression animal models that the observed (albeit non-significant) GxE pattern is different in these two populations (Fig. 4.2).

Phenotypic variation in clutch size (elevation) reaction norms did not differ between HV and WW ($\chi^2_1 = 3.39$, $P = 0.065$), and when comparing the additive

genetic variance in reaction norm elevation for clutch size between the populations there was also no suggestion of a significant difference ($\chi^2_1 = 0.15$, $P = 0.702$). Furthermore, the reaction norm plasticity matrices for clutch size in the two populations were not significantly different, at the phenotypic level ($\chi^2_3 = 4.60$, $P = 0.204$), or at the genetic level ($\chi^2_3 = 2.40$, $P = 0.494$).

4.5 Discussion

We explored and compared the multivariate genetic basis of variation in phenotypic plasticity in response to changing environmental conditions in two long running individual-based study populations of great tits. Very few studies have compared the multivariate patterns of plasticity, and our study is the first to do so in natural populations using the random regression animal model framework. We found that, although both populations exhibited similar population-level trends in laying date (see Fig. 4.1 e and f) and clutch size (Fig. 4.1 g and h), they differed in laying date IxE pattern, although when we partitioned this further the GxE pattern was marginally non-significant.

The bivariate random regression models showed little indication that individuals that were plastic for laying date also showed plasticity in clutch size: thus in these two populations there was little evidence for phenotypic integration. Interestingly, the multivariate analysis indicated that the genetic correlation between clutch size and laying date in the two populations was of opposite sign, despite a similar phenotypic correlation. A significant negative phenotypic and genetic correlation was found in the WW population, but in the HV population the phenotypic correlation was negative and the genetic correlation positive (although non-significant).

Furthermore, when we compared the reaction norm patterns for laying date in the two populations there was a significant difference on the phenotypic level (IxE); at the genetic level (GxE) this was marginally non-significant. For clutch size, however, we did not find any significant difference in reaction norm patterns at the phenotypic (IxE) or at the genetic level (GxE), although the population-specific univariate models suggested IxE and GxE in the WW population and in the HV population neither was significant (Table 4.2b). Hence, our results highlight the need to be cautious about extrapolating results from one population to other populations of the same species when predicting responses to climate change.

4.5.1 Population-level response

Although there has been a similar increase in spring temperatures over the study period in both populations (Fig. 4.1a and b), there was a stronger relationship between onset of laying and spring temperatures in WW (Fig. 4.1d) than in the HV (Fig. 4.1c) population, as well as more rapid advancement in mean laying dates in WW (Fig. 4.1f) than in HV (Fig. 4.1e); these results agree with previous analyses (McCleery and Perrins 1998; Visser et al. 1998; Gienapp et al. 2006; Garant et al. 2008). Spring temperature has been shown to have a profound impact on seasonal timing of reproduction in birds in general (reviewed in Slagsvold 1976; Dunn 2004), as well as in these two populations in particular (Visser et al. 1998; Charmantier et al. 2008), and so represents a reasonable environmental variable with which to examine phenotypic plasticity in laying date. Although it is may be less clear that this is a good measure with which to examine plasticity in clutch size, we emphasize that we are concerned here with the effect increasing spring temperatures have on general plasticity patterns. However, this clearly does not mean that clutch size could not

respond to other environmental factors (e.g., density, Both et al. 2000; Wilkin et al. 2006).

4.5.2 Individual-level variation in plasticity

In common with other studies that have estimated components of variance in laying date (e.g. Sheldon et al. 2003; Brommer et al. 2005), we found that females differed significantly in their average (phenotypic) laying date (significant V_I component, Table 4.2a) and that a significant amount of this variation was due to additive genetic effects (V_A, Table 4.2a). The estimated heritability for laying date in HV and WW (Table 4.3) correspond well with what has been shown previously for these two populations (Gienapp et al. 2006; Garant et al. 2008).

We found that there was significant between-individual variation in phenotypic plasticity (IxE) for laying date in both populations (see Table 4.2a), indicating that females differ in how they adjust their laying date in relation to the spring temperature. This supports the findings from an earlier study in the HV population (Nussey et al. 2005), but is in contrast to a recent study in the WW population that did not find statistical support for IxE (Charmantier et al. 2008). There are several possibilities as to why our result is different from those of Charmantier et al. (2008), some of which we can exclude. For instance, we used a heterogeneous error structure (see Methods) whereas Charmantier and colleagues used a homogenous error structure, but re-running the models with a homogenous error structure gave the same conclusion of IxE (although $P = 0.016$ compared to $P = 0.008$ with a heterogeneous error structure). The number of years included in this study is also different (1960 – 2008 vs 1973 – 2006 in our study), but again this is unlikely to be the cause of the difference, unless birds from the period 1960 – 1973

were much less plastic than individuals from the later part, which seems unlikely. Furthermore, Charmantier et al. (2008) only used females that bred three times or more whereas we used all breeding females (i.e. also those that only bred once); although this should not influence the estimate of variance in plasticity itself, as it is only females with at least two breeding records that provide information on plasticity, it may still influence the statistical power to detect IxE because including all females will increase sample size and thus the precision of the estimated variation in elevations. Finally, as mentioned above, we used a different environmental measure (mean temperatures) whereas Charmantier et al (2008) used the 'warmth sum' (sum of daily maximum temperatures during the period 1st March – 25th April). Repeating the analysis using 'warmth sum' instead of mean temperature over the period 15th February – 25th April (i.e. the same period as for mean temperature used in this study) gave identical results to those reported in Charmantier et al. (2008), i.e. no support for any IxE interaction ($\chi^2 = 0.02$, d.f.=2, $P = 0.99$, compared to a standard animal model) and the estimated slope variance was essentially zero ($\sigma^2_s < 0.0001$). Furthermore, using the 'warmth sum' over the period 1st March – 25th April (as used by Charmantier and colleagues) yielded again no support for IxE and estimated slope variance close to zero in agreement with that reported by Charmantier et al. (2008). Thus, it is very likely that the use of mean temperatures instead of maximum temperatures is the reason for the different conclusions reached between our study and that by Charmantier et al. (2008). Note however, that we did find evidence for differing degrees of plasticity (IxE) in the two populations.

In many ways the different conclusions about IxE we reach using the two different (but still highly correlated, $r_s = 0.963$, $P < 0.001$) environmental variables are a cause for concern. Although it is clear that plasticity is only defined in relation

to a particular environment (Scheiner 1993), it also raises the question of how we can draw general conclusions from different studies that use different environmental measures. This is just as much a concern for laboratory-based studies as precise replication of environments is extremely difficult: more work needs to be carried out assessing the sensitivity of random regression models to detect patterns of IxE (and GxE) for different environmental variables if we are to be able to generalize plasticity patterns across populations and species.

Although population-level plasticity in clutch size is frequently reported (e.g. Both et al. 2000), it is less common to look at individual-level variation in this plasticity (but see Przybylo et al. 2000). In this study we found support for between-individual variation in clutch size plasticity, and a genetic basis to the variation in plasticity, in the WW population, but not in the HV population (Table 4.2b, Table 4). We are only aware of one other study looking at between-individual variation in clutch size plasticity, Przybylo et al. (2000) found that collared flycatchers (*Ficedula albicollis*) differed in their adjustment of clutch size in relation to the NAO-index (North Atlantic Oscillation), a high NAO value (indicating warm moist winters in Scandinavia) resulted in higher clutch sizes. The individual variation in clutch size plasticity found in the collared flycatcher population and in the great tits in WW, contrasts with the lack of such plasticity in the great tits in the HV population, suggesting that inter-population differences, for example due to characteristics of the experienced environmental conditions, may be important.

Although we have shown here that both populations show IxE variation in laying date, there is no *a priori* reason to expect this for other populations and species. For instance, Reed et al. (2006) found no between-individual variation in

plasticity in common guillemots (*Uria aalge*) although there was a clear population-level plasticity in relation to spring temperature.

4.5.3 Genetic basis to phenotypic plasticity variation

Both populations showed IxE for laying date, but when we tried to separate the IxE variation into its genetic (GxE) and environmental (PExE) effects we found that such a model was not significantly better (Table 4.2a), suggesting that we might not have the statistical power to separate the two. Nonetheless, this partitioning revealed that the majority of the variation in plasticity is due to additive genetic effects in the HV population whereas this was not so in the WW population (Table 4.S1), and this difference was also apparent when we visualised the change in V_A with increasing spring temperature (Fig. 4.2) using the character state approach. Whereas there was a non-significant increase in additive genetic variance with spring temperature for laying date in the HV population (Fig. 4.2a), there was no such change for the WW population (Fig. 4.2b). Thus our finding in the HV population is similar to the conclusions reached by Brommer et al. (2008) investigating the genetic basis of variation in laying date plasticity in a population of common gulls (*Larus canus*), who found IxE but no statistical support for GxE.

Our result of no GxE for laying date plasticity in the HV population is in contrast to the findings from Nussey et al. (2005) who estimated the genetic basis of variation in plasticity using a 'two step approach'. The 'two step approach' is different to a random regression approach in that one first runs a linear mixed effect model on the phenotypic values and extracts the 'best linear unbiased predictors' (BLUPs) for elevation and slope, and then use these estimates in an animal model to estimate the heritability and thus genetic basis of laying date elevation and laying

date slope. This approach ignores the large uncertainty associated with the BLUP estimates and is considered to be less robust than running a random regression model where this uncertainty is taken into account (Nussey et al. 2007; Brommer et al. 2008). For instance, Nussey et al. (2005) failed to find a significant heritability of elevation, only for slope, suggesting that laying date itself is not heritable, but only its plasticity is. A similar result, using the same approach, was reported by Brommer et al. (2005) in a long term study of collared flycathers breeding on Gotland, Sweden, where there was no heritable basis for elevation (or slope). This lack of laying date heritability contrasts with previous findings in the same population (Merilä and Sheldon 2000; Sheldon et al. 2003). Our results (Table 4.2a) clearly suggest a heritable basis of laying date elevation in the HV (and WW) population, as has previously been shown (van Noordwijk et al. 1981; Gienapp et al. 2006). However, our finding that most of the variation in plasticity is due to additive genetic effects (Table 4.S1) does lend some support to Nussey and co-workers` findings, and, when repeating the 'two step approach' on our dataset (and environmental variable), we also found statistical support for GxE (unpublished results). Thus, the lack of a GxE when using the random regression approach could be because of lack of power to detect a GxE interaction, even if present, in this system. We agree with Brommer et al. (2008), however, that we need to be careful about dismissing non-significant GxEs because, not only may the power to detect them in natural populations be low, but also changes in VA (and possibly heritability) can be very different if we assume there is no GxE.

Interestingly, there was not only IxE variation but also GxE variation, and thus a heritable basis of plasticity, for clutch size in the WW population (Table 4.2b). However, the additive genetic slope variance and genetic covariance was small and

so there was relatively little change in V_A with temperature (Fig. 4.3). An overview of the results obtained from this study compared to the findings of Nussey et al. (2005) and Charmantier et al. (2008) is in Table 4.5.

4.5.4 Multivariate plasticity patterns

The four between-trait covariances determine the degree to which the two traits and the plasticity in these traits are correlated and thus the populations' multivariate reaction norm pattern. Our results indicate little evidence of any significant correlation between plasticity in the two traits. Indeed the only significant between-trait correlations were between elevations of the two traits with a significant negative correlation between laying date and clutch size both at the phenotypic (both populations) and at the genetic level (only WW). The presence of a negative phenotypic correlation between clutch size and laying date is commonly found in a number of different populations and species of birds (e.g. Klomp 1970; Perrins and McCleery 1989; Winkler and Allen 1996). Interestingly, although both the HV and WW population showed a significant negative correlation between clutch size and laying date, the genetic correlation was positive (although non-significant) in the HV population and significantly negative in the WW population. There are surprisingly few studies that have estimated the genetic correlation between clutch size and laying date, and so far there seems to be no emergent pattern. A negative genetic correlation has been found in collared flycatchers and great tits (Sheldon et al. 2003; Garant et al. 2008), whereas a previous study in the HV population also found the genetic correlation to be non-significant (Gienapp et al. 2006). It is difficult to speculate as to the causes behind the divergent correlations observed, but genetic correlations are sensitive to allele frequencies and subject to rather large sampling variance and so

can differ substantially between populations (Falconer and Mackay 1996). Nevertheless, it is interesting that strong correlational selection involving these two traits has previously been demonstrated for the WW population, and the genetic correlation estimated here agrees with the direction of the axis of correlational selection (see Garant et al. 2007).

Very few studies have investigated if the phenotypic, genetic and/or environmental correlation changes with environmental conditions, which would be the case if one or more of the covariances between elevation and slope, or between slopes in two traits would be significant. Robinson et al. (2009) studied how the phenotypic, additive genetic and environmental correlation between horn length and body weight, horn length and parasite load and between body weight and parasite load changed with environmental conditions in a wild population of Soay sheep (*Ovis aries*). In that population the genetic correlation between horn length and body weight, and that between horn length and parasite load both showed a significant decline in more favourable environmental conditions, whereas the phenotypic correlation between horn length and body weight decreased with increasing environmental conditions, and that between horn length and parasite load showed a positive increase. Thus there seems to be no clear expectation as to how the environment should influence phenotypic and genetic correlations in the wild, it will most likely be trait and population specific.

Although the concept of between-trait correlations in plasticity was first suggested over twenty years ago by Schlichting (1986), it has, perhaps due to experimental and statistical difficulties, received relatively little attention. For instance, we are unaware of other studies on the multivariate patterns of plasticity in natural populations and so it is difficult to suggest what the case might be in other

populations. Interestingly, some experimental work by Newman (1994) suggests that the correlation between-trait plasticities can depend on the environmental variable used to study it. Newman (1994) collected families of spadefoot toads (*Scaphiopus couchii*) and raised them in the lab with different temperature and food availability regimes and showed that plasticity in size and plasticity in larval period were negatively correlated for the food regimes, but positively correlated under temperature variation. This clearly shows that even if we do find a between-trait correlation in plasticity this may be subject to change depending on the environmental variables we use in the context of studying plasticity. It is clear that the possibility of between-trait correlations in plasticity, or so-called phenotypic integration (Schlichting 1986), in natural populations requires more studies.

Table 4.5. Summary table of results from this study compared to previous studies on the same populations.

This study	Laying date		Clutch size	
	HV	WW	HV	WW
Population-level plasticity	Yes	Yes	n.s	n.s
IxE	Yes	Yes	n.s	Yes
GxE	n.s	n.s	n.s	Yes
Previous studies[1,2]	HV	WW	HV	WW
Population-level plasticity	Yes	Yes	-	-
IxE	Yes	n.s	-	-
GxE	Yes	n.s	-	-

Note that both previous studies used a two-step model (see Discussion), whereas the results from this study uses a random regression animal model. [1]Nussey et al. (2005) for HV population, [2]Charmantier et al. (2008) for WW population. n.s. = not significant, - = not tested.

4.5.5 Between-population comparison

Although there was significantly more phenotypic variance in laying date in WW than in HV, there was no indication that the amount of additive genetic variance was significantly different. However, when we compared the phenotypic IxE patterns, there was a clear difference between the two populations and this was also reflected at the genetic level by the near-significant difference in the genetic reaction norm patterns (GxE pattern). The between-population comparison is interesting for several reasons: firstly, it demonstrates that although both populations showed IxE, there is considerable more within-population variance in the reaction norm slopes in HV than in WW. Secondly, this difference was also present, although marginally non-significant, when comparing the genetic reaction norm patterns, and this support the impressions from the separate univariate analysis (see Table 4.2a and Fig. 4.2). It is unclear why there is more variation in plasticity between individuals in the HV population than in WW. A plausible reason may be that it is due to differences in environmental heterogeneity in these two populations. If the HV population has a more heterogeneous environment than the WW population this may lead to such (environmentally-induced) variation. Unfortunately we did not have the statistical power to separate the IxE into its genetic and environmental part, although the population comparison did suggest a difference also in the genetic reaction norm patterns.

The population comparison also showed that the GxE pattern for clutch size that we detected in WW is not significantly different from that in the HV population. This probably reflects the fact that, although significant, plasticity in clutch size was low also in the WW population. No comparison between the two populations was actually significant for clutch size, suggesting similarity rather than difference, contrary to what would be inferred from the univariate population-specific analyses.

In conclusion, we have demonstrated that changing environmental conditions may not have the same consequences in different populations, even if it is the same species and rates of change in environmental conditions are similar. Using a multivariate approach, we have shown that plasticity in laying date and clutch size, two key life-history traits, was not significantly associated in these two populations. Furthermore, we found that the additive genetic correlation was of opposite sign in HV and WW, providing further support that it is important to be careful when extrapolating quantitative genetic parameters measured in a particular population onto other, unmeasured, populations.

Although our findings do not, as is the case in all quantitative genetics studies, allow us to dissect the molecular genetic basis of phenotypic plasticity they do suggest that, given the limited evidence for additive genetic variance in plasticity found in these two populations (and in WW in particular), a QTL approach (Lynch and Walsh 1998) to studying plasticity in natural populations might prove challenging, despite the recent development of linkage maps for such populations (Backstrom et al. 2006; van Bers et al. submitted).

Table 4.S1. The full variance-covariance matrices from the bivariate random regression animal model in the two populations. Variances are on the diagonal, covariances on the lower triangular matrix and correlations on the upper triangular matrix.

Table 4.S1a: Permanent environment variance-covariance matrices (PExE)

	HV population					WW population			
	σ_{LDe}	σ_{LDs}	σ_{CSe}	σ_{CSs}		σ_{LDe}	σ_{LDs}	σ_{CSe}	σ_{CSs}
σ_{LDe}	12.510	0.532	-0.703	-0.121	σ_{LDe}	10.270	0.822	-0.386	-0.524
σ_{LDs}	2.448	1.689	-0.217	0.132	σ_{LDs}	2.092	0.631	-0.389	1.377
σ_{CSe}	-3.467	-0.218	1.920	0.324	σ_{CSe}	-1.238	-0.309	1.001	0.923
σ_{CSs}	-0.013	0.003	0.007	0.0002	σ_{CSs}	-0.456	0.297	0.251	0.074

Table 4.S1b: Additive genetic variance-covariance matrices (GxE)

	HV population					WW population			
	σ_{LDe}	σ_{LDs}	σ_{CSe}	σ_{CSs}		σ_{LDe}	σ_{LDs}	σ_{CSe}	σ_{CSs}
σ_{LDe}	5.266	0.717	0.688	0.421	σ_{LDe}	11.940	-0.428	-0.310	0.151
σ_{LDs}	2.996	3.315	0.051	-0.167	σ_{LDs}	-1.354	0.837	0.550	-1.802
σ_{CSe}	1.615	0.095	1.047	0.525	σ_{CSe}	-1.323	0.621	1.525	-0.542
σ_{CSs}	0.131	-0.041	0.073	0.018	σ_{CSs}	0.176	-0.554	-0.225	0.113

CHAPTER 5

SPEEDING UP MICRO-EVOLUTION: CLIMATE CHANGE INDUCES A POSITIVE ASSOCIATION BETWEEN ANNUAL STRENGTH OF SELECTION AND HERITABILITY IN A WILD BIRD POPULATION

Husby, A, Visser, M.E. & Kruuk, L.E.B. submitted to *PLoS Biology*

5.1 Abstract

The heritability of a trait and the strength of selection acting on that trait are the two key parameters which determine any evolutionary response to selection. Despite substantial evidence that, in natural populations, both parameters may vary across environmental conditions, very little is known about the extent to which the two parameters covary in response to environmental heterogeneity. Here we show that, in a wild bird population, the strength of the standardized directional selection differentials on timing of breeding increased with increasing spring temperatures due to climate change, and that genotype by environment interactions predicted an increase in additive genetic variance, and heritability, of timing of breeding with increasing spring temperature. Consequently, there was a positive correlation between the strength of selection and the heritability of the trait in each year, something that may speed up the rate of micro-evolution.

5.2 Introduction

Predicting an evolutionary response to selection pressures requires knowledge of the strength of selection acting on a trait and of its heritability (Falconer and Mackay 1996). Although it has long been recognised that the strength, and direction, of selection often vary with environmental conditions (e.g. Grant et al. 1976) widespread recognition of the fact that heritability may also change with environmental conditions has been more recent (Hoffmann and Parsons 1991; Charmantier and Garant 2005). Taken together, these observations generate an expectation of an environmentally-driven covariance between the two parameters that, in theory, has the potential to either enhance (positive covariance) or constrain (negative covariance) any response to selection. Surprisingly, however, to our

knowledge only one study to date has quantified the association between annual estimates of selection and heritability in a heterogeneous environment (Wilson et al. 2006). In this paper, we present data from a long-term study of a great tit (*Parus major*) population known to be experiencing substantial shifts in climatic conditions, and test for the effects of the novel environmental conditions on the heritability and selection of a key life history trait, breeding time.

Many studies have found that selection is often strongest when environmental conditions are adverse (e.g. Grant and Grant 1995; Milner et al. 1999; Grant and Grant 2002; Garant et al. 2004; Wilson et al. 2006), and there is a clear indication that 'perturbed or stressed populations' have larger standardized selection differentials than 'undisturbed populations' (Endler 1986, p 208). For example, Garant and co-workers (2004) examined selection on fledgling body mass in a population of great tits and found that selection differentials were greater in years when average body mass was low and when the proportion of individuals surviving to recruitment was low, both conditions of which are characteristic of poor/adverse environmental conditions. In general, therefore, it seems clear that the strength of selection is often stronger when environmental conditions are adverse.

Conclusions regarding the effects of good versus adverse environmental conditions on the expression of additive genetic variance are more mixed however. Laboratory studies investigating the effect of environmental conditions have generally found a weak tendency for additive genetic variance and heritability to increase in stressful environments (reviewed in Hoffmann and Merilä 1999). This pattern, however, is in contrast to most studies from natural populations which find that, at least for morphological traits, additive genetic variance and heritability is generally low in unfavourable conditions (Hoffmann and Merilä 1999; Merilä and

Sheldon 2001; Charmantier and Garant 2005). Explanations for why heritability may change in different environmental conditions are reviewed by Hoffmann & Merilä (1999) and include changes in additive genetic variance (V_A), for example due to genotype-environment interactions, increased mutation and recombination rates, removal of alleles with low fitness by selection, environment specific effects on mutations, or, change in environmental variance (V_E). This may lead to either an increase or a decrease in heritability depending on which of the above factors are involved (Hoffmann and Merilä 1999).

The realisation that both the heritability of a trait and the strength of selection acting on the trait may vary with environmental conditions is a key insight, as such environmentally-induced variation may be important in determining the evolutionary dynamics of natural populations. In particular, the observation of a general increase in heritability of morphological traits in favourable conditions in natural populations (Hoffmann and Merilä 1999; Merilä and Sheldon 2001; Charmantier and Garant 2005) and that selection if often weak when environmental conditions are good (Milner et al. 1999; Wilson et al. 2006), leads to a expectation of a negative correlation between the heritability and the strength of selection, such that selection is strongest in years in which the expression of additive genetic variance is least. This covariance would severely constrain a response to selection, and could provide one explanation for the frequently observed stasis in natural populations (Wilson et al. 2006).

In contrast to morphometric traits, life history traits do not appear to show a clear indication of higher heritability in stressful environments (Charmantier and Garant 2005). This makes it more difficult to predict how, or if, heritability and selection of life history traits may covary in a heterogeneous environment.

Surprisingly, despite the potential importance of environmentally-induced selection - heritability correlations, studies examining this are extremely scarce, and we are only aware of one previous study that has examined this. Wilson et al. (2006) found that the strength of selection on body weight in a free-living population of Soay sheep (*Ovis aries*) in a given year was negatively correlated with the heritability of body weight, suggesting a possible constraint on the potential for evolution of body weight in this species. However, so far no study has, to our knowledge, examined this in a life history trait. Hence, we do not know if such relationships are common in nature, and whether they are generally negative, which may constrain an evolutionary response, or whether there are examples of positive covariance between strength of selection and heritability, something that would speed up an evolutionary response.

Here we use data from an exceptionally long term study population of great tits (*Parus major*) in the Netherlands to investigate how selection and heritability of a key life-history trait (timing of breeding, or "laying date") vary in relation to rapid changes in environmental conditions (spring temperature). This system is particularly well suited to an exploration of the association between selection and heritability in a variable environment because phenotypic data, pedigree data and a thorough understanding of how environmental conditions influence laying date are available. Previous studies in this population have shown that warm spring temperatures lead to, firstly, earlier laying dates (Visser et al. 1998). Furthermore, warmer temperatures lead to reproduction being mistimed relative to the food peak (Visser et al. 1998), which results in a decrease in both the number and size of fledglings (Visser et al. 2006), and in the proportion of females producing a second clutch (Husby et al. 2009). Spring temperatures are thus not only directly related to observed variation in laying dates, but can also be used as a measure of environmental quality in the

population. In addition, spring temperatures have now increased significantly above what the population has previously experienced (Husby et al. in review), providing an ideal opportunity to study how novel environmental conditions may influence evolutionary dynamics.

5.3 Materials and methods

5.3.1 Study system and data collection

The data were collected in the Hoge Veluwe National Park, the Netherlands, during the period 1973 to 2007. Nest boxes were visited at least once every week during the breeding season (April – June). The laying date of the first egg of a female's clutch (laying date, LD) was calculated from the number of eggs found during the weekly checks, assuming that one egg was laid per day. Both parents were caught and individually marked on the nest using a spring trap when the young were 7 – 10 days old. Laying dates are presented as the number of days after 31st March (day 1 = 1st April, day 31 = 1st May). We only used information on the first clutch, defined as any clutch started within 30 days of the first laid egg in any given year. Replacement and second clutches (which currently compromise < 5% of breeding attempts, Husby et al. 2009) were thus excluded from the analysis. In total therefore, we had information on about 3852 breeding records from 2394 females. More details about the study population can be found in van Balen (1973).

Temperature data was obtained from the De Bilt weather station of the Royal Dutch Meteorological Institute (www.knmi.nl/klimatologie/daggegevens) and was calculated as the daily average temperature over the period 13th March – 20th April, which is the period that was found to best predict the onset of laying using a sliding window approach (see Husby et al. in review for more details).

5.3.2 Selection analysis

To test for a relationship between spring temperature and the strength of selection on laying date we took two approaches. Firstly, we used a generalized linear mixed effects model (GLMM) with a Poisson error link fitted in ASREML-R (Gilmour et al. 2006) to model the relationship between number of recruits (fitness) and standardized laying date (see below), and to test its dependence on spring temperature (as measured by the interaction term between standardized laying date and spring temperature). Individual identity and year were included as random effects to account for repeated measures on the same individuals and on years. Secondly, we estimated annual standardized selection differentials, $S = \mathrm{Cov}(\omega, z)$, as the covariance between a female's relative fitness (ω) and her standardized laying date using the Land-Arnold approach (Lande and Arnold 1983). Thus, laying date was standardized to have zero mean and unit variance (creating *z*-scores) within each year and yearly fitness values (number of recruits produced per year) were divided by the mean number of recruits produced in the given year to give relative fitness scores (ω) for each individual. We then regressed the annual standardized selection differentials against the environmental values using a least-squares regression in R 2.8.0 (RDevelopmentCoreTeam 2007).

5.3.3 Pedigree structure

Quantitative genetic analyses require knowledge about the relationships among individuals within a population. Here, a pedigree was constructed where all ringed females known to have bred were assigned to their mother and father as determined from observational data. In cases where brood manipulation experiments had been carried out and chicks had been moved between nests, we assigned the genetic parent

rather than the social parent. If only one parent was known, we 'dummy coded' the missing parent to preserve sibship information (note that we did not assign a phenotype to this parent). The extra-pair paternity (EPP) rate is unknown in this population, but is generally found to be low (3 - 9 %) in other populations of great tits (Verboven and Mateman 1997; Lubjuhn et al. 1999) and as extra pair paternity rates of less than 20% have been shown to have a negligible impact on heritability estimates (Charmantier and Reale 2005) using a social pedigree is unlikely to be problematic.

5.3.4 Quantitative genetic analyses

Phenotypic trait variances can be separated into genetic and environmental causes of variation using an 'animal model' (Henderson 1950; Lynch and Walsh 1998; Kruuk 2004). However, here, we were more interested in whether the variance components changed with environmental conditions and to examine this we used a 'random regression animal model' (Meyer 1998). Random regression models use covariance functions (Kirkpatrick et al. 1990) to explicitly fit variance components as a function of the environment and hence allow a detailed examination of how environmental heterogeneity may influence genetic architecture. Thus our model was:

$$\mathbf{LD}_i = \mathbf{Xb}_i + \mathbf{Z_1}\varphi(\mathbf{a}_i,n_1,T) + \mathbf{Z_2}\varphi(\mathbf{pe}_i,n_1,T) + \mathbf{Z_3yr}_i + \mathbf{e}_i \quad (1),$$

where $\mathbf{LD}_i$ is the vector of the individual laying dates and $\mathbf{X}$, $\mathbf{Z_1}$, $\mathbf{Z_2}$ and $\mathbf{Z_3}$ are the design and incidence matrices relating to the fixed and random effects of the additive genetic ($\mathbf{a}_i$), permanent environment ($\mathbf{pe}_i$) and year ($\mathbf{yr}_i$) observations respectively. Fixed effects ($\mathbf{b}_i$ vector) included age as a two level factor (first year breeder or older), to correct for the fact that laying date generally is later in young birds compared to older birds in great tits (Wilkin et al. 2006), and spring temperature to

account for the population-level response in mean trait value. Year (**yr** vector) was included as a random effect in order to model variation between years not explained by spring temperature and a permanent environment effect ($\mathbf{pe}_i$ vector) was fitted because of the repeated sampling of the same individuals; this also reduces inflation of estimates of the additive genetic variance due to environmental factors (Kruuk and Hadfield 2007). The error term (**e** vector) was partitioned into three decade- specific (1973-1984, 1985-1996, 1997- 2007) groups, thus allowing residual errors to vary between decades. $\varphi(\mathbf{a}_i,n_1,T)$ is the random regression function of order n_1 of the additive genetic effect of individual *i,* and, similarly, $\varphi(\mathbf{pe}_i,n_2,T)$, is the random regression function of order n_2 of the permanent environment effect.

Because we were only interested in whether the two variance components (and particularly V_A) changed with the environment we only fitted two models. The first model was a zero order function ($n_1=n_2=0$) for both V_A and V_{PE} and so variance components are constant across the environment. In the second model we fitted a first order polynomial ($n_1=n_2=1$) for both V_A and V_{PE}, thus allowing both variance components to vary across the environment. These two models were then compared using a likelihood-ratio test by calculating two times the difference in log likelihood which is chi-square distributed with degrees of freedom equal to the difference in degrees of freedom between the two models (Pinheiro and Bates 2000). As the model where both variance components were allowed to vary was significantly better than a model in which they were assumed to be constant (see results) we used the estimates from the first order polynomial model to predict the change in V_A (and V_{PE}) across the environment. The environment-specific additive genetic covariance matrix, **G**, was then obtained as $\mathbf{G}= \mathbf{zQz}^T$, where **z** is the vector of orthogonal polynomials evaluated at standardized temperature values and **Q** is the additive genetic variance-

covariance matrix of the random regression parameters. Approximate standard errors for the (co)variance components of **G** as a function of the temperature values were calculated according to Fischer et al. (2004), with confidence intervals defined as twice the standard errors. Finally, environment-specific heritability estimates were calculated as the environment-specific V_A estimate divided by the environment-specific V_P estimate from the model in which both V_A and V_{PE} varied with the environment.

For more information about the use of random regression animal models in natural populations see Wilson et al. (2005) and Husby et al. (provisionally accepted). All animal models were fitted using REML methods implemented in ASReml v 2.0 (Gilmour et al. 2006).

5.4 Results

5.4.1 Environmental dependent strength of selection

We found, firstly, an overall strong selection on standardized laying date with early breeding birds having higher fitness than late breeding individuals (Table 5.1). Indeed, 29 out of the 35 estimates of annual selection were negative (Fig. 5.1), reflecting general selection for earlier breeding as has previously been shown in this population (Visser et al. 1998; Gienapp et al. 2006). Secondly, the interaction between standardized laying date and standardized spring temperature was significantly negative (Table 5.1), which demonstrates that with increasing spring temperatures the relationship (slope) between fitness and standardized laying date became more negative (i.e. slope steeper in warmer years). Consequently, selection for breeding early was significantly stronger in warm years than in cold years, i.e. the strength of selection on lay date varied with environmental conditions. This result

was confirmed by regressing the annual standardized selection differentials against temperature, there was a significant increase in the (absolute) magnitude of the standardized selection differentials with increasing temperatures (β = -0.26, se = 0.09, t_{33} = -2.804, *P* = 0.008, Fig. 5.1).

Table 5.1. Mixed model selection analysis for effects of standardized laying date and mean centred spring temperatures on number of offspring recruited to the breeding population each year. The model was fitted in ASREML-R using Poisson error structure (log link function) with individual identity and year included as random effects. Significance of fixed effects was assessed based on their Wald test statistic.

Effect	$\beta \pm$ SE	Wald-statistics	*P* - value	Variance (SE)
Random: Individual identity				0.295 (0.044)
Year				0.0636 (0.167)
Fixed: sLD	-0.215 ± 0.030	47.424	< 0.001	
TEMP	-0.018 ± 0.147	0.001	0.987	
sLD X TEMP	-0.109 ± 0.034	10.488	0.001	

Fig. 5.1 Standardized selection differentials on laying date regressed against standardized spring temperature. Each point is the estimated standardized selection differential (±1 s.e) and the solid line represents the least squares regression line of selection differentials on spring temperature.

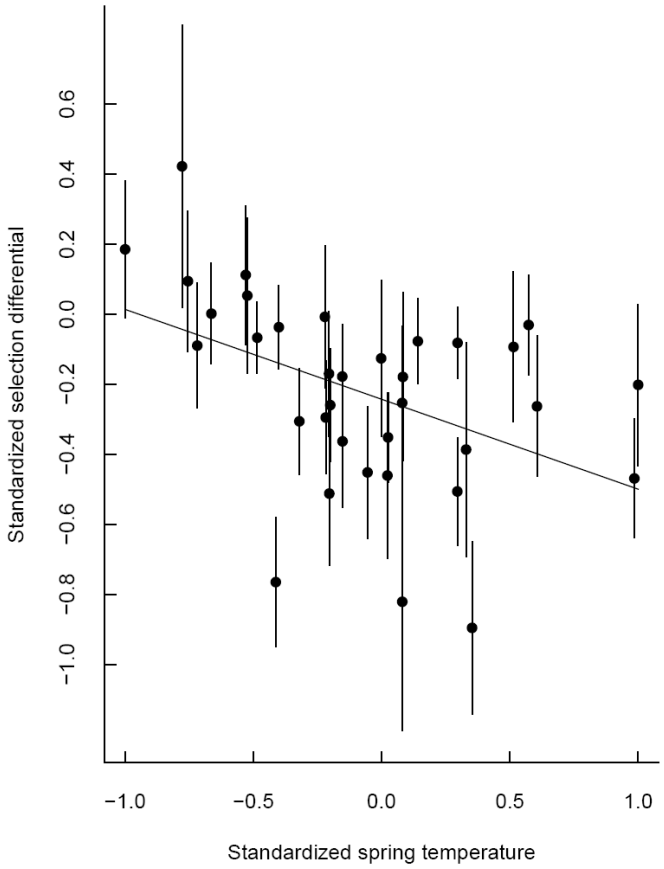


Fig 1

5.4.2 Environmental dependence of additive genetic variance

Comparing a model where V_A and V_{PE} were constant across the environment to a model where V_A and V_{PE} could vary with the environment gave strong support for environmental dependence of V_A and V_{PE} ($\chi^2_4 = 74.90$, $P < 0.001$). Consequently, we used the predictions from the model in which the two variance components varied with spring temperature to generate the predicted change in V_A and h^2 and see what

consequences this may have for a potential covariance with the selection differentials.

The estimated environment-specific G-matrix predicted a rapid increase in V_A with increasing standardized spring temperatures (Fig. 5.2a). Equally importantly, when we calculated the year specific heritability estimates there was a corresponding rapid increase with increasing temperature (Fig. 5.2b, each point represents a year (and thus temperature) specific h^2 estimate).

Fig. 5.2 a) Estimated change in additive genetic variance (V_A) with 95% confidence interval against standardized spring temperature as predicted from the model in which V_A and V_{PE} vary with temperature. b) Estimated change in heritability across spring temperature as predicted from a model where V_A and V_{PE} changed with standardized spring temperature, each point represent the year specific (and thus temperature specific) h^2 estimate.

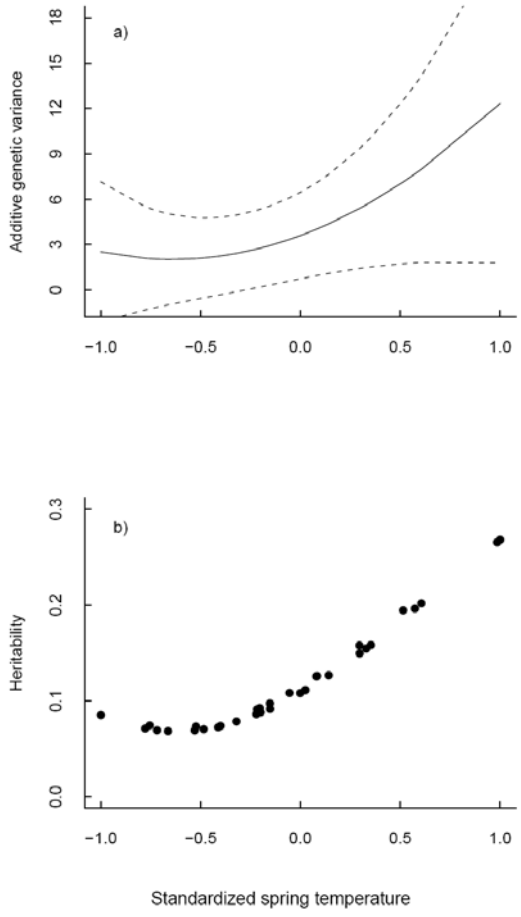


Fig 2

5.4.3 Association between strength of selection and heritability

As both the strength of the standardized selection differentials and heritability for laying date increased with increasing spring temperatures, this consequently generated a significant correlation between the strength of selection and the annual estimates of heritability (r_s = -0.47, P = 0.005, note that because there is selection for early breeding selection differentials are negative, but there is a *positive* association between the strength of selection and heritability). Hence, in years when selection on laying date was strong, estimated heritability was higher than in years when selection was weak (Fig. 5.3). This result was robust against removal of the selection differential outliers (excluding S > -0.6, r_s = -0.54, P = 0.002; excluding S > 0.25, r_s = -0.43: P = 0.012), the two extreme heritability estimates (excluding h^2 > 0.25, r_s = -0.46, P = 0.007) and all six outliers (i.e. only retaining points with criteria: -0.6 < S < 0.25 and h^2 < 0.25, r_s = -0.48, P = 0.009) as no exclusions changed the direction or significance of the correlation.

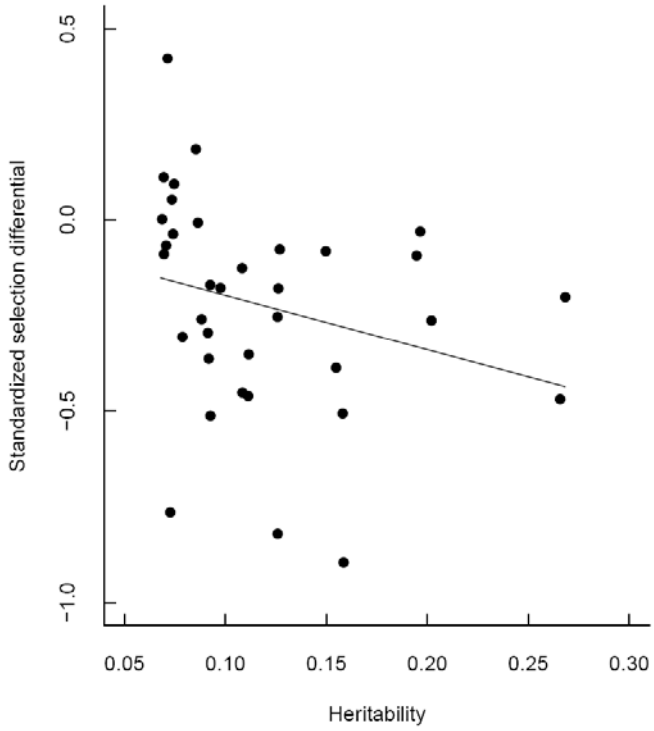


Fig 3

Fig. 5.3 Annual standardized selection differentials against annual estimated heritability with the estimated regression line from a least squares regression (regression slope: $\beta = -0.257 \pm 0.091$). Removal of outliers (see main text) did not change the direction or significance of this correlation. Note that because there is selection for early breeding selection differentials are negative, but there is a *positive* association between the strength of selection and heritability.

5.5 Discussion

We have demonstrated here a positive association between the strength of selection and the heritability of an important life history trait in a wild bird population. In years when spring temperatures were highest, selection was strongest, and estimated heritability highest. This positive correlation between the strength of selection and heritability may considerably advance the expected response to selection.

As has generally been found in selection studies on laying date in birds (Perrins 1965; Van Noordwijk et al. 1995; Svensson 1997; Visser et al. 1998; Gienapp et al. 2006) the standardized selection differentials were generally negative, indicating that early breeding individuals had higher fitness than late breeding individuals. Furthermore, the strength of selection was strongest when temperatures were highest (Fig. 5.1). It has previously been shown in this population that when temperatures are high this leads to mistimed reproduction and strong selection for early breeding (Visser et al. 1998; Gienapp et al. 2006) and thus high temperatures are generally associated with adverse environmental conditions, and indeed, fitness has declined (Nussey et al. 2005) as temperatures have increased. Hence, our results confirm the general trend found in natural populations of stronger selection in adverse environmental conditions (Endler 1986). It is important to point out however, that high temperatures are not necessarily associated with adverse environmental conditions in other systems. For example, a population of great tits in England has also experienced increasing temperatures, but recruitment rate in this population has increased over time (Charmantier et al. 2008).

In general, we would expect additive genetic variance to decrease when environmental conditions are stressful (Hoffmann and Merilä 1999; Charmantier and Garant 2005), but we found that additive genetic variance and heritability of laying date increased rather than decreased (Fig. 5.2). Although there was substantial evidence that V_A and V_{PE} changed with environmental conditions (see results), the change in V_A itself was not significant (Husby et al. in review), something that is reflected in the large standard errors in Fig 5.2a. Nevertheless, the increase in V_A is very large and 81.4% of the total change in V_P is due to the increase in additive genetic effects. Furthermore, the power to detect significant changes in additive

genetic variance using a random regression animal model approach is limited and no study has yet convincingly shown a significant change in V_A with the environment using this approach.

The increase in V_A with spring temperature found here (Fig. 5.2a), is, for example, much larger than the increase in maternal genetic variance (V_M) found for birth weight in Soay sheep (Wilson et al. 2006). Note also that in the Soay sheep analysis, maternal environmental effects were not fitted with same order polynomials as the maternal genetic effects, so that some of the increase in maternal genetic effects could potentially be driven by environmental rather than genetic effects (in the same way as permanent environment variance will inflate additive genetic variance if not fitted explicitly, (Kruuk and Hadfield 2007)).

One possible explanation for why we found an increase in V_A with higher temperatures (and thus stressful conditions) rather than a decrease can be that in this study the extreme range of the environmental variable constitutes not only a stressful, but also a novel environment. For example, 2005 and 2007 had the highest recorded spring temperatures since this population study began back in 1955. It has been suggested that V_A should increase in novel environments because of the expression of new genes that have not previously been under selection in the ancestral environment (Holloway et al. 1990) something that has been generally confirmed in empirical studies (Guntrip et al. 1997; Laugen et al. 2005a).

More generally, our finding add support to the idea that predicting the direction in which V_A should change with environmental conditions is complicated when environmental changes may also lead to novel conditions, as, for example, is often the case with human-induced changes (Charmantier and Garant 2005).

The increase in V_A and heritability with increasing spring temperature and the fact that the strength of selection was strongest when temperatures were highest meant that there was a significant correlation between the strength of selection on laying date and heritability of laying date (Fig. 5.3). Thus, when temperatures were high, heritability was highest as was the strength of selection, something that may considerably advance the expected response to selection. Very few studies have simultaneously examined how environmental factors influence genetic expression and selection and their covariance. The only study we are aware of to date is a study of birth weight in Soay sheep were strength of selection and heritability were negatively correlated, something which could contribute to the apparent evolutionary stasis of birth weight in this species (Wilson et al. 2006). Another example where there may be a negative correlation between the strength of selection and heritability is for juvenile growth rates in North American red squirrels (*Tamiasciurus hudsonicus*). Although the covariance between selection and heritability was not explicitly considered, McAdam & Boutin (2003) found that V_A and maternal genetic variance increased in years with low cone abundance (poor environment) but because viability selection in this system is stronger when cone abundance is high (due to competition for territories, McAdam and Boutin 2003) this should generate a negative covariance between selection and total genetic variance, something that may hamper a response to selection.

Our results thus demonstrate a relatively unexplored mechanism that may allow for increased speed of adaptation to climate change. As temperatures are expected to continue to increase (IPCC 2007) a positive correlation between the strength of selection and heritability of laying date may prove an important factor allowing this population to adapt to the rapid environmental conditions experienced.

As it is ultimately this rate of adaptation which is crucial for species to cope with climate change (Visser 2008) our finding, that climate change itself may fuel evolution, may potentially also alter the predictions from models linking population viability to climate change which ignore such evolutionary processes.

CHAPTER 6

General discussion

6.1 Discussion

In this thesis I have provided novel insights into how environmental change influences morphological (Chapter 3) and life history traits (Chapter 2, 4, 5) using long-term studies of great tit populations. I have explored patterns of trait variation (Chapter 2, 3, 4), micro-evolution (Chapter 3), phenotypic plasticity and its genetic basis (Chapter 4), environmentally dependent expression of genetic variance (Chapter 4, 5), selection (Chapter 2, 3, 5) and how environmental factors can cause an interaction between strength of selection and heritability (Chapter 5).

In this final chapter I will firstly summarise the findings from the chapters presented in this thesis and then outline some future questions that have arisen from the work I have undertaken in this thesis. As the results of my findings have been put into a wider perspective and relevant context in their respective chapters, I will here focus more on what I see as interesting issues still to be resolved and, hopefully, provide some ideas and guidelines as to how these issues could be examined.

6.2 Genetic basis of multiple brooding?

In Chapter 2 I examined the causes behind the dramatic decline in the proportion of females producing a second clutch in four long-term study populations in the Netherlands. I showed that this decline particularly strong in the later years when the temperature increase has been stronger. The probability to produce a second clutch is closely related to the timing of breeding in relation to the caterpillar peak (Fig. 2.2a) and declined over time (Fig 2.2b). Thus, birds were less likely to produce a second clutch in the later part of the study compared to birds experiencing a similar mistiming in the early part. Although there was overall no selection against double

brooded individuals, the number of chicks recruiting from the second clutch also declined (Fig. 2.3).

The observed decline in the proportion of females producing a second clutch over time has been rapid, with about 50% of females double brooding in the early part of the study to less than 5% in the last few years (Fig. 2.1). As yet we do not know how general this pattern is and, to my knowledge, the only published study available for comparison is from a population of barn swallows in Denmark where Møller showed that there was no change in double brooding patterns over a 35 year period (1971-2005) (Møller 2007). However, these two species have very different ecology which makes it difficult to make a particularly insightful comparison.

Interestingly, it seems that coal tits (*Parus ater*) and great tits in Lingen, Germany have also experienced similar declines in the proportion of second broods produced (Wolfgang Winkel & Tim Schmoll, personal communication). This suggests that the decline in proportion of second broods may be related to large scale patterns, such as a general increase in the mistiming as I discussed in Chapter 2. It is possible that the mistiming has also increased in these populations as the degree of synchrony between bud burst and caterpillar emergence are mainly driven by large-scale climatic conditions (Visser and Holleman 2001). Indeed, estimates of the lag between caterpillar emergence and oak bud burst from the Netherlands (Visser and Holleman 2001) were positively correlated with annual mean body condition index, tarsus length and fledgling success in collared flycatchers (*Ficedula albicollis*) on Gotland, Sweden (Merilä et al. 2001a) as well as with body mass in great tits from Wytham Woods, Oxford, UK (Garant et al. 2004).

Experimental removal of second cluthches have shown that producing a second clutch can incur survival costs, particularly if environmental conditions are

harsh and may also increase the number of recruits produced from the first clutch (Verhulst 1998). Consequently, changes in environmental conditions, as induced for example by an increase in the mistiming, will have dramatic effects on the proportion of females producing a second clutch.

Whether the decline in double brooding activity is due to micro-evolutionary change or a phenotypic response to the changing environmental conditions is an open question. However, the proportion of females producing a second clutch can vary considerably from year to year (see Fig 2.1) which suggests that plasticity may be an important mechanism and the results presented in chapter 2 suggest that plasticity is the more likely mechanism causing the observed decline in number of second clutches in these populations.

Of course, if the observed decline was due to micro-evolutionary change then double brooding behaviour must be heritable and under selection. To my knowledge no study has as yet examined the genetic basis of double brooding behaviour. However, we do have some unpublished results from a multi-state mark recapture analysis suggesting that individual females are consistent in whether they produce one or two clutches (Sæther, S.A, Husby, A & Visser, M.E, ms in prep). Of course this does not tell us anything about the genetic basis of double brooding behaviour as it may well be that this repeatability is due to an environmental rather than genetic effect, for example females in good condition may be able to occupy the better habitats and may therefore be able to produce two clutches more frequently than females in lower condition. Nevertheless, it would be interesting to examine the genetic basis of double brooding behaviour in this system, ideally using a Bayesian animal model (e.g. Hadfield et al. in press), since the 'standard' animal model (Henderson 1950; Kruuk 2004) is not well suited to deal with binary traits as the

estimation of variance components with binary traits is difficult and can lead to biased results (Gilmour et al. 2006).

6.3 Phenotypic change due to evolutionary change or not?

In Chapter 3 I showed that adult body mass and tarsus length (except in Vlieland) have declined over 36 years in three Dutch great tit study populations. As temperatures have increased during the same period this result seemed to confirm Bergmann's rule (Bergmann 1847) and agree with many subsequent studies (e.g. Freckleton et al. 2003) which have found that body size in general declines with increasing temperature. Very few studies have however examined if this decline is due to phenotypic adjustment or micro-evolutionary change. Using the animal model I examined the quantitative genetic basis of body mass and tarsus length and found, as is generally the case for avian morphological traits (see review by Merilä and Sheldon 2001), that both traits showed moderately high heritability. Furthermore, body mass, but not tarsus length, was under significant positive selection in HV and Oosterhout when examined across all years, although there was large inter-annual variation in the selection differentials (Table 3.4 and 3.5).

Despite positive selection on and a heritable basis to body size I did not find any indication of microevolution (as assessed using a Bayesian approach, see below), which suggests that the decline was due to a phenotypic adjustment. This adjustment did not seem to be due to temperature *per se* however, but rather driven by the time of caterpillar peak abundance and the relative timing between caterpillar abundance and laying date of the birds.

Surprisingly few studies have examined whether temporal changes in body size are a result of phenotypic adjustment or micro-evolution and there is as yet no

consensus on this issue. I tested for micro-evolutionary change in body mass and tarsus length using the estimated breeding values (EBVs) that the animal model can generate (Chapter 1). A change in EBVs over time reflects a genetic change over time and hence micro-evolution (Kruuk 2004), something that has traditionally been examined in a linear mixed model framework (Merilä et al. 2001a, b; Charmantier et al. 2004). Very recently however, it has been pointed out that this method will tend to overestimate significance levels and thus be anti-conservative (Hadfield et al. in press). Because of this I used both the standard method (mixed model approach) and the more recently developed Bayesian approach of Hadfield and colleagues.

Interestingly I found a significant change in EBVs for body mass over time in the predicted direction in Oosterhout and Vlieland (Table 3.6) using the mixed model approach, but when I repeated the analysis using the Bayesian animal model this was not significant (Table 3.6). It is interesting to note that this situation, with a decline in phenotype and a significant positive change in EBVs, closely parallel the 'cryptic evolution' pattern found for body condition index in collared flycatchers on Gotland (Merilä et al. 2001a). In our case this pattern is however unlikely to be a case of 'cryptic evolution' as when I used the Bayesian method the genetic change was no longer significant (Table 3.6).

Additionally, when comparing the estimated slope for change in EBVs over time using the linear mixed model approach and the Bayesian approach (Table 3.6) I found that, although the slope estimates were in some cases identical, there were also incidences were the slopes were different, although never significantly so (Table 3.6). Differences in slope estimates of the genetic trend from the mixed model approach and Bayesian approach are in contrast to what has been found when re-analysing two previous studies that have examined temporal trends in EBVs. Using

the Bayesian animal model approach Hadfield et al. (in press) found that both the previously shown genetic change in fledgling mass in great tits (Garant et al. 2004) and in body mass in Soay sheep (Wilson et al. 2007) were not significant when using the Bayesian approach, similar to my findings. However, unlike what I found, the slope of change in EBVs over time did coincide with the estimate from the linear mixed model analysis in both of these cases (Hadfield et al. in press). It is unclear why the two approaches give different estimates in our case, but it may be related to pedigree structure and differences in sampling variance of the variance components. An important future challenge therefore, will be to compare the LMM approach and the Bayesian approach for assessing micro-evolutionary trends in more studies, as well as to re-analyse existing studies that have used the mixed model approach so as to confirm if these studies do indeed provide an example of micro-evolution or not. Perhaps, as Hadfield et al. (in press) point out, is rapid micro-evolutionary change less frequent than the literature from long term studies suggest at the present time. It is becoming increasingly clear that even with very long term dataset and large pedigrees the power to detect evolutionary change in natural populations may be low. Indeed, in a recent review Gienapp et al. (2008) conclude that very few studies have demonstrated that the observed phenotypic change is due to a genetic change.

6.4 Issues of power when detecting GxE interactions in natural populations

In Chapter 4 I examined the multivariate patterns of phenotypic plasticity in laying date and clutch size and its genetic basis in the great tit populations of Hoge Veluwe (HV) and Wytham Woods (WW) and also compared the patterns of plasticity between the two populations. Our results showed that population level patterns for laying date and clutch size were similar in HV and WW, with both populations

showing plasticity for laying date, but not clutch size, in relation to temperature. Examining the individual variation in plasticity (or lack of it in the case of clutch size) revealed that individuals differed significantly in their laying date plasticity in both populations (i.e. IxE), although there was significantly more between-individual variation in HV than in WW. This variation did not have a significant genetic basis, although genetic variance for plasticity was higher in HV than in WW. In contrast, there was significant between-individual phenotypic and genetic plasticity variation for clutch size in WW, but no such variation (phenotypic or genetic) in HV.

When I examined the multivariate patterns I found a significant negative phenotypic correlation between clutch size and laying date in both populations, but the genetic correlation was significantly negative in WW and non-significantly *positive* in HV. In neither population did I find any indication that plasticity in one trait was correlated with plasticity in the other and thus the phenotypic and genetic correlations did not change significantly with the environmental conditions.

As is evident from Chapter 4, the analytical methods available to examine between-individual variation in phenotypic plasticity and its genetic basis are well developed (see review by Nussey et al. 2007). Less so is our understanding of the ecological variables that are important for the decision making process, which will determine how plastic individuals are expected to be. For most traits this is something we only have a vague idea about at the moment, but laying date is perhaps an exception as temperature not only explains a huge (in ecological terms) amount of variance in laying date (van Balen 1973; Perrins and McCleery 1989), but it has also been shown experimentally that timing of breeding is causally related to temperature (Visser et al. 2009a). Obtaining similar understanding of which ecological factors are important for other traits will require an experimental approach, and even then it may

be difficult to find a particular variable that explain as much of the variance as temperature does for laying date. For example, in Chapter 4 I used temperature as an environmental axis to explore plasticity in clutch size, as I was interested in how plasticity varies under changing temperature conditions. However, it is unlikely that temperature *per se* is important in determining clutch size and to explore plasticity in clutch size other environmental variables, such as population density, should be explored.

An alternative to measuring and comparing the results for many different ecological variables that are likely to influence how individuals adjust their trait, is to use a combined measure like principal component analysis. This would allow us to summarise the major axis of environmental variation whatever variables or variable composition may be important (although there may still be other important variables we have failed to measure that will not be incorporated). This is related, at least conceptually, to the approach used by some authors (e.g. Wilson et al. 2006; Robinson et al. 2009) where yearly 'population survival' (e.g. proportion of juveniles surviving) is used as an indicator of environmental conditions in a specific year.

More effort needs to be directed towards understanding the optimal environmental axis for the trait we study and how, what to us seems, small changes in the environmental axis influence our ability to detect a GxE. For instance, when I repeated the plasticity analysis using the Wytham temperature period (but using the De Bilt temperature data) in the HV population this changed the *P*-values for the GxE for laying date much more dramatically (WW period: $P = 0.96$; HV period: $P = 0.15$) than did using a longer time series (1960-2006, $P = 0.25$, unpublished). That the estimated GxE can change so much with, what seems, very small changes in the

environmental factor is worrying. Certainly, it makes comparison between studies difficult.

In addition to the problem of deciding the environmental axis there is also an issue as to how much differences in data amount and pedigree structure influences our ability to pick up GxE patterns in natural populations. Somewhat surprisingly there is little discussion in quantitative genetic studies on natural populations about how differences in pedigree structure and data may influence quantitative genetic estimates (but see Charmantier and Reale 2005; Quinn et al. 2006; Kruuk and Hadfield 2007; Pemberton 2008), yet this is important knowledge when comparing studies where data amount and pedigree quality may differ. Although there has been a great deal of work in the animal breeding literature examining the effects of pedigree structure on quantitative genetic estimates, it is not clear to what extent these results are directly transferable to studies in natural populations which have very different pedigree structure (Quinn et al. 2006).

To my knowledge, only one study has examined the effect of data structure (depth and completeness) on genetic variances and covariances within a natural population setting. Quinn and colleagues (2006) manipulated data completeness (number of records within years) and amount (number of records) in two long-term bird populations with contrasting demographic characteristic: one great tit population where maximum lineage length was 30 generations and mean lineage length 6.4, and a mute swan population with maximum lineage length of 6 generations and mean lineage length of 2.8. Importantly, the pedigree was not manipulated in any of the systems such that the effect of data amount and completeness could be evaluated. Their results showed that estimates of heritability were robust against data amount and completeness in both systems, but, as expected

(Kruuk 2004), standard errors did increase with smaller data amount. Interestingly there was a tendency for the heritability estimates in the mute swan system to be more easily perturbed than was the case in the great tit system, suggesting that the effect of data structure on heritability estimates may be dependent on the demographic characteristic of the population under study (and thus on pedigree structure).

Unfortunately, the results from this study are probably not directly relevant for our understanding of detecting GxE in natural populations. This is because one needs to have repeated measures on individuals, and their relatives, across the environmental range and Quinn et al. (2006) sampled individuals without replacement when manipulating data completeness. As such these results do not provide any information for detecting GxE interactions. Nevertheless, in general I believe that data amount will be less important for the ability to demonstrate GxE in natural populations than data completeness. Consequently, polygynous mating systems with little immigration may offer the best systems in which to detect GxE interactions in natural populations.

Estimating quantitative genetics parameters using the animal model relies on having a good pedigree, but errors in the pedigree can be introduced if there is extra-pair paternity (EPP). This can be a problem if the pedigree is based on observations (so-called social pedigree) as EPP rates can be high even in socially monogamous species, but is also a problem where paternity has been determined using molecular markers as these studies often have genotyping error rates similar to EPP levels (Charmantier and Reale 2005). Such errors will generally lead to a downward bias in V_A because both paternal and sib relationships will be wrong and thus phenotypes differ more than expected from their presumed genetic relationship. Simulation

studies show however that quantitative genetic estimates are surprisingly robust even when EPP rates are as high as 20 %, at least when sample size was large (Charmantier and Reale 2005). This robustness was dependent on the heritability however, such that when heritability was high (40%) social pedigrees underestimated h^2 more than if heritability was low. Another important point arising from these simulation studies was that the deeper pedigrees were not more affected by the EPP rate than the pedigrees spanning fewer generations.

In addition to pedigree structure also the number of individuals in the pedigree is important for obtaining an accurate estimate of the variance components. In Chapter 4, for example, the pedigree for Wytham Woods (WW, n = 11,117) contains almost twice as many individuals as that of the Hoge Veluwe (HV, n = 6,907). As immigration rates and other demographic characteristics in these two populations are similar, the pedigree structure itself should be similar except for the increased sample size in WW. Presumably then this should give better power to detect a significant GxE in WW than in HV, although we do not know exactly how much more power increased pedigree size (or depth) give. The significant GxE for clutch size in WW (and lack of GxE in laying date for HV) provides some support for the need for a large pedigree when detecting GxE interactions.

As an additional example, in Chapter 4 I used data from Hoge Veluwe collected over the years 1973 – 2006, whereas in Chapter 5 I used data from 1973 – 2007, something that resulted in an additional 195 individuals in the pedigree and, presumably, increased power. The P – value for the significance test of GxE was reduced from P = 0.153 to P = 0.145, but it is unclear whether this change was because of the extra data-information or the increase in pedigree size. For more

discussion about pedigree quality and its implications on quantitative genetic estimates see the recent review by Pemberton (2008).

In my view two important questions need to be addressed with regards to the use of random regression animal models in natural populations. Firstly, we need to know what power random regression models have to demonstrate GxE in natural populations, and secondly, we need better understanding of how variation in (the same) environmental variables influence GxE estimates. These are important issues to consider as there is so far only a single study (Chapter 4) that has convincingly demonstrated a significant GxE using the random regression animal model method in natural populations. It may be that this is due to power issues but we need to know to what extent this is the case and also whether most power is to be gained from increased pedigree size/structure or data amount. Therefore simulation studies examining this would be a highly valuable addition to the literature.

As the number of studies examining GxE in natural populations continue to grow (e.g. Brommer et al. 2005; Nussey et al. 2005; Brommer et al. 2008; Husby et al. provisionaly accepted) it is natural to explore the generality of these findings. However, before studies are in place examining how data and pedigree influence our ability to detect GxEs, I believe that any meta-analysis exploring the generality of GxE's in natural populations will be futile.

6.5 Environmental coupling between selection and heritability

The results from Chapter 4 suggested that additive genetic variance for laying date increased with increasing spring temperature in the Hoge Veluwe population (Fig. 4.2a). In Chapter 5 I further examined how environmental change may lead to changes in the expression of genetic variance (and in heritability) as well as to

change in selection pressure and, in particular, how selection and heritability covaried.

Using an additional year of data in comparison with Chapter 4, I found, not surprisingly, the same pattern of increasing V_A with spring temperature (Fig. 5.2a) but also heritability increased with spring temperature (Fig. 5.2b). Again, the increase in V_A itself was not significant although comparing a model in which both V_A and V_{PE} were constant to a model in which both were allowed to vary with the environment was highly significant. Nevertheless, the increase in V_A is very rapid and over 80% of this increase is due to genetic rather than environmental effects (see section above for a discussion about the power to detect GxE in natural populations). As high spring temperatures are associated with large mistiming in this population (Visser et al. 1998) this suggests that V_A is higher under stressful environments compared to when environmental conditions are good. Although in contrast to some other studies in natural populations (Wilson et al. 2006; Robinson et al. 2009) this is similar to what has been found, for example, in a study of North American red squirrels (McAdam and Boutin 2003). Furthermore, in common gulls (*Larus canus*) V_A also increases with increasing spring temperatures, although it is not known if high temperatures are stressful or not in this system (Brommer et al. 2008).

Some indirect support for the observed increase in V_A during stressful environmental conditions in our studies (Chapter 4, 5) comes from a between-population comparison of V_A and heritability of laying date. In his PhD thesis Erik Postma (Postma 2005, p 74) compared heritability of clutch size and laying date across eight Dutch great tit populations and found that V_A for laying date was higher in populations where fledglings mass was lowest ('poor environment') than in populations where fledgling mass was high ('good environment').

It is interesting to note that in both the great tits (Chapter 5) and common gulls (Brommer et al. 2008) V_A increases with increasing spring temperature, and it is tempting to conclude that this is due to the fact that higher spring temperatures represents a novel environment. It has been hypothesised, and also confirmed empirically, that V_A should increase in novel environments (see Holloway et al. 1990; Charmantier and Garant 2005), perhaps as a result of released variance for alleles that have previously been under strong directional selection and are no longer subject to it in the new environment. Although in this system a novel environment is also stressful because of the increased mistiming this is not necessarily the case in other systems. This makes it difficult to predict in what direction V_A is expected to change in relation to changing environmental conditions (see Charmantier and Garant 2005).

It is generally accepted that the strength of selection is often strongest when environmental conditions are hard (Endler 1986) and our results confirm this, selection differentials were more negative when temperatures were higher (and birds more out of synchrony with the caterpillar peak).

Although many studies have examined how the expression of genetic variance may change with environmental conditions and other studies how strength of selection can vary with environmental factors, very few studies have examined how strength of selection and heritability covary. Indeed, I am only aware of one previous study that has examined this. Wilson and colleagues (2006) showed that, in Soay sheep, strength of selection and heritability of birth weight are negatively correlated, something that provided one possible explanation for the observed stasis in this species despite substantial heritability and selection on birth weight. Interestingly, I found the opposite result: annual strength of selection on laying date

and heritability of laying date were positively associated (Fig. 5.3) something that provide an important counter-example to the negative correlation between selection and heritability found in Soay sheep. Although variation around the h^2 and selection differentials estimates are not taken into account when examining the correlation between the two this is unlikely to be a problem because the uncertainty in either parameter was not dependent on the other, that is, large SE's of h2 was not associated with small/large SE's of the selection differentials. Consequently, the correlation estimate is unlikely to be biased even if variation in heritability estimates and selection differentials have not been taken into account.

Clearly, such environmentally – induced correlations between selection and heritability in natural populations need to be further explored in other systems and for other traits before any conclusions can be drawn about the generality of these findings. Nevertheless, both studies (Wilson et al. 2006 and Chapter 5) provide unique examples of the importance of incorporating environmental heterogeneity when examining evolutionary dynamics in the wild.

6.6 The future of genetic studies in natural populations?

In this discussion I have introduced some of the questions that have arisen as a result of the work presented in this thesis and suggested some implications and ideas for future studies. These ideas are framed in the context of quantitative genetics and so far I have ignored the fact that there is currently great interest in exploring the interface between molecular and quantitative genetics in natural populations. Although I believe quantitative genetic studies will continue to be an important tool for studying the genetics of natural populations the increasing availability of molecular genetic tools and genomic data in natural populations offers new and

exiting approaches to studying evolution in the wild and I will therefore briefly discuss some of these opportunities here.

Quantitative genetics only deals with the *expected* genetic relationship between individuals. However, because recombination events are relatively infrequent (Lynch and Walsh 1998) this implies that segments of the chromosomes ('haplotype blocks'), rather than independent genes, are inherited between parents and offspring, something that creates variation around the expected genetic relationship. The rapidly decreasing cost of genotyping means that it is now possible to genotype individuals for thousands of markers and *realized* rather than predicted genetic relationship between individuals can be calculated (Kruuk et al. 2008). This approach should also resolve some of the problems with marker-based estimates of relatedness (Garant and Kruuk 2005; Thomas 2005; Pemberton 2008) and provide researchers with the possibility to examine quantitative genetic relationship in populations where pedigree data does not exist.

Another interesting aspect is that molecular markers will also allow us to identify loci of adaptive significance through the use of either linkage mapping (Lynch and Walsh 1998) where the co-segregation of marker alleles and phenotypes is studied, or association mapping (Slate et al. 2009) which can bypass the need for a linkage map. The QTL approach does not pinpoint the actual genes underlying the trait, but identifies regions of the genome where the genes are likely to reside whereas an association study is more likely to pick up the actual genetic variant or to be closely linked to it (Slate et al. 2009).

At present linkage maps are only available in a few organisms, although this is expected to change soon (Kruuk et al. 2008), and so QTL (Quantitative Trait Loci: a region of the genome that influences trait variation) studies in natural populations

are currently rare (Slate et al. 2009). Ultimately, this approach has potential to see how allele frequencies change over time and hence provide important insight into evolutionary responses and will also allow comparison with the current approach of looking at changes in breeding values (Kruuk et al. 2008).

The increasing uptake of genomic tools for wild populations will rapidly advance our understanding of the relationship between DNA sequence variation and variation in phenotypes, something that will yield unique insights into the genetic basis of phenotypic traits and how evolutionary changes take place.

6.7 Concluding remarks

An important observation in evolutionary biology is that individuals are variable and ecological factors are among the strongest and most important selective agents responsible for the genetic variation observed in natural populations. In this thesis I have presented empirical evidence that changes in ecological factors can lead to rapid phenotypic change and I explored some of the mechanisms that may be responsible for the observed phenotypic changes. Understanding these mechanisms will help us understand how populations adapt to environmental changes and provide important insight into the evolutionary process.

LITTERATURE

Adams, D. C., and J. O. Church. 2008. Amphibians do not follow Bergmann's rule. Evolution 62:413-420.

Åkesson, M., S. Bensch, D. Hasselquist, M. Tarka, and B. Hansson. 2008. Estimating heritabilities and genetic correlations: comparing the 'animal model' with parent-offspring regression using data from a natural population. Plos One 3:e1739.

Arnold, S. J., and M. J. Wade. 1984a. On the Measurement of Natural and Sexual Selection - Applications. Evolution 38:720-734.

Arnold, S. J., and M. J. Wade. 1984b. On the Measurement of Natural and Sexual Selection - Theory. Evolution 38:709-719.

Ashton, K. G., M. C. Tracy, and A. de Queiroz. 2000. Is Bergmann's rule valid for mammals? American Naturalist 156:390-415.

Backstrom, N., M. Brandstrom, L. Gustafsson, A. Qvarnstrom, H. Cheng, and H. Ellegren. 2006. Genetic mapping in a natural population of collared flycatchers (Ficedula albicollis): Conserved synteny but gene order rearrangements on the avian Z chromosome. Genetics 174:377-386.

Baldwin, J. M. 1896. A new factor in evolution. American Naturalist 30:441-451;536-553.

Barbraud, C., and H. Weimerskirch. 2006. Antarctic birds breed later in response to climate change. Proceedings of the National Academy of Sciences of the United States of America 103:6248-6251.

Barton, N. H., and P. D. Keightley. 2002. Understanding quantitative genetic variation. Nature Reviews Genetics 3:11-21.

Bergmann, C. 1847. Über die verhältnisse der värmeökonomie der thiere zu ihrer grösse Göttinger Studien 3:595-708.

Björklund, M. 1996. Similarity of growth among Great tits (Parus major) and Blue tits (P-caeruleus). Biological Journal Of The Linnean Society 58:343-355.

Blakey, J. K. 1994. Genetic-Evidence for Extra-Pair Fertilizations in a Monogamous Passerine, the Great Tit Parus-Major. Ibis 136:457-462.

Blanckenhorn, W. U., and V. Llaurens. 2005. Effects of temperature on cell size and number in the yellow dung fly Scathophaga stercoraria. Journal of Thermal Biology 30:213-219.

Blows, M. W. 2007. A tale of two matrices: multivariate approaches in evolutionary biology. Journal of Evolutionary Biology 20:1-8.

Both, C., S. Bouwhuis, C. M. Lessells, and M. E. Visser. 2006. Climate change and population declines in a long-distance migratory bird. Nature 441:81-83.

Both, C., J. M. Tinbergen, and M. E. Visser. 2000. Adaptive density dependence of avian clutch size. Ecology 81:3391-3403.

Brinkhof, M. W. G., A. J. Cave, S. Daan, and A. C. Perdeck. 2002. Timing of current reproduction directly affects future reproductive output in European coots. Evolution 56:400-411.

Brommer, J. E., L. Gustafsson, H. Pietiainen, and J. Merila. 2004. Single-generation estimates of individual fitness as proxies for long-term genetic contribution. American Naturalist 163:505-517.

Brommer, J. E., J. Merila, B. C. Sheldon, and L. Gustafsson. 2005. Natural selection and genetic variation for reproductive reaction norms in a wild bird population. Evolution 59:1362-1371.

Brommer, J. E., K. Rattiste, and A. J. Wilson. 2008. Exploring plasticity in the wild: laying date-temperature reaction norms in the common gull Larus canus. Proceedings of the Royal Society B-Biological Sciences 275:687-693.
Bryant, D. M. 1979. Reproductive costs in the house martin (Delichon-Urbica). Journal of Animal Ecology 48:655-675.
Charmantier, A., and D. Garant. 2005. Environmental quality and evolutionary potential: lessons from wild populations. Proceedings of the Royal Society B-Biological Sciences 272:1415-1425.
Charmantier, A., L. E. B. Kruuk, J. Blondel, and M. M. Lambrechts. 2004. Testing for microevolution in body size in three blue tit populations. Journal of Evolutionary Biology 17:732-743.
Charmantier, A., R. H. McCleery, L. R. Cole, C. Perrins, L. E. B. Kruuk, and B. C. Sheldon. 2008. Adaptive phenotypic plasticity in response to climate change in a wild bird population. Science 320:800-803.
Charmantier, A., and D. Reale. 2005. How do misassigned paternities affect the estimation of heritability in the wild? Molecular Ecology 14:2839-2850.
Chown, S. L., and C. J. Klok. 2003. Altitudinal body size clines: latitudinal effects associated with changing seasonality. Ecography 26:445-455.
Clutton-Brock, T. H., K. E. Rose, and F. E. Guinness. 1997. Density-related changes in sexual selection in red deer. Proceedings Of The Royal Society Of London Series B-Biological Sciences 264:1509-1516.
Coltman, D. W., P. O'Donoghue, J. T. Hogg, and M. Festa-Bianchet. 2005. Selection and genetic (CO)variance in bighorn sheep. Evolution 59:1372-1382.
Coltman, D. W., P. O'Donoghue, J. T. Jorgenson, J. T. Hogg, C. Strobeck, and M. Festa-Bianchet. 2003. Undesirable evolutionary consequences of trophy hunting. Nature 426:655-658.
Cook, L. M. 2003. The rise and fall of the Carbonaria form of the peppered moth. Quarterly Review of Biology 78:399-417.
Crawley, M. J. 2002. Statistical computing. An introduction to data analysis using S-Plus. John Wiley & Sons Ltd, Sussex, England.
Cresswell, W., J. A. Clark, and R. Macleod. 2009. How climate change might influence the starvation-predation risk trade-off response. Proceedings of the Royal Society B-Biological Sciences 276:3553-3560.
Crick, H. Q. P., C. Dudley, D. E. Glue, and D. L. Thomson. 1997. UK birds are laying eggs earlier. Nature 388:526-526.
Damuth, J. 1981. Population density and body size in mammals. Nature 290:699-700.
Darwin, C. R. 1859. On the Origin of Species. John Murray, London.
Drent, R. H., and S. Daan. 1980. The prudent parent - energetic adjustments in avian breeding. Ardea 68:225-252.
Dunn, P. 2004. Breeding dates and reproductive performance *in* A. P. Moller, W. Fiedler, and P. Berthold, eds. Birds and climate change. Elsevier Academic Press, Oxford.
Endler, J. A. 1986. Natural selection in the wild. Princeton University Press.
Falconer, D. S. 1952. The Problem of Environment and Selection. American Naturalist 86:293-298.
Falconer, D. S., and T. F. C. Mackay. 1996. Introduction to quantitative genetics. Longman Group Limited, Essex.

Fischer, T. M., A. R. Gilmour, and J. H. J. van der Werf. 2004. Computing approximate standard errors for genetic parameters derived from random regression models fitted by average information REML. Genetics Selection Evolution 36:363-369.
Ford, E. B. 1964. Ecological genetics. Methuen, London.
Franks, S. J., S. Sim, and A. E. Weis. 2007. Rapid evolution of flowering time by an annual plant in response to a climate fluctuation. Proceedings of the National Academy of Sciences of the United States of America 104:1278-1282.
Freckleton, R. P., P. H. Harvey, and M. Pagel. 2003. Bergmann's rule and body size in mammals. American Naturalist 161:821-825.
Garant, D., J. D. Hadfield, L. E. B. Kruuk, and B. C. Sheldon. 2008. Stability of genetic variance and covariance for reproductive characters in the face of climate change in a wild bird population. Molecular Ecology 17:179-188.
Garant, D., and L. E. B. Kruuk. 2005. How to use molecular marker data to measure evolutionary parameters in wild populations. Molecular Ecology 14:1843-1859.
Garant, D., L. E. B. Kruuk, R. H. McCleery, and B. C. Sheldon. 2004. Evolution in a changing environment: A case study with great tit fledging mass. American Naturalist 164:E115-E129.
Garant, D., L. E. B. Kruuk, R. H. McCleery, and B. C. Sheldon. 2007. The effects of environmental heterogeneity on multivariate selection on reproductive traits in female great tits. Evolution 61:1546-1559.
Ghalambor, C. K., J. K. McKay, S. P. Carroll, and D. N. Reznick. 2007. Adaptive versus non-adaptive phenotypic plasticity and the potential for contemporary adaptation in new environments. Functional Ecology 21:394-407.
Gienapp, P., L. Hemerik, and M. E. Visser. 2005. A new statistical tool to predict phenology under climate change scenarios. Global Change Biology 11:600-606.
Gienapp, P., E. Postma, and M. E. Visser. 2006. Why breeding time has not responded to selection for earlier breeding in a songbird population. Evolution 60:2381-2388.
Gienapp, P., C. Teplitsky, J. S. Alho, J. A. Mills, and J. Merila. 2008. Climate change and evolution: disentangling environmental and genetic responses. Molecular Ecology 17:167-178.
Gilmour, A. R., B. J. Gogel, B. R. Cullis, S. J. Welham, and R. Thompson. 2006. ASReml user guide. Relase 2.0. VSN International., Hemel Hempstead, U.K.
Gilmoure, A. R., B. J. Gogel, B. R. Cullis, S. J. Welham, and R. Thompson. 2006. ASReml user guide. Relase 2.0. VSN International., Hemel Hempstead, U.K.
Gosler, A. G. 1993. The Great Tit. Hamlyn Limited, London.
Grant, P. R., and B. R. Grant. 1995. Predicting microevolutionary responses to directional selection on heritable variation. Evolution 49:241-251.
Grant, P. R., and B. R. Grant. 2002. Unpredictable evolution in a 30-year study of Darwin's finches. Science 296:707-711.
Grant, P. R., B. R. Grant, J. N. M. Smith, I. J. Abbott, and L. K. Abbott. 1976. Darwins finches - population variation and natural selection. Proceedings of the National Academy of Sciences of the United States of America 73:257-261.

Gratten, J., A. J. Wilson, A. F. McRae, D. Beraldi, P. M. Visscher, J. M. Pemberton, and J. Slate. 2008. A localized negative genetic correlation constrains microevolution of coat color in wild sheep. Science 319:318-320.
Guillaumet, A., J. B. Ferdy, E. Desmarais, B. Godelle, and P. A. Crochet. 2008. Testing Bergmann's rule in the presence of potentially confounding factors: a case study with three species of Galerida larks in Morocco. Journal of Biogeography 35:579-591.
Guntrip, J., R. M. Sibly, and G. J. Holloway. 1997. The effect of novel environment and sex on the additive genetic variation and covariation in and between emergence body weight and development period in the cowpea weevil, Callosobruchus maculatus (Coleoptera, Bruchidae). Heredity 78:158-165.
Gustafsson, L., A. Qvarnstrom, and B. C. Sheldon. 1995. Trade-Offs between Life-History Traits and a Secondary Sexual Character in Male Collared Flycatchers. Nature 375:311-313.
Hadfield, J. D. 2008. Estimating evolutionary parameters when viability selection is operating. Proceedings of the Royal Society B-Biological Sciences 275:723-735.
Hadfield, J. D. in revision. MCMC methods for Multi-response Generalised
Linear Mixed Models: The MCMCglmm R Package Journal of Statistical software.
Hadfield, J. D., A. J. Wilson, D. Garant, B. C. Sheldon, and L. E. B. Kruuk. in press. The misuse of BLUP in Ecology and Evolution. American Naturalist.
Henderson, C. R. 1950. Estimation of Genetic Parameters. Annals of Mathematical Statistics 21:309-310.
Hill, W. G., M. E. Goddard, and P. M. Visscher. 2008. Data and theory point to mainly additive genetic variance for complex traits. Plos Genetics 4.
Hoffmann, A. A., and J. Merilä. 1999. Heritable variation and evolution under favourable and unfavourable conditions. Trends in Ecology & Evolution 14:96-101.
Hoffmann, A. A., and P. A. Parsons. 1991. Evolutionary genetics and environmental stress. Oxford University Press, Oxford.
Holloway, G. J., S. R. Povey, and R. M. Sibly. 1990. The effect of new environment on adapted genetic architecture. Heredity 64:323-330.
Houghton, J. T., Y. Ding, D. J. Griggs, M. Noguer, P. J. van der Linden, and D. Xiaogu, eds. 2001. Climate change 2001: The scientific basis. Contribution of working group I to the third assessment reposrt of the Intergovermental panel on climate change (IPCC).
Houle, D. 1992. Comparing Evolvability and Variability of Quantitative Traits. Genetics 130:195-204.
Huisman, A. E., R. F. Veerkamp, and J. A. M. Van Arendonk. 2002. Genetic parameters for various random regression models to describe the weight data of pigs. Journal of Animal Science 80:575-582.
Husby, A., L. E. B. Kruuk, and M. E. Visser. 2009. Decline in the frequency and benefits of multiple brooding in great tits as a consequence of a changing environment. Proceedings of the Royal Society B-Biological Sciences 276:1845-1854.
Husby, A., D. H. Nussey, M. E. Visser, A. J. Wilson, B. C. Sheldon, and L. E. B. Kruuk. in review. Contrasting patterns of phenotypic plasticity in reproductive traits in two great tit populations.

Husby, A., D. H. Nussey, M. E. Visser, A. J. Wilson, B. C. Sheldon, and L. E. B. Kruuk. provisionaly accepted. Contrasting patterns of phenotypic plasticity in reproductive traits in two great tit populations. Evolution.

Husby, A., B. E. Saether, H. Jensen, and T. H. Ringsby. 2006. Causes and consequences of adaptive seasonal sex ratio variation in house sparrows. Journal of Animal Ecology 75:1128-1139.

IPCC. 2007. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change. Pp. 996 *in* S. Solomon, D. Qin, M. Manning, Z. Chen, M. Marquis, K. B. Averyt, M. Tignor, and H. L. Miller, eds, Cambridge, United Kingdom & New York, USA.

Jensen, H., B. E. Sæther, T. H. Ringsby, J. Tufto, S. C. Griffith, and H. Ellegren. 2003. Sexual variation in heritability and genetic correlations of morphological traits in house sparrow (Passer domesticus). Journal Of Evolutionary Biology 16:1296-1307.

Jensen, H., B. E. Sæther, T. H. Ringsby, J. Tufto, S. C. Griffith, and H. Ellegren. 2004. Lifetime reproductive success in relation to morphology in the house sparrow Passer domesticus. Journal Of Animal Ecology 73:599-611.

Johnston, R. F., and R. K. Selander. 1964. House sparrows - rapid evolution of races in North America. Science 144:548-&.

Kettlewell, H. B. D. 1973. The Evolution of Melanism. Oxford University Press, Oxford, UK.

Kingsolver, J. G., H. E. Hoekstra, J. M. Hoekstra, D. Berrigan, S. N. Vignieri, C. E. Hill, A. Hoang, P. Gibert, and P. Beerli. 2001. The strength of phenotypic selection in natural populations. American Naturalist 157:245-261.

Kirkpatrick, M., D. Lofsvold, and M. Bulmer. 1990. Analysis of the Inheritance, Selection and Evolution of Growth Trajectories. Genetics 124:979-993.

Klomp, H. 1970. Determination of clutch-size in birds - a review. Ardea 58:1-&.

Kluijver, H. N. 1951. The population ecology of the great tit, *Parus m. major L.* Ardea 39:1-135.

Kruuk, L. E. B. 2004. Estimating genetic parameters in natural populations using the 'animal model'. Philosophical Transactions of The Royal Society of London Series B-Biological Sciences 359:873-890.

Kruuk, L. E. B., T. H. Clutton-Brock, J. Slate, J. M. Pemberton, S. Brotherstone, and F. E. Guinness. 2000. Heritability of fitness in a wild mammal population. Proceedings of the National Academy of Sciences of the United States of America 97:698-703.

Kruuk, L. E. B., and J. D. Hadfield. 2007. How to separate genetic and environmental causes of similarity between relatives. Journal of Evolutionary Biology 20:1890-1903.

Kruuk, L. E. B., J. Merilä, and B. C. Sheldon. 2001. Phenotypic selection on a heritable size trait revisited. American Naturalist 158:557-571.

Kruuk, L. E. B., J. Slate, and A. J. Wilson. 2008. New Answers for Old Questions: The Evolutionary Quantitative Genetics of Wild Animal Populations. Annual Review of Ecology Evolution and Systematics 39:525-548.

Lack, D. 1968. Ecological adaptations for breeding in birds. Methuen, London.

Lande, R. 1979. Quantitative genetic analysis of multivariate evolution, applied to brain body size allometry. Evolution 33:402-416.

Lande, R. 2009. Adaptation to an extraordinary environment by evolution of phenotypic plasticity and genetic assimilation. Journal of Evolutionary Biology 22:1435-1446.
Lande, R., and S. J. Arnold. 1983. The Measurement of Selection on Correlated Characters. Evolution 37:1210-1226.
Laugen, A. T., L. E. B. Kruuk, A. Laurila, K. Rasanen, J. Stone, and J. Merila. 2005a. Quantitative genetics of larval life-history traits in Rana temporaria in different environmental conditions. Genetical Research 86:161-170.
Laugen, A. T., A. Laurila, K. I. Jonsson, F. Soderman, and J. Merila. 2005b. Do common frogs (Rana temporaria) follow Bergmann's rule? Evolutionary Ecology Research 7:717-731.
Lewontin, R. C. 1974. The genetic basis of evolutionary change. Columbia University Press, New York.
Lindèn, M. 1988. Reproductive trade-off between 1st and 2nd clutches in the great tit Parus major - an experimental study. Oikos 51:285-290.
Lubjuhn, T., S. Strohbach, J. Brun, T. Gerken, and J. T. Epplen. 1999. Extra-pair paternity in great tits (Parus major) - A long term study. Behaviour 136:1157-1172.
Lynch, M., and B. Walsh. 1998. Genetics and Analysis of Quantitative Traits. Sinauer, Sunderland,MA.
Mackay, T. F. C., and R. R. H. Anholt. 2007. Ain't misbehavin'? Genotype-environment interactions and the genetics of behaviour. Trends in Genetics 23:311-314.
Mackay, T. F. C., E. A. Stone, and J. F. Ayroles. 2009. The genetics of quantitative traits: challenges and prospects. Nature Reviews Genetics 10:565 - 577.
Mayr, E. 1956. Geographical character gradients and climatic adaptation. Evolution 10:105-108.
McAdam, A. G., and S. Boutin. 2003. Effects of food abundance on genetic and maternal variation in the growth rate of juvenile red squirrels. Journal of Evolutionary Biology 16:1249-1256.
McCleery, R. H., and C. M. Perrins. 1998. ... temperature and egg-laying trends. Nature 391:30-31.
Meiri, S., Y. Yom-Tov, and E. Geffen. 2007. What determines conformity to Bergmann's rule? Global Ecology and Biogeography 16:788-794.
Merilä, J., L. E. B. Kruuk, and B. C. Sheldon. 2001a. Cryptic evolution in a wild bird population. Nature 412:76-79.
Merilä, J., L. E. B. Kruuk, and B. C. Sheldon. 2001b. Natural selection on the genetical component of variance in body condition in a wild bird population. Journal of Evolutionary Biology 14:918-929.
Merilä, J., and B. C. Sheldon. 2000. Lifetime reproductive success and heritability in nature. American Naturalist 155:301-310.
Merilä, J., and B. C. Sheldon. 2001. Avian quantitative genetics *in* J. Van Nolan, and E. D. Ketterson, eds. Current Ornithology. Kluwer academic, New York.
Merilä, J., B. C. Sheldon, and L. E. B. Kruuk. 2001c. Explaining stasis: microevolutionary studies in natural populations. Genetica 112:199-222.
Meyer, K. 1998. Estimating covariance functions for longitudinal data using a random regression model. Genetics Selection Evolution 30:221-240.

Millien, V. 2004. Relative effects of climate change, isolation and competition on body-size evolution in the Japanese field mouse, Apodemus argenteus. Journal of Biogeography 31:1267-1276.
Millien, V., and J. Damuth. 2004. Climate change and size evolution in an island rodent species: New perspectives on the island rule. Evolution 58:1353-1360.
Milner, J. M., S. D. Albon, A. W. Illius, J. M. Pemberton, and T. H. Clutton-Brock. 1999. Repeated selection of morphometric traits in the Soay sheep on St Kilda. Journal of Animal Ecology 68:472-488.
Mitchell-Olds, T., and R. G. Shaw. 1987. Regression-Analysis of Natural-Selection - Statistical-Inference and Biological Interpretation. Evolution 41:1149-1161.
Møller, A. P. 2007. Interval between clutches, fitness, and climate change. Behavioral Ecology 18:62-70.
Nagy, L. R., and R. T. Holmes. 2005. Food limits annual fecundity of a migratory songbird: An experimental study. Ecology 86:675-681.
Newman, R. A. 1994. Genetic variation for phenotypic plasticity in the larval life-history of spadefoot toads (*Scaphiopus couchii*). Evolution 48:1773-1785.
Nussey, D. H., E. Postma, P. Gienapp, and M. E. Visser. 2005. Selection on heritable phenotypic plasticity in a wild bird population. Science 310:304-306.
Nussey, D. H., A. J. Wilson, and J. E. Brommer. 2007. The evolutionary ecology of individual phenotypic plasticity in wild populations. Journal of Evolutionary Biology 20:831-844.
Orr, H. A. 2009. Fitness and its role in evolutionary genetics. Nature Reviews Genetics 10:531 - 540.
Orr, H. A., and R. L. Unckless. 2008. Population extinction and the genetics of adaptation. American Naturalist 172:160-169.
Parejo, D., and E. Danchin. 2006. Brood size manipulation affects frequency of second clutches in the blue tit. Behavioral Ecology and Sociobiology 60:184-194.
Parmesan, C. 2006. Ecological and evolutionary responses to recent climate change. Annual Review of Ecology Evolution and Systematics 37:637-669.
Parmesan, C. 2007. Influences of species, latitudes and methodologies on estimates of phenological response to global warming. Global Change Biology 13:1860-1872.
Parmesan, C., and G. Yohe. 2003. A globally coherent fingerprint of climate change impacts across natural systems. Nature 421:37-42.
Partridge, L., and J. A. Coyne. 1997. Bergmann's rule in ectotherms: Is it adaptive? Evolution 51:632-635.
Pemberton, J. 2008. Wild pedigrees: the way forward. Proceedings of the Royal Society B-Biological Sciences 275:613-621.
Perdeck, A. C., M. E. Visser, and J. H. Van Balen. 2000. Great Tit Parus major survival, and the beech-crop cycle. Ardea 88:99-108.
Perrins, C. M. 1965. Population Fluctuations and Clutch-Size in the Great Tit, Parus-Major L. Journal of Animal Ecology 34:601-647.
Perrins, C. M. 1970. Timing of birds breeding seasons. Ibis 112:242 -
Perrins, C. M., and R. H. McCleery. 1989. Laying Dates and Clutch Size in the Great Tit. Wilson Bulletin 101:236-253.
Pinheiro, J. C., and D. Bates. 2000. Mixed-effects models in S and S-Plus. Springer Verlag, New York.

Postma, E. 2005. Evolutionary genetics of life-history traits in a structured environment. Understanding variation in clutch size and laying date in great tits (*Parus major*). Pp. 204. Utrecht University/Netherlands Institute for Ecology (NIOO), Utrecht.
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., and A. Charmantier. 2007. What 'animal models' can and cannot tell ornithologists about the genetics of wild populations. Journal of Ornithology 148:S633-S642.
Price, T. D., A. Qvarnstrom, and D. E. Irwin. 2003. The role of phenotypic plasticity in driving genetic evolution. Proceedings of the Royal Society of London Series B-Biological Sciences 270:1433-1440.
Przybylo, R., B. C. Sheldon, and J. Merila. 2000. Climatic effects on breeding and morphology: evidence for phenotypic plasticity. Journal of Animal Ecology 69:395-403.
Quinn, J. L., A. Charmantier, D. Garant, and B. C. Sheldon. 2006. Data depth, data completeness, and their influence on quantitative genetic estimation in two contrasting bird populations. Journal of Evolutionary Biology 19:994-1002.
RDevelopmentCoreTeam. 2007. R: A language and environment for statistical computing, R Foundation for Statistical Computing,
Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.
Reale, D., M. Festa-Bianchet, and J. T. Jorgenson. 1999. Heritability of body mass varies with age and season in wild bighorn sheep. Heredity 85:596-603.
Reale, D., A. G. McAdam, S. Boutin, and D. Berteaux. 2003. Genetic and plastic responses of a northern mammal to climate change. Proceedings of the Royal Society of London Series B-Biological Sciences 270:591-596.
Reed, T. E., S. Wanless, M. P. Harris, M. Frederiksen, L. E. B. Kruuk, and E. J. A. Cunningham. 2006. Responding to environmental change: plastic responses vary little in a synchronous breeder. Proceedings of the Royal Society B-Biological Sciences 273:2713-2719.
Reznick, D. 1985. Costs of reproduction - an evaluation of the empirical evidence. Oikos 44:257-267.
Robinson, M. R., A. B. Wilson, J. G. Pilkington, T. Clutton-Brock, J. Pemberton, and L. E. B. Kruuk. 2009. The impact of environmental heterogeneity on genetic architecture in a wild population of Soay sheep. Genetics 181:1639 - 1648.
Sæther, B. E., and O. Bakke. 2000. Avian life history variation and contribution of demographic traits to the population growth rate. Ecology 81:642-653.
Schall, R. 1991. Estimation in generalized linear models with random effects. Biometrika 78:719-727.
Scheiner, S. M. 1993. Genetics and Evolution of Phenotypic Plasticity. Annual Review of Ecology and Systematics 24:35-68.
Scheiner, S. M., and R. F. Lyman. 1989. The Genetics of Phenotypic Plasticity .1. Heritability. Journal of Evolutionary Biology 2:95-107.
Schlichting, C. D. 1986. The evolution of phenotypic plasticity in plants. Annual Review of Ecology and Systematics 17:667-693.
Schlichting, C. D., and M. Pigliucci. 1998. Phenotypic evolution: a reaction norm perspective. Sinauer Associates, Inc., Sunderland, Massachusetts.

Schluter, D. 2000. The ecology of adaptive radiation. Oxford University Press, New York.
Schluter, D., and J. N. M. Smith. 1986. Natural selection on beak and body size in the song sparrow. Evolution 40:221-231.
Shaw, R. G. 1991. The Comparison of Quantitative Genetic-Parameters between Populations. Evolution 45:143-151.
Sheldon, B. C., L. E. B. Kruuk, and J. Merilä. 2003. Natural selection and inheritance of breeding time and clutch size in the collared flycatcher. Evolution 57:406-420.
Simons, L. S., and T. E. Martin. 1990. Food limitation of avian reproduction - an experiment with the cactus wren. Ecology 71:869-876.
Slagsvold, T. 1976. Annual and Geographical Variation in Time of Breeding of Great Tit Parus-Major and Pied Flycatcher Ficedula-Hypoleuca in Relation to Environmental Phenology and Spring Temperature. Ornis Scandinavica 7:127-145.
Slate, J., J. Gratten, D. Beraldi, J. Stapley, M. Hale, and J. Pemberton. 2009. Gene mapping in the wild with SNPs: guidelines and future directions. Genetica 136:97-107.
Smith, F. A., J. L. Betancourt, and J. H. Brown. 1995. Evolution of body-size in the woodrat over the past 25,000 years of climate change. Science 270:2012-2014.
Sokal, R. R., and F. J. Rohlf. 1995. Biometry. W.H. Freeman and company, New York.
Spear, L. B., and D. G. Ainley. 1999. Migration routes of sooty shearwaters in the Pacific Ocean. Condor 101:205-218.
Stearns, S. C. 1992. The evolution of life histories. Oxford University Press, Oxford.
Stenseth, N. C., A. Mysterud, G. Ottersen, J. W. Hurrell, K. S. Chan, and M. Lima. 2002. Ecological effects of climate fluctuations. Science 297:1292-1296.
Stillwell, R. C., G. E. Morse, and C. W. Fox. 2007. Geographic variation in body size and sexual size dimorphism of a seed-feeding beetle. American Naturalist 170:358-369.
Svensson, E. 1997. Natural selection on avian breeding time: Causality, fecundity-dependent, and fecundity-independent selection. Evolution 51:1276-1283.
Teplitsky, C., J. A. Mills, J. S. Alho, J. W. Yarrall, and J. Merilä. 2008. Bergmann's rule and climate change revisited: Disentangling environmental and genetic responses in a wild bird population. Proceedings of the National Academy of Sciences of the United States of America 105:13492-13496.
Thomas, S. C. 2005. The estimation of genetic relationships using molecular markers and their efficiency in estimating heritability in natural populations. Philosophical Transactions Of The Royal Society B-Biological Sciences 360:1457-1467.
Tinbergen, J. M., and C. Both. 1999. Is clutch size individually optimized? Behavioral Ecology 10:504-509.
Tinbergen, J. M., J. H. van Balen, and H. M. van Eck. 1985. Density dependent survival in an isolated great tit population - Kluyvers data reanalysed. Ardea 73:38-48.
Turesson, G. 1922. The genotypical response of the plant species to the habitat. Hereditas 3:211-350.

van Balen, J. H. 1973. Comparative study of breeding ecology of great tit (*Parus major*) in different habitats. Ardea 61:1-93.
van Bers, N. E. M., K. van Oers, H. H. D. Kerstens, B. W. Dibbits, R. P. M. A. Crooijmans, M. E. Visser, and M. A. M. Groenen. submitted. Genome-wide SNP detection in the great tit, *Parus major*, using high throughput sequencing. Molecular Ecology.
Van Noordwijk, A. J., R. H. McCleery, and C. M. Perrins. 1995. Selection for the Timing of Great Tit Breeding in Relation to Caterpillar Growth and Temperature. Journal of Animal Ecology 64:451-458.
van Noordwijk, A. J., J. H. van Balen, and W. Scharloo. 1981. Genetic-Variation in the Timing of Reproduction in the Great Tit. Oecologia 49:158-166.
Verboven, N., and A. C. Mateman. 1997. Low frequency of extra-pair fertilizations in the Great Tit Parus major revealed by DNA fingerprinting. Journal of Avian Biology 28:231-239.
Verboven, N., J. M. Tinbergen, and S. Verhulst. 2001. Food, reproductive success and multiple breeding in the Great Tit Parus major. Ardea 89:387-406.
Verboven, N., and S. Verhulst. 1996. Seasonal variation in the incidence of double broods: The date hypothesis fits better than the quality hypothesis. Journal of Animal Ecology 65:264-273.
Verboven, N., and M. E. Visser. 1998. Seasonal variation in local recruitment of great tits: the importance of being early. Oikos 81:511-524.
Verhulst, S. 1998. Multiple breeding in the Great Tit, II. The costs of rearing a second clutch. Functional Ecology 12:132-140.
Verhulst, S., J. M. Tinbergen, and S. Daan. 1997. Multiple breeding in the Great Tit. A trade-off between successive reproductive attempts? Functional Ecology 11:714-722.
Via, S., R. Gomulkiewicz, G. Dejong, S. M. Scheiner, C. D. Schlichting, and P. H. Vantienderen. 1995. Adaptive Phenotypic Plasticity - Consensus and Controversy. Trends in Ecology & Evolution 10:212-217.
Visser, M. E. 2008. Keeping up with a warming world: assessing the rate of adaptation to climate change. Proceedings of the Royal Society B-Biological Sciences 275:649-661.
Visser, M. E., F. Adriaensen, J. H. van Balen, J. Blondel, A. A. Dhondt, S. van Dongen, C. du Feu, E. V. Ivankina, A. B. Kerimov, J. de Laet, E. Matthysen, R. McCleery, M. Orell, and D. L. Thomson. 2003. Variable responses to large-scale climate change in European Parus populations. Proceedings of the Royal Society of London Series B-Biological Sciences 270:367-372.
Visser, M. E., and L. J. M. Holleman. 2001. Warmer springs disrupt the synchrony of oak and winter moth phenology. Proceedings of the Royal Society of London Series B-Biological Sciences 268:289-294.
Visser, M. E., L. J. M. Holleman, and S. P. Caro. 2009a. Temperature has a causal effect on avian timing of reproduction. Proceedings of the Royal Society B-Biological Sciences 276:2323-2331.
Visser, M. E., L. J. M. Holleman, and P. Gienapp. 2006. Shifts in caterpillar biomass phenology due to climate change and its impact on the breeding biology of an insectivorous bird. Oecologia 147:164-172.
Visser, M. E., A. C. Perdeck, J. H. van Balen, and C. Both. 2009b. Climate change leads to decreasing bird migration distances. Global Change Biology 15:1859-1865.

Visser, M. E., A. J. van Noordwijk, J. M. Tinbergen, and C. M. Lessells. 1998. Warmer springs lead to mistimed reproduction in great tits (Parus major). Proceedings of the Royal Society of London Series B-Biological Sciences 265:1867-1870.

Weggler, M. 2006. Constraints on, and determinants of, the annual number of breeding attempts in the multi-brooded Black Redstart Phoenicurus ochruros. Ibis 148:273-284.

White, T. C. R. 2008. The role of food, weather and climate in limiting the abundance of animals. Biological Reviews 83:227-248.

Wilkin, T. A., D. Garant, A. G. Gosler, and B. C. Sheldon. 2006. Density effects on life-history traits in a wild population of the great tit Parus major: analyses of long-term data with GIS techniques. Journal of Animal Ecology 75:604-615.

Williams, G. C. 1966. Natural selection costs of reproduction and a refinement of Lacks principle. American Naturalist 100:687-&.

Wilson, A. B. 2009. Fecundity selection predicts Bergmann's rule in syngnathid fishes. Molecular Ecology 18:1263-1272.

Wilson, A. J. 2008. Why h^2 does not always equal V_A/V_P? Journal of Evolutionary Biology 21:647-650.

Wilson, A. J., L. E. B. Kruuk, and D. W. Coltman. 2005. Ontogenetic patterns in heritable variation for body size: Using random regression models in a wild ungulate population. American Naturalist 166:E177-E192.

Wilson, A. J., J. M. Pemberton, J. G. Pilkington, T. H. Clutton-Brock, D. W. Coltman, and L. E. B. Kruuk. 2007. Quantitative genetics of growth and cryptic evolution of body size in an island population. Evolutionary Ecology 21:337-356.

Wilson, A. J., J. M. Pemberton, J. G. Pilkington, D. W. Coltman, D. V. Mifsud, T. H. Clutton-Brock, and L. E. B. Kruuk. 2006. Environmental coupling of selection and heritability limits evolution. Plos Biology 4:1270-1275.

Winkler, D. W., and P. E. Allen. 1996. The seasonal decline in tree swallow clutch size: Physiological constraint or strategic adjustment? Ecology 77:922-932.

Yom-Tov, Y. 2001. Global warming and body mass decline in Israeli passerine birds. Proceedings Of The Royal Society Of London Series B-Biological Sciences 268:947-952.

Yom-Tov, Y., and S. Yom-Tov. 2006. Decrease in body size of Danish goshawks during the twentieth century. Journal of Ornithology 147:644-647.

Yom-Tov, Y., S. Yom-Tov, J. Wright, C. J. R. Thorne, and R. Du Feu. 2006. Recent changes in body weight and wing length among some British passerine birds. Oikos 112:91-101.