Exploring the causes of heterozygosity-fitness correlations in the blue tit

Dissertation

zur Erlangung des akademischen Grades
des Doktors der Naturwissenschaften
an der Universität Konstanz
Fachbereich Biologie

vorgelegt von
Juanita Olano Marín

Tag der mündlichen Prüfung: 6 Oktober 2011

Referent: Prof. Dr. Martin Wikelski
Referent: Prof. Dr. Bart Kempenaers

"Indeed the evidence for greater fitness of heterozygotes is stronger than that for most other kinds of natural selection"

J.B.S. Haldane, The biochemistry of genetics (1954)

Table of contents

General introduction

Decades of studies have established that natural populations contain abundant genetic variation and that selection is an ubiquitous process (Mitton 1997). Understanding how these observations relate to each other and, specifically, how selection influences the abundant genetic variation, is a major goal in evolutionary biology. This task can be addressed through the direct examination of the phenotypic consequences (e.g. measured as morphological, physiological, behavioral and/or life-history traits) of molecular variation. Quantitative trait locus (QTL) linkage analyses, for example, study the correlated inheritance patterns of molecular markers and phenotypic traits in order to identify the causal genetic basis of the traits. QTL linkage analyses, however, are not easily implemented in studies of natural populations, as they require that large pedigrees or extensive pairs of siblings can be sampled and that the phenotypes of the individuals can be measured (Ellegren and Sheldon 2008). An alternative, less demanding approach, explores the statistical associations between molecular markers and phenotypic traits. Correlations between individual heterozygosity at multiple loci and traits related to fitness, commonly known as heterozygosity-fitness correlations or HFC, are one of such associations. HFC have been investigated extensively in natural populations of several species with the fundamental aim of studying natural selection in the wild.

Heterozygosity-fitness correlations

Numerous studies have reported significant correlations between marker-based measures of individual heterozygosity and phenotypic traits in natural populations. The evidence of HFC is, however, an assortment of positive, negative and null results, with high variability in the effect sizes both between and within species. Nevertheless, meta-analyses indicate that although HFC are weak (explaining around 1% of the variation in the phenotypic traits), they are on the whole significant (Britten 1996; Coltman and Slate 2003; Chapman *et al.* 2009). Historical changes in the types of molecular markers used to measure individual heterozygosity (i.e. allozymes vs. microsatellites) and considerable differences in the population history of the species studied have opened a long-

standing debate about the causes of HFC (David 1998; Hansson and Westerberg 2002; Szulkin *et al.* 2010). Two main explanations (and three hypotheses) for the occurrence of HFC exist: inbreeding effects across the whole genome (referred to as the general effect hypothesis) or localized effects at single loci (referred to as the direct and the local effect hypotheses) (Mitton 1997; David 1998; Hansson and Westerberg 2002) (Table 1.1).

Table 1.1.
Current hypotheses to explain heterozygosity-fitness correlations

	General effect	Localized effect (local or direct)
Variance in inbreeding and identity disequilibrium required?	Yes	No
Best conditions to detect it:		
- Study design	Individuals with different levels of inbreeding	Individuals with contrasting phenotypes, sharing the same level of inbreeding (e.g. full-sibs)
- Type of markers	Neutral	Neutral (but linked to functional, for local effects) / Functional (for direct effects)
- Heterozygosity measure	Multilocus	Single-locus
Expected effects	+ (inbreeding depression) / - (outbreeding depression)	+ / - / Neutral (depending on marker-specific characteristics and type of selection)

Inbreeding and the general effect hypothesis

Inbreeding can have a considerable negative impact on the fitness of inbred individuals relative to outbred ones, a phenomenon commonly known as inbreeding depression (Crnokrak and Roff 1999; Keller and Waller 2002). By causing a reduction in individual heterozygosity across the genome, inbreeding increases the risk of expressing recessive deleterious alleles and decreases the occurrence of beneficial overdominant effects (Charlesworth and Willis 2009). Inbreeding arises via two different, not mutually exclusive, processes occurring within and among populations (Keller and Waller 2002). Within-population inbreeding occurs when related individuals mate. Detailed pedigree data can be

used to estimate the individual inbreeding coefficient, *F* (Wright 1922). Inbreeding depression is shown by the lower performance of individuals with higher *F*. Among-population inbreeding, on the other hand, occurs when populations are subdivided into small and isolated groups where genetic drift occurs. As a consequence, different deleterious alleles become fixed in the different subpopulations. In this case, inbreeding occurs even with random mating within the subpopulations and can be detected by estimating the amount of population differentiation with pedigree (Wright 1965) or genetic data (Pritchard *et al.* 2000; Balloux and Lugon-Moulin 2002). Inbreeding depression due to among-population inbreeding can only be detected through crosses between members of different subpopulations. The individuals produced from such crosses would mask the effects of the different subsets of deleterious alleles fixed in each subpopulation and, therefore, would show a higher fitness than "pure" individuals (Keller and Waller 2002). The opposite phenomenon, outbreeding depression, occurs when progeny produced by crosses between individuals from genetically differentiated populations have lower fitness than progeny from crosses between individuals from the same population. The decline in fitness in this case is attributed to a breakup of coadapted gene complexes or favorable epistatic interactions (Lynch 1991).

Multilocus heterozygosity (MLH) and related indexes of heterozygosity at multiple loci (e.g. standardized heterozygosity or SH: Coltman *et al.* 1999; internal relatedness or IR: Amos *et al.* 2001; homozygosity by locus or HL: Aparicio *et al.* 2006) have been used as correlates of the individual inbreeding coefficient. Thus, positive correlations between these indexes and fitness-related traits have been interpreted as evidence of inbreeding depression (e.g. Ledig *et al.* 1983; Slate and Pemberton 2002; Lesbarreres *et al.* 2005; Charpentier *et al.* 2008). Less frequently reported, negative or quadratic correlations between heterozygosity indexes and phenotypes are interpreted as outbreeding depression (e.g. Marshall and Spalton 2000; Neff 2004). The underlying assumption for these interpretations is that marker heterozygosity reflects a more general state of the genome, that is genome-wide heterozygosity. This assumption has been questioned in several theoretical and empirical studies (Ohta and Kimura 1970; Ohta 1971; Houle 1989; Whitlock 1993; Balloux *et al.* 2004). First, a small number of loci cannot reliably rank the individuals for their level of heterozygosity in the whole genome (Chakraborty 1981; Balloux *et al.* 2004; Slate *et al.* 2004; DeWoody and DeWoody 2005). Second, identity disequilibrium (ID), defined as the

correlation in heterozygosity and/or homozygosity accross loci, is necessary for marker loci to indicate genome-wide heterozygosity (Weir and Cockerham 1973; Szulkin *et al.* 2010). For ID to arise, on the other hand, specific population processes (e.g. frequent matings between close relatives and/or population subdivision combined with admixture or immigration), causing a high variance in individual inbreeding levels must occur (Balloux *et al.* 2004; Szulkin *et al.* 2010). Third, selection acting on functional markers may prevent those markers to accurately reflect population processes affecting the whole genome, like inbreeding. Therefore, in order to detect general effects of inbreeding it is necessary to 1) measure heterozygosity at multiple loci with a high number of neutral markers, to 2) sample a large number of individuals with different inbreeding levels, to 3) document the population processes that contribute to inbreeding and to 4) estimate the level of ID in the population. Without meeting these conditions, interpreting positive or negative HFC as the result of general effects across the whole genome could be misleading.

Heterozygosity at single loci and the direct and local effect hypotheses

There is an increasing tendency to explain HFC by localized effects of functional loci or of neutral markers linked to such loci, especially when the population studied is not expected to exhibit a high variance in the level of individual inbreeding (Lieutenant-Gosselin and Bernatchez 2006; Luikart *et al.* 2008; Da Silva *et al.* 2009; Hoffman *et al.* 2010a). Under the direct effect hypothesis, the effects of heterozygosity on fitness are caused by the scored loci *per se* (Mitton 1997; David 1998; Hansson and Westerberg 2002). This hypothesis is thus potentially important when non-neutral markers (e.g. allozymes, functional microsatellites, nsSNPs) are used. The local effect hypothesis, on the other hand, proposes that effects of heterozygosity at marker loci are the result of the effects of heterozygosity at closely linked functional loci (Mitton 1997; David 1998; Hansson and Westerberg 2002). Very strong linkage disequilibrium (i.e. non-random association of alleles at different loci in gametes) between marker and fitness loci is thus required (Hansson and Westerberg 2002; Szulkin *et al.* 2010). In both cases, the direction of the effects depends on the function of the fitness loci, on the allelic dominance of the genes affecting the phenotype, on allele frequencies, and on the type of selective pressures acting on the fitness loci, among others (Mueller *et al.* 2010). As positive, negative and neutral effects can be expected (Lieutenant-Gosselin and Bernatchez 2006; Mueller *et al.* 2010), a

common approach to investigate local (or direct) effects is by correlating heterozygosity at each single locus with fitness traits. This approach increases the probability of Type I errors (i.e. the error of rejecting the null hypothesis when it is actually true, also referred to as "false positives"), especially when many markers are tested for associations with many fitness-related traits (Chapman *et al.* 2009). Moreover, biological interpretations of statistically significant local or direct effects require some knowledge about the location of the marker loci with respect to genes and about the function of the marker loci and linked genes. Since variation in the level of inbreeding is not required to generate direct or local effects of heterozygosity, studying HFC in individuals with contrasting phenotypes but with the same level of inbreeding (e.g. full-sibs) is useful to discriminate between direct/local effects and the general effects of inbreeding (e.g. Hansson *et al.* 2001; Hansson *et al.* 2004). Therefore, in order to study local or direct effects of heterozygosity it is important to 1) consider the neutrality or functionality of the markers employed, to 2) map marker loci in the genome and identify their position with respect to genes, to 3) control for multiple testing and to 4) employ a study design that allows to control for the level of inbreeding.

In Chapters 2 and 3 of my thesis I investigated HFC in an open population of blue tits (*Cyanistes caeruleus*), following the guidelines proposed above, to establish the importance of inbreeding and localized effects as causes of significant HFC.

Molecular markers

Linking molecular variation to fitness ultimately requires the use of polymorphic markers. HFC studies originated in the late 1960s with the development of molecular tools to survey protein polymorphisms in population samples (Hubby and Lewontin 1966; Lewontin and Hubby 1966). Until the late 1990s, allozymes were commonly used to measure individual heterozygosity in HFC studies (Mitton and Grant 1984; Zouros and Foltz 1987; Britten 1996; Mitton 1997). Later on, protein polymorphisms were replaced by DNA polymorphisms, and microsatellites became the markers of choice for measuring heterozygosity (Coltman and Slate 2003; Chapman *et al.* 2009). The advantages of microsatellites over other molecular markers used for population genetic studies include their abundance in the genome, level of polymorphism, assumed neutrality,

codominant nature and methodological convenience (e.g. the small amount of tissue or blood samples required and the ease and speed of genotyping and scoring) (Ashley and Dow 1994; Scribner and Pearce 2000). The initial development of microsatellites by cloning and sequencing is, however, costly and time consuming (Zane *et al.* 2002). Moreover, the markers designed with this approach are, in many cases, species-specific or their cross-amplification success is restricted to closely related species (Moore *et al.* 1991; Primmer *et al.* 1996).

Recently, expression libraries and whole-genome sequences for an increasing number of organisms have become available. They provide useful tools for exploring the genome of almost any species. EST (Expressed Sequence Tags) collections of model organisms have been used as a source to develop *in silico* molecular markers for related non-model species of interest to ecologists and evolutionary biologists (Kantety *et al.* 2002; Rexroad *et al.* 2005; Karaiskou *et al.* 2008; reviewed in Bouck and Vision 2007). The mining of SSR (Simple Sequence Repeats) from EST sequence databases has proven to be a useful, inexpensive and fast approach to identify microsatellite loci in several species (e.g. Perez *et al.* 2005; Kong *et al.* 2007; Slate *et al.* 2007; Kim *et al.* 2008; Tang *et al.* 2008). Another advantage of EST-SSRs over markers designed by cloning and sequencing is their higher transferability between species, and even between genera when compared to anonymous microsatellites (Cordeiro *et al.* 2001; Bouck and Vision 2007; Karaiskou *et al.* 2008; Dawson *et al.* 2010). Publicly available whole-genome sequences and bioinformatic tools for sequence-alignment have even opened the possibility to determine the chromosomal location of molecular markers in the genome of model species. This allows, for example, to establish the relative position of microsatellites with respect to genes and to ensure the use of potentially independent or linked markers in studies that require so.

Microsatellites are usually considered evolutionarily neutral (Jarne and Lagoda 1996). However, evidence is accumulating about the functionality of repeats located in expressed regions of the genome and within protein-coding genes (reviewed in Li *et al.* 2004). Specifically, microsatellites within exons, 3′-UTRs, 5′-UTRs and introns could cause disruptions in reading frames, changes in mRNA stability, fluctuations in gene expression, and inactivation and/or change of function of genes, which ultimately can lead to phenotypic change (Li *et al.* 2004). Therefore, variation in microsatellites' repeat numbers can cause

phenotypic variation on which natural selection can act. Moreover, it has been proposed that microsatellites located within genes may provide a molecular basis for adaptation to environmental changes (Trifonov 2004). These observations can have important implications in HFC studies, since the interpretation of the effects of heterozygosity at multiple and single loci on the phenotype depends strongly on the functionality of the markers considered. The possibility of anchoring microsatellite loci in the genome of model species can provide clues about their functionality. Microsatellite loci located within a coding gene would be expected to be more evolutionarily constrained than markers located in intergenic regions. Similarly, microsatellites located in expressed regions of the genome (e.g. markers developed from EST libraries) can be used as a source of potentially functional loci (Vasemagi *et al.* 2005; Oliveira *et al.* 2009).

In Chapter 1 of my thesis I developed a panel of EST-SSRs and characterized a comprehensive set of microsatellites for the blue tit. I anchored this set of markers in the genome of the two sequenced avian species: the chicken (*Gallus gallus*) and the zebra finch (*Taeniopygia guttata*).

Study species

The blue tit (formerly *Parus caeruleus*), is a common hole-nesting passerine bird of deciduous forests in the western Palearctic. It has been the focus of numerous ecological and behavioral studies in a vast array of topics (e.g. Dhondt *et al.* 1992; Kempenaers *et al.* 1997; Doutrelant *et al.* 1999; Pulido and Diaz 2000; Tremblay *et al.* 2003; Dauwe *et al.* 2005; Smallegange *et al.* 2010; Steinmeyer *et al.* 2010). The main reasons for the popularity of the blue tit as a study species in natural environments are its wide distribution across Europe and the fact that it readily breeds and roosts in artificial nestboxes where researchers have easy access to adults and offspring. The seasonal reproductive output of individual birds can easily be recorded and has been used to investigate causes and correlates of variation in individual quality (e.g. Kempenaers *et al.* 1992; Norris 1993; Przybylo *et al.* 2001; Delhey *et al.* 2003; Foerster *et al.* 2003; Doutrelant *et al.* 2008; Garcia-Navas *et al.* 2009).

Female blue tits generally lay one clutch per breeding season, with an average of 11 eggs per clutch; second clutches are produced occasionally in some populations. The female incubates alone and is fed by the male during this

period. Although socially monogamous, extra-pair paternity in the blue tit is common (Kempenaers *et al.* 1997; Krokene and Lifjeld 2000; Foerster *et al.* 2003; Charmantier *et al.* 2004) and increases the variance in male reproductive success (Kempenaers *et al.* 1992; Kempenaers *et al.* 1997). During the nestling period and until about two weeks after fledging, both parents feed their progeny. The local recruitment rate (i.e. the proportion of the breeding population that is locally born, Matthysen *et al.* 2001) is low (between 2.7% - 4.5%) and male-biased, and most fledglings are presumed to disperse far away from their natal areas. Adults, on the other hand, are mostly sedentary (86% of males and 61% of females remain on breeding grounds the whole year, Glutz von Blotzheim and Bauer 1993). Post-breeding dispersal has been recorded and occurs mostly in females (Valcu and Kempenaers 2008). The adult mortality rate is high, with 25% - 50% of the birds recorded from one breeding season to the next (Glutz von Blotzheim and Bauer 1993).

HFC in blue tits have been investigated previously. In a well-studied Austrian blue tit population, Foerster *et al.* (2003) reported significant positive effects of female multilocus heterozygosity on egg production and survival, and of male multilocus heterozygosity on success in raising young and in plumage ornamentation. Garcia-Navas *et al.* (2009) studied HFC in blue tits breeding in central Spain and reported significant positive effects of female multilocus heterozygosity on egg production and egg quality, and of male multilocus heterozygosity on chick provisioning and plumage ornamentation. In these studies heterozygosity at multiple loci was measured with a panel of 5-14 anonymous microsatellite loci, limiting the possibility to test the alternative hypotheses for the occurrence of HFC. In order to fill this gap and to shed light on the causes of HFC in open, natural populations, I investigated HFC in the population of blue tits studied by Foerster *et al.* (2003) using a large panel of molecular markers. This nestbox population was established in Kolbeterberg, Austria (48°13'N, 16°20'E) in an area of 42 - 50 ha of mixed deciduous woodland, and was carefully monitored between 1998 and 2006.

Outline of this thesis

In Chapter 1 I present a comprehensive set of 106 polymorphic microsatellite loci for blue tits. I assigned the location of all the markers on the genome of the two

avian species that have been completely sequenced (a non-passerine, the chicken, and a passerine, the zebra finch), in order to have some approximation to their position in the genome of the blue tit. I assessed the suitability of the markers for population genetic studies after measuring Hardy-Weinberg equilibrium, linkage disequilibrium and the frequency of null alleles.

In Chapter 2 I present an analysis of the correlation between individual heterozygosity and estimates of annual reproductive success, and explored its causes. For this purpose, 794 breeding birds were genotyped with 79 microsatellites described in Chapter 1, covering 25 autosomal chromosomes of the zebra finch. I classified the markers as potentially functional or neutral and identified the (closest) genes where the microsatellites were located. I tested for the effect of overall heterozygosity, as well as for the effect of functional and neutral heterozygosity on two direct measures of individual reproductive output and on three traits related to offspring success in nests of adult blue tits. Moreover, I investigated within-population inbreeding by reconstructing the pedigrees in the population with molecular markers. In order to investigate among-population inbreeding, I analyzed the occurrence of a cryptic population structure in the study area. I evaluated the potential of multilocus heterozygosity to reflect the individual level of inbreeding by quantifying identity disequilibrium. Finally, with the aim of exploring local and direct effects, I analyzed the effect of heterozygosity at single loci on the five measures of reproductive success and tested for differences between the effects of single "functional" and "neutral" markers.

In chapter 3 I present an analysis on the effect of individual heterozygosity on measures of survival of blue tits. A total of 1496 birds were genotyped with 79 microsatellites described in Chapter 1. To analyze the effect of heterozygosity on early survival (i.e. of embryos, nestlings and fledglings) I used a full-sibling approach and statistical models that allow to control for the level of inbreeding in the comparisons. For the analysis of adult survival I used one-year-old birds with different levels of inbreeding. I tested for the effects of multilocus heterozygosity measured with presumably functional or neutral loci, as well as for the effects of heterozygosity at single loci. To interpret the results I considered the differences in the markers and in the study designs.

Chapter 1

A genome-wide set of 106 microsatellite markers for the blue tit (*Cyanistes caeruleus*)

Juanita Olano-Marin, Deborah A. Dawson, Alexander Girg, Bengt Hansson, Marcus Ljungqvist, Bart Kempenaers, Jakob C. Mueller

Abstract

We have characterized a set of 106 microsatellite markers in 26-127 individual blue tits (Cyanistes caeruleus), and assigned their location on the zebra finch (Taeniopygia guttata) and on the chicken (Gallus gallus) genome based on sequence homology. Thirty-one markers are newly designed from zebra finch EST sequences, 22 markers were developed by others from EST sequences using different methods, and the remaining 53 loci were previously designed or modified passerine markers. The 106 microsatellite markers are distributed over 26 and 24 chromosomes in the zebra finch and in the chicken genome, respectively, and the number of alleles varies between two and 49. Eight loci deviate significantly from Hardy-Weinberg equilibrium and show a high frequency of null alleles, and three pairs of markers located in the same chromosome appear to be in linkage disequilibrium. With the exception of these few loci, the polymorphic microsatellite markers presented here provide a useful genome-wide resource for population and evolutionary genetic studies of the blue tit, in addition to their potential utility in other passerine birds.

Published as: Olano-Marin J, Dawson DA, Girg A, Hansson B, Ljungqvist M, Kempenaers B, Mueller JC (2010) A genome-wide set of 106 microsatellite markers for the blue tit (*Cyanistes caeruleus*). *Molecular Ecology Resources* 10:516-532.

Research on passerine birds has provided important insights to evolutionary biology. For instance, quantitative genetic studies on this group have enhanced our knowledge on the associations between heritability, selection and microevolution of fitness-related traits (Merilä *et al.* 2001; MacColl and Hatchwell 2003; Raberg *et al.* 2003; Charmantier *et al.* 2004; Garant *et al.* 2005; Postma and van Noordwijk 2005). Recent advances in avian genomics, in particular the release of the genome sequence assembly of the red jungle fowl *Gallus gallus* (International Chicken Genome Sequencing Consortium 2004), and the zebra finch *Taeniopygia guttata* genome assembly (Warren *et al.* 2010), are promising for the development of genetic and genomic resources for passerines. In this context, the design of a large set of genome-wide distributed polymorphic markers, which are anchored in the avian reference genomes, would be of particular interest for commonly studied passerine species. Such a marker set opens up the possibility to build linkage maps, conduct comparative genomics, map quantitative trait loci (QTL), understand heterozygosity–fitness correlations and reveal the underlying genetic basis of phenotypic variation and adaptive evolution.

In the past few years, EST (Expressed Sequence Tags) collections of model organisms have been used as a source to develop molecular markers and to explore the genome of related non-model species of interest to ecologists and evolutionary biologists (Kantety *et al.* 2002; Rexroad *et al.* 2005; Karaiskou *et al.* 2008; reviewed in Bouck and Vision 2007). The mining of SSR (Simple Sequence Repeats) from EST sequence databases has proven to be a useful, inexpensive and fast approach to identify microsatellite loci in several species (e.g. Perez *et al.* 2005; Kong *et al.* 2007; Slate *et al.* 2007; Kim *et al.* 2008; Tang *et al.* 2008). Another advantage of EST-SSRs over markers designed by cloning and sequencing, is their higher transferability between species, and even between genera when compared to anonymous microsatellites (e.g. Cordeiro *et al.* 2001; Bouck and Vision 2007; Karaiskou *et al.* 2008; Dawson *et al.* 2010).

An additional valuable resource for the development of markers in an array of bird species, is the high numbers of passerine microsatellite loci deposited on public sequence databases (e.g. EMBL, GenBank). Details for 550 passerine microsatellite loci were compiled by Dawson *et al.* (2006), and the Passerine BIRDMARKER Database presents data of the cross-species utility of a large number of markers in a wide range of passerine species (http://www.shef.ac.uk/misc/groups/molecol/deborah-dawson-birdmarkers.html). Many of these markers are polymorphic only in the species where they were isolated,

although cross-species amplification success generally increases as the genetic distance between the species decreases (Primmer *et al.* 1996; Dawson *et al.* 2000; Galbusera *et al.* 2000). Indeed, microsatellites originally isolated in a diverse set of bird species were used for the construction of the first linkage map in any passerine, the great reed warbler *Acrocephalus arundinaceus* (Hansson *et al.* 2005).

The blue tit, *Cyanistes caeruleus*, is a common European passerine bird that has been the focus of numerous ecological and behavioral studies in the wild (e.g. Doutrelant *et al.* 1999; Foerster *et al.* 2003; Tremblay *et al.* 2003; Valcu and Kempenaers 2008). Around 20 polymorphic microsatellite markers have been previously described for this species, either by isolation from blue tit genomic libraries (Dawson *et al.* 2000), by testing markers originally isolated in other bird species (Primmer *et al.* 1996; Galbusera *et al.* 2000; Richardson *et al.* 2000; Johannessen *et al.* 2005; Poesel *et al.* 2006; see also Passerine BIRDMARKER Database http://www.shef.ac.uk/misc/groups/molecol/deborah-dawson-birdmarkers.html), or by searching for potentially functional polymorphisms in candidate genes (Johnsen *et al.* 2007; Steinmeyer *et al.* 2009).

In the present study, we present a set of 106 polymorphic microsatellite markers for the blue tit, and anchor them to the zebra finch and the chicken genome. We designed 31 new markers from zebra finch EST-SSR. The other loci tested were developed by others from EST-SSR sequence using different methods or previously described for other bird species, including 10 markers that were originally isolated in the blue tit (Table 1.2). Approximately 60% of the 53 previously described loci had been assigned a location in the chicken genome based on sequence homology (Dawson *et al.* 2006). In order to assign chromosome locations for as many loci as possible, we compared all sequences (and/or the zebra finch homologs of these sequences) against the zebra finch and chicken genomes (following Dawson *et al.* 2006; Dawson *et al.* 2007). This large set of *in silico* mapped polymorphic markers complements the limited genetic resources available for the blue tit, and constitutes a potential resource of molecular markers for other passerine birds.

Material and Methods

Microsatellite markers

In a previous study, Slate *et al.* (2007) identified simple sequence repeats in a collection of zebra finch EST sequences deposited in GenBank, and predicted their location in the chicken genome. We aligned these zebra finch EST sequences with homologous sequences of multiple species, and then designed conserved primer pairs flanking the repeat region. First, we aligned the zebra finch EST bearing SSR with its chicken genome sequence homolog and, if available, with other bird species' sequence homolog. Homologs were identified using the BLAT search function at the University of California Santa Cruz (UCSC) Chicken Genome Browser (http://genome.ucsc.edu/cgi-bin/hgGateway), and/or the BLAST search function against "Aves" sequences on the nucleotide collection (nr/nt) or the non-human, non-mouse ESTs (est_others) databases of the National Center for Biotechnology Information (NCBI) (http://www.ncbi.nlm.nih.gov). Second, we designed primers with the program PrimaClade (http://www.umsl.edu/services/kellogg/primaclade.html): primer length was between 17-27 bp, with a maximum of 3 degenerated positions, and with an expected product size between 100-400 bp. We tested the primers in one zebra finch and 7-14 presumably unrelated blue tit individuals from a breeding population at Kolbeterberg, Vienna (Austria). Genomic DNA was extracted from blood with the GFX Genomic Blood DNA Purification Kit (GE Healthcare Europe, Freiburg, Germany), following the manufacturer's protocol. Each 10 µl PCR reaction contained 1x PCR buffer (Fermentas), 2mM $MgCl_2$, 0.2mM dNTPs, 0.5 µM of each primer, 0.25 U Taq DNA Polymerase (recombinant, Fermentas), and between 20–40 ng of genomic DNA. We used a touchdown PCR program as follows: 94°C for 5 min; 8 cycles of 94°C for 30 sec, annealing temperature (Ta, Table 1.2) + 4°C (reducing 1°C per cycle) for 1 min, and 72°C for 1 min; 22 cycles of 94°C for 30 sec, Ta for 1 min, and 72°C for 1 min; and 70°C for 15 min. The products were visualized on a 4.5–10% native polyacrylamide gel stained with ethidium bromide. When bands of different sizes were observed, we confirmed the presence of a polymorphism by separating the fragments on an ABI 3130 *xl* Genetic Analyzer with the GeneScan™ 500 LIZ® Size Standard (Applied Biosystems), using fluorescently labeled primers in the PCR reactions. Raw data were analyzed with GeneMapper 4.0.

Additionally, to make the list of markers more exhaustive, we also tested and validated all markers known to be of utility in blue tit and other passerine species. We tested 80 loci including many for which conserved primer sets had been developed and proven to be of high utility in a wide range of passerine and non-passerine birds (Dawson 2007; Dawson *et al.* 2010 & unpublished data; G.N. Hinten, unpublished data).

We checked that the microsatellite loci listed in Table 1.2 were not represented multiple times due to description of the same repeat region in different species and/or via different methods. To identify redundant sequences we used a similar approach to the one described in Dawson *et al.* (2006): a file of the compiled sequences of all the microsatellite loci was compared against itself using the NCBI specialized BLAST to align two (or more) sequences (bl2seq) (http://blast.ncbi.nlm.nih.gov/Blast.cgi), with the sequence accession records of all microsatellites as both the input file and the reference database. A hit was defined as significant with an E value $< 1 \times 10^{-10}$ and an alignment score > 180.

Genotyping

Multiplex PCR reactions were optimized for 82 of the markers (Table 1.2). For all loci we genotyped either a sample of 127 adult blue tits from the 2002 breeding population at Kolbeterberg, Vienna (Austria), or 26-73 unrelated adults from a population at Revinge (Sweden). The PCR conditions were as described in Haas & Hansson (2008), or when Qiagen Multiplex PCR Master Mix was used, following the manufacturer's recommendations: each 10 µl multiplex PCR contained 15-40 ng DNA, 5 µl of the 2x Qiagen Multiplex PCR Master Mix and 1 µl of the primer mix. Unless otherwise stated in Table 1.2, cycling conditions were: 95°C for 15 min; 24 cycles of 94°C for 30 sec, annealing temperature for 90 sec, and 72°C for 90 sec; and 60°C for 30 min. PCR products were separated and detected on an ABI 3130 *xl* Genetic Analyzer with the GeneScan™ 500 LIZ® Size Standard (Applied Biosystems), and analyzed with GeneMapper 4.0.

Zebra finch and chicken chromosome location assignments

In order to rule out spurious chromosome location assignments due to potential vector contamination in both the microsatellite loci and the genome assemblies (as identified by Dawson *et al.* 2006 & unpublished data), all sequences containing microsatellites were screened for vector sequences with the VecScreen (build 5.1)

web tool in the NCBI (http://www.ncbi.nlm.nih.gov/VecScreen/VecScreen.html).

The location of the zebra finch EST-SSR sequences (from which the markers were designed) on the zebra finch genome could be found by typing the accession number of the sequence in the University of California Santa Cruz (UCSC) Genome Browser (http://genome.ucsc.edu/cgi-bin/hgGateway). To assign locations for the remaining markers in the zebra finch, we performed a BLAT search of each sequence containing a microsatellite against the zebra finch genome at the UCSC browser (http://genome.ucsc.edu/cgi-bin/hgGateway). We confirmed the position of the markers by performing a BLASTN search of each sequence containing a microsatellite against the zebra finch genome in the GSC BLAST server at the Washington University in St. Louis (WUSTL) School of Medicine (http://genome.wustl.edu/tools/blast/): we used the DUST/SEG filter and the RepeatMasker (v 3.2.7), and accepted matches with an E value of $< 1\times10^{-8}$; when more than one significant match occurred, the best hit was accepted when it had an E value $< 1\times10^{-8}$ and the next hit was more than one order of magnitude weaker. When a sequence matched both a named chromosome and the chromosome Unknown with the criteria described above, the position in the named chromosome was assigned (method as Dawson *et al.* 2006). Microsatellite loci that could not be directly located in the zebra finch genome were assigned chromosome locations using zebra finch sequence data, following Dawson *et al.* (2007): homologous zebra finch sequences were identified by performing a cross-species megaBLAST search of the loci against the NCBI's zebra finch WGS database (http://www.ncbi.nlm.nih.gov/projects/genome/seq/BlastGen/BlastGen.cgi?pid=12898). The homologous zebra finch sequences were much longer (ca. 700–900 bp) than the original sequences bearing the microsatellite, and therefore when these homologs were compared against the WUSTL zebra finch genome, many additional loci could be assigned a location. When one marker was assigned to a chromosome position in only one of the two BLA(S)T servers, we confirmed its location by performing a BLAST search against the zebra finch reference genome in the NCBI. When the position of one marker differed between the zebra finch genome servers, we report the location in the UCSC, which in all cases could be confirmed by performing a BLAST search of the locus against the zebra finch reference genome in the NCBI.

The microsatellite markers newly designed from zebra finch EST-SSR were assigned chromosome locations in the chicken genome by Slate and collaborators (2007). To assign the remaining loci to a position in the chicken genome we followed the approach of Dawson *et al.* (2007). In short, we performed a BLAST search of each microsatellite locus against the chicken genome with the Ensembl WU-BLAST software (Gish W. 1996–2004; http://blast.wustl.edu) using the Ensembl chicken genome browser and the 'distant homologies' search setting, that is optimized for detecting homology between divergent taxa which uses a word length of nine (http://www.ensembl.org/Gallus_gallus/blastview). We accepted matches with an E value of $< 1\times10^{-10}$; when more than one significant match occurred, the best hit was accepted when it had an E value $< 1\times10^{-10}$ and the next hit was more than one order of magnitude weaker. Furthermore, microsatellite loci that could not be directly located in the chicken genome, were assigned chromosome locations using the homologous zebra finch sequence for each locus following Dawson *et al.* (2007) as described above, but matched against the 'WASHU 2.1' chicken genome assembly in the final step.

Population genetics and linkage disequilibrium

Allele frequencies and observed and expected heterozygosity were calculated with GENEPOP 4.0 (Rousset 2008). To test for Hardy-Weinberg equilibrium (HWE), we used the exact tests available in GENEPOP 4.0 (Rousset 2008): the p-value for the markers with 4 or less alleles was calculated with the complete enumeration algorithm, and for the markers with 5 or more alleles with the Markov-chain algorithm with the default parameters. Estimated null allele frequencies were calculated for the loci that deviated from HWE due to heterozygote deficiency, using the maximum likelihood method with the EM algorithm of Dempster *et al.* (1977) available in GENEPOP 4.0. The loci assigned to the Z chromosome were confirmed to be Z-linked from a comparison of the genotype data of males and females. Sex-linkage was assessed in 56 females and 71 males (sex-type assigned using the P2-P8 primers, Griffiths *et al.* 1998). At all Z-linked loci, females, the heterogametic sex, amplified only a single allele, whereas males were either homozygotes or heterozygotes. For the allele frequency and HWE analyses of Z-linked markers we used the data of the males only, the homogametic sex in birds. All loci amplified in males (ZZ), indicating that none of the markers is linked to the W chromosome in the blue tit.

The degree of linkage disequilibrium (LD) was estimated between all pairs of loci assigned to the same chromosome in the zebra finch (Table 1.2), using the data of the genotyped birds from the study population of Kolbeterberg. The markers located in a "chr_random" are known to be located on the named chromosome, but their exact position still has to be defined; therefore they were included in the linkage analysis of their chromosome. The LD correlation coefficient (*r*LD) between alleles at different loci was computed with the program LinkDos on the web (http://genepop.curtin.edu.au/linkdos.html, Garnier-Gere and Dillmann 1992). To test the significance of *r*LD we used the exact genotypic disequilibrium test available in GENEPOP 4.0 (Rousset 2008). This test evaluates whether genotypes at one locus are independent from genotypes at the other loci using a Markov-chain algorithm on contingency tables. Sequential Bonferroni and False Discovery Rate corrections were used to account for multiple testing with a routine implemented in R (R Development Core Team 2008, script by H. Schielzeth).

Results

Microsatellite markers

We characterized a total of 106 polymorphic microsatellite loci in the blue tit (Table 1.2). From these, 31 markers are newly designed from zebra finch EST-SSR sequences, 22 were isolated by others from EST-SSR sequences using different methods, 43 were previously designed or modified passerine markers, and the remaining 10 are markers previously designed for the blue tit (see references in Table 1.2). From 101 primer pairs designed from zebra finch EST-SSR for this study, amplification was successful with 75 (74%), and among the amplifying loci, 36 (48%) were polymorphic (details not shown).

The BLAST of the sequences of the loci against themselves did not reveal significant homology between any of the sequences, confirming that the list of markers presented here is composed of 106 unique loci.

Table 1.2.
Characteristics of 106[a] polymorphic microsatellite markers for the blue tit (*Cyanistes caeruleus*)

Primer set (Locus name[b])	EMBL / GenBank Accesion No.	Primer sequence (5' - 3')	Reference for primer set development (and original sequence isolation)	Primer set designed from Tgu ESTs	Tgu / Gga chr	Tgu / Gga Start	Ta (°C)	PCR multiplex set (F primer 5' fluoro label)	No. of BT typed[l]	No. of alleles	Allele size range (bp)	BT motif length[n]	Max/Min/Null[o] allele freq	Ho	*He*
ADCYAP1	FJ464427.1	F: GATGTGAGTAACCAGCCACT R: ATAACACAGGAGCGGTGA	Steinmeyer *et al.* 2009	No	2 / 2	107,400,835 / 104,980,255	TD 60-53[i]	16 (6-FAM)	126 (K)	10	158-173	1	0.45 / 0.004	0.63	0.64
ApCo46-ZEST (ApCo46)	AF520885.1	F: GCTGCCAGCACTCTGAATGTC R: GATTCAGCAAAATAGGGGTCAGAAG	Dawson unpublished (Stenzler and Fitzpatrick 2002)	Yes	1 / 1	104,151,755 / 96,690,969	57	13 (NED)	125 (K)	4	210-222	(2)	0.63 / 0.004	0.52	0.52
Ase04-ZFGgaM (Ase4)	AJ287387.1	F: CTCCATCATCACCACAAAGC R: TTCGTTAGCAAGGGTATTTAGTTG	Dawson unpublished (Richardson *et al.* 2000)	No	2 / 2	53,379,036 / 33,278,425	56	-	73 (R)	4	211-217	2	0.62 / 0.01	0.48	0.50
Ase18	AJ276375.1	F: ATCCAGTCTTCGCAAAAGCC R: TGCCCCAGAGGGAAGAAG	Richardson *et al.* 2000	No	3 / 3	13,906,080 / 23,864,832	58	15 (VIC)	125 (K)	15	170-238	2	0.20 / 0.004	0.87	0.88
Ase46-ZFM (Ase46)	AJ276775.1	F: CTGGCTGTATCTTGGTGTGC R: GCTAACTTTCCATTGAACTGTCC	Dawson unpublished (Richardson *et al.* 2000)	No	Z / Z	32,967,367 / 2,104,168	56	-	65[m] (R)	10	135-159	2	0.35 / 0.01	0.80	0.79
Ase60-ZFS (Ase60)	AJ276789.1	F: GGCTTGCTTTTATTGTATCATGC R: CAGGACTGGCATATTAGAAATGTTTAC	Dawson unpublished (Richardson *et al.* 2000)	No	3 / 3	54,084,231 / 53,187,565	56	-	32 (R)	11	204-226	2	0.30/0.03	0.81	0.82
Asμ15-ZEST (Asμ15)	AY172993.1	F: AATAGATTCAGGTGCTTTTTCC R: GGTTTTTGAGAAAATTATACTTTCAG	Dawson unpublished (Bulgin *et al.* 2003)	Yes	5[d] / 5	30,043,672 (30,046,841) / 32,869,016	TD 60-53[i]	16 (6-FAM)	126 (K)	6	97-121	2	0.57 / 0.004	0.45	0.53
BF08-ZEST (BF08)	AB091049.1	F: ACTGACCACTGTTCATGCTG R: GTGACAACCAAAGCAGGTAAG	Dawson unpublished (Yodogawa *et al.* 2003)	Yes	2 / 2	65,741,602 / 47,623,440	55	-	68 (R)	5	101-117	2	0.56 / 0.02	0.60	0.62
Calex-05	AM072453.1	F: TCCAGCTGAAGTCTTCCGTGAAT R: GTTTCTTTCCACACCTGTTCGACAGTTCAATA	Kupper *et al.* 2007	No	2 / 2	55,225,113 / 35,085,671	58	-	66 (R)	8	179-187	(2)	0.57 / 0.02	0.73	0.63
CcaTgu1	DV953173.1	F: AGAGCCCTGTTYATRGCTGT R: CCACCATGCAAACAYCAR	This study	Yes	1 / -	93,427,832 / -	56	19 (NED)	122 (K)	4	283-298	1	0.83 / 0.004	0.25	0.28
CcaTgu2	DV950317.1	F: CAGCMSACAAATGCATCTAC R: GAAGGYGAARTGCTGTCTCTT	This study	Yes	1 / 1	102,481,135 / 94,986,105	56	4 (NED)	125 (K)	2	160-174	3	0.98 / 0.02	0.03	0.03
CcaTgu3	DV580602.1	F: CAAGTGCRGAAAGGAAAGTG R: TCAACTGCATCAGACTTCAAAA	This study	Yes	1 / 1	117,775,285 / 88,668,098	56	19 (VIC)	121 (K)	5	171-185	1	0.75 / 0.01	0.37	0.40
CcaTgu4	CK304284.1	F: TCTTTGCATGCACTCYRG R: CAATTTGCACAAAACATKGC	This study	Yes	1A / 1	53,257,466 / 55,403,554	52	6 (6-FAM)	125 (K)	3	116-124	1	0.82 / 0.02	0.33	0.31
CcaTgu5[c]	DV950249.1	F: AACGAGATCCAGAGGGTCATC R: YTTGGAGCTGCACTGCG	This study	Yes	2 / 2[c]	517,295 / 461,011	56	4 (6-FAM)	64[c] (K)	2	344-350	4	0.73 / 0.27 / 0.35	0.16	0.39*
CcaTgu6	CK235244.1	F: ACAATTGCTAACAAGTGGCAAG R: AAGTGAAATCTKCTTGGGKC	This study	Yes	2 / 2	16,466,870 / 14,817,407	55	2 (NED)	127 (K)	4	95-109	3	0.91 / 0.004	0.17	0.18
CcaTgu7	DV961773.1	F: TTTTTCAGGAAARGGAAACA R: CAAGCTTTTACAGTGCTAWT	This study	Yes	2 / 2	75,982,449 / 56,801,228	52	6 (VIC)	124 (K)	8	274-292	2	0.39 / 0.004	0.72	0.73
CcaTgu8	CK309161.1	F: GCTCTGCASCAGCCMGAT R: CCATCAGTTTCCTGTTGCTG	This study	Yes	2 / 2	121,250,236 / 119,338,034	58	15 (VIC)	124 (K)	9	396-436	(3)	0.47 / 0.004	0.61	0.57
CcaTgu9	DV949447.1	F: ACTGGCAGATGGAGGATTTC R: GGGGATTTTTCAGTCWAAAGA	This study	Yes	2_random / 2	1,385,550 / 5,170,626	55	2 (6-FAM)	127 (K)	5	109-126	3	0.79 / 0.004	0.37	0.34
CcaTgu10	DV575298.1	F: TGAAAAATCTTGCCTTTTTGGTA R: TACAAAGCACTGTGGAGCA	This study	Yes	3 / 3	28,478,877 / 2,475,751	52	6 (NED)	126 (K)	3	121-131	2	0.87 / 0.03	0.23	0.23
CcaTgu11	CK235701.1	F: TGCTTAGGAAATAGGAAGCACA R: CTGCAACTTAAGCARRGTTATGA	This study	Yes	3 / 3	62,483,485 / 62,758,279	53	5 (NED)	124 (K)	6	188-203	2	0.31 / 0.03	0.81	0.76
CcaTgu12	DV582287.1	F: TGTTTACCAACTATTGCTTTGTTGA R: ACGCAGCTTGAAAGGCATAC	This study	Yes	3 / 3	88,019,185 / 88,427,734	53	5 (VIC)	126 (K)	2	117-128	5	0.98 / 0.04	0.04	0.04
CcaTgu13	DV949158.1	F: GAGGAACRTTGTACAGAACACCT R: GTAGCTTGTCACYSTAAGCCA	This study	Yes	4 / 4	65,195,621 / 88,229,339	TD 61-53[j]	1 (PET)	127 (K)	8	127-160	3	0.34 / 0.004 / 0.06	0.67	0.73*

Table 1.2 cont.

Primer set (Locus name[b])	EMBL / GenBank Accesion No.	Primer sequence (5' - 3')	Reference for primer set development (and original sequence isolation)	Primer set designed from Tgu ESTs	Tgu / Gga chr	Tgu / Gga Start	Ta (°C)	PCR multiplex set (F primer 5' fluoro label)	No. of BT typed[l]	No. of alleles	Allele size range (bp)	BT motif length[n]	Max/Min/Null[o] allele freq	Ho	*He*
CcaTgu14	DV960138.1	F: GTTGTTCYAATTCCAATGC R: CTAAAAATAGCAGTAAAAATACAYAAA	This study	Yes	5 / 5	40,815,020 / 42,880,050	51	2 (VIC)	127 (K)	17	117-178	3	0.31 / 0.004	0.83	0.80
CcaTgu15	DV952837.1	F: TTAATCCTAGGGTGTYGAGAGAAC R: CCTTTTTCCTTAAATTAKCTCAGCTT	This study	Yes	5 / 5	61,276,973 / 59,347,278	60	20 (NED)	125 (K)	7	124-138	1	0.36 / 0.004	0.68	0.68
CcaTgu16	DV947938.1	F: AGCGAATTTCACACATGCC R: YTAAGTACAGTCTCACRCAGCAG	This study	Yes	7 / 7	38,131,923 / 36,638,086	TD 61-53[j]	1 (6-FAM)	124 (K)	5	233-246	3	0.87 / 0.01	0.23	0.24
CcaTgu17	DV952014.1	F: GGGTCTCTGRKTTGGAGC R: CARAGTCCCCTCTGCCTTC	This study	Yes	8_random / 8[c]	1,710,707 / 4,812,635	58	3 (PET)	126 (K)	3	129-140	3	0.63 / 0.06	0.51	0.50*
CcaTgu18	CK308660.1	F: GGTTTCCTGGAGTCCTTTTG R: CAAACCRACAAAACCAACCC	This study	Yes	8_random / 8[c]	2,367,489 / 5,413,143	56	4 (PET)	126 (K)	3	111-119	3	0.86 / 0.004	0.24	0.25
CcaTgu19	DV579042.1	F: CTGGACCATGACTGCAAGATT R: CAGTGGCAAAKCAGCACCT	This study	Yes	10 / 10	6,370,604 / 7,689,780	TD 61-53[j]	1 (PET)	124 (K)	21	234-299	3	0.12 / 0.004	0.94	0.93
CcaTgu20	DV956700.1	F: AGACKRAAGAAGTCACCAAGTAAAR R: TTTCCTGAAGGTRCTTCWCAG	This study	Yes	10 / 10	6,994,363 / 8,313,724	55	7 (6-FAM)	126 (K)	2	105-110	3	0.92 / 0.07	0.15	0.14
CcaTgu21	DV961016.1	F: GGCAGACATGATTGCATCC R: TCTCAGTGGTCATTGGAAAGTG	This study	Yes	11 / 11	12,285,425 / 21,546,491	TD 61-53[j]	1 (VIC)	126 (K)	8	195-225	3	0.84 / 0.004	0.29	0.28
CcaTgu22	CK309186.1	F: TGTGCTCTGCYAAAYCTCTC R: GGAAAYAGAGAATTGTCGTCATC	This study	Yes	12 / 12	5,239,036 / 3,600,137	56	4 (VIC)	125 (K)	2	99-109	9	0.95 / 0.05	0.08	0.09
CcaTgu23	CK316117.1	F: CAAGGMYCATGCCAAAATAA R: CCCTYCCTCCCTTCAGTTTT	This study	Yes	14 / 14	11,485,613 / 4,122,846	53	5 (PET)	126 (K)	8	123-145	3	0.85 / 0.004	0.24	0.26
CcaTgu24	CK316202.1	F: ATCTTGAGTYATGACCTTAAAARTCT R: ATAGCTACAGAAACCTACTTGGGA	This study	Yes	17 / 17	9,423,028 / 8,902,297	55	12 (VIC)	126 (K)	3	155-160	1	0.56 / 0.21	0.66	0.59
CcaTgu25	CK305580.1	F: CCAGGTYRGCATCAGTACAC R: TAAACCACRCGTGGGCAC	This study	Yes	18 / 18	7,447,751 / 4,041,446	55	12 (PET)	124 (K)	22	133-178	1	0.22 / 0.004	0.83	0.90
CcaTgu26	DV577718.1	F: AAATTCATTCATRTGATTGGTGG R: CATATGTACAAAGGCTGMAAAGT	This study	Yes	18 / 18	9,768,597 / 5,152,373	52	6 (PET)	125 (K)	4	133-149	1	0.71 / 0.05	0.46	0.46
CcaTgu27	DV947660.1	F: ARACAGGGCGAAGTTTCTGAR R: GCAGATTCATGAGATGATGAGAGA	This study	Yes	19 / 19	10,535,947 / 8,737,881	58	3 (6-FAM)	127 (K)	4	159-170	3	0.97 / 0.004	0.06	0.05
CcaTgu28	DV955772.1	F: TCTGGACTCTTGGCACCTG R: GCTTAAGGAGAAAAYAATCCTTCAC	This study	Yes	23_random / 23	246,698 / 5,315,319	60	20 (6-FAM)	123 (K)	11	230-247	1	0.37 / 0.004	0.76	0.75
CcaTgu29	CK305172.1	F: CTTAGGCTTTCCATAGCTGTATT R: CCAATCATTCAGTTTTCTTGCTT	This study	Yes	26 / 26	802,761 / 3,742,885	57	13 (6-FAM)	126 (K)	16	174-201	1	0.18 / 0.004 / 0.04	0.84	0.91*
CcaTgu30	DV954489.1	F: TCYCACGCCTCACCTTTYG R: GTGCCTTCRTTYAGCTCCTCTTT	This study	Yes	28 / 28	2,572,470 / 3,840,622	58	3 (NED)	124 (K)	2	163-170	6	0.996 / 0.004	0.01	0.01[p]
CcaTgu31	CK311793.1	F: AGCCCAGATTKGAAATRAACT R: TGTAGCYYTGTAGAATTAGCAAA	This study	Yes	Z / Z	44,295,294 / 13,067,682	55	7 (NED)	70[m] (K)	5	148-160	2	0.36 / 0.01	0.71	0.69
Cdi31-ZM (Cdi31)	AB089172.1	F: GAACTTCTGCATTTGTTCCTCTC R: GAGAGCGTGCTGAATGAGTG	Dawson unpublished (Otsuka *et al.* 2003)	No	7 / 7	4,201,012 / 29,397,214	58	17 (6-FAM)	125 (K)	15	143-205	2	0.71 / 0.004	0.49	0.49
ClkpolyQcds	AY338427.1	F: TTTTCTCAAGGTCAGCAGCTTGT R: CTGTAGGAACTGTTGYGGKTGCTG	Johnsen *et al.* 2007	No	4 / 4	43,440,144 / 66,976,142	58	23 (6-FAM)	121 (K)	4	267-283	3	0.53 / 0.03	0.52	0.60
CREB1	FJ464428.1	F: CTCTGAGAGCTTTTAAGTCAGGA R: ATAAGCGCTAGAATAACGCAGC	C. Steinmeyer unpublished	No	7 / 7	21,321,109 / 13,253,758	53	5 (6-FAM)	126 (K)	8	172-196	2	0.85 / 0.004	0.29	0.27
Cuμ04	AF122891.1	F: AATTGCATAAATGTGATCCAC R: AAATGAAATGTGGTAGAATTCC	Gibbs *et al.* 1999	No	5 / 5	30,459,232 / 33,265,884	55	-	67 (R)	12	124-154	2	0.26 / 0.01	0.81	0.83

Table 1.2 cont.

Primer set (Locus name[b])	EMBL / GenBank Accesion No.	Primer sequence (5' - 3')	Reference for primer set development (and original sequence isolation)	Primer set designed from Tgu ESTs	Tgu / Gga chr	Tgu / Gga Start	Ta (°C)	PCR multiplex set (F primer 5' fluoro label)	No. of BT typed[l]	No. of alleles	Allele size range (bp)	BT motif length[n]	Max/Min/Null[o] allele freq	Ho	*He*
DkiB102-ZEST (DkiB102)	AY769673.1	F: TTGCAACAGGAGGACAAGG R: CAGCAGCACTTCCCAATACA	Dawson unpublished (King *et al.* 2005)	Yes	15 / 15	8,166,844 / 660,477	58	15 (PET)	125 (K)	22	189-269	3	0.13 / 0.004	0.92	0.94
DkiB119-CEST (DkiB119)	AY769677.1	F: CATACAACTTCATGACTACCATAGCAC R: TCCATAGTGACATAGAACGAGCTG	Dawson unpublished (King *et al.* 2005)	Yes	18_random / 18	93,096 / 8,033,076	58	15 (6-FAM)	125 (K)	5	228-236	1	0.51 / 0.02	0.60	0.61
DkiD12-ZF Chr9 (DkiD12)	AY769684.1	F: GCTTGGCAATTAAAAACTCAA R: CAAGACACTGAGGCATCAAA	Dawson unpublished (King *et al.* 2005)	No	9 / 9	9,399,880 / 9,417,033	55	12 (NED)	122 (K)	9	169-245	(2)	0.53 / 0.01 / 0.22	0.38	0.63*
Gf06	AF081930.1	F: GCTATTGAGCTAACTAAATAAACAACT R: CACAAATAGTAATTAAAAGGAAGTACC	Petren 1998	No	1A[c] / 1	31,887,783 / 34,375,129	50	-	123 (K)	6	140-168	(2)	0.27 / 0.004	0.76	0.80
LEI160	AM159172.1	F: GCAGACAGCCGTTAATATATGCG R: AACCAAAACACAAGCTCTTGCA	Gibbs *et al.* 1997	No	1 / 1	109,699,352 / 102,460,313	58	17 (PET)	124 (K)	4	157-165	2	0.84 / 0.03 / 0.07	0.23	0.28*
LOX1	Y16820.2	F: ATGATGGTAAGTCTAATGAAAGC R: CCACACACATTCACTCTATTG	Piertney *et al.* 1998	No	2 / 2	132,067,279 / 130,451,221	54	-	67 (R)	3	264-266	1	0.80 / 0.01	0.40	0.33
Mcyμ4	U82388.1	F: ATAAGATGACTAAGGTCTCTGGTG R: TAGCAATTGTCTATCATGGTTTG	Double *et al.* 1997	No	5 / 5	31,893,652 / 34,635,556	53 / 55[g]	-	125 (K)	18	145-196	2	0.32 / 0.004	0.78	0.86
MJG1	U82673.1	F: CCCGGGAAAGGCTTCGTCTTC R: GGAGATTTTATATCGGTGGC	Li *et al.* 1997	No	4A / 4	297,614 / 4,559,773	TD 60-50[k]	-	54 (R)	10	116-128	1	0.46 / 0.01	0.76	0.70
MSLP2-Gga5 (MSLP2)	AB031374.1	F: TYRTATCTRTAACTACAGCCAKTTAG R: AAAGTTGAKRTAAATGARGTTACTGG	Hinten *et al.* unpublished (Ishibashi *et al.* 2000)	No	5 / 5	47,677,905 / 49,013,616	53	-	32 (R)	5	71-89	2	0.61/0.02	0.50	0.58
MSLP4-ZEST (MSLP4)	AB031376.1	F: CAGTGCAAGCCTCTACTCCTCTGC R: CTTGGCTCTGCCGCACCTC	Dawson unpublished (Ishibashi *et al.* 2000)	Yes	9 / -	4,935,546 / -	60	20 (VIC)	122 (K)	8	141-159	2	0.60 / 0.004	0.57	0.59
NPAS2	FJ464429.1	F: CTGTGGTAAATTTGATGATTCTGA R: ACACCAAGTTCTTTGCACAATG	Steinmeyer *et al.* 2009	No	1[d] / 1	30,322,977 (30,327,831) / 137,500,260	55	2 (PET)	127 (K)	6	170-193	3	0.36 / 0.004	0.77	0.74
PAT MP 2-43	AM056063.1	F: ACAGGTAGTCAGAAATGGAAAG R: GTATCCAGAGTCTTTGCTGATG	Otter *et al.* 1998	No	2[e] / 2	47,487,652 / 27,856,022	58	23 (6-FAM)	125 (K)	8	122-156	2	0.66 / 0.004	0.57	0.52
Pca3	AJ279805.1	F: GGTGTTTGTGAGCCGGGG R: TGTTACAACCAAAGCGGTCATTTG	Dawson *et al.* 2000	No	4 / 4	64,610,435 / 87,603,353	TD 60-53[i]	21 (PET)	127 (K)	26	161-251	1	0.13 / 0.004	0.95	0.93
Pca4	AJ279806.1	F: AATGTCTTACAGGCAAAGTCCCCA R: AACTTGAAGCTTCTGGCCTGAATG	Dawson *et al.* 2000	No	8[e] / 8	22,128,537 / 25,300,802	58	23 (NED)	123 (K)	14	148-197	2	0.31 / 0.004	0.83	0.81*
Pca7	AJ279809.1	F: TGAGCATCGTAGCCCAGCAG R: GGTTCAGGACACCTGCACAATG	Dawson *et al.* 2000	No	1 / 1	81,091,062 / 191,155,658	TD 60-53[i]	21 (6-FAM)	127 (K)	17	112-142	2	0.23 / 0.004	0.91	0.89
Pca8	AJ279810.1	F: ACTTCTGAAACAAAGATGAAATCA R: TGCCATCAGTGTCAAACCTG	Dawson *et al.* 2000	No	2 / -	38,606,108 / -	57	-	57 (R)	43	157-384	(2)	0.11 / 0.01	0.98	0.97
Pca9	AJ279811.1	F: ACCCACTGTCCAGAGCAGGG R: AGGACTGCAGCAGTTTGTGGG	Dawson *et al.* 2000	No	7[e] / 7	10,442,876 / 23,851,120	56 / 57[h]	22 (VIC)	127 (K)	13	103-133	2	0.33 / 0.004	0.83	0.83
Pdo47-Gga5 (Pdo47)	AM159027.1	F: TCCARCAKGARCWGMGAATA R: YSWCMCAGCCTTSCTARAAGTG	Hinten *et al.* unpublished (Dawson *et al.*submitted)	No	5 / 5	54,452,634 / 55,151,044	53	-	32 (R)	2	251-252	1	0.66/0.34	0.31	0.46
Pdoμ5	Y15126.1	F: GATGTTGCAGTGACCTCTCTTG R: GCTGTGTTAATGCTATGAAAATGG	Griffith *et al.* 1999	No	4[d] / 4	48,504,861 (48,501,861) / 71,692,281	54	-	32 (R)	20	233-289	1	0.17/0.02	0.88	0.92
Phtr3	AM056070.1	F: ATTTGCATCCAGTCTTCAGTAATT R: CTCAAAGA AGTGCATAG AGATTTCAT	Fridolfsson *et al.* 1997	No	Z[f] / Z	13,746,684 / 57,675,142	TD 60-50[k]	-	54[m] (R)	11	125-145	2	0.23 / 0.01	0.87	0.85
PiJ14-23-CEST (PiJ14)	AB204825.1	F: ATCTGGCATKGAAAACTTGG R: CTCCCTGCACCCCAAAC	Dawson unpublished (Saito *et al.* 2005)	No	7 / 7	20,810,935 / 13,700,295	57	13 (VIC)	126 (K)	14	156-222	2	0.25 / 0.004	0.88	0.88
PK12	AF041466.1	F: CCTCCTGCAGTTGCCTCCCG R: CGTGGCCATGTTTATAGCCTGGCACTAAGAAC	Tanner *et al.* unpublished	No	5[f] / -	22,671,767 / -	58	23 (VIC)	126 (K)	19	168-220	2	0.29 / 0.004	0.83	0.85

Table 1.2 cont.

Primer set (Locus name[b])	EMBL / GenBank Accesion No.	Primer sequence (5' - 3')	Reference for primer set development (and original sequence isolation)	Primer set designed from Tgu ESTs	Tgu / Gga chr	Tgu / Gga Start	Ta (°C)	PCR multiplex set (F primer 5' fluoro label)	No. of BT typed[l]	No. of alleles	Allele size range (bp)	BT motif length[n]	Max/Min/Null[o] allele freq	Ho	*He*
Pma303	AB094110.1	F: CCCACAGCAATCTCCCTCCA R: GGTGGCTTTTCTCTGCACAC	Kawano 2003	No	20[c] / 20	7,188,252 / 6,607,187	58	17 (VIC)	124 (K)	10	155-185	2	0.44 / 0.004	0.72	0.70
PmaC25	AY260526.1	F: CGTCCTGCTGTTTGTATTTCTG R: CCATGAACCATTTTTAGGGTG	Saladin *et al.* 2003	No	12[d] / 12	1,378,621 (1,384,963) / 5,637,982	55	14 (PET)	124 (K)	7	315-351	3	0.39 / 0.02	0.74	0.75
PmaD22	AY260527.1	F: GATCAGAGCTTGCCTCAACAC R: TCTGGGCTGAAATACCTACCC	Saladin *et al.* 2003	No	1 / -	112,574,836 / -	57	-	57 (R)	19	370-482	2	0.13 / 0.01	0.86	0.92
PmaGAn11	AY260531.1	F: GCTTCTGCCTCCATTAAGAGTC R: GAAAAATCACCCACTCAGCC	Saladin *et al.* 2003	No	2[d] / 2	107,642,995 (107,623,746) / 105,186,697	57	-	67 (R)	2	106-108	2	0.65 / 0.35	0.49	0.45
PmaGAn27	AY260532.1	F: TATAAACCACAGCCACACGC R: CACAACCACAGAGGCATGAG	Saladin *et al.* 2003	No	6 / 6	36,100,371 / 37,223,385	55	14 (VIC)	126 (K)	32	107-188	(2)	0.32 / 0.004	0.84	0.84
PmaGAn30-ZEST (PmaGAn30)	AY260534.1	F: GTCAGTAAAAATGCTGTGTTTGTC R: AGATCCCCAGGTATCTGCTC	Dawson unpublished (Saladin *et al.* 2003)	Yes	20[d] / 20	14,965,400 (14,971,064) / 13,784,540	56	11 (VIC)	126 (K)	4	106-118	2	0.72 / 0.004	0.48	0.44
PmaGAn40	AY260536.1	F: CGTTCCTCCTTTGCTTTCTG R: AATGGCACAACACCTTCTCC	Saladin *et al.* 2003	No	21 / 21	3,523,674 / 1,609,479	58	17 (NED)	122 (K)	6	403-428	(2)	0.61 / 0.01	0.61	0.57
PmaTAGAn71	AY260537.1	F: TCAGCCTCCAAGGAAAACAG R: GCATAAGCAACACCATGCAG	Saladin *et al.* 2003	No	7 / 7	19,953,067 / 14,495,694	57	-	31 (R)	16	199-283	4	0.21/0.02	0.87	0.90
PmaTAGAn86	AY260538.1	F: AAAACAAGGCCACTTAGAGCTG R: ACTCCTCCAGGTCACACAGG	Saladin *et al.* 2003	No	1 / 1	34,807,980 / 133,108,205	57	-	55 (R)	9	112-168	2	0.45 / 0.01 / 0.17	0.44	0.68*
PmaTGAn42	AY260540.1	F: ACTTCCACATGCCAGTTTCC R: TGTTAAGGCAGAGAGGTGGG	Saladin *et al.* 2003	No	2 / 2	70,321,386 / 62,574,465	57	-	67 (R)	12	257-297	4	0.18 / 0.01	0.93	0.88
PmaTGAn45	AY260541.1	F: CCCCTGGCTCTTTCATATCC R: GACAGGTGTTGGCACAAGG	Saladin *et al.* 2003	No	9[d] / 9	802,719 (810,337) / 6,795,561	55	14 (6-FAM)	120 (K)	29	301-384	1	0.10 / 0.004	0.89	0.94
POCC1	U59113.1	F: TTCTGTGCTGCAATCACACA R: GCTTCCAGCACCACTTCAAT	Bensch *et al.* 1997	No	7 / 7	1,646,806 / 6,224,071	56 / 57[h]	22 (6-FAM)	127 (K)	14	219-251	1	0.19 / 0.004	0.89	0.88
POCC6	U59117.1	F: TCACCCTCAAAAACACACACA R: ACTTCTCTCTGAAAAGGGGAGC	Bensch *et al.* 1997	No	2[c] / 2	44,644,858 / 66,037,942	TD 60-53[i]	21 (VIC)	127 (K)	25	202-254	2	0.23 / 0.004	0.89	0.90
Pte24-CEST (Pte24)	DQ234870.1	F: AACAAAGGACGCCGAGTAG R: TCATTTAATGGCTYTACTTCATACAT	Dawson unpublished (Blackmore *et al.* 2006)	Yes	2 / 2	107,400,963 / 104,980,341	55	-	55 (R)	3	236-240	2	0.74 / 0.07	0.44	0.41
TG01-000	CK314156.1	F: TTGCTACCARAATGGAATGT R: TCCTAACCATGAGAAGCAGA	Dawson *et al.* 2010	Yes	1A / 1	201,308 / 725,367	56	9 (PET)	124 (K)	3	221-233	1	0.76 / 0.04 / 0.17	0.24	0.38*
TG01-040	DV576233.1	F: TGGCAATGGTGAGAAGTTTG R: AGAATTTGTACAGAGGTAATGCACTG	Dawson *et al.* 2010	Yes	1A / 1	42,620,504 / 45,170,264	55	7 (VIC)	126 (K)	5	289-299	1	0.77 / 0.004	0.37	0.38
TG01-124	CK306631.1	F: AGTACTACTTGCCTGCAGAGTTTAT R: TGTGTATGGCAGCATTTACAA	Dawson *et al.* 2010	Yes	1 / 1	32,303,776 / 135,576,414	55	7 (6-FAM)	123 (K)	7	395-409	2	0.87 / 0.004	0.22	0.23
TG02-088	DV579347.1	F: TGTGTGTTGACAGTATTCTCTTGC R: TTTAAACCTAATAAACGTCACACAGTC	Dawson *et al.* 2010	Yes	2 / 2	93,538,047 / 90,692,704	55	12 (6-FAM)	123 (K)	6	254-264	1	0.76 / 0.004	0.41	0.40
TG03-098	DV573670.1	F: TTTGCCTTAATTCTTACCTCATTTG R: TTGCAACCTCTGTGGAAGC	Dawson *et al.* 2010	Yes	3 / 3	102,966,495 / 103,032,357	55	8 (VIC)	124 (K)	6	238-247	1	0.49 / 0.02	0.63	0.62
TG04-004	DV946288.1	F: CTGGAGCAGTATTTATATTGATCTTCC R: GAAGATGTGTTTCACAGCATAACTG	Dawson *et al.* 2010	Yes	4A / 4	6,997,361 / 4,185,987	55	8 (NED)	125 (K)	3	162-168	1	0.94 / 0.01	0.11	0.11
TG04-061	CK235034.1	F: GACAATGGCTATGAAATAAATTAGGC R: AGAAGGGCATTGAAGCACAC	Dawson *et al.* 2010	Yes	4 / 4	20,910,674 / 61,990,507	56	-	26 (R)	4	193-196	1	0.58 / 0.10	0.65	0.60
TG05-046	DV957774.1	F: AAAACATGGCTTACAAACTGG R: GCTCAGATAAGGGAGAAAACAG	Dawson *et al.* 2010	Yes	5 / 5	50,735,925 / 51,923,094	55	7 (NED)	124 (K)	2	329-338	1	0.58 / 0.42	0.45	0.49

Table 1.2 cont.

Primer set (Locus name[b])	EMBL / GenBank Accesion No.	Primer sequence (5' - 3')	Reference for primer set development (and original sequence isolation)	Primer set designed from Tgu ESTs	Tgu / Gga chr	Tgu / Gga Start	Ta (°C)	PCR multiplex set (F primer 5' fluoro label)	No. of BT typed[l]	No. of alleles	Allele size range (bp)	BT motif length[n]	Max/Min/Null[o] allele freq	Ho	*He*
TG05-053	CK314425.1	F: GCATCATCTGGTTGAACTCTC R: ACCCTGTTTACAGTGAGGTGTT	Dawson *et al.* 2010	Yes	5 / 5	61,275,929 / 59,348,595	55	8 (PET)	125 (K)	8	202-219	1	0.35 / 0.01	0.78	0.75
TG11-011	CK308096.1	F: ACAAACTAAGTACATCTATATCTGAAG R: TAAATACAGGCAACATTGG	Dawson *et al.* 2010	Yes	11 / 11	19,380,799 / 11,693,857	56	9 (VIC)	126 (K)	4	210-219	2	0.82 / 0.02	0.34	0.30
TG12-015	DV953675.1	F: ACAACAGTGGCTTTACTGTGTGA R: TACAGCAGCTGCAGCAAAGT	Dawson *et al.* 2010	Yes	12 / 12	16,288,963 / 15,500,052	55	10 (PET)	121 (K)	3	279-285	2	0.88 / 0.01	0.25	0.22
TG13-009	DV948691.1	F: TGTGGTGGGATAGTGGACTG R: CTGTAAAATGTGCAAGTAACAGAGC	Dawson *et al.* 2010	Yes	13 / 13	3,672,471 / 10,457,270	55	10 (NED)	124 (K)	2	189-198	2	0.996 / 0.004	0.01	0.01[p]
TG13-017	CK313422.1	F: GCTTTGCATCTTGCCTTAAA R: GGTAACTACAACATTCCAACTCCT	Dawson *et al.* 2010	Yes	13 / 13	18,208 / 18,850,515	55	7 (PET)	125 (K)	15	349-377	1	0.22 / 0.004	0.82	0.85
Tgu01	DV952809.1	F: CACATCAAAGGATGTGGTGG R: CAATGTGAATTGCAGGGTCA	Slate *et al.* 2007	Yes	7[c] / 7	11,806,628 / 22,761,104	55	8 (6-FAM)	126 (K)	3	187-195	2	0.83 / 0.02	0.28	0.28
Tgu02	DV952125.1	F: TGGATTACCTGTCTGAAAGACC R: TTCACTGTCTAGTCCAACCCTGT	Slate *et al.* 2007	Yes	7 / 7	13,280,326 / 21,195,721	56	11 (NED)	126 (K)	4	182-191	1	0.98 / 0.004	0.03	0.03
Tgu05	DV946651.1	F: CACAGAAAAGTGAGTGCATTCC R: TGGGAAAACATCTTTACCATCA	Slate *et al.* 2007	Yes	1A / 1	64,236,986 / 71,055,819	58	9 (NED)	122 (K)	3	274-280	2	0.82 / 0.02	0.34	0.31
Tgu07	DV948303.1	F: CTTCCTGCTATAAGGCACAGG R: AAGTGATCACATTTATTTGAATAT	Slate *et al.* 2007	Yes	6 / 6	22,807,094 / 24,725,850	57	13 (PET)	126 (K)	7	94-114	2	0.48 / 0.004	0.71	0.68
Tgu09	CK307658.1	F: GGGGACGTTTTATCTGTTACC R: GCAGTGCCCAAGTTCAGAGT	Slate *et al.* 2007	Yes	Z / Z	70,750,908 / 61,092,726	56	9 (6-FAM)	66[m] (K)	4	195-204	2	0.84 / 0.04 / 0.14	0.21	0.29*
TguEST09-005	DV954446.1	F: AACCCAACCAACAAAATTGG R: CCAACTATCAGTTTTACAAGGCATAC	Dawson *et al.* unpublished	Yes	9 / 9	5,196,903 / 5,467,408	56	11 (PET)	125 (K)	5	159-171	2	0.85 / 0.02	0.29	0.27
TguEST09-021	CK313022.1	F: TGCTGAGGTATGATCTCTACCAAC R: GACCTACAGCCTGAACACACC	Dawson *et al.* unpublished	Yes	9 / 9	23,079,052 / 21,494,997	56	11 (6-FAM)	126 (K)	2	106-111	1	0.96 / 0.04	0.09	0.08
Titgata02	AY792958.1	F: ATTGCTTGATATTTGAAAGCATA R: TTGTTGCAACTACAAGACATTTGA	Wang *et al.* 2005	No	2 / 2	108,368,370 / 105,859,386	TD 60-50[k]	-	66 (R)	21	213-306	(4)	0.21 / 0.01	0.92	0.88
Titgata68	AY792960.1	F: TTTGCATCCACAAGGGC R: CATCCAGTTACTTATATCTGAGAA	Wang *et al.* 2005	No	5 / 5	14,783,223 / 15,929,389	60	20 (PET)	124 (K)	11	228-289	4	0.32 / 0.004	0.82	0.79
Titgata79	AY792962.1	F: TCTTTGCTCCATAAAGGAAT R: ATAGAAACCAGAATCAGCAA	Wang *et al.* 2005	No	6 / 6	17,783,403 / 19,791,554	TD 60-50[k]	-	70 (R)	49	293-470	1	0.11 / 0.01	0.97	0.96
VeCr02	AY542875.1	F: AATAGGCTTTGAGGAGGAATCC R: AGCCCCAAAGTGCTGAAATA	Stenzler *et al.* 2004	No	12 / 12	10,572,128 / 10,434,859	55	14 (NED)	125 (K)	5	243-260	1	0.45 / 0.004	0.62	0.66
Z-037	DV945670.1	F: AAAACACCTTGTAATTTAAAACTGG R: CATAGATACATATCAATACAGCACATTC	Dawson 2007	Yes	Z / Z	54,801,985 / 37,359,037	55	10 (6-FAM)	71[m] (K)	2	160-167	2	0.85 / 0.15	0.21	0.25
Z-040	DV949035.1	F: AAAAGTCTTTCTGGACTGTGCT R: AAAATACAACAGACATAGGCATACA	Dawson 2007	Yes	Z / Z	52,307,759 / 39,991,371	55	10 (VIC)	70[m] (K)	6	129-143	2	0.39 / 0.02	0.81	0.75

Tgu, zebra finch (Taeniopygia guttata); Gga, chicken (Gallus gallus); EST, expressed sequence tag; Ta, annealing temperature; BT, blue tit (Cyanistes caeruleus); Ho, observed heterozygosity; *He*, expected heterozygosity.

[a]: Loci TG04-012, TG04-012A and TG07-022 from Dawson *et al.* (2009) were tested and showed polymorphism in four blue tit individuals. They are not included in the table due to the low number of individuals typed.

[b]: When different from the primer set name, the locus name is stated in parenthesis.

[c]: The marker is possibly linked to the Z chromosome in the blue tit. Therefore, the data presented in the table is from males only.
[d]: The sequence of this marker also displays homology to a zebra finch sequence with an E value < 1x10-8 in another position of the same chromosome. Second position is in parenthesis in the "Start" column.
[e]: The sequence of this marker also displays homology to the zebra finch "Unknown chromosome".
[f]: The sequence of the marker displays homology to a zebra finch sequence with an E value > 1×10-8.
[g]: PCR program: 95°C for 15 min; 18 cycles of 94°C for 30 sec, 53°C for 90 sec, and 72°C for 90 sec; 11 cycles of 94°C for 30 sec, 55°C for 90 sec, and 72°C for 90 sec; and 60°C for 30 min.
[h]: PCR program: 95°C for 15 min; 14 cycles of 94°C for 30 sec, 56°C for 90 sec, and 72°C for 90 sec; 11 cycles of 94°C for 30 sec, 57°C for 90 sec, and 72°C for 90 sec; and 60°C for 30 min.
[i]: PCR program: 95°C for 15 min; 15 cycles of 94°C for 30 sec, 60°C for 90 sec, and 72°C for 90 sec, decreasing the Ta 0.3°C per cycle; 11 cycles of 94°C for 30 sec, 53°C for 90 sec, and 72°C for 90 sec; and 60°C for 30 min.
[j]: Touchdown PCR program: 95°C for 15 min; 8 cycles of 94°C for 30 sec, 60°C for 90 sec, and 72°C for 90 sec, decreasing the Ta 1°C per cycle; 22 cycles of 94°C for 30 sec, 53°C for 90 sec, and 72°C for 90 sec; and 60°C for 30 min.
[k]: Touchdown PCR program: 95°C for 15 min; 11 cycles of 94°C for 30 sec, 60°C for 90 sec, and 72°C for 90 sec, decreasing the Ta 1°C per cycle; 25 cycles of 94°C for 30 sec, 50°C for 90 sec, and 72°C for 90 sec; and 60°C for 30 min.
[l]: The origin of the genotyped individuals is in parenthesis: (K) from Kolbeterberg (Austria) and (R) from Revinge (Sweden).
[m]: For Z-linked markers we present data from males only, the homogametic sex in birds.
[n]: Indicates the apparent motif length in the blue tit from the genotypic data, and corresponds to the bp difference between consecutive alleles at the locus. In parenthesis are "imperfect motifs", where the bp difference between all the consecutive alleles is not consistent.
[o]: Max. allele frequency: frequency of the most common allele of the locus in the blue tit population typed; Min. allele frequency: frequency of the least common allele of the locus in the blue tit population typed; Null allele frequency: Estimated null allele frequency, calculated only for the loci that deviate from Hardy-Weinberg equilibrium due to an homozygote excess.
[p]: Deviation from Hardy-Weinberg equilibrium could not be assessed due to extremely low variability (the second of the two alleles appeared just in one individual).
*: Significant deviation from Hardy-Weinberg equilibrium ($p<0.05$).

Zebra finch and chicken chromosome location assignments

The genome distribution of the loci comprises a total of 26 zebra finch chromosomes and 24 chicken chromosomes, including the Z chromosome (Table 1.2, Figure 1.2). Curiously, locus *CcaTgu5* was assigned to chromosome 2 in zebra finch and chicken based on sequence homology, but based on genotyping of individuals of known sex, appears to be Z-linked in the blue tit. Heterozygosity at locus *CcaTgu5* was observed only in males (Fisher's exact test $p < 0.01$), although there was a high estimated null allele frequency. Although the markers *PK12* and *Phtr3* were assigned a location on chromosome 5 and chromosome Z in the zebra finch with a weak E value ($> 1\times10^{-8}$), we present these positions because the chromosome assignment in the chicken (this study) and pedigree mapping in the blue tit (Hansson *et al.* 2010b), indicates that *PK12* is indeed located in chromosome 5, and *Phtr3* in chromosome Z. Moreover, locus *Phtr3* has been reported as Z-linked in several other passerine species (e.g. Fridolfsson *et al.* 1997; Forstmeier *et al.* 2002; Hansson *et al.* 2005). Seven markers were assigned to two different positions on the same zebra finch chromosome (Table 1.2). It is possible that these markers are duplicated in the zebra finch; however, since this is not observed in the chicken chromosomes, it is possible that the double matches are due to errors in the first draft assembly of the zebra finch. Six markers in the zebra finch, and two in the chicken, display homology to both a named chromosome and the chromosome Unknown (i.e. an assembly of sequences that have not yet been assigned to a named chromosome). Again, it is possible that the markers are duplicated in the genome or, as suggested by Dawson *et al.* (2007), that some sequences already assigned to named chromosomes were not removed from the Unknown chromosome when the genome sequence was compiled. Revised assemblies of the zebra finch and the chicken genomes would clarify this. Markers *Titgata02*, *Titgata68* and *Titgata79* show contamination with vector sequences in the flanking regions; however, removal of these sequences did not change the assigned chromosome location of the markers in the zebra finch or in the chicken genome.

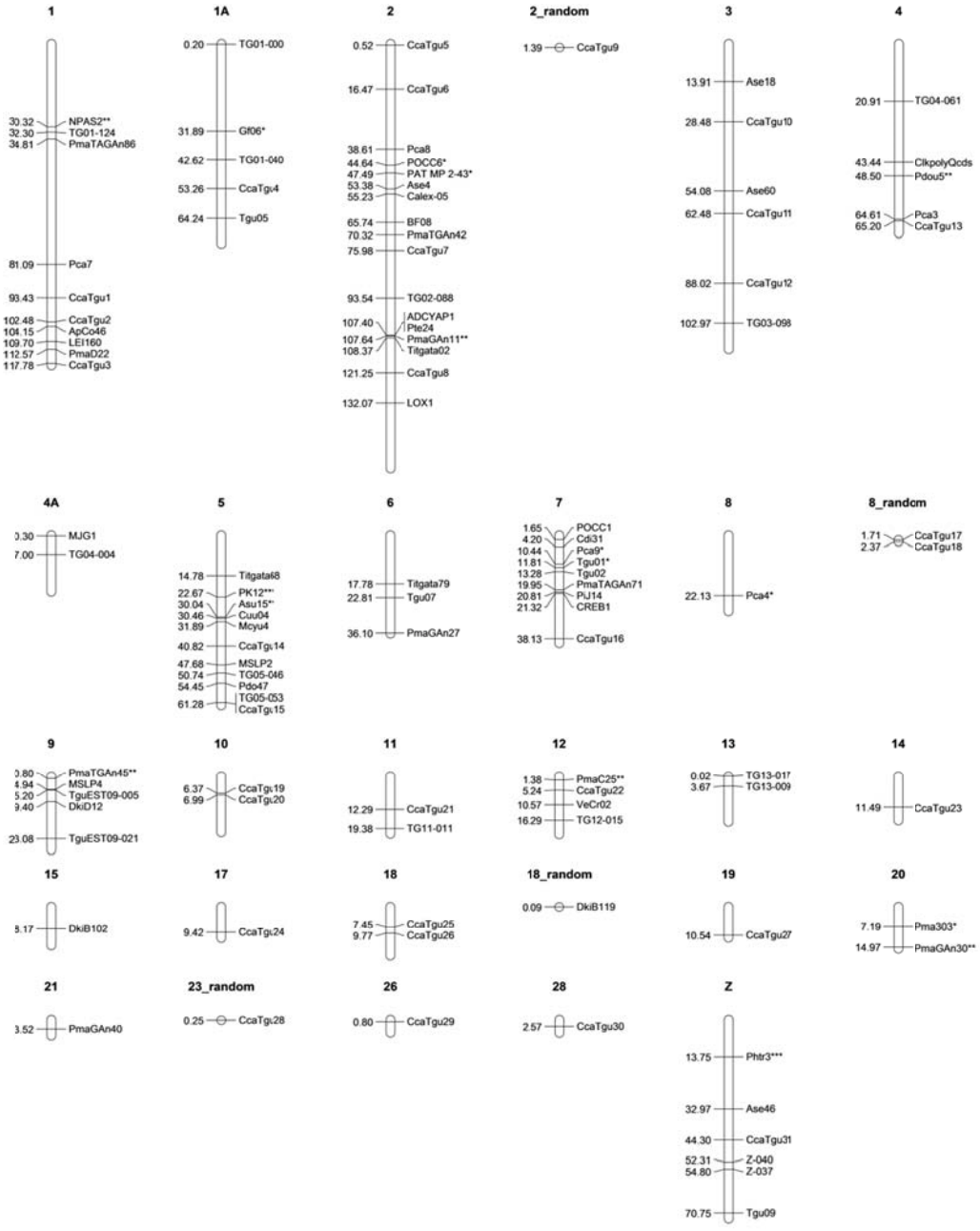

Figure1.2

Predicted zebra finch map of the 106 microsatellite markers found to be polymorphic in the blue tit. The Mbp position of the loci is indicated on the left side of their names. *The marker was also assigned to the Unknown chromosome. **The marker was also assigned to another position of the same chromosome. ***The marker's chromosome location was assigned with an E value $>1 \times 10^{-8}$.

Population genetics and linkage disequilibrium

The number of alleles per locus ranged from 2 to 49, with a mean of 9.2. The average expected heterozygosity is 0.56 ± 0.28 s.d. Ten loci deviated significantly from HWE ($p < 0.05$), eight of them due to heterozygote deficiency (Table 1.2). The estimated null allele frequencies for the latter eight loci varied between 4% and 35% (Table 1.2). The calculation of the population genetic parameters for locus *CcaTgu5*, which is autosomal in zebra finch and chicken but possibly *Z*-linked in blue tit, did not change substantially when genotypes of males only were considered.

Twenty two of 173 pairs of markers located in the same zebra finch chromosome showed significant linkage disequilibrium (LD) correlation coefficients in the blue tit population from Kolbeterberg (data not shown). This number is higher than would be expected by chance due to multiple comparisons (8.7 cases at $\alpha = 0.05$). After correction for multiple testing, three pairs of loci are still in significant LD: *Gf06 (gf2.40)* and *CcaTgu4* on chromosome 1A, *CcaTgu15* and *TG05-053* on chromosome 5, and *PmaTGAn45* and *DkiD12ZF Chr9* on chromosome 9. For all pairs of loci, *r*LD ranges between 0.03 and 0.19; for the three pairs of loci with significant LD after correction for multiple testing, *r*LD ranges between 0.07 and 0.19.

Discussion

The use of EST data as a source of microsatellite loci in avian species was originally utilized to develop Galliform microsatellites (Ruyter-Spira *et al.* 1998; Dranchak *et al.* 2003; Mannen *et al.* 2005). Primer sets designed from zebra finch EST-SSR sequence have been found to display higher cross-utility than anonymous microsatellite loci (Karaiskou *et al.* 2008). Avian microsatellite cross-utility has been shown to be the highest when the most conserved EST microsatellite loci are identified and zebra finch-chicken 100% consensus primer sets are developed (Dawson *et al.* 2010). Dawson and colleagues assessed 34 loci in 17 passerine species and found 100% amplification and a mean of 48% of the amplifying loci polymorphic. In this study, we found that zebra finch EST-SSR sequences are useful as molecular markers for the blue tit with a similar picture: amplification was successful with 75 out of 101 (74%) primer pairs designed, and

36 (48%) of the amplifying loci were polymorphic (details not shown). Therefore, for groups of organisms with preexisting EST libraries, the primer design strategy presented here is efficient to find polymorphic microsatellite markers (see also Kantety *et al.* 2002; Perez *et al.* 2005; Bouck and Vision 2007; Tang *et al.* 2008; Dawson *et al.* 2010). Another advantage of using EST-SSRs is the higher potential applicability of the markers across different species. First, as the EST-SSR are presumed to be under a high selective constraint due to their exonic location, it is expected that the primer binding region would diverge at a slower rate than that of intergenic markers (Bouck and Vision 2007; Kim *et al.* 2008). Second, as the primers for this study were designed from sequence data of different bird species (zebra finch, chicken, and others), but amplified and showed polymorphism in the blue tit, it is likely that they will amplify and show polymorphism in a broad range of bird species.

Microsatellite markers previously designed for other bird species, and with a high degree of cross-species applicability in other passerines (D.A. Dawson unpublished data), were also a good resource to find polymorphic loci in the blue tit. It is important to note that the primers for many of these markers were modified, which successfully increased their utility in a wide range of passerine species (Dawson 2007; see also Klein *et al.* 2009; Lee *et al.* 2009; Martin-Galvez *et al.* 2009; Simeoni *et al.* 2009; Dawson *et al.* 2010; D.A. Dawson, unpublished data; G.N. Hinten, unpublished data). The 43 markers used here were originally isolated in 22 species, 20 of them being passerines. However, just three of these species (comprising 15 markers) belong to the same family as blue tit (Paridae). Loci that show cross-family and even cross-order amplification and polymorphism success in the blue tit, are strong candidates for amplification/polymorphism success in other passerine birds (Primmer *et al.* 2005).

Cytogenetic analyses and comparative gene mapping between chicken and other bird species have revealed a high degree of conservation in the genome organization of the avian lineage (Shields 1982; Derjusheva *et al.* 2004; Dawson *et al.* 2006; Dawson *et al.* 2007; Griffin *et al.* 2007; Backstrom *et al.* 2008; Hale *et al.* 2008; Stapley *et al.* 2008; Hansson *et al.* 2010b). Some exceptions have been noted, such as the atypical karyotype observed in Falconiformes (Bed'Hom *et al.* 2003; de Oliveira *et al.* 2005). However, the distribution of genetic markers in the zebra finch and in the chicken genome is a good approximation for their chromosome location in the blue tit and in other passerine species. Although we find support

for conserved synteny between the zebra finch and the chicken, the marker order within a chromosome is moderately conserved. This is further supported by comparative gene mapping studies between the chicken and the great reed warbler (Hansson *et al.* 2005; Dawson *et al.* 2007), the zebra finch (Hale *et al.* 2008; Stapley *et al.* 2008), the collared flycatcher (*Ficedula albicollis*, Backstrom *et al.* 2008), and the blue tit (Hansson *et al.* 2010b). Thus, it appears that the intra-chromosomal rearrangements are slightly different between different passerine species, which means that the chromosomal position of the markers in the blue tit cannot be directly inferred from their position in the zebra finch. Physical mapping of the markers in the chromosomes of the species of interest cannot be replaced by location assignment *in silico* on the genome of related species, as inter and intrachromosomal rearrangements cannot be detected. This is potentially illustrated by the possible Z-linkage of marker *CcaTgu5* in the blue tit which was assigned to chromosome 2 based on sequence homology.

The set of markers presented here is suitable for population genetic studies in the blue tit. First, the number of alleles and the level of polymorphism of the loci, although variable, is high on average. Second, with the exception of 8 loci that deviate from HWE and have an estimated frequency of null alleles between 4-35%, the markers do not show any evidence of amplification problems or scoring errors. The two markers that deviate from HWE due to a heterozygote excess are interesting to investigate in the context of balancing selection or amplification errors. Third, just three pairs of loci located on the same chromosome appear to be in linkage disequilibrium. From these pairs, only *CcaTgu15* and *TG05-053* are closely located on chromosome 5 in the zebra finch. The other two pairs (*Gf06 (gf2.40)* and *CcaTgu4*, and *PmaTGAn45* and *DkiD12ZF Chr*) do not appear to be closely located in the zebra finch or chicken chromosomes; however, their genomic location in the blue tit may be different to the locations assigned based on sequence homology. In studies where it is important to assure independence of the loci, one of the markers in each of these pairs might be discarded. After accounting for these considerations, 95 markers with a high potential to be used in studies of blue tit population, ecological and evolutionary genetics remain. The high number of loci plus the large proportion of the genome represented, make this marker set a valuable genomic resource for the blue tit and other passerine birds.

Chapter 2

Correlations between heterozygosity and reproductive success in the blue tit (*Cyanistes caeruleus*): an analysis of inbreeding and single locus effects

Juanita Olano-Marin, Jakob C Mueller, Bart Kempenaers

Abstract

In order to understand the mechanisms behind heterozygosity-fitness correlations (HFC), it is necessary to employ large numbers of markers with known function and to independently estimate the variation in inbreeding in the population. Here we genotyped 794 blue tits with 79 microsatellites that were distributed across 25 chromosomes and that were classified either as "functional" (N=58) or "neutral" (N=21). We found a positive effect of individual heterozygosity at multiple loci on clutch size, on the number of eggs sired by males, and on the number of recruits produced by males and females. We documented the occurrence of some consanguineous matings and found evidence for a cryptic genetic population structure that can contribute to the occurrence of inbreeding. As the set of "neutral" loci provided more power to detect HFC and identity disequilibrium, we argue that "neutral" markers are better predictors of the effects of inbreeding. The number of significant effects at single loci did not exceed the expected number of false positives and no strong effects were associated with heterozygosity at "functional" markers. Thus, HFC found here can be attributed to small effects of many loci, rather than to strong effects of few loci.

Manuscript in press in *Evolution*

The relationship between genetic diversity and fitness has been recognized as an important issue in different areas of evolutionary biology, as it allows us to identify the forces that maintain genetic diversity in natural populations (Charlesworth and Charlesworth 1987; Ellegren and Sheldon 2008). The importance of understanding the relationship between genetic diversity and fitness has also been acknowledged in a conservation biology context, as it allows us, for example, to evaluate the evolutionary potential of the populations and to predict the consequences of a reduction in their genetic diversity (Hedrick and Kalinowski 2000; Reed and Frankham 2003).

Heterozygosity-fitness correlations (HFC) have been used to study the relationship between genetic diversity and fitness at the individual level in a variety of organisms (reviewed in Britten 1996; David 1998; Coltman and Slate 2003; Chapman *et al.* 2009). Most HFC studies in animal populations report a linear, positive relationship between measures of individual heterozygosity and fitness-related traits (i.e. the higher the heterozygosity the better, for a review see Chapman *et al.* 2009). Although less common, negative effects of heterozygosity are also reported (reviewed in Chapman *et al.* 2009), and they can reflect quadratic relationships between heterozygosity and fitness (i.e. intermediate levels of heterozygosity are better than the two extremes, e.g. Marshall and Spalton 2000; Neff 2004). Small, non-significant and negative effects are underrepresented in the HFC literature (Coltman and Slate 2003; Chapman *et al.* 2009). Fortunately, this publication bias has lessened in the last years and does not affect considerably the conclusions of meta-analysis about the existence of HFC (Chapman *et al.* 2009). Historical changes in the approaches and techniques used to study HFC and considerable differences in the population history of the species studied have opened a long-standing debate about the causes of HFC (Szulkin *et al.* 2010). Two main explanations for the occurrence of HFC are commonly put forward: inbreeding effects across the whole genome (referred to as the general effect hypothesis), or localized effects at single loci (referred to as the local and direct effect hypotheses) (David 1998; Hansson and Westerberg 2002). However, the interpretation of significant effects of heterozygosity on fitness-related traits has been hindered by methodological constraints, such as difficulties in estimating the amount of variation in inbreeding in natural populations, the generally low number of genetic markers used to measure heterozygosity, and the lack of knowledge about the genomic location and the functionality of the loci employed.

Inbreeding can have a considerable negative impact on the fitness of inbred individuals relative to outbred ones, a phenomenon commonly known as inbreeding depression (Crnokrak and Roff 1999; Keller and Waller 2002). By causing a reduction in individual heterozygosity across the genome, inbreeding increases the risk of expressing recessive deleterious alleles and decreases the occurrence of beneficial overdominant effects (Charlesworth and Willis 2009). Inbreeding arises via two different, not mutually exclusive, processes occurring within and among populations (Keller and Waller 2002). Within-population inbreeding occurs when related individuals mate. Detailed pedigree data can be used to estimate the individual inbreeding coefficient, *F* (Wright 1922). Inbreeding depression is shown by the lower performance of individuals with higher *F*. Among-population inbreeding, on the other hand, occurs when populations are subdivided into small and isolated groups where genetic drift occurs and deleterious alleles become fixed. In this case, inbreeding occurs even with random mating within the subpopulations and can be detected by estimating the amount of population differentiation with pedigree (Wright 1965) or genetic data (Pritchard *et al.* 2000; Balloux and Lugon-Moulin 2002). Inbreeding depression due to among-population inbreeding can only be detected through crosses between members of different subpopulations. The individuals produced from such crosses would mask the effects of the different subsets of deleterious alleles fixed in the different subpopulations and, therefore, would show a higher fitness than "pure" individuals (Keller and Waller 2002). Despite the importance of population differentiation as a cause of inbreeding and inbreeding depression, even in the absence of matings between relatives, its occurrence has rarely been considered in HFC studies (for some exceptions see von Hardenberg *et al.* 2007; Garcia-Navas *et al.* 2009; Hoffman *et al.* 2010b).

Obtaining a direct and accurate measure of the individual level of inbreeding caused by within- and among-population processes can be challenging, particularly for free-living, open populations. Therefore, other, more indirect measures have been used to estimate it. Multilocus heterozygosity (MLH) and related indexes of heterozygosity at multiple loci (e.g. SH: Coltman *et al.* 1999; IR: Amos *et al.* 2001; HL: Aparicio *et al.* 2006) are frequently used as correlates of the individual inbreeding coefficient. Thus, positive correlations between these indexes and fitness have been interpreted as evidence of inbreeding depression (e.g. Slate and Pemberton 2002; Hoffman *et al.* 2004; Lesbarrères *et al.* 2005; Charpentier *et al.* 2008). However, for measures of multilocus heterozygosity to

reflect genome-wide heterozygosity and therefore inbreeding, specific population processes (e.g. frequent matings between close relatives and/or population subdivision combined with admixture or immigration), causing a high variance in individual inbreeding levels must occur (Balloux *et al.* 2004; Szulkin *et al.* 2010). In addition, a large number of marker loci should be used (Balloux *et al.* 2004; Slate *et al.* 2004; DeWoody and DeWoody 2005). In most HFC studies, however, the population processes that contribute to inbreeding are generally unknown and heterozygosity is measured with a small number of markers (median number of markers from the studies considered in Chapman *et al.* 2009 ~ 10). Therefore, in many cases, positive HFC cannot be clearly attributed to the effects of inbreeding across the whole genome.

As the correlation between measures of heterozygosity at multiple loci and genome-wide heterozygosity and inbreeding has been challenged, there is an increasing tendency to explain HFC by localized (direct or indirect) effects at single loci (e.g. Lieutenant-Gosselin and Bernatchez 2006; Luikart *et al.* 2008; Da Silva *et al.* 2009; Hoffman *et al.* 2010b). Direct effects occur at functional loci, whereas indirect effects (commonly known as local effects in the HFC literature) are detected with neutral markers linked to functional loci. As with the general effect hypothesis, these explanations for HFC are not free of controversy. The detection of local effects with anonymous, randomly distributed neutral markers in populations with no inbreeding would imply that the levels of linkage disequilibrium in the genome are very high (Balloux *et al.* 2004; Szulkin *et al.* 2010), that a high number of loci show overall heterozygote advantage (Mueller *et al.* 2010), and/or that the effects of linked loci are very strong (Szulkin *et al.* 2010). In addition, tests for local (and direct) effects suffer from important statistical caveats (Chapman *et al.* 2009; Szulkin *et al.* 2010). More importantly, biological interpretations of statistically significant effects at single loci require some knowledge about the function and/or location of the marker loci in the genome; however, for the microsatellite loci employed in a great number of HFC studies, this knowledge is lacking (e.g. Acevedo-Whitehouse *et al.* 2006; Da Silva *et al.* 2006; Blanchet *et al.* 2009).

Microsatellites, probably the most common genetic markers for measuring heterozygosity in HFC studies, are usually considered evolutionarily neutral (Jarne and Lagoda 1996). However, evidence is accumulating about the functionality of repeats located in expressed regions of the genome and within genes (reviewed in Li *et al.* 2004). This functionality can have implications for the

detection and interpretation of HFC. General effects of inbreeding are more likely to be detected with loci that are not functionally constrained, whereas localized effects are caused either directly by functional loci or by neutral markers closely linked to functional loci (including local effects through associative overdominance, Ohta and Kimura 1969; Ohta and Kimura 1970). The possibility of anchoring microsatellite loci in the genome of model species can provide clues about their functionality. Microsatellite loci located within a coding gene would be expected to be more evolutionarily constrained than markers located in intergenic regions. Similarly, microsatellites located in expressed regions of the genome (e.g. markers developed from libraries of Expressed Sequence Tags or ESTs) can be used as a source of potentially functional loci (Vasemagi *et al.* 2005; Oliveira *et al.* 2009).

Significant HFC have been found in different populations of blue tits, *Cyanistes caeruleus* (Foerster *et al.* 2003; Garcia-Navas *et al.* 2009). In a previous study on an Austrian population, Foerster *et al.* (2003) found significant correlations between heterozygosity and several fitness-related traits. Since Foerster *et al.* (2003) only used 5-7 anonymous microsatellite loci, they could not clearly establish the causes behind the HFC. In an extended dataset from the same population, we investigated correlations between individual heterozygosity and estimates of annual reproductive success, and explored its causes. For this purpose, we genotyped all the breeding birds with a panel of 79 microsatellites distributed across 25 chromosomes in a model passerine (the zebra finch *Taeniopygia guttata*). We classified the markers as potentially functional or neutral and identified the (closest) genes where the microsatellites were located. Our specific objectives were: 1) to test for the overall effect of heterozygosity at multiple loci (measured as HL and SH) and of heterozygosity measured with presumably functional or neutral loci on two direct measures of reproductive output and on three traits related to offspring success of male and female blue tits; 2) to investigate within- and among-population sources of inbreeding as causes of HFC and the potential of multilocus heterozygosity to reflect the individual level of inbreeding; 3) to investigate the effect of heterozygosity at single loci on the five reproductive success traits and test for differences in the effects of single "functional" and "neutral" markers.

Material and Methods

Study system and general procedures

A population of blue tits breeding in nestboxes at Kolbeterberg (Vienna, Austria, 48°13'N, 16°20'E) was studied between 1998 and 2006. Nestboxes were installed at a uniform density of about 4/ha across 42 - 50 ha of mixed deciduous woodland. Birds that were roosting in the nestbox (December – March) or feeding nestlings (May) were caught, banded with a metal band and a unique combination of three color bands, bled for DNA extraction, and aged (as one year or older) according to plumage characteristics (Svensson 1992). During the breeding season, starting before nest building and continuing until the chicks fledged, the social pairing and the breeding success of the population was monitored and recorded. All unhatched eggs and dead nestlings were collected and the surviving nestlings were bled before fledging, in order to extract DNA for parentage analyses. Parentage was assigned with the methodology described in Foerster *et al.* (2003) and Delhey *et al.* (2003), using a combination of five to eight highly polymorphic microsatellite markers. An analysis of paternity with a group of c.a. 1000 chicks genotyped with the 79 loci used in this study confirmed the previous parentage assignments in 100% of the cases. The sex of the adults and the chicks was assigned by amplifying the loci CHD with the primers P2-P8 (Griffiths *et al.* 1998).

Reproductive success components

We estimated the reproductive success of the blue tits with different measures recorded during the breeding season of each study year. We used clutch size as a measure of female fecundity. The total number of sired eggs (number of offspring sired in the social female nest + number of extra-pair offspring, i.e. sired in nests of females other than the social mate) was used as an integral measure of male reproductive success. Hatching, fledging and recruitment success of a nest (measured as the proportion of hatchlings, fledglings or recruits over the clutch size, number of hatchlings or number of fledglings, respectively) were used as estimates of the reproductive performance of both parents. For these analyses, we first only used data from nests without extra-pair young (EPY), as these traits can also be influenced by the genetic quality of the extra-pair male. Second, to compare our results with the ones reported by Foerster *et al.* (2003), we repeated the analysis of fledging success including nests with EPY. For all analyses we

excluded data from re-nesting attempts and from nestboxes that were used for field experiments (mostly in years 2005 and 2006). For the hatching and fledging success analyses we excluded data from nestboxes where none of the eggs hatched or none of the chicks fledged, as complete brood failure is more likely due to predation or other stochastic events.

Microsatellite markers and classification of loci

We genotyped the adult blue tit breeding population (*N*=794 individuals, 385 males and 411 females) with a panel of 80 autosomal microsatellites (Table 1.3). Details on markers' characteristics, chromosome location in the zebra finch (Warren *et al.* 2010) and the chicken (International Chicken Genome Sequencing Consortium 2004) genomes, and amplification and genotyping in the blue tit can be found in Olano-Marin *et al.* (2010). We tested for deviations from Hardy-Weinberg equilibrium at each locus, as well as for linkage disequilibrium (LD) between all pairs of syntenic loci (defined by homology to the zebra finch chromosomes), using the genotype data from the adult population of each study year as described in Olano-Marin *et al.* (2010). We corrected *p*-values for false discovery rate (FDR, Benjamini and Hochberg 1995), with the number of tests made per year of study.

We classified the markers as potentially functional or neutral by considering whether the genomic region where the markers are located is transcribed to RNA and therefore expressed. Thus, loci that were designed based on or showed homology to zebra finch expressed sequence tags (ESTs) were considered as and will be referred to as functional, whereas markers designed using traditional cloning methods and with no homology to avian ESTs were considered as and will be referred to as neutral (Table 1.3, details in Olano-Marin *et al.* 2010). To test some predictions about the effect of heterozygosity at single loci, we also considered if the markers were located within a coding vertebrate gene. For this purpose, we performed a BLAT search of each sequence containing a microsatellite against the zebra finch (Warren *et al.* 2010) genome at the UCSC browser (http://genome.ucsc.edu/cgi-bin/hgGateway) (details in Olano-Marin *et al.* 2010). The "RefSeq Genes" and the "Other RefSeq" options in the UCSC results window were fully displayed with the purpose of identifying zebra finch coding genes or other vertebrate exons with homology to the region where the microsatellite markers were located. Following this procedure, for the loci that were not located within a reference zebra finch gene or an exon from another

vertebrate species, we determined the approximate distance to the nearest region with an annotated or an aligned vertebrate gene (Table 1.3).

Table 1.3.
Microsatellite markers used to estimate individual heterozygosity in a population of breeding blue tits. The loci are ordered according to their position in the zebra finch genome. Loci with significant effects of heterozygosity at single loci (HSL) on at least one estimate of reproductive success are in bold. References for the primer sets for each locus can be found in Olano-Marin *et al.* (2010)

Locus	Location in the ZF genome (Chromosome / Start)	Category	Gene / nearest gene (Kb distance)	Number of significant effects of HSL (fitness trait[o]: effect direction[p])
NPAS2	1 / 30,322,977	Functional	NPAS2	
TG01-124	1 / 32,303,776	Functional	GABRB3 (0.01)	
Pca7[q]	1 / 81,091,062	Neutral	FAT3 (145.7)	1 (FSm: +)
CcaTgu1	1 / 93,427,832	Functional	IGSF11 (0.5)	
CcaTgu2	1 / 102,481,135	Functional	CD247 (1.2)	
ApCo46	1 / 104,151,755	Functional	CADM2 (4.3)	
LEI160[q]	1 / 109,699,352	Neutral	MIR125B (2.6)	
CcaTgu3	1 / 117,775,285	Functional	PCNP	1 (RSm: -)
TG01-000[l]	1A / 201,308	Functional	UBE2H	
Gf06[n, q]	1A / 31,887,783	Neutral	USP15 (343.8)	
TG01-040[m]	1A / 42,620,504	Functional	DUSP6	
CcaTgu4[m, n, q]	1A / 53,257,466	Functional	BTBD11 (10.7)	1 (RSf: +)
Tgu05	1A / 64,236,986	Functional	MPPED1 (0.03)	
CcaTgu6	2 / 16,466,870	Functional	KIAA1462 (5.1)	
Pocc6[q]	2 / 44,644,858	Neutral	LY86 (38.7)	
PAT MP 2-43[q]	2 / 47,487,652	Neutral	TMEM195 (78.2)	
CcaTgu7	2 / 75,982,449	Functional	TPPP (7.8)	1 (CS: +)
TG02-088	2 / 93,538,047	Functional	GABBR2 (2.6)	
ADCYAP1[a, q]	2 / 107,400,835	Functional	ADCYAP1	1 (RSm: -)
CcaTgu8[q]	2 / 121,250,236	Functional	CRH	
CcaTgu9[a]	2_random / 1,385,550	Functional	POP1 (60.0)	1 (FSf: -)
Ase18[q]	3 / 13,906,080	Neutral	SERTAD4 (5.6)	
CcaTgu10	3 / 28,478,877	Functional	RTN4	
CcaTgu11[q]	3 / 62,483,485	Functional	NKAIN2 (6.4)	
CcaTgu12	3 / 88,019,185	Functional	PTP4A1	1 (RSf: +)
TG03-098	3 / 102,966,495	Functional	VSNL1	
ClkpolyQcds	4 / 43,440,144	Functional	CLOCK	1 (SE: -)
Pca3	4 / 64,610,435	Neutral	MAEA (5.4)	
CcaTgu13	4 / 65,195,621	Functional	FGFRL1	
TG04-004	4A / 6,997,361	Functional	MMGT1	
Titgata68[b]	5 / 14,783,223	Neutral	MUC2 (79.9)	1 (HSm: +)
PK12[c, j]	5 / 22,671,767	Neutral	CRY2	
Asμ15	5 / 30,043,672	Functional	FMN1 (6.0)	
Mcyμ4[l, b]	5 / 31,893,652	Neutral	GJD2 (1.1)	1 (HSm: +)
CcaTgu14[c, d, q]	5 / 40,815,020	Functional	NRXN3	
TG05-046	5 / 50,735,925	Functional	PPP2R5C	1 (RSm: +)
TG05-053[j, l, q]	5 / 61,275,929	Functional	MDGA2	1 (SE: -)
CcaTgu15[l, d, l, q]	5 / 61,276,973	Functional	MDGA2 (0.1)	

Table 1.3 cont.

Locus	Location in the ZF genome (Chromosome / Start)	Category	Gene / nearest gene (Kb distance)	Number of significant effects of HSL (fitness trait[o]: effect direction[p])
Tgu07[q]	6 / 22,807,094	Functional	TRIM8 (0.3)	1 (CS: +)
PmaGAn27	6 / 36,100,371	Neutral	INPP5A	
Pocc1[1, e]	7 / 1,646,806	Neutral	HDAC4	2 (CS: +, RSf: -)
Cdi31	7 / 4,201,012	Neutral	KALRN	2 (FSm: -, RSm: +)
Pca9[f]	7 / 10,442,876	Neutral	IHH (11.8)	3 (SE: +, FSf: +, RSf: -)
Tgu01	7 / 11,806,628	Functional	SLC4A10	1 (RSm: +)
Tgu02[g]	7 / 13,280,326	Functional	GALNT3 (4.5)	2 (CS: -, SE: +)
PiJ14[e, f, g, q]	7 / 20,810,935	Functional	INO80D (0.4)	
CREB1	7 / 21,321,109	Functional	CREB1	1 (RSf: +)
CcaTgu16	7 / 38,131,923	Functional	RND3	
Pca4[1, q]	8 / 22,128,537	Neutral	PODN (4.7)	1 (RSf: +)
CcaTgu17[1]	8_random / 1,710,707	Functional	VAMP4 (1.2)	
CcaTgu18	8_random / 2,367,489	Functional	PBX1 (3.4)	1 (CS: -)
PmaTGAn45[1, h]	9 / 802,719	Neutral	FOXL2 (43.4)	
MSLP4[h]	9 / 4,935,546	Functional	ILKAP (1.3)	
TguEST09-005	9 / 5,196,903	Functional	IGF2BP2 (0.5)	
DkiD12[8]	9 / 9,399,880	Neutral	SERPINE2 (21.0)	-
TguEST09-021	9 / 23,079,052	Functional	MYNN (2.4)	1 (FSm: -)
CcaTgu19	10 / 6,370,604	Functional	FAM81A	2 (FSm: +, RSm:-)
CcaTgu20	10 / 6,994,363	Functional	CGNL1	
CcaTgu21[k]	11 / 12,285,425	Functional	ZFHX3	
TG11-011[k]	11 / 19,380,799	Functional	KIAA0355 (0.4)	
PmaC25[q]	12 / 1,378,621	Neutral	IQSEC (154.2)	
CcaTgu22	12 / 5,239,036	Functional	CENPP	
VeCr02[q]	12 / 10,572,128	Neutral	PLXNA (189.3)	1 (FSm: +)
TG12-015	12 / 16,288,963	Functional	FAM19A1	
TG13-017[1, q]	13 / 18,208	Functional	EGR1	2 (SE: +, FSf: +)
TG13-009	13 / 3,672,471	Functional	RNF44	
CcaTgu23[1]	14 / 11,485,613	Functional	TNRC18	
DkiB102	15 / 8,166,844	Functional	HIRA	
CcaTgu24	17 / 9,423,028	Functional	CAMSAP1 (1.1)	
CcaTgu25[2, i, q]	18 / 7,447,751	Functional	TNRC6C	
CcaTgu26[i]	18 / 9,768,597	Functional	CA10 (0.07)	
DkiB119	18_random / 93,096	Functional	MAP2K6 (0.2)	
CcaTgu27	19 / 10,535,947	Functional	MSI2	
Pma303[q]	20 / 7,188,252	Neutral	TGM3 (8.4)	
PmaGAn30	20 / 14,965,400	Functional	TMEM189	1 (RSf: +)
PmaGAn40[q]	21 / 3,523,674	Neutral	SKI (43.2)	
CcaTgu28[q]	23_random / 246,698	Functional	PTP4A2	
CcaTgu29	26 / 802,761	Functional	CSDE1	1 (HSm: +)
CcaTgu30	28 / 2,572,470	Functional	TMEM38A	
PK11[q]	- / -	Neutral	-	1 (CS: +)

[1, 2, 8]: The superscript indicates the number of study years where the loci showed significant deviations from Hardy-Weinberg equilibrium. The locus with significant deviations in eight study years (DkiD12) was discarded for further HFC analyses.

[a-n]: Pairs of loci with the same letter showed significant linkage disequilibrium (LD) in at least one study year ([a-i]: LD in one year, [j-k]: LD in two years, [l]: LD in three years, [m]: LD in six years, [n]: LD in nine years)

[o]: CS: Clutch size, SE: Number of sired eggs, HSm: Hatching success (male effect), FSf / FSm: Fledging success (female / male effect), RSf / RSm: Recruitment success (female / male effect).

[p]: +: positive effect, -: negative effect

[q]: loci with this superscript were selected for a subset of functional and neutral loci with similar levels of diversity (see Methods for details)

Heterozygosity measures

A total of 794 individuals (99.7%, 384 males and 410 females) were successfully scored with 75% - 100% of the markers that were used for further analyses. We calculated standardized heterozygosity (SH) (Coltman *et al.* 1999) and homozygosity by locus (HL) (Aparicio *et al.* 2006) as measures of individual heterozygosity at multiple loci, using the R function GENHET (Coulon 2010). SH is a simple measure that is insensitive to the fact that not all individuals were genotyped at all loci (Coltman *et al.* 1999). This index also gives the same weight to all the markers, independent of the number of alleles and the frequencies of these alleles. HL is a more complex estimator that gives a higher weight to more informative loci (e.g. loci with more alleles which are more evenly distributed). In simulated populations subjected to migration and admixture, HL correlates better with the inbreeding coefficient and with genome-wide heterozygosity than other heterozygosity indexes, reducing the sample size required to detect HFC due to inbreeding (Aparicio *et al.* 2006). We calculated these two heterozygosity indexes for all the individuals using the genotype data from all markers, and separately for each of the subgroups of functional and neutral loci. Heterozygosity at single loci (HSL) was coded as a binary variable, with "1" representing a heterozygous state and "0" a homozygous one.

Effects of heterozygosity at multiple and single loci on reproductive success

To analyze the effect of individual heterozygosity on estimates of reproductive success we used mixed effects models (Pinheiro and Bates 2000) in the free software R (R Development Core Team 2008). These models account for non-independence of observations due to individuals breeding in the study area across

several years. All the models included the reproductive success measure as the response variable, heterozygosity (SH, HL or HSL) and its squared term (for SH and HL only) as predictors, individual age as non-genetic cofactor, and year and individual identity as random effects. As quadratic effects of heterozygosity were non-significant for all reproductive success traits (all $p > 0.05$), we excluded this term from all the analyses. Similarly, non-significant age effects (all $p > 0.1$) were excluded from the analyses of hatching, fledging and recruitment success. We ran separate models for multilocus heterozygosity measured with functional and neutral markers, and then compared the goodness of fit of the models using differences in the Akaike's Information Criterion (AIC) (Burnham and Anderson 2002). We used general linear models to analyze clutch size (female trait) and the number of sired eggs (male trait) with the add-on R package *nlme* (Pinheiro *et al.* 2009), fitting individual identity within year as random intercepts. In addition, we estimated the effect of heterozygosity of males and females on the success of the broods they produced and raised. Consequently, for hatching, fledging and recruitment success we fitted heterozygosity of both parents as fixed effects in the same model. We used clutch size, number of hatchlings and number of fledglings, respectively, as the binomial denominator in generalized linear models with binomial error structure and logit link function with the package *lme4* (Bates and Maechler 2010). We calculated confidence intervals for the model estimates of each reproductive success measure with the add-on R package *multcomp* (Hothorn *et al.* 2008). We standardized the effect size of heterozygosity on each reproductive success estimate by converting it to *r*, the equivalent of the Pearson correlation coefficient. For this purpose, we used the *t*-statistic from the linear models or the *z*-value from the generalized linear models, as described in Coltman and Slate (2003). We tested if the differences in the HFC obtained with neutral and functional markers could be attributed to the difference in the degree of diversity of the markers in the two categories. For this purpose, we chose a subset of 12 neutral and 12 functional markers (that will be distinguished from the original categorization of markers with the prime (') symbol) that were matched for their number of alleles (*A*) and expected heterozygosity (*He*) (*t*-tests for the difference in *A* and *He* between neutral' and functional' loci: both *p*-values > 0.95). We repeated the analysis of the effect of heterozygosity at multiple loci (measured as HL) on clutch size and on the number of sired eggs with the subset of functional' and neutral' markers.

Table 2.3.
Effect of individual heterozygosity at multiple loci (measured as HL) on measures of reproductive success in a blue tit population. Clutch size and number of sired eggs: general mixed effects model with normal error structure and identity link function. Hatching, fledging and recruitment success: generalized mixed effects model with binomial error structure and logit link function. All the models included season and individual ID as random effects, and HL, HL^2 and age as fixed effects. Non-significant effects of HL^2 and age were removed. SE denotes standard error. Significant effects of HL and the model(s) with the lowest AIC value for each trait are shown in bold. Coefficients of determination (r^2) were calculated for terms with significant effects.

Reproductive success trait	Model AIC	Factor	Estimate ± SE	Test statistic	*p*-value	r^2
Clutch size[a] (*N* = 408 nests of 297 females)	**1675.0**	Intercept	12.842 ± 0.567	t = 22.630	0.000	-
		HL_{all}	**-4.835 ± 1.654**	**t = -2.924**	**0.004**	**0.021**
		Age	0.528 ± 0.185	t = 2.853	0.005	0.020
	1676.0	Intercept	11.934 ± 0.296	t = 40.316	0.000	-
		$HL_{neutral}$	**-3.069 ± 1.055**	**t = -2.910**	**0.004**	**0.021**
		Age	0.505 ± 0.185	t = 2.729	0.007	0.018
	1681.8	Intercept	12.051 ± 0.557	t = 21.647	0.000	-
		$HL_{functional}$	-1.958 ± 1.329	t = -1.473	0.141	-
		Age	0.540 ± 0.186	t = 2.895	0.004	0.021
Number of sired eggs[b] (*N* = 364 observations of 265 males)	1941.3	Intercept	11.745 ± 1.105	t = 10.630	0.000	-
		HL_{all}	-5.380 ± 3.430	t = -1.568	0.118	-
		Age	1.545 ± 0.366	t = 4.222	0.000	0.048
	1936.0	Intercept	11.311 ± 0.487	t = 23.243	0.000	-
		$HL_{neutral}$	**-6.289 ± 2.124**	**t = -2.960**	**0.003**	**0.024**
		Age	1.523 ± 0.363	t = 4.197	0.000	0.047
	1944.0	Intercept	9.732 ± 1.101	t = 8.837	0.000	-
		$HL_{functional}$	0.829 ± 2.776	t = 0.299	0.765	-
		Age	1.554 ± 0.367	t = 4.232	0.000	0.048

Table 2.3 cont.

Reproductive success trait	Model AIC	Factor	Estimate ± SE	Test statistic	*p*-value	r^2
Hatching success[c] (*N* = 148 nests of 128 females / 126 males)	192.3	Intercept	2.332 ± 0.935	z = 2.494	0.013	-
		Male HL_{all}	1.801 ± 2.093	z = 0.860	0.390	-
		Female HL_{all}	-0.335 ± 2.151	z = -0.156	0.876	-
	192.0	Intercept	2.462 ± 0.370	z = 6.659	0.000	-
		Male $HL_{neutral}$	1.238 ± 1.316	z = 0.941	0.347	-
		Female $HL_{neutral}$	0.393 ± 1.366	z = 0.287	0.774	-
	192.7	Intercept	2.835 ± 0.948	z = 2.989	0.003	-
		Male $HL_{functional}$	0.612 ± 1.746	z = 0.351	0.726	-
		Female $HL_{functional}$	-0.733 ± 1.642	z = -0.447	0.655	-
Fledging success[d] (*N* = 139 nests of 120 females / 118 males)	**216.2**	Intercept	3.094 ± 1.430	z = 2.164	0.031	-
		Male HL_{all}	-0.192 ± 3.144	z = -0.061	0.951	-
		Female HL_{all}	0.421 ± 3.303	z = 0.128	0.899	-
	216.2	Intercept	3.146 ± 0.582	z = 5.408	0.000	-
		Male $HL_{neutral}$	-0.152 ± 1.963	z = -0.077	0.938	-
		Female $HL_{neutral}$	0.243 ± 2.051	z = 0.118	0.906	-
	216.2	Intercept	3.128 ± 1.420	z = 2.203	0.028	-
		Male $HL_{functional}$	-0.061 ± 2.611	z = -0.023	0.981	-
		Female $HL_{functional}$	0.167 ± 2.467	z = 0.068	0.946	-
Recruitment success[e] (*N* = 139 nests of 120 females / 118 males)	165.4	Intercept	-0.309 ± 1.361	z = -0.227	0.820	-
		Male HL_{all}	-3.892 ± 2.956	z = -1.317	0.188	-
		Female HL_{all}	-5.180 ± 3.108	z = -1.667	0.096	-
	160.5	Intercept	-2.596 ± 0.514	z = -5.050	0.000	-
		Male $HL_{neutral}$	**-5.251 ± 1.779**	**z = -2.951**	**0.003**	**0.063**
		Female $HL_{neutral}$	1.951 ± 1.677	z = 1.163	0.245	-
	162.2	Intercept	-1.317 ± 1.326	z = -0.993	0.321	-
		Male $HL_{functional}$	1.274 ± 2.354	z = 0.541	0.589	-
		Female $HL_{functional}$	**-6.084 ± 2.343**	**z = -2.596**	**0.009**	**0.048**

[a]: 83.7% and 4.3% of the total variance in the random effects for clutch size were explained by female identity and season, respectively (100% = variance female identity + variance season + residual variance).

[b]: 100% of the total variance in the random effects for number of sired eggs was explained by male identity (100% = variance male identity + variance season + residual variance).

[c]: 100% of the variance in the random effects for hatching success was explained by female identity (100% = variance female identity + variance male identity + variance season).

[d]: 100% of the variance in the random effects for fledging success was explained by female identity (100% = variance female identity + variance male identity + variance season).

e: 58% and 42% of the variance in the random effects for recruitment success were explained by male identity and season, respectively (100% = variance female identity + variance male identity + variance season)

We analyzed the effect of HSL by fitting one model per locus per reproductive success measure, as described above. We investigated the predominance of negative versus positive effects of HSL with cumulative binomial tests assuming equal probabilities (0.5) for both types of associations. We also used cumulative binomial tests to examine if the probability of obtaining significant effects exceeded the expected probability of false positive associations, a (with 79 loci, four reproductive traits per sex and $a = 0.05$, the expected number of false positives per reproductive trait per sex is *ca*. 4). We tested if the effect of heterozygosity at single loci was associated with its putative function, as localized effects are expected to be stronger at the functional loci themselves than at linked markers. Therefore, we used χ^2 tests or Fisher exact tests to investigate whether the number of positive / negative or significant / non-significant effects differed between the groups of functional and neutral loci, or between the groups of markers located within or outside a vertebrate gene; we also compared the effect size of heterozygosity of loci located within and outside a vertebrate gene with *t*-tests. We further tested whether *He*, used as an integral measure of marker diversity, explained variation in the absolute effect size of HSL by fitting linear models. We tested for localized (local and direct) effects with the approach described in Szulkin *et al.* (2010), by 1) regressing the reproductive success measure on MLH (*m1*), 2) regressing the reproductive success measure on all "normalized" single locus heterozygosities in one model (*m2*), and 3) testing for differences in the variance explained by the two models with an *F*-ratio test. We included year and age as co-factors in *m1* and *m2* for the analysis of clutch size

and number of sired eggs. We kept the largest possible sample size for each reproductive success measure by considering all the individual observations as independent. The relatively small sample size and the complex structure of the binomial model including all single loci of males and females resulted in clearly overfitted *m2* models for hatching, fledging and recruitment success. Therefore, for these traits, we did not test for localized effects using the procedure described in Szulkin *et al.* (2010).

Within-population inbreeding

We reconstructed the pedigree of the whole blue tit population according to the genetic parentage analyses, using the R package *pedigree* (Coster 2009). We visualized the full pedigree and verified the individual inbreeding coefficients (*F*) with the program Pedigree Viewer (Kinghorn and Kinghorn 2006). The detection of inbreeding by estimation of *F* depends strongly on the depth of the pedigree (Marshall *et al.* 2002). In our pedigree, where a high proportion (~80%) of the breeding individuals are founders with unknown ancestral history, it is not possible to detect all the events of mating between relatives. Although we identified some inbred offspring, none of them recruited locally as adult breeders. Therefore, the inbreeding coefficient of all the adult individuals genotyped for this study was 0 and we could not test for differences in heterozygosity of individuals with different values of *F*.

Genetic population structure

Among-population inbreeding occurs when the population is subdivided into small, discrete groups. We investigated the occurrence of cryptic genetic substructure using the nonspatial clustering program STRUCTURE (Pritchard *et al.* 2000; Falush *et al.* 2003) as well as the spatial model of clustering implemented in the program GENELAND (Guillot *et al.* 2005; Guillot 2008). For the analysis in STRUCTURE, we followed the recommendations in the program's user's manual (Pritchard *et al.* 2010) and chose the admixture model and the option of correlated allele frequencies between populations; we let the parameter alpha (the degree of admixture between subpopulations) be inferred from the data and set lambda (the allele frequency prior) to 1. We conducted 10 independent replicate analyses for values of *K* (number of subpopulations or genetic clusters) between 1 - 14 with a burn-in period of 10.000 iterations and 10.000 Markov

chain Monte Carlo (MCMC) cycles. We used the STRUCTURE HARVESTER (Earl 2010) to compile and visualize the results from the STRUCTURE runs.

For the analysis in GENELAND, we defined the geographic location of the individuals within the study area with the GPS coordinates of the nestboxes where the birds were breeding; if the individuals bred across several seasons in different nestboxes, we only considered the coordinates of the first breeding nestbox. Uncertainty in the coordinates was set to $1*10^{-5}$ in order to allow individuals sharing the same coordinate to be assigned to different populations, and the maximum number of populations was set to 14. We used the correlated allele frequency model as described in Guillot (2008), for runs with 100.000 MCMC iterations. 10 independent runs were made in order to look for convergence in the number of estimated *K*.

To determine the most likely or "best" value of *K* we employed different approaches, depending on the program used for the analysis. For the STRUCTURE results we used the approach proposed by Evanno *et al.* (2005), which estimates ΔK (an *ad hoc* quantity based on the second order rate of change of the likelihood function with respect to *K*). In GENELAND we considered the convergence of the different runs in the number of estimated *K*, the average posterior probability of the runs and the percentage of iterations per run where the modal *K* was supported.

When admixture of different subpopulations occurs, individuals of mixed ancestry would be more heterozygous and have a higher fitness than individuals from a "pure" origin. Therefore, we tested for the influence of the level of admixture of the individuals on heterozygosity at multiple loci (measured as HL) and on one reproductive trait for which we found significant HFC (clutch size). For these analyses we used the results from the replicate run of STRUCTURE with the highest value of Ln P(D) for the "best" *K* (*K*=3, Ln P(D) = -152.219,8), as well as the "best" GENELAND run (*K*=5). We used the maximal individual *q* value from these analyses (*maxq*) as a measure of the level of admixture. *q* values indicate the membership of the individuals in each of the subpopulations, and each individual has as many *q* values as the number of subpopulations. Individual *q* values range from 0 to 1 for each subpopulation and sum up to 1 across subpopulations. Therefore, a $maxq \approx 1$ would indicate that an individual is a pure member of one genetic cluster, whereas a $maxq \approx 1/K$ would indicate that an individual is highly admixed. To test for the effect of the individual level of

admixture on heterozygosity we calculated a Spearman correlation between *maxq* and HL. To test for the effect of admixture on clutch size we fitted a general mixed model with clutch size as the response variable, *maxq* and female age as predictors, and individual identity nested within year as random intercepts.

Table 3.3.
Results of the analysis in GENELAND to estimate population genetic structure. Runs were made with the correlated allele frequencies model, 100.000 iterations (20.000 burn-in), $K_{max} = 14$ and $1^{*}10^{-5}$ uncertainty in the individual geographical location. *K* denotes the number of estimated subpopulations and Ln P(D) the average posterior probability of the runs. Run 1 (in bold) was used for subsequent analyses.

Run	*K*	Ln P(D)	% iterations supporting *K*
1	**5**	**-234.921**	**-95.75%**
2	6	-223.687	-91.75%
3	5	-235.068	-93.75%
4	7	-218.943	-100%
5	7	-227.217	-100%
6	5	-226.063	-53.87%
7	8	-212.122	-61.62%
8	6	-230.75	-67.25%
9	8	-213.291	-99.37%
10	5	-235.299	-97.38%

Identity disequilibrium (ID)

When within- and/or among- population inbreeding causes enough variation in the individual inbreeding coefficient, identity disequilibrium, defined as the correlation in heterozygosity and/or homozygosity across loci within individuals (Weir and Cockerham 1973; Szulkin *et al.* 2010), should occur. We tested the significance of ID by following the procedure described in Balloux *et al.* (2004), referred to as heterozygosity-heterozygosity correlation (HHC). In short, we randomly chose 10.000 combinations of two halves of the loci and computed the correlation of SH or HL measured with the first and second half using a routine implemented in R (R Development Core Team 2008, script provided by A. Cohas). However, this procedure yields a complicated distribution of non-independent HHC coefficients (Szulkin *et al.* 2010). We therefore calculated the

parameter *g2*, originally described by David *et al.* (2007), and put in the context of heterozygosity-fitness correlations (HFC) by Szulkin *et al.* (2010). *g2* quantifies the central measure of ID, which is the excess of multiple heterozygotes (and multiple homozygotes) at many loci relative to the expectation under random association, standardized by average heterozygosity (Szulkin *et al.* 2010). We calculated *g2* with the software RMES (described in David *et al.* 2007). We used all the loci and the subset of functional and neutral markers separately to estimate HHC and *g2*. Additionally, *g2* was estimated with the subgroups of neutral' and functional' loci that were matched for their degree of diversity.

Results

General descriptive genetic parameters

The average expected heterozygosity (*He*) for the 80 loci was 0.52 ± 0.29 (range: 0.02 - 0.94). The number of alleles per locus (*A*) ranged from 2 to 43, with a mean of 11.5. The number of rare alleles per locus (*R*, alleles with frequency < 0.01, Kimura 1983) ranged between 0 and 26 (which accounts for between 0% and 80% of the alleles in a locus), with a mean of 5 (39%). After FDR correction, 69 loci showed no significant deviation from HWE across all years, 10 loci showed deviations in one or two years, and one locus showed deviations in eight years. The latter marker (DkiD12) was discarded for further HFC analyses. After FDR correction with the number of tests made per year of study, nine out of 162 pairs of syntenic loci showed significant LD in one study year, two pairs in two years, and three pairs in three to nine years (Table 1.3). As the LD correlation coefficient for these pairs of loci was small (0.05 to 0.23, median 0.1, details not shown), none of the markers were discarded.

From the 79 loci that were used for further analyses, 58 belonged to the functional and 21 to the neutral category of markers. The characteristics of the markers were clearly different for the two groups of loci (Figure 1.3). Diversity of neutral loci (measured either as *He*, *A* or *R*) was significantly higher than diversity of functional ones (median *He* / *A* / *R* for neutral loci = 0.81 / 18 / 7, median *He* / *A* / *R* for functional loci = 0.43 / 6 / 3, Mann-Whitney tests: all *p*-values < 0.001). The median distance of the markers to a vertebrate gene in the zebra

finch genome was significantly larger for neutral loci (median distance (Kb) to a gene for neutral / functional loci = 10.1 / 0, Mann-Whitney test: $p < 0.001$).

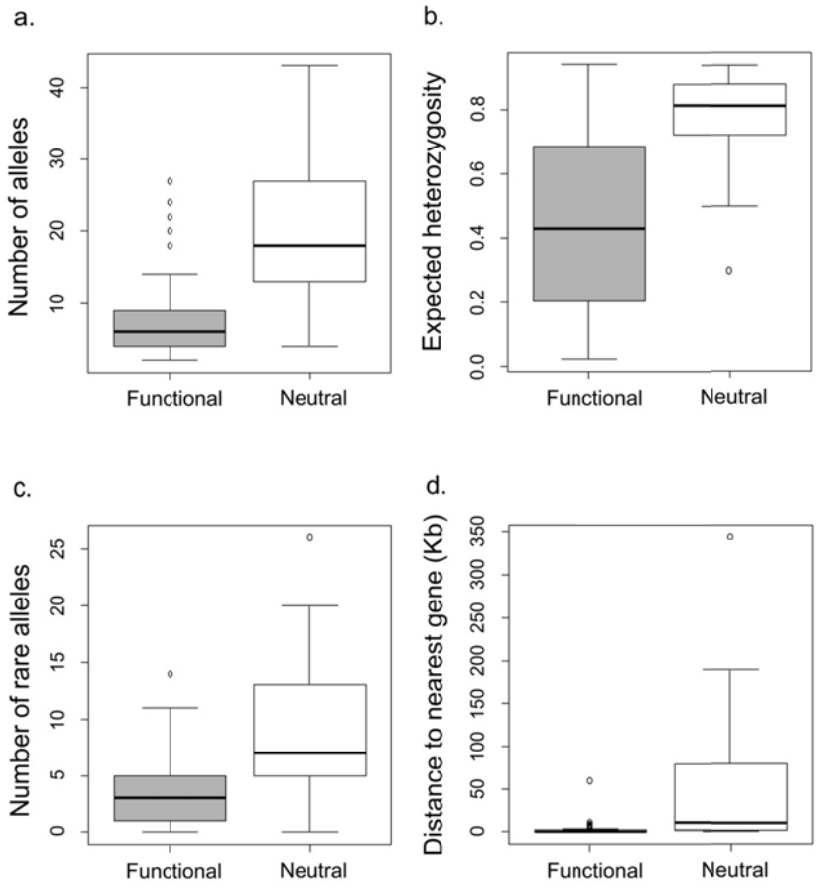

Figure 1.3.
Characteristics of presumably functional (in gray) and neutral (in white) markers (see text for definition). a. Number of alleles, Mann-Whitney U-test, $U = 167$, $p < 0.001$; b. Expected heterozygosity, Mann-Whitney U-test, $U = 201$, $p < 0.001$; c. Number of rare alleles, Mann-Whitney U-test, $U = 207$, $p < 0.001$; d. Distance to the nearest gene (in Kb units), Mann-Whitney U-test, $U = 212$, $p < 0.001$.

Multilocus heterozygosity measures

Individual SH and HL were highly correlated ($r = -0.917$, $p < 0.001$; N = 794). SH measured with all markers (SH_{all}) ranged from 0.675 to 1.327 (mean 0.994), and HL measured with all markers (HL_{all}) ranged from 0.145 to 0.546 (mean 0.319). The correlation between SH or HL measured with the subsets of functional and

neutral markers was low and non-significant (r $SH_{functional} - SH_{neutral} = 0.043$, $p = 0.16$; r $HL_{functional} - HL_{neutral} = 0.035$, $p = 0.26$). This confirmed the relevance of conducting subsequent analyses with the two subsets of markers separately.

Effect of heterozygosity at multiple loci on reproductive success

The results from the analyses using SH and HL as measures of heterozygosity at multiple loci led, in general, to similar conclusions. As HL is a better predictor of genome-wide heterozygosity and inbreeding in populations subjected to migration and admixture (Aparicio *et al.* 2006), we present statistics and figures only for this measure. We found significant positive linear effects of HL measured with different subsets of markers on estimates of male and female reproductive success (Table 2.3, Figure 2.3). Clutch size and the number of sired eggs were significantly influenced by female and male HL estimated with the neutral loci. The recruitment success of broods without extra-pair young also depended on HL: for neutral loci, we only found a significant effect of male HL, whereas for functional loci only female HL had a significant effect. We did not find any significant effects of male or female HL on hatching and fledging success of their offspring in nests without extra-pair young (EPY). However, when nests with EPY were also considered, we found a significant positive effect of male $HL_{neutral}$ on fledging success in the first years of study (details not shown). When significant effects of HL on the reproductive success estimates were found, the models including $HL_{neutral}$ fitted the data better (e.g. had lower AIC values) than the models with $HL_{functional}$. Significant HL effects explained around 2% of the variance in clutch size and number of eggs sired, and >4% of the variance in recruitment success. The results of the tests for the effects of HL with the subsets of neutral and functional markers that were matched for their degree of diversity (defined as functional' and neutral') were similar to the ones obtained with the complete sets of loci (clutch size: slope of $HL_{functional'} = -0.119$, 95% CI = (-1.628, 1.389), $p = 0.877$; slope of $HL_{neutral'} = -2.332$, 95% CI = (-3.881, -0.7826), $p = 0.003$; number of sired eggs: slope of $HL_{functional'} = -0.282$, 95% CI = (-3.250, 2.686), $p = 0.852$; slope of $HL_{neutral'} = -2.811$, 95% CI = (-5.733 - 0.110), $p = 0.003$).

Effects of heterozygosity at single loci on reproductive success

We observed between 33 and 48 positive effects of HSL (out of *ca.* 79 per sex) on each reproductive success trait (Table 1.3). The probability of obtaining positive (or negative) effects, independent of its significance, was not different from 0.5

for each trait. For most traits, the direction of the effects did not differ significantly for the functional or neutral set of loci. The exception was recruitment success, where more positive effects were associated with the functional markers of females (χ^2 test: $p = 0.02$). For each reproductive trait, the direction of the effects and the mean absolute effect size of HSL did not differ significantly for loci located within or outside a vertebrate gene. We observed between 0 and 7 significant (positive and negative) effects of HSL on each of the traits related to reproductive success. These numbers did not significantly exceed the expected number of false positives (*ca.* 4 per reproductive trait per sex). Significant effects of HSL were not associated more with the functional or neutral set of loci. However, for males, all the loci with significant effects on recruitment success were located within genes (Fisher's test: $p < 0.01$). Locus diversity (measured as *He*) was positively and significantly correlated with the effect size of HSL for both clutch size and number of eggs sired by males (linear models: *p*-values < 0.01 and 0.05, respectively). There was no such correlation for the other traits. Overall, heterozygosity of a total of 28 loci showed significant effects for at least one reproductive trait (Table 1.3); half of these loci were located within a gene region. Following the procedure described in Szulkin *et al.* (2010), we did not find a significant difference in the variance explained by the model with all single loci (*m2*) compared to the model with MLH (*m1*) (clutch size: $F_{[399,321]} = 1.07$, $p = 0.34$; number of sired eggs: $F_{[355,277]} = 0.92$, $p = 0.65$); therefore, the partial regression coefficients of all the loci are not different from each other.

Within-population inbreeding

In 601 out of 852 matings between known individuals (71%), both pair members were founders (with *F* assumed to be 0), while in the other 251 cases at least one member of the pair had a known ancestor. From the latter, we identified four cases (2%) of consanguineous pairs: two mother-son pairs (offspring $F = 0.25$), one pair of half sibs breeding in two consecutive years (offspring $F = 0.125$) and one pair of "aunt-nephew" (sharing a common mother/grandmother, offspring $F = 0.0625$). From these pairings, 43 inbred chicks (2% of all chicks from the 251 matings mentioned above) were produced. None of the inbred individuals recruited in the population.

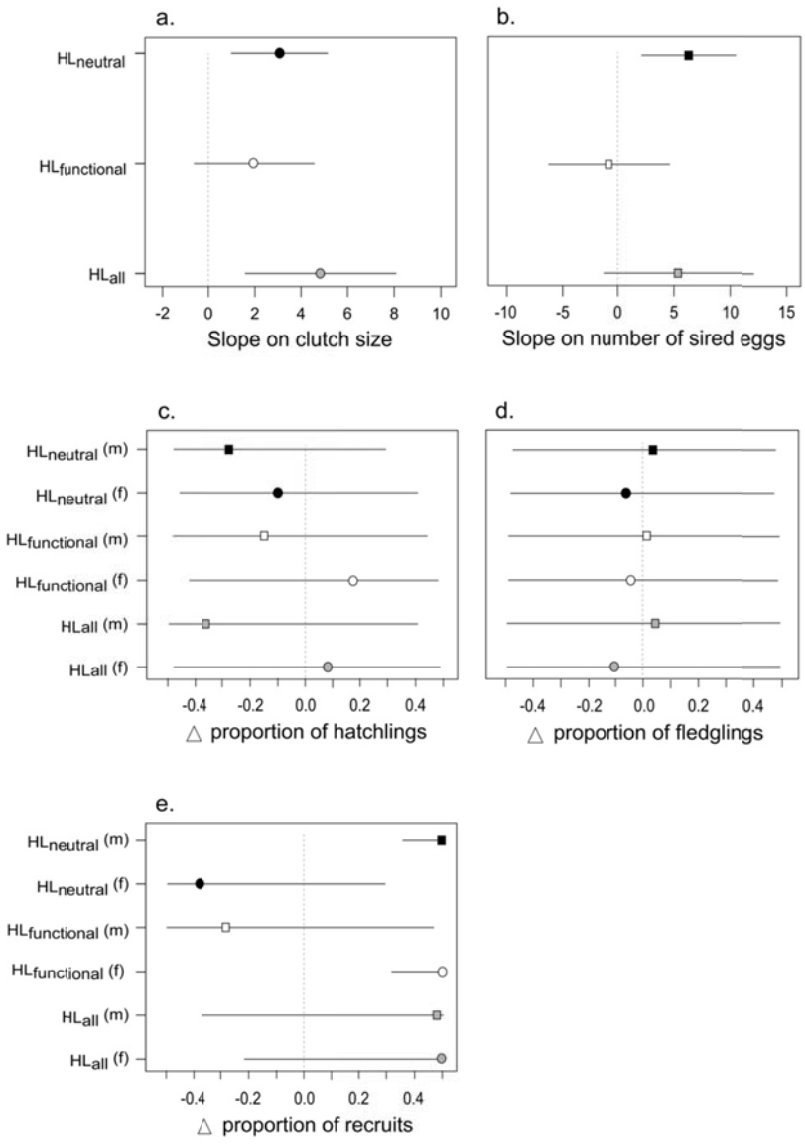

Figure 2.3.
Effect sizes of individual heterozygosity at multiple loci (measured as HL) on estimates of reproductive success of blue tits. Effect sizes are estimated from the "Estimate" value of HL from Table 2.3, and are represented with their 95% confidence intervals. As HL increases with decreasing heterozygosity, we changed the sign of the effects. For clutch size and number of sired eggs, effect size is the slope of HL on the trait. For hatching, fledging and recruitment success, effect sizes are the back-transformed "Estimate" value, and represent a change in proportion (of hatchlings / number of eggs, fledglings / number of hatchlings or recruits / number of fledglings) due to an increase in heterozygosity. Black, white and gray symbols for $HL_{neutral}$, $HL_{functional}$ and HL_{all}, respectively. Squares and circles for males (m) and females (f), respectively. a. Clutch size; b. Number of sired eggs; c. Hatching success; d. Fledging success; e. Recruitment success

Genetic population structure

We found evidence for a genetic substructure within our blue tit population, using both spatial (GENELAND) and non-spatial (STRUCTURE) clustering methods (Figure 3.3a, Table 3.3). In the STRUCTURE analysis, the mean value of the model choice criterion (Ln P(D)) for detecting the number of genetic clusters (K) steadily increased for values of K between 1 and 11, and fluctuated for values of $K > 11$. However, as the variance in Ln P(D) notably increased for values of $K > 7$, we only considered the runs with K between 1 and 6 for the estimation of the "real" number of genetic clusters. As the value of ΔK (Evanno *et al.* 2005) peaked at $K=3$ (Figure 3.3b), we chose 3 as the number of populations in our sample according to the STRUCTURE analyses. The 10 runs of the analysis in GENELAND did not converge in a unique number of estimated K, which varied between 5 and 8. However, since 5 was the modal K, we chose the run with the highest average posterior probability where 5 populations were supported in more than 90% of the iterations. Although the distribution of the individual level of admixture (measured as *maxq*) obtained with the two programs was very different (Figure 4.3), we obtained similar results in the subsequent analyses: we found neither significant correlations between the level of admixture and HL (Spearman rank correlations for the results from GENELAND / STRUCTURE: rho = -0.066 / 0.005, p = 0.064 / 0.893), nor a significant influence of the level of admixture on clutch size (General mixed models for the results from GENELAND / STRUCTURE: Slope of *maxq* on clutch size = 0.204 / 0.841, Std. error = 0.519 / 0.872, p = 0.695 / 0.335) or on the number of sired eggs (General mixed models for the results from GENELAND / STRUCTURE: Slope of *maxq* on the number of eggs sired = 1.022 / 1.309, Std. error = 1.075 / 1.372, p = 0.343 / 0.341).

Identity disequilibrium (ID)

The within-individual HHC (Balloux *et al.* 2004) were low, but always positive. However, we obtained different results depending on whether multilocus heterozygosity was estimated as SH or HL. When SH was measured either with all the markers or with the neutral subset of loci, the HHC were marginally significant; when the functional loci were used, the HHC were lower and non-significant (mean $r \pm$ s.d.; all loci: $r = 0.045 \pm 0.031$, $p = 0.07$; neutral loci: $r = 0.046 \pm 0.029$, $p = 0.05$; functional loci: $r = 0.022 \pm 0.031$, $p = 0.25$). In the case of HL, the highest HHC were obtained when all markers were used, they were

marginally significant with the functional loci and not significant with the neutral loci (all loci: $r = 0.054 \pm 0.030$, $p = 0.03$; functional loci: $r = 0.050 \pm 0.031$, $p = 0.05$; neutral loci: $r = 0.022 \pm 0.030$, $p = 0.23$). The parameter *g2* (David *et al.* 2007; Szulkin *et al.* 2010) was significant when measured with all the markers or with the neutral ones (all loci: $g2 = 0.00088$, $p = 0.03$; neutral loci: $g2 = 0.00165$, $p = 0.019$; functional loci: $g2 = 0.00082$, $p = 0.18$). *g2* estimated with the groups of neutral' and functional' loci that were matched for their level of diversity was not different from 0 (neutral' loci: $g2 = 0.00207$, $p = 0.110$; functional' loci: $g2 = 0.00043$, $p = 0.382$).

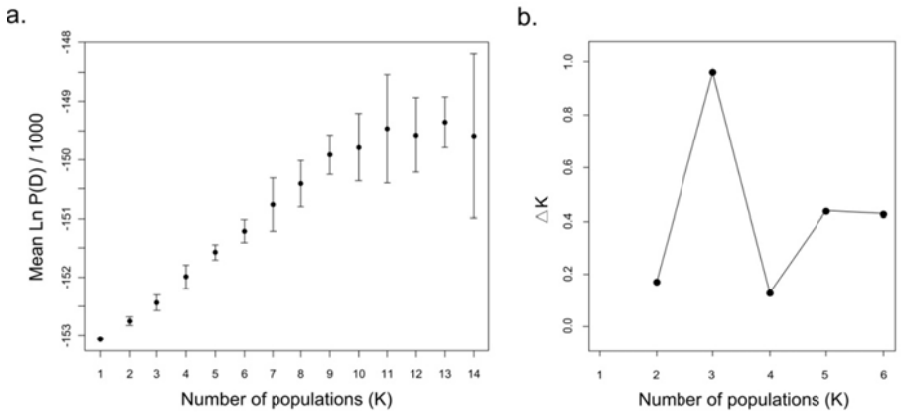

Figure 3.3.
Results from a population structure analysis using the software STRUCTURE. a. Mean Ln P(D) (± standard deviation) for the different number of subpopulations (*K*) considered. b. *ΔK* method for estimating the number of populations; the value of *K* for which *ΔK* peaks is considered to be the number of genetically differentiated groups in the sample.

Discussion

We found strong positive effects of individual heterozygosity at multiple loci (measured either as HL or SH) on three estimates of the annual reproductive output of male and female blue tits. Significant effects of multilocus heterozygosity were detected more often with the panel of presumably neutral markers, even though the effects of functional and neutral loci were not significantly different (see overlapping confidence intervals in Figure 2.3). The

characteristics of the markers in the two groups and the level of identity disequilibrium measured with them suggest that the set of neutral markers (N = 21) has more power than the panel of functional loci (N = 58) to detect HFC and individual differences in the level of inbreeding. Moreover, the detection of consanguineous matings and a cryptic genetic substructure supports a scenario where variation in the levels of inbreeding in the population may cause detectable HFC. We did not identify single loci with disproportionate effects on reproductive success and, therefore, our results are better explained by small effects of many loci.

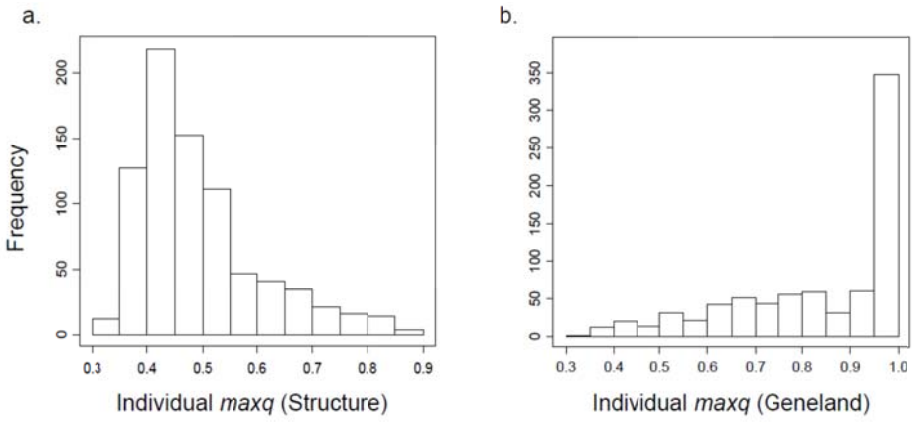

Figure 4.3.
Histogram of the individual level of admixture (*maxq*) from the replicate run of STRUCTURE with the highest value of Ln P(D) for K = 3 (a) and the "best" GENELAND run (K = 5) (b). *maxq* values ≈ 1 are for pure individuals whereas *maxq* values ≈ 1 / K are for admixed individuals.

Effect of heterozygosity at multiple loci

We found a significant effect of female multilocus heterozygosity on clutch size, of male multilocus heterozygosity on the number of eggs sired and of male and female multilocus heterozygosity on recruitment success of their offspring. Significant effects of multilocus heterozygosity on similar traits have been found in a variety of species (clutch size: Ortego *et al.* 2009; male mating success: Hoglund *et al.* 2002; Johnson *et al.* 2010; offspring survival: Amos *et al.* 2001; Brouwer *et al.* 2007). The estimated effect of heterozygosity at multiple loci (measured as HL) on the reproductive traits of blue tits in this study (2% - 6% of the variance explained) was higher than the average (less than 1%), but within

the range of the effect sizes reported in a recent meta-analysis of HFC in animal populations (Chapman *et al.* 2009). It is expected that traits directly related to fitness (e.g. life-history traits) which are under strong directional selection would show greater levels of inbreeding depression and a stronger relationship with heterozygosity than traits which are not directly linked to fitness (e.g. morphological traits) (Crnokrak and Roff 1995; DeRose and Roff 1999). This could explain the relatively strong effects of multilocus heterozygosity on measures of reproductive success found here. Our estimation of the effect of heterozygosity at multiple loci on the reproductive traits should be rather precise, due to the large number of individuals and markers assayed (Chapman *et al.* 2009).

Positive HFC in blue tits have also been found in a different population. Garcia-Navas *et al.* (2009) investigated HFC in blue tits breeding in 2007 and 2008 in central Spain and reported significant positive effects of female multilocus heterozygosity on clutch size and eggshell spotting distribution, and of male multilocus heterozygosity on feeding rates and crown coloration brightness. From the 14 microsatellites used in the study of Garcia-Navas *et al.* (2009), 11 were included in our study. All of them belonged to the neutral category. Therefore, the positive effect of neutral heterozygosity on clutch size of blue tits seems to be widespread. In a previous study on the same population used here, Foerster *et al.* (2003) used 5-7 presumably neutral microsatellite markers to estimate multilocus heterozygosity and investigated its relationship with reproductive success measures recorded between 1998 and 2001. Foerster *et al.* (2003) reported a positive correlation between female multilocus heterozygosity and clutch size, and between male multilocus heterozygosity and the number of recruits produced. Thus, for these traits, a small number of neutral microsatellites was sufficient to detect an effect similar to the one obtained here with a 3- to 4-fold enlarged panel of presumably neutral loci genotyped in an enlarged dataset. Moreover, a further increase in the number of markers with a set of presumably functional loci did not increase the power to detect the effects of heterozygosity at multiple loci for these traits. Foerster *et al.* (2003) did not find an effect of multilocus heterozygosity on the number of eggs sired by males. The almost three-fold increase in the number of males sampled and the use of a larger set of neutral markers in this study may have increased the power to detect the effect of neutral heterozygosity on this trait. On the other hand, the effect of male multilocus heterozygosity on fledging success reported by Foerster *et al.*

(2003) was not found here with any subset of markers. For the latter analysis, Foerster *et al.* (2003) included data from all the nests during the first three study years, while in this study we only considered nests without extra-pair young (EPY) across seven breeding seasons. An analysis of the data including all nests revealed that the positive effect of male heterozygosity at neutral loci on fledging success occurred only in the first years of study.

We found positive effects of heterozygosity at multiple loci on an individual's reproductive output (e.g. clutch size and number of sired eggs), as well as on fitness of the offspring (recruitment success). HFC on individual traits arise via mechanisms directly affecting the phenotype of the individual. Offspring fitness, on the other hand, can be influenced by effects of heterozygosity or inbreeding on direct or environmental parental effects (through somatic effects, hormones, provisioning and parental care) (Brouwer *et al.* 2007; Garcia-Navas *et al.* 2009), or by effects of heterozygosity or inbreeding on the offspring themselves (Hansson *et al.* 2001; Foerster *et al.* 2003; Taylor *et al.* 2010). The mechanism through which female and male multilocus heterozygosity was associated with recruitment success of their offspring in this study remains to be resolved. However, since the significant effects were found with different subsets of markers for males and females, we hypothesize that they could arise via general effects of inbreeding on male traits affecting, for example, provisioning and parental care (Garcia-Navas *et al.* 2009) as well as through effects of heterozygosity at single loci of females affecting maternal effects, offspring feeding or other forms of maternal care (Brouwer *et al.* 2007).

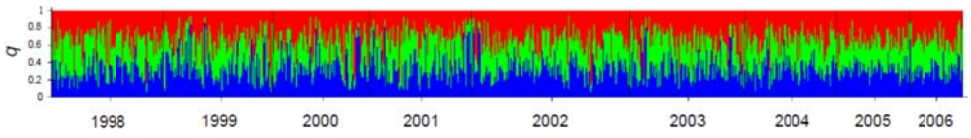

Figure 5.3.
Individual membership coefficients to the different subpopulations (q values) from the replicate run of STRUCTURE with the highest value of Ln P(D) for $K = 3$, grouped by year of study. Each individual is represented by a vertical segment partitioned in three colored segments. These segments represent the estimated membership coefficient of the individuals in each of the three subpopulations.

Inbreeding as a cause of HFC

In studies of populations where the occurrence of inbreeding is assumed or easily detected, the most parsimonious explanation for positive multilocus HFC is inbreeding depression (Coltman *et al.* 1999; Slate and Pemberton 2002; Fitzpatrick and Evans 2009; Mainguy *et al.* 2009; Blomqvist *et al.* 2010). In contrast, in populations where the occurrence of inbreeding is neither expected nor directly measurable, inbreeding is usually disregarded as the cause of HFC and other mechanisms are claimed (e.g. local effects). As pointed out by Szulkin *et al.* (2010), HFC provide indirect information on inbreeding, even in populations where it is not anticipated (e.g. David *et al.* 1997). Therefore, testing directly for the occurrence of inbreeding can shed light on the causes of HFC.

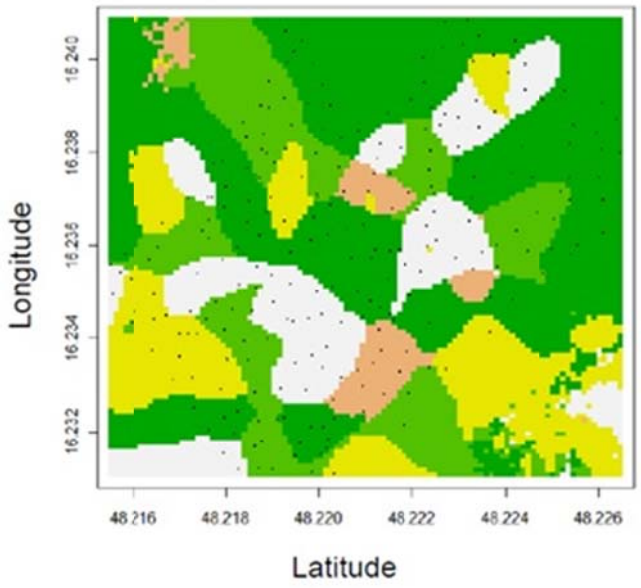

Figure 6.3.
Map of estimated population membership from the "best" GENELAND run (K=5). Individuals are represented with black dots located in the geographical coordinates of their (first) breeding nestbox. The different colors represent the five estimated subpopulations.

We tested for the incidence of within-population inbreeding in our study system by constructing the pedigree, and estimated that in 2% of the matings the individuals in the pair were related. If the observed HFC are caused by this source of inbreeding, we would expect individuals produced by related parents to

show lower estimates of multilocus heterozygosity than the ones produced by non-related ones. Unfortunately, as none of the progeny produced by consanguineous pairs was breeding in our population, we cannot provide direct evidence of HFC to be caused by lower heterozygosity and reproductive success values of "pedigree inbred" individuals. It is important to note here that the estimation of the number of consanguineous matings in the population is conservative, as the origin of the majority of the birds breeding in the study area is unknown. Therefore, the number of individuals with inbreeding coefficient higher than zero is likely to be underestimated.

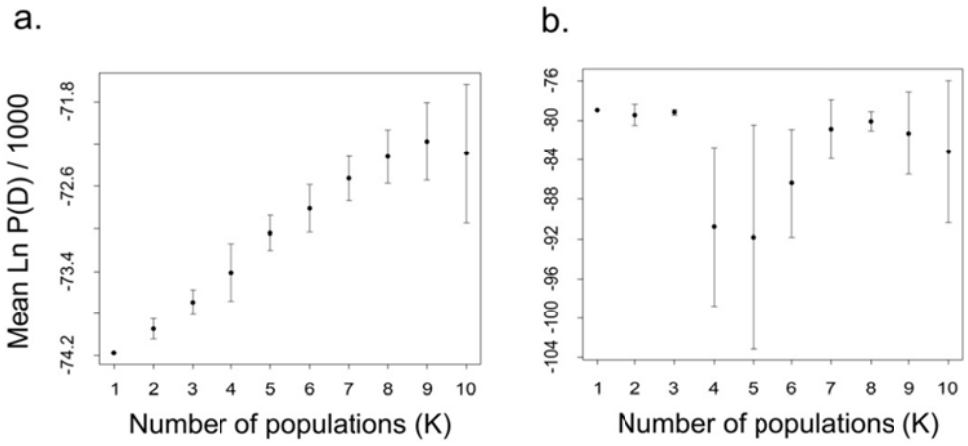

Figure 7.3.
Results from a population structure analysis for a) males and b) females using the software STRUCTURE.

Inbreeding can also occur through population subdivision and genetic drift (Keller and Waller 2002). Therefore, we investigated the occurrence of genetic structure in our study population using both spatial (Guillot *et al.* 2005; Guillot 2008) and nonspatial (Pritchard *et al.* 2000; Falush *et al.* 2003) Bayesian clustering methods. Interestingly, both methods indicate that the sample of birds breeding in the study area is likely to be composed of more than one genetically differentiated group. However, defining the number of populations (*K*) in our sample of breeding individuals was not straightforward. A discussion of the problems associated with estimating *K* would go beyond the purpose of this paper (see e.g. Evanno *et al.* 2005; Latch *et al.* 2006; Waples and Gaggiotti 2006; Guillot 2008; Pritchard *et al.* 2010).

Understanding how population structure can arise and be maintained in an open population of a mobile species in an area with no obvious barriers to gene flow is important to predict its consequences for inbreeding and HFC. A visual inspection of the STRUCTURE's summary plot of the individual *q*-values grouped by year (Figure 5.3) does not show yearly differences in the genetic composition of the population. The map of population membership of the individuals obtained with GENELAND (Figure 6.3) reveals that the detected subpopulations are not spatially confined. Therefore, genetic structure in this blue tit population is caused neither by differences in the groups of individuals breeding in the study area across the years of study, nor by spatial isolation. However, in an analysis of population structure for each sex separately we detected a genetic differentiation in males only (Figure 7.3). This type of genetic substructure has been described previously in birds (Double *et al.* 2005; Temple *et al.* 2006; Lee *et al.* 2009) and is attributable to the lower rates and distances of dispersal of males, the philopatric sex (Greenwood 1980). Parent-offspring relationships in the population, driven by the local recruitment of males, are thus an important cause of the structure found. In fact, a STRUCTURE analysis with founder individuals only (i.e. excluding data from offspring) resulted in no clear structuring of the population (details not shown). This kind of structure does not necessarily contribute to inbreeding, as long as the dispersal pattern of females and/or the mechanisms of kin recognition are enough to prevent the occurrence of matings between relatives. However, we have two lines of evidence against this possibility. First, female fledglings also recruit locally (26% of all recruits in the population are females) and adult females usually remain in the study area after breeding (Valcu and Kempenaers 2008), creating the opportunity for matings between related individuals. In fact, in the detected cases of inbreeding by pedigree, a male recruit bred with an older relative or with a half-sib. Second, in a study on the same population, Foerster *et al.* (2006) found that pairs with local recruits were more related than expected by chance in some of the study years. Therefore, the detected population structure can contribute to the occurrence of inbreeding. However, since this type of structure does not arise as the result of genetic isolation of groups, but instead is caused by the local settlement of males born in the study area, it does not alone explain the HFC found. In studies of HFC where population structure is the result of genetic isolation of groups, correlating the individual level of admixture with

heterozygosity or with phenotypic traits could be useful to investigate between-population inbreeding as a cause of HFC.

The detection of consanguineous matings and a genetic structure that contributes to the occurrence of mating between relatives indicates that, in this blue tit population, there is higher heterogeneity in the individual level of inbreeding than previously thought. This heterogeneity is the prime material for HFC to arise and for multilocus heterozygosity measured with a handful of genetic markers to reflect individual inbreeding coefficients (Balloux *et al.* 2004; Slate *et al.* 2004; Szulkin *et al.* 2010).

Functional vs. neutral markers and effect of heterozygosity at single loci

We found that variation in inbreeding in the population is detected by measures of heterozygosity at neutral loci and at neutral and functional loci together, but not by measures of heterozygosity at functional loci alone. This is expected based on the difference in marker diversity between neutral and functional markers, as it has been reported that the use of more variable loci significantly improves the correlation between heterozygosity and the inbreeding coefficient (Slate *et al.* 2004; Aparicio *et al.* 2006; Hansson *et al.* 2010a; Ljungqvist *et al.* 2010). However, we found that the power to detect significant HFC with neutral markers was not only due to their higher degree of diversity, because significant HFC were still detected with a subset of the neutral markers that showed a similar diversity as the functional loci.

Our results are in contrast with the many sources that recommend increasing the number of microsatellite markers in order to obtain a better estimation of the inbreeding level (Balloux *et al.* 2004; Slate *et al.* 2004; DeWoody and DeWoody 2005). However, this recommendation is based on the assumption that microsatellites are selectively neutral (Jarne and Lagoda 1996). We *a priori* classified the microsatellite markers used in this study as "neutral" or "functional" based on whether they were located in expressed regions of the genome. This classification is not free of assumptions due to, for example, differences in expression between different tissues or lack of complete information on gene expression at the organismal level. However, even with these limitations, our marker classification captured differences that would be expected to occur between groups of loci with genuine differences in functionality (e.g. number of alleles, expected heterozygosity, degree of HHC).

Therefore, the interpretation of HFC are highly dependent on the characteristics and functionality of the markers used to measure multilocus heterozygosity. In particular, care should be taken when interpreting non-significant HFC, as inbreeding effects reflected by some loci can be masked by the addition of markers with a weaker relation with inbreeding.

The effects of heterozygosity at multiple loci on fitness must be caused by direct effects of functional loci or by neutral markers linked with such loci. A question to resolve is if heterozygosity effects are caused by a small number of loci with strong effects or a large number of loci with rather weak effects. Expectations for the effects of heterozygosity at single loci depend on the location and linkage of the markers with functional genes (in the case of neutral loci), on the function of the markers themselves (for functional loci), on the allelic dominance of the genes affecting the phenotypes, and on allele frequencies (Mueller *et al.* 2010), among others. Therefore, knowledge about the location of the markers with respect to functional genes are important to interpret effects of heterozygosity at single loci, especially when many markers are tested and the risk of Type I errors increases. Our analyses of HSL on traits reflecting reproductive success of blue tits revealed an assortment of positive, negative, significant and non-significant effects that differed between traits (details not shown). We found evidence neither for the occurrence of larger absolute effects in markers located within a gene, nor for associations of significant effects with markers within genes or with expressed loci. Moreover, we did not find more significant effects than expected by chance, even with the large number of markers tested. Although significant effects are overrepresented on chromosome 7 (Table 1.3, Fisher's exact test for the association of significant effects with markers on chromosome 7 vs. markers on all the other chromosomes: $p < 0.01$), neither the HFC nor the ID analyses in this study are driven only by markers on chromosome 7 (details not shown). From the 28 loci with at least one significant effect, 6 (21%) loci showed two to three significant effects on different reproductive traits of males and females. However, with the exception of locus TG13-017, these multiple effects were always in opposite directions, either between sexes, or within a sex on different traits. This reveals the limitations of generalizing the effects of single loci when only a few fitness traits are investigated. General tests for localized (i.e. local and direct) effects (Szulkin *et al.* 2010) require sample sizes considerably larger than the number of loci tested in order to reach the statistical power to detect differences in the effects of single

loci. This is a limitation in studies of wild populations when a high number of genetic markers are used.

For the above mentioned reasons, the effects of multilocus heterozygosity found in this study can most likely be attributed to the interaction of small positive and negative effects of many loci. If the general goal in a study is to find strong, widely-applicable effects of single genes, further progress could be made by implementing a candidate-gene approach (Fitzpatrick *et al.* 2005). The number of test loci would then be reduced and the interpretation of significant effects of HSL would be more straightforward.

Implications

The influence of multilocus heterozygosity on important reproductive success traits of male and female blue tits has several important implications. Overall, they emphasize the relevance of mechanisms of inbreeding avoidance (Pusey and Wolf 1996) and the importance of non-additive genetic effects for the reproductive success of males and females (e.g. Seddon *et al.* 2004; Tomiuk *et al.* 2007). In the context of mate choice, they emphasize the benefits of choosing a heterozygote individual as a social partner (reviewed in Kempenaers 2007), since heterozygosity of both males and females increases the probability of raising recruits. Moreover, our finding of higher siring success of more heterozygous males provides indirect evidence that females may actively choose more heterozygote males as partners. This choice might be based on visual cues, because male heterozygosity has also been correlated with plumage characteristics in this and other blue tit populations (Foerster *et al.* 2003; Garcia-Navas *et al.* 2009).

In a conservation genetics context, the effects reported with presumably neutral markers can be interpreted as an indirect estimate of the deleterious effects of inbreeding. Similar correlations of neutral heterozygosity with different traits have been reported in other populations of blue tits (Garcia-Navas *et al.* 2009), indicating that the effect may be widespread. This highlights the importance of genetic diversity, even for common and abundant species where the occurrence of strong inbreeding is rare. Moreover, the HFC, consanguineous matings and genetic population structure found here reflect heterogeneity in the level of inbreeding in a species expected to be homogeneous.

Chapter 3

Heterozygosity and survival in blue tits (*Cyanistes caeruleus*): contrasting effects of presumably functional and neutral loci

Juanita Olano-Marin, Jakob C Mueller, Bart Kempenaers

Abstract

The relationship between genetic diversity and fitness has important implications in evolutionary and conservation biology. This relationship has been widely investigated at the individual level in studies of heterozygosity-fitness correlations (HFC). General effects due to inbreeding and/or local effects at single loci have been used as explanations of HFC, but the debate about the causes of HFC in open, natural populations is still ongoing. Study designs that allow to control for variation in the inbreeding level of the individuals, and knowledge on the function and location of the markers used to measure heterozygosity are fundamental to understand the causes of HFC. Here we investigated correlations between individual heterozygosity and estimates of survival at different life-history stages in an open population of blue tits (Cyanistes caeruleus). For survival at the egg, nestling and fledgling stage we used a full-sibling approach, i.e. we controlled for the level of inbreeding. We genotyped 1496 individuals with 79 microsatellites distributed across 25 chromosomes in another passerine and classified either as potentially functional (58 loci) or neutral (21 loci). We found different effects of heterozygosity at multiple loci (measured as SH): $SH_{functional}$ had a negative effect on the probability of hatching and local recruitment of females, whereas $SH_{neutral}$ had a positive effect on adult survival. The negative effects of functional loci are better explained by local effects, whereas the positive effects of neutral markers could reflect inbreeding effects in the population. Our results highlight the importance of considering the characteristics of the markers used in HFC studies and confirm the mixed effects of heterozygosity in different contexts (e.g. sex and life-history stage).

Manuscript accepted for publication in *Molecular Ecology*, pending minor revisions

The influence of genetic diversity on the fitness of individuals and populations has important implications for evolutionary and conservation biology. Heterozygosity-fitness correlations (HFC) have been used to study the relationship between genetic diversity and fitness-related traits at the individual level in a variety of organisms (reviewed in Britten 1996; David 1998; Coltman and Slate 2003; Chapman *et al.* 2009).

HFC are commonly explained by inbreeding effects across the whole genome (referred to as the general effect hypothesis), or by localized effects at single loci (referred to as the local/direct effect hypotheses) (David 1998; Hansson and Westerberg 2002). Inbreeding causes a reduction in individual heterozygosity across the genome, and thereby increases the risk of expressing recessive deleterious alleles and decreases the occurrence of beneficial overdominant effects, a phenomenon commonly known as inbreeding depression (Crnokrak and Roff 1999; Keller and Waller 2002; Charlesworth and Willis 2009). Under this scenario, a positive relationship between individual heterozygosity at multiple loci and fitness is expected. The opposite phenomenon, outbreeding depression, occurs when progeny produced by crosses between individuals from genetically differentiated populations have lower fitness than progeny from crosses between individuals from the same population. The decline in fitness in this case is attributed to a breakup of coadapted gene complexes or favorable epistatic interactions (Lynch 1991). Under this scenario, negative or quadratic multilocus heterozygosity-fitness correlations can be detected (e.g. Marshall and Spalton 2000; Neff 2004). Local or direct effects of heterozygosity, on the other hand, are caused by the effect of individual functional loci or of neutral markers linked to such loci. Under this scenario, positive, negative or neutral HFC can be detected (e.g. Lieutenant-Gosselin and Bernatchez 2006; Mueller *et al.* 2010). Despite the numerous HFC studies in animal populations (reviewed in Chapman *et al.* 2009), the interpretation of significant HFC has been hindered by methodological constraints, such as difficulties in estimating individual inbreeding in natural populations (Pemberton 2004), the generally low number of genetic markers used to measure heterozygosity (Balloux *et al.* 2004; Slate *et al.* 2004; DeWoody and DeWoody 2005), statistical caveats in the analysis of local (and direct) effects (Szulkin *et al.* 2010), and the lack of knowledge about the genomic location and the functionality of the loci employed, among others.

Microsatellites, the most widely used genetic markers for HFC studies in the last years, are commonly considered evolutionarily neutral (Jarne and Lagoda 1996). This view has been challenged by accumulating evidence on the functionality of repeats located within genes and expressed regions of the genome (reviewed in Li *et al.* 2004). The functionality of the markers used to measure heterozygosity in HFC studies can have important implications for formulating predictions and interpreting results. General effects of inbreeding, for example, are more likely to be detected with neutral loci, direct effects can only be caused by functional loci and local effects are caused by neutral markers closely linked with functional loci. The public release and annotation of whole genome sequences (e.g. in the NCBI and UCSC genome browsers in the web) has opened the possibility of anchoring microsatellites within the genome of model species (e.g. Dawson *et al.* 2006; Olano-Marin *et al.* 2010) and with respect to genes. This can be useful to predict the functionality of the markers. Similarly, markers developed from libraries of Expressed Sequence Tags can be used as a source of potentially functional loci (Vasemagi *et al.* 2005; Oliveira *et al.* 2009).

In this study we investigated the influence of individual heterozygosity on estimates of survival of blue tits (*Cyanistes caeruleus*) from an Austrian population studied between 1998 - 2006. For this purpose, we used a panel of 79 microsatellites distributed across 25 chromosomes in a passerine (the zebra finch *Taeniopygia guttata*). We classified the markers as potentially functional or neutral and identified the (closest) genes where the microsatellites were located. We tested for the effect of multi- and single-locus heterozygosity measured with presumably functional or neutral loci on 1) measures of early survival in the nest (i.e. survival until hatching and until fledging), 2) an estimate of survival after fledging (local recruitment) and 3) an estimate of survival of one-year-old breeding birds until the next breeding season. In all analyses, we considered possible differences between the sexes and controlled for year effects. For the analysis of early survival ((1) and (2)) we used a full-sibling approach in order to control for the level of inbreeding.

Material and Methods

Study system and general procedures

A nestbox population of blue tits was studied for 9 years (1998 - 2006) in 42 - 50 ha of a mixed deciduous forest at Kolbeterberg (Vienna, Austria, 48°13'N, 16°20'E). Birds were caught while roosting in the nest box in winter (between December and March), or while feeding nestlings (May). After catching, each bird was banded with a metal band and three colored rings, aged (as one year or older, Svensson 1992), and bled by brachial venipuncture for DNA extraction. All nesting attempts were carefully monitored until the chicks fledged. Unhatched eggs and dead nestlings were collected for DNA extraction. At day fourteen after hatching, nestlings were measured, weighed, banded with a metal band and bled. Parentage was assigned by genotype data of adults and chicks at 5 - 8 microsatellite loci, as described in Foerster *et al.* (2003) and Delhey *et al.* (2003). The primers P2-P8 (Griffiths *et al.* 1998) were used to amplify the loci CDH for assigning sex to adults and chicks.

Survival measures

We recorded the survival of eggs and nestlings as a binary variable per individual. Thus, survival until hatching was coded "0" for all unhatched eggs and "1" for all hatched eggs. Similarly, survival until fledging was coded "0" for nestlings found dead in the nest box or for those known to have hatched but that disappeared before the estimated fledging date, and "1" for all nestlings that fledged. We also recorded the local recruitment of fledglings as an estimate of local survival only, due to the lack of information about post-natal dispersers that left the study site. Fledglings that started a breeding attempt in the study area in subsequent years were coded as "1", whereas all fledglings that did not breed in the study area were coded as "0". For all offspring survival estimates we used data from individuals born between 1998 and 2005. We also recorded the "survival" of one-year-old birds (yearlings) breeding in the study area between 1998 - 2004. One-year-old breeding individuals that were caught in subsequent breeding seasons were coded as "1", whereas birds that bred only in one season as yearlings were coded as "0". This is a reasonably accurate measure of survival for male blue tits that show high fidelity to their breeding territory (Valcu and Kempenaers 2008). For females, this survival estimate is less accurate because of more frequent post-breeding dispersal (Valcu and Kempenaers 2008).

Table 1.4.
Microsatellite markers used to estimate individual heterozygosity. The loci are ordered according to their position in the zebra finch (ZF) genome. Loci with significant effects of heterozygosity at single loci (HSL) on at least one survival estimate are in bold. References for the primer sets for each locus can be found in Olano-Marin *et al.* (2010)

Locus	Location in the ZF genome (Chromosome / Start)	Category	Gene / nearest gene (Kb distance)	Number of significant effects of HSL (survival trait[a] / effect direction[b])
NPAS2	1 / 30,322,977	Functional	NPAS2	
TG01-124	1 / 32,303,776	Functional	GABRB3 (0.01)	
Pca7	1 / 81,091,062	Neutral	FAT3 (145.7)	
CcaTgu1	1 / 93,427,832	Functional	IGSF11 (0.5)	
CcaTgu2	1 / 102,481,135	Functional	CD247 (1.2)	
ApCo46	1 / 104,151,755	Functional	CADM2 (4.3)	
LEI160	1 / 109,699,352	Neutral	MIR125B (2.6)	
CcaTgu3	1 / 117,775,285	Functional	PCNP	
TG01-000	1A / 201,308	Functional	UBE2H	1 (H* / m: +, f: -)
Gf06	1A / 31,887,783	Neutral	USP15 (343.8)	1 (Y / +)
TG01-040	1A / 42,620,504	Functional	DUSP6	
CcaTgu4	1A / 53,257,466	Functional	BTBD11 (10.7)	
Tgu05	1A / 64,236,986	Functional	MPPED1 (0.03)	
CcaTgu6	2 / 16,466,870	Functional	KIAA1462 (5.1)	
Pocc6	2 / 44,644,858	Neutral	LY86 (38.7)	1 (R* / m: +, f: -)
PAT MP 2-43	2 / 47,487,652	Neutral	TMEM195 (78.2)	
CcaTgu7	2 / 75,982,449	Functional	TPPP (7.8)	
TG02-088	2 / 93,538,047	Functional	GABBR2 (2.6)	
ADCYAP1	2 / 107,400,835	Functional	ADCYAP1	2 (H*/ m: +, f: -; Y / +)
CcaTgu8	2 / 121,250,236	Functional	CRH	
CcaTgu9[a]	2_random / 1,385,550	Functional	POP1 (60.0)	
Ase18	3 / 13,906,080	Neutral	SERTAD4 (5.6)	
CcaTgu10	3 / 28,478,877	Functional	RTN4	
CcaTgu11	3 / 62,483,485	Functional	NKAIN2 (6.4)	
CcaTgu12	3 / 88,019,185	Functional	PTP4A1	
TG03-098	3 / 102,966,495	Functional	VSNL1	
ClkpolyQcds	4 / 43,440,144	Functional	CLOCK	
Pca3	4 / 64,610,435	Neutral	MAEA (5.4)	
CcaTgu13	4 / 65,195,621	Functional	FGFRL1	1 (R / -)
TG04-004	4A / 6,997,361	Functional	MMGT1	
Titgata68	5 / 14,783,223	Neutral	MUC2 (79.9)	
PK12	5 / 22,671,767	Neutral	CRY2	

Table 1.4 cont.

Locus	Location in the ZF genome (Chromosome / Start)	Category	Gene / nearest gene (Kb distance)	Number of significant effects of HSL (survival trait[a] / effect direction[b])
Asμ15	5 / 30,043,672	Functional	FMN1 (6.0)	
Mcyμ4	5 / 31,893,652	Neutral	GJD2 (1.1)	
CcaTgu14	5 / 40,815,020	Functional	NRXN3	1 (R / +)
TG05-046	5 / 50,735,925	Functional	PPP2R5C	
TG05-053	5 / 61,275,929	Functional	MDGA2	
CcaTgu15	5 / 61,276,973	Functional	MDGA2 (0.1)	
Tgu07	6 / 22,807,094	Functional	TRIM8 (0.3)	
PmaGAn27	6 / 36,100,371	Neutral	INPP5A	
Pocc1	7 / 1,646,806	Neutral	HDAC4	
Cdi31	7 / 4,201,012	Neutral	KALRN	
Pca9	7 / 10,442,876	Neutral	IHH (11.8)	
Tgu01	7 / 11,806,628	Functional	SLC4A10	
Tgu02	7 / 13,280,326	Functional	GALNT3 (4.5)	
PiJ14	7 / 20,810,935	Functional	INO80D (0.4)	
CREB1	7 / 21,321,109	Functional	CREB1	
CcaTgu16	7 / 38,131,923	Functional	RND3	
Pca4	8 / 22,128,537	Neutral	PODN (4.7)	
CcaTgu17	8_random / 1,710,707	Functional	VAMP4 (1.2)	1 (H*/ m: +, f: -)
CcaTgu18	8_random / 2,367,489	Functional	PBX1 (3.4)	
PmaTGAn45	9 / 802,719	Neutral	FOXL2 (43.4)	
MSLP4	9 / 4,935,546	Functional	ILKAP (1.3)	1 (Y / +)
TguEST09-005	9 / 5,196,903	Functional	IGF2BP2 (0.5)	
TguEST09-021	9 / 23,079,052	Functional	MYNN (2.4)	1 (H*/ m: +, f: -)
CcaTgu19	10 / 6,370,604	Functional	FAM81A	
CcaTgu20	10 / 6,994,363	Functional	CGNL1	
CcaTgu21	11 / 12,285,425	Functional	ZFHX3	
TG11-011	11 / 19,380,799	Functional	KIAA0355 (0.4)	
PmaC25	12 / 1,378,621	Neutral	IQSEC (154.2)	1 (F / +)
CcaTgu22	12 / 5,239,036	Functional	CENPP	
VeCr02	12 / 10,572,128	Neutral	PLXNA (189.3)	
TG12-015	12 / 16,288,963	Functional	FAM19A1	
TG13-017	13 / 18,208	Functional	EGR1	
TG13-009	13 / 3,672,471	Functional	RNF44	
CcaTgu23	14 / 11,485,613	Functional	TNRC18	
DkiB102	15 / 8,166,844	Functional	HIRA	
CcaTgu24	17 / 9,423,028	Functional	CAMSAP1 (1.1)	

Table 1.4 cont.

Locus	Location in the ZF genome (Chromosome / Start)	Category	Gene / nearest gene (Kb distance)	Number of significant effects of HSL (survival trait[a] / effect direction[b])
CcaTgu25	18 / 7,447,751	Functional	TNRC6C	
CcaTgu26	18 / 9,768,597	Functional	CA10 (0.07)	
DkiB119	18_random / 93,096	Functional	MAP2K6 (0.2)	1 (R*/ m: +, f: -)
CcaTgu27	19 / 10,535,947	Functional	MSI2	
Pma303	20 / 7,188,252	Neutral	TGM3 (8.4)	
PmaGAn30	20 / 14,965,400	Functional	TMEM189	
PmaGAn40	21 / 3,523,674	Neutral	SKI (43.2)	1 (R*/ m: +, f: -)
CcaTgu28	23_random / 246,698	Functional	PTP4A2	
CcaTgu29	26 / 802,761	Functional	CSDE1	
CcaTgu30	28 / 2,572,470	Functional	TMEM38A	
PK11	- / -	Neutral	-	

[a] H: Survival until hatching, F: Survival until fledging, R: Local recruitment, Y: Survival of breeding yearlings.

[b] +: positive effect, -: negative effect

* Significant effect of HSL in interaction with sex.

m: direction of the effect in males, f: direction of the effect in females.

Microsatellite markers and loci classification

We genotyped all individuals (N = 1496) with a panel of 79 autosomal microsatellites (Table 1.4). Details on marker characteristics, chromosome location in the zebra finch (Warren et al. 2010) and the chicken (International Chicken Genome Sequencing Consortium 2004) genomes, amplification and genotyping in the blue tit, and Hardy-Weinberg equilibrium and linkage disequilibrium in the blue tit population used in this study can be found in Olano-Marin *et al.* (2010) and Chapter 2 of this thesis.

We classified the markers as presumably functional or neutral, as described in Chapter 2 of this thesis. In short, loci that were designed or showed homology to zebra finch expressed sequence tags (ESTs) were considered functional, whereas markers designed with traditional cloning methods and no homology to avian ESTs were considered neutral. The characteristics of the markers in the functional and neutral categories were clearly different (Chapter 2). Diversity of

neutral loci, measured either as expected heterozygosity (*He*), number of alleles (*A*) or number of rare alleles (*R*, defined as alleles with frequency < 0.01, Kimura 1983), was significantly higher than diversity of functional ones (median *He* / *A* / *R* for neutral loci = 0.81 / 18 / 7, median *He* / *A* / *R* for functional loci = 0.43 / 6 / 3, Mann-Whitney tests: all *p*-values < 0.001). To test some predictions about the effect of heterozygosity at single loci, we also considered if the markers were located within a coding vertebrate gene, as revealed by the UCSC zebra finch genome browser (http://genome.ucsc.edu/cgi-bin/hgGateway) following the procedure described in Olano-Marin *et al.* (2010) and Chapter 2 of this thesis.

Table 2.4
Effect of individual standardized heterozygosity (SH) on estimates of survival of embryos, nestlings, fledglings and yearlings. All survival traits were analyzed with generalized mixed effects models with binomial error structure and logit link function. All the models included year as random effect, and SH, SH^2, sex, SH*sex and SH^2*sex as fixed effects. SH was centered in models with interactions. Sex was coded "0" for males and "1" for females to have males as the reference sex. Models for survival until hatching, until fledging and local recruitment included nestbox as an additional random effect. Non-significant interactions, quadratic terms and sex effects were removed. Factors in italics are presented to clarify significant interactions of SH and sex, and were calculated by fitting a model with females coded as "0" and males as "1". SE denotes standard error. Significant effects of SH and interactions with sex are shown in bold. Coefficients of determination (r^2) were calculated for terms with significant effects.

Survival measure	Model AIC	Factor	Estimate ± SE	*z*-statistic	*p*-value	r^2
Hatching (*N* = 264 eggs in 25 nests, 225 hatchlings)	222.5	Intercept (males)	1.521 ± 0.225	6.749	<0.001	-
		Intercept (females)	*2.246 ± 0.325*	*6.909*	*<0.001*	-
		$SH_{functional}$ (males)	0.068 ± 0.229	0.296	0.767	-
		$SH_{functional}$ (females)	*-0.803 ± 0.291*	*-2.760*	***0.006***	*0.029*
		Sex	0.725 ± 0.396	1.834	0.067	-
		$SH_{functional}$ * sex	-0.871 ± 0.370	-2.353	**0.019**	0.021
	229.1	Intercept	1.591 ± 1.456	1.093	0.275	-
		$SH_{neutral}$	0.160 ± 1.449	0.110	0.912	-

Table 2.4 cont.

Survival measure	Model AIC	Factor	Estimate ± SE	*z*-statistic	*p*-value	r^2
Fledging (*N* = 427 hatchlings in 45 nests, 334 fledglings)	411.8	Intercept	3.625 ± 1.331	2.724	0.006	-
		$SH_{functional}$	-1.952 ± 1.275	-1.531	0.126	-
	411.8	Intercept	-0.415 ± 1.333	-0.311	0.756	-
		$SH_{neutral}$	2.056 ± 1.336	1.539	0.124	-
Local recruitment (*N* = 609 fledglings from 61 nests, 89 recruits)	484.3	Intercept (males)	-1.327 ± 0.143	-9.279	<0.001	-
		Intercept (females)	*-2.859 ± 0.286*	*-9.988*	*<0.001*	-
		$SH_{functional}$ (males)	0.020 ± 0.131	0.154	0.877	-
		$SH_{functional}$ (females)	*-0.529 ± 0.262*	*-2.021*	***0.043***	*0.007*
		Sex	-1.532 ± 0.310	-4.935	**<0.001**	0.040
		$SH_{functional}$ * sex	-0.549 ± 0.292	-1.880	0.06	-
	484.8	Intercept	-2.571 ± 1.003	-2.564	0.010	-
		$SH_{neutral}$	1.231 ± 0.981	1.254	0.210	-
		Sex	-1.449 ± 0.291	-4.973	**<0.001**	0.041
Yearling survival (*N* = 540 breeding yearlings, 276 bred in subsequent years)	748.2	Intercept	0.008 ± 0.689	0.011	0.991	-
		$SH_{functional}$	0.269 ± 0.676	0.398	0.691	-
		Sex	-0.486 ± 0.174	-2.795	**0.005**	0.014
	744.9	Intercept	-1.160 ± 0.778	-1.491	0.136	-
		$SH_{neutral}$	1.443 ± 0.769	1.875	0.061	-
		Sex	-0.473 ± 0.174	-2.711	**0.007**	0.014

Data analysis

We calculated standardized heterozygosity (SH) (Coltman *et al.* 1999) as a measure of individual heterozygosity at multiple loci, using the R function GENHET (Coulon 2010). We used the genotype data from all markers, and separately from each of the subgroups of functional and neutral loci to calculate this index. Heterozygosity at single loci (HSL) was coded as a binary variable, with "1" representing a heterozygous state and "0" a homozygous one.

We used a full-sibling comparison to examine the relationship between heterozygosity and egg, nestling and fledgling survival. To this end, we genotyped all the chicks and embryos from all nests that (a) had no extra-pair young and (b) showed partial brood mortality until at least one of the analyzed life-stages. We used generalized linear mixed models with binomial error structure and logit link function (Pinheiro and Bates 2000) in the free software R (R Development Core Team 2008) with the add-on R package *lme4* (Bates and Maechler 2010). All models included survival as the response variable, nest identity and year as random effects, and SH, SH^2, sex, SH*sex and SH^2*sex as predictors. Interactions, quadratic effects of SH and main effects of sex were removed in a stepwise backwards procedure and were retained only when significant. We ran separate models for SH measured with functional and neutral markers. We standardized the effect size of SH on each survival measure by converting it to *r*, the equivalent of the Pearson product moment correlation coefficient (Coltman and Slate 2003). By using only full-siblings, our aim was to compare the effect of heterozygosity on the survival of individuals with the same level of inbreeding by pedigree. However, as mixed models do not analyze exclusively the variation within the random factor (in our case, within-nests), we used the approach proposed by van de Pol and Wright (2009) to investigate within- and between-nest effects of heterozygosity on the survival traits where significant effects of SH were found. As the significant effects of SH on early survival occurred with functional markers in females only, we only used data of $SH_{functional}$ of females for these analyses. To this end, we fitted 1) a model with survival as the response variable, $SH_{functional}$ as predictor, and nest identity and year as random effects (model 1), and 2) a model with the individually centered values of $SH_{functional}$ within a nest (for within-nest effects) and the mean $SH_{functional}$ of each nest (for between-nest effects) as predictors, and nest identity and year as random effects (model 2, van de Pol and Wright 2009). A stronger statistical

support of the second model (e.g. a lower AIC value) would indicate that the effects of SH are differentially driven by within- and between-nest effects.

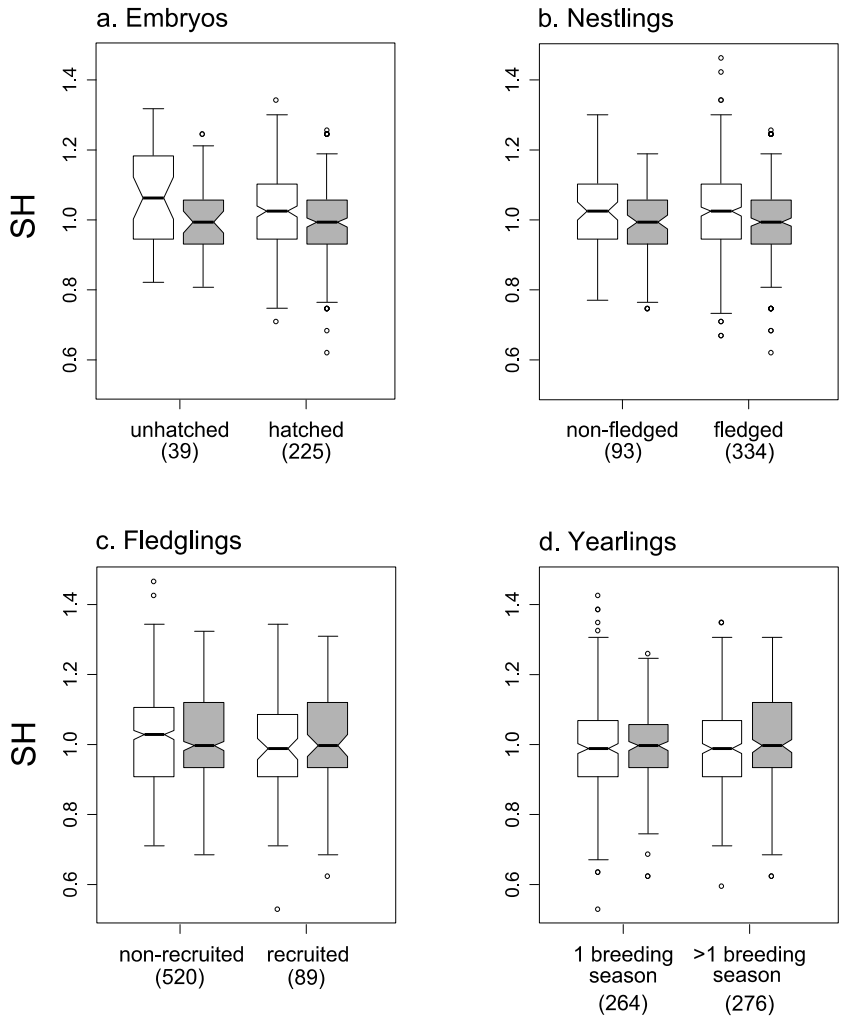

Figure 1.4.
Standardized heterozygosity of the a. embryos, b. nestlings, c. recruits and d. yearlings. White boxes for $SH_{functional}$ and gray ones for $SH_{neutral}$. Numbers in parenthesis indicate the sample size.

To analyze the effect of heterozygosity on survival of one-year-old breeding birds we used generalized linear mixed models with binomial error structure and logit link function. In contrast to the analyses of early survival, we did not use a full-sibling comparison and included all local and immigrant one-year-old

breeding birds. As above, full models included survival as the response variable, year as random effect, and SH, SH^2, sex, SH*sex and SH^2*sex as predictors. We simplified the full model as described above.

We analyzed the effect of HSL on the survival of eggs, nestlings, fledglings and one-year-old blue tits by fitting one model per locus. Each survival estimate was used as the response variable, HSL as predictor, and year and nest identity (the latter only for the analyses of eggs, nestlings and fledglings) as random effects, adding up to a total of 79 models of HSL per survival measure. In the analysis of egg survival until hatching and survival of yearlings, where a significant interaction of SH and sex was found, we fitted models of HSL with and without the interaction term. We investigated the predominance of negative versus positive effects of HSL, and negative vs. positive interactions of HSL and sex with cumulative binomial tests assuming equal probabilities (0.5) for both types of associations. In order to correct for multiple testing, we used cumulative binomial tests to examine if the probability of obtaining significant effects or significant interactions of HSL with sex exceeded the expected probability of false positive associations, a (with 79 loci, the expected number of false positives per survival measure is about 4). We used χ^2 tests or Fisher exact tests to investigate associations of positive / negative or significant / non-significant effects and interactions of HSL and sex with the functional or neutral category of loci. We tested for the influence of marker diversity on the effect size of HSL or the interaction of HSL and sex by fitting linear models with effect size as the response variable and the loci's *He* as predictor. We tested for local effects on survival of one-year-old breeding birds with the approach described in Szulkin *et al.* (2010), by 1) regressing survival on MLH (*m1*), 2) regressing survival on all "normalized" single locus heterozygosities in one model (*m2*), and 3) testing for differences in the variance explained by the two models with an *F*-ratio test. We included year as co-factor in *m1* and *m2*. Since the number of independent observations needs to be considerably larger than the number of loci, the sample of full-siblings used for the analysis of early survival was not suitable for this test.

Results

Multilocus heterozygosity measures

$SH_{functional}$ and $SH_{neutral}$ of the embryos, nestlings, recruits and yearlings used for this study are shown in Figure 1.4. Within nests ($N = 98$), the mean SH measured with all markers ranged from 0.8683 to 1.1470, with standard deviations varying between 0.032 and 0.123 (mean = 0.068). Within chicks, the correlation of SH measured with the subset of functional and neutral markers was low and non-significant ($r = 0.026$, $p = 0.404$). For the one-year-old breeding blue tits ($N = 540$), the correlation coefficient of SH measured with the two subsets of markers was 0.046 ($p = 0.283$).

Table 3.4.
Within- (w-n) and between-nests (b-n) effects of $SH_{functional}$ on survival of females until hatching and local recruitment. See Material and Methods for further details.

Survival measure	Model	AIC	Factor	Estimate ± SE	*z*-statistic	*p*-value
Hatching (*N* = 130 eggs in 25 nests, 115 hatchlings)	Model 1	92.66	Intercept	8.891 ± 2.580	3.446	<0.001
			$SH_{functional}$	-6.476 ± 2.347	-2.760	**0.006**
	Model 2	93.45	Intercept	11.390 ± 3.475	3.278	0.001
			$SH_{functional}$ w-n	-4.151 ± 3.142	-1.321	0.186
			$SH_{functional}$ b-n	-8.899 ± 3.238	-2.749	**0.006**
Recruitment (*N* = 264 fledglings from 60 nests, 16 recruits)	Model 1	123.5	Intercept	1.111 ± 2.149	0.517	0.605
			$SH_{functional}$	-4.086 ± 2.193	-1.864	0.062
	Model 2	125.4	Intercept	0.451 ± 3.616	0.125	0.901
			$SH_{functional}$ w-n	-4.501 ± 2.904	-1.550	0.121
			$SH_{functional}$ b-n	-3.441 ± 3.602	-0.955	0.339

Effects of heterozygosity at multiple loci on survival

We found different effects of SH measured with functional and neutral markers on the early survival of male and female blue tits in our population (Table 2.4, Figure 2.4). Sex influenced significantly the probability of local recruitment, with male fledglings more likely to recruit. The probability of hatching and fledging, on the other hand, was not different for male and female embryos or nestlings. $SH_{functional}$ showed significant negative correlations with survival until hatching and with the probability of local recruitment of females. Survival until hatching and the probability of local recruitment of males, on the other hand, were not significantly influenced by SH. Survival of male and female nestlings until fledging was not significantly influenced by heterozygosity at functional or neutral markers. The centered models, fitted to discriminate effects of heterozygosity within a nest (e.g. between full-siblings, sharing the same inbreeding coefficient) or between nests (e.g. between individuals with different levels of inbreeding) were not better supported by lower AIC values than the non-centered models (Table 3.4). Therefore, separating the effects of $SH_{functional}$ between full-siblings and between individuals with different inbreeding coefficients is not justified.

As expected, survival of breeding yearlings was significantly influenced by sex, with males having a higher probability of breeding in the population as adults in subsequent years (Table 2.4). $SH_{neutral}$ had a positive, almost significant effect on survival of both males and females. Interestingly, when heterozygosity was measured as homozygosity by locus (HL) (Aparicio *et al.* 2006), an estimator that correlates better with the inbreeding coefficient and with genome-wide heterozygosity, the effect of heterozygosity at neutral markers became significant ($p = 0.03$; details not shown). SH at functional loci had no significant effects on survival of one-year-old birds (Table 2.4).

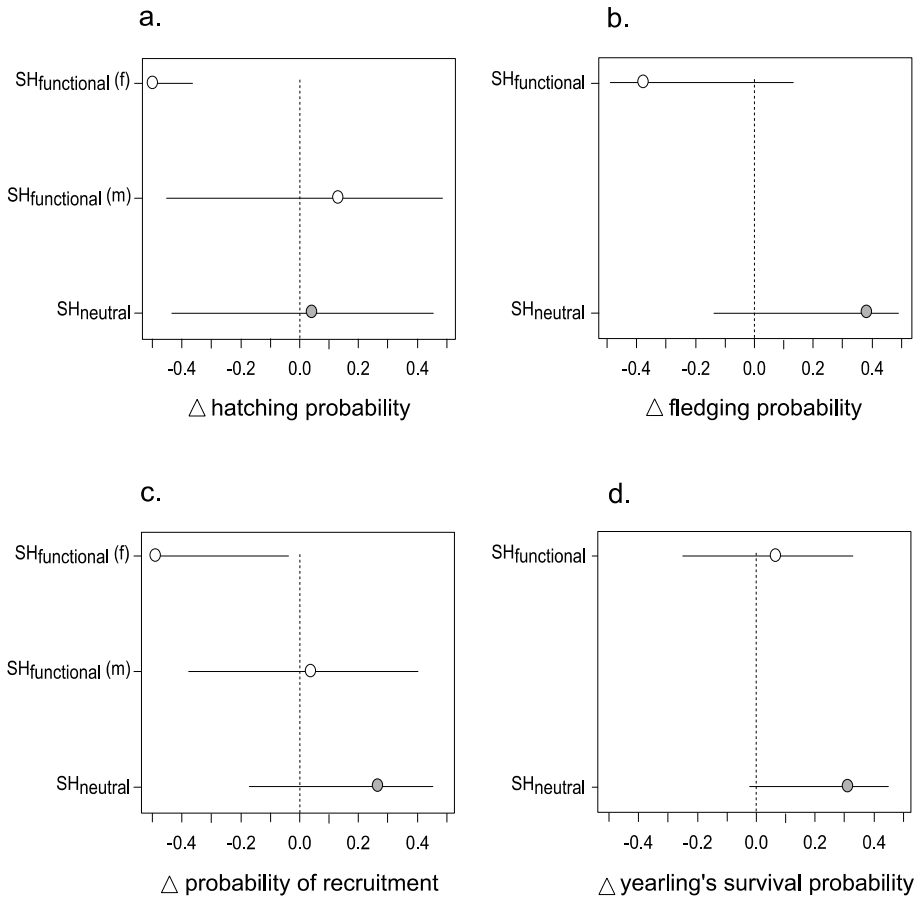

Figure 2.4.
Effect sizes of individual standardized heterozygosity (SH) on estimates of early and adult survival. Effect sizes are the back-transformed "Estimate" values of SH from Table 2.4 with their 95% confidence intervals, and represent a deviation from the random probability of survival (0.5) until a certain life-stage due to an increase in one unit of SH. Effect sizes of markers with a significant interaction of SH and sex are shown separately for males (m) and females (f). a. Survival until hatching; b. Survival until fledging; c. Local recruitment; d. Survival of breeding yearlings.

Effects of heterozygosity at single loci (HSL) on survival

We observed between 34 and 40 positive effects of HSL (out of *ca.* 79) on the survival estimates (details not shown). The probability of obtaining positive or negative effects, independent of its significance, was not different from 0.5. Negative effects on local recruitment were associated with functional loci (χ^2 test: $p = 0.02$), whereas for the other survival estimates the direction of the effects did not differ significantly between functional and neutral loci. For all survival traits,

the direction of the effects and the mean absolute effect size of HSL did not differ significantly for loci located within or outside a vertebrate gene. We observed between 0 and 3 significant effects of HSL on each measure of survival, which is lower than the expected number of false positives (*ca.* 4 / survival trait). For survival until hatching and local recruitment, we observed 4 and 3 significant interactions of heterozygosity and sex, respectively; in all cases, heterozygosity in females had a stronger negative effect. Significant main effects of HSL or significant interactions of HSL with sex were not associated with the functional or neutral category of the markers. The diversity of the loci (measured as *He*) was not correlated with the effect size of HSL or with its interaction with sex. Following the procedure described in Szulkin *et al.* (2010) for the analysis of survival of breeding yearlings, we did not find a significant difference in the variance explained by the model with HSL for all loci (*m2*) compared to the model with MLH (*m1*) ($F_{[399,321]} = 0.74$, $p = 0.95$); therefore, the partial regression coefficients of all the loci on survival of one-year-old birds were not different from each other.

Discussion

We found different effects of SH (as a measure of heterozygosity at multiple loci) on survival of blue tits, depending on marker type, life-history stage and sex. On the one hand, $SH_{functional}$ had a negative effect on early survival of females, but not of males. $SH_{neutral}$, on the other hand, had a positive effect on survival of adult birds of both sexes. Due to the characteristics of the markers and to the fact that we controlled for the level of inbreeding in the comparisons, the negative effects of functional loci on early survival are better explained by local effects, whereas the positive effects of neutral markers on survival of yearlings could reflect inbreeding effects in the population.

Functional vs. neutral markers

We *a priori* classified the microsatellite markers used in this study as "neutral" or "functional" based on whether they were located in expressed regions of the genome. Interestingly, this classification captured differences that would be expected to occur between groups of loci with genuine differences in functionality (e.g. number of alleles, expected heterozygosity). Neutral markers

are expected to reflect general phenomena affecting the whole genome (like inbreeding or outbreeding), whereas functional markers can be constrained to do so, possibly due to selection acting on them. In fact, we showed previously that variation in inbreeding in this population of blue tits is reflected by heterozygosity of neutral loci, but not by heterozygosity of functional loci alone (Chapter 2). Moreover, the individual information given by multilocus heterozygosity measured with the two groups of markers and its relationship with fitness-related traits is different (Chapter 2). Therefore, it is interesting to consider the characteristics of the markers used to measure multilocus heterozygosity when interpreting the results of HFC (e.g. Kupper *et al.* 2010).

Effect of heterozygosity on early survival

We found negative effects of $SH_{functional}$ on the probability of hatching and local recruitment of female embryos and fledglings, respectively, using a full-sibling approach. As we analyzed these HFC with mixed models (fitting nest as a random factor), we cannot attribute the effects solely to the comparison between surviving and non-surviving nest mates, but instead have to include the comparison between surviving and non-surviving birds from different nests (Fossoy *et al.* 2009; van de Pol and Wright 2009). Within-nest effects (i.e. between individuals sharing the same inbreeding coefficient) are independent of the level of inbreeding and are based on differences in heterozygosity between full-siblings that result from the random assortment of parental chromosomes (Hansson *et al.* 2001; Hansson *et al.* 2004). In this case, local/direct effects are used to explain significant HFC (Hansson and Westerberg 2008). Between-nests effects, on the other hand, occur between individuals with different levels of inbreeding and therefore, can be attributed to genome-wide effects (i.e. inbreeding or outbreeding depression). In this study, modelling within- and between-nest effects in the relationship between $SH_{functional}$ on survival of female chicks was not considerably better than considering both effects jointly. Therefore, the negative HFC could be attributed to effects at single loci or to outbreeding depression. However, for HFC to reflect the effects of outbreeding, multilocus heterozygosity measured with marker loci should represent genome-wide heterozygosity and the level of individual inbreeding (or outbreeding). We have demonstrated that heterozygosity at multiple loci measured with our set of functional loci does not reflect genome-wide heterozygosity in this blue tit population (Chapter 2). Therefore, the negative effects of $SH_{functional}$ on the early

survival of female blue tits are better explained by direct or local effects at single loci rather than by the general negative effects of outbreeding. As we did not find more significant effects than expected by chance in our analyses of HSL, we would attribute the observed effect of SH to small contributions of many loci.

In the absence of outbreeding depression and underdominance, loci under directional selection can produce signals of negative HFC via codominant or recessive allele advantage (Mueller *et al.* 2010). These mechanisms are expected to be important as causes of multilocus HFC if populations of large effective size are subject to major recent ecological changes or to strong fluctuating selection (Bell 2010; Mueller *et al.* 2010). Blue tits in natural populations have shown signals of adaptation to a rise in temperature in recent years (Smallegange *et al.* 2010). Furthermore, density-dependent effects caused by habitat heterogeneity (Dhondt *et al.* 1992) can result in fluctuating selection in this species. In this scenario, functional loci under selective pressures are expected to give stronger HFC signals than neutral markers (Mueller *et al.* 2010). This supports our findings of significant negative effects of $SH_{functional}$, but no effects of $SH_{neutral}$, on hatching and recruitment, and the association of negative effects with functional markers in the analysis of HSL on local recruitment.

Our results are in strong contrast with other bird studies, including some in blue tits, that have reported positive effects of individual multilocus heterozygosity, level of inbreeding or genetic similarity between the parents on hatching (Kempenaers *et al.* 1996; Krokene and Lifjeld 2000; Blomqvist *et al.* 2010; Taylor *et al.* 2010) and recruitment (Hansson *et al.* 2001; Foerster *et al.* 2003; Hansson *et al.* 2004; Markert *et al.* 2004; Jensen *et al.* 2007). Negative effects of heterozygosity on early survival are, in fact, scarce in the literature (Coulson *et al.* 1998; Marshall and Spalton 2000; Dibattista *et al.* 2008) and, to our knowledge, have not been reported yet for any bird species. This can be due to a publication bias towards positive results (Coltman *et al.* 1999; Chapman *et al.* 2009), to the type of markers used to measure heterozygosity, to demographic factors of the populations studied and/or to the study design. Kupper *et al.* (2010) recently proposed that the preferential use of non-exonic, highly polymorphic markers may contribute to the prevalence of positive results. In fact, in the great majority of HFC studies, heterozygosity is measured with markers that were used for paternity analysis and were thus selected for their level of polymorphism. Supporting this statement, Dibattista *et al.* (2008) reported significant negative effects of heterozygosity on survival of lemon sharks (*Negaprion brevirostris*) with

their least variable loci. Here, negative effects of SH on early survival of blue tits were also found with our least variable group of loci. The demographic history of the populations and the study design are important factors to consider when interpreting HFC (Chapman *et al.* 2009). The publication of mostly positive HFC can be due to a bias towards study species where high variance in individual inbreeding occurs and/or to the selection of individuals with varying levels of inbreeding (e.g. Jensen *et al.* 2007; Taylor *et al.* 2010). In these situations, positive effects of heterozygosity due to inbreeding depression are likely to be detected. Full-sibling designs, as the one used here for the analysis of early survival, are suitable for the detection of local effects, which are not constrained to be positive. In summary, the negative effects of multilocus heterozygosity reported here do not contradict or invalidate previous findings of positive effects of genetic diversity on early survival of blue tits (Kempenaers *et al.* 1996; Krokene and Lifjeld 2000; Foerster *et al.* 2003) and other bird species. More studies in which the level of inbreeding is controlled for and heterozygosity is measured with less variable loci are needed to establish the importance of these factors for the detection of negative HFC.

In a previous study on the same population of blue tits used here, Foerster *et al.* (2003) reported that fledglings that recruited locally were more heterozygous than their non-recruited nestmates. In this case, the differences with our results can be attributed to the reduced dataset ($N = 26$ pairs of recruited vs. non-recruited fledglings, mostly males), the smaller panel of neutral microsatellites used to measure multilocus heterozygosity (5 - 7 loci), and the inclusion of extra-pair young (which adds variance in the individual inbreeding coefficient within nests) in the study of Foerster *et al.* (2003). It is important to note that local recruitment is a measure that can be affected by both survival and natal dispersal. Therefore, the negative effects of heterozygosity on the probability of local recruitment can be interpreted either as negative effects on survival, positive effects on dispersal, or a combination of both. We cannot distinguish between these effects due to the methodological difficulties associated with estimating survival and dispersal of fledglings.

We found that SH had a negative effect on early survival of female blue tits, but no effect was detected in males. Differences in the strength and direction of HFC between sexes have been reported previously (Coulson *et al.* 1999; Rossiter *et al.* 2001; Foerster *et al.* 2003). Similarly, sex-differential effects of inbreeding have been found in a variety of species (Saccheri *et al.* 2005; Charpentier *et al.*

2006; Fox *et al.* 2006; Reid *et al.* 2007; Rioux-Paquette *et al.* 2011). In most cases, negative effects of inbreeding (or positive effects of outbreeding) were found and, with the exception of the studies by Rossiter *et al.* (2001) and Saccheri *et al.* (2005), females were more likely to be affected than males. Sex-related differences in traits such as early mortality (Fox *et al.* 2006), expression of deleterious alleles (Fox *et al.* 2006), trade-offs in allocation of resources (Fox *et al.* 2006; Reid *et al.* 2007), maternal investment (Charpentier *et al.* 2006), food acquisition (Charpentier *et al.* 2006) and growth strategies (Coulson *et al.* 1999) have been proposed as mechanisms to explain the differences in the effects of heterozygosity or inbreeding between males and females. The mechanism by which heterozygosity affects negatively the early survival of females, but not males, in our population remains to be explored.

Effect of heterozygosity on yearling survival

We found a positive, close to significant effect of heterozygosity at multiple loci, measured as SH, with neutral markers on survival of one-year-old breeding blue tits. Interestingly, when heterozygosity was measured as homozygosity by locus (HL) (Aparicio *et al.* 2006), the effect of neutral loci on survival became significant. It has been described that, in simulated populations subjected to migration and admixture, HL is an estimator that correlates better with the inbreeding coefficient and with genome-wide heterozygosity (Aparicio *et al.* 2006). Since we used individuals with different levels of inbreeding for this analysis, measuring multilocus heterozygosity as HL increased the power to detect HFC. Females appeared to have lower survival probabilities because, compared to males, they are more likely to disperse (in this population, 25% of the surviving females dispersed after breeding vs. only 3% of the males, Valcu and Kempenaers 2008). Foerster *et al.* (2003) analyzed the effect of heterozygosity measured with 5-7 presumably neutral markers on survival of one-year-old breeding males and females from the same population used here. Multilocus heterozygosity had a positive effect on female, but not on male survival (Foerster *et al.* 2003). In contrast, the effect on survival we reported here occurs in both females and males. Therefore, with an extended dataset and a larger panel of neutral loci we confirmed the effect of multilocus heterozygosity on survival of females reported previously (Foerster *et al.* 2003) and, additionally, extended it to males. However, a further increase in the number of markers with a set of presumably functional loci did not increase the power to detect the effect of

multilocus heterozygosity on survival. The same pattern was found when the effect of heterozygosity on reproductive success traits was investigated in the same blue tit population (Chapter 2): heterozygosity at neutral loci, but not at functional ones, had a positive significant effect on clutch size and on the number of eggs sired by males. As neutral markers seem to give some information on genome-wide heterozygosity (Chapter 2), positive effects found with neutral heterozygosity can be interpreted as an indirect measure of the effects of inbreeding in this population.

We investigated if the effect of heterozygosity at multiple loci on survival of yearling breeding birds was due to large effects at few loci or to small effects at many loci. The analyses of HSL did not result in more significant effects than expected by chance and did not reveal larger absolute effects in markers located within a gene, nor for associations of significant effects with markers within genes or with expressed loci. Moreover, the test for local effects proposed by Szulkin *et al.* (2010) does not support a scenario where effect sizes of some loci are significantly different from the effects at other loci. As expected with inbreeding effects, our results are better explained by the combined effect of many loci.

Conclusion and implications

We identified both local and general effects of heterozygosity on survival of embryos, nestlings, fledglings and yearling blue tits in our population. The contrasting effects of heterozygosity of functional and neutral markers at different life stages reveals the limitations of generalizing the results when only few fitness traits are investigated and when the characteristics of the markers are not considered. As the overall effects of heterozygosity are complex (i.e. both positive and negative effects can occur) and context-dependent (i.e. differing between sexes, environments, traits and loci) it is difficult to predict the optimal heterozygosity for individuals at different loci in different environmental settings. This can have important implications in studies of mate choice, genotype-phenotype associations and conservation strategies based on genetic data.

General discussion

In this thesis I thoroughly investigated the causes of correlations between heterozygosity and fitness-related traits in blue tits. I developed one of the most comprehensive microsatellite-based genotypic datasets (Chapter 1) for a long-term study population. The amount and characteristics of the genotypic and phenotypic data and the versatile study designs and analysis methods that I employed allowed me to test the main hypotheses proposed to explain the occurrence of heterozygosity-fitness correlations (HFC) (Chapters 2 and 3). In Table 1.5 I present a summary of the traits considered, the study designs employed, and the main findings and conclusions of this work.

Table 1.5
Overview of study design, fitness-traits considered, main results and proposed explanations for HFC in the blue tit

Individuals sampled	Design	Fitness-related trait		Markers with significant multilocus HFC (direction)	Most supported explanation
Breeding birds	Without control for inbreeding level	Reproductive success	Clutch size	Neutral (+)	General effects (inbreeding)
			Number of sired eggs	Neutral (+)	
			Hatching success	None	
			Fledging success	None	
			Recruitment success	Neutral (+) / Functional (+)*	
		Survival of yearlings		Neutral (+)	
Embryos, nestlings, fledglings	Full-sibling (controlling for inbreeding level)	Survival	of embryos	Functional (-)*	Localized effects at multiple loci
			of nestlings	None	
		Local recruitment		Functional (-)*	

*: sex-differences in the effects

The importance of measuring individual heterozygosity with a high number of genetic markers is well acknowledged in the HFC literature (Balloux *et al.* 2004; Slate *et al.* 2004; DeWoody and DeWoody 2005), but rarely taken into

practice (Coltman and Slate 2003; Chapman *et al.* 2009). Similarly, although it is recognized that microsatellites can be functional (Li *et al.* 2004), neutrality of the markers used in HFC studies is usually assumed, even though the power to detect effects and/or the interpretation of the results (Hansson and Westerberg 2002) may be influenced by the function of the loci. These facts motivated and justified the development of a large panel of molecular markers and the consideration of their putative function for investigating HFC. Therefore, I presented a set of at least 95 polymorphic markers with utility for the blue tit and with potential use for other bird species (Chapter 1). I demonstrated that EST libraries are useful for developing molecular markers, and that whole-genome sequences from related species can be used to roughly predict the genomic position of the loci in the species of interest. The source of a marker can provide hints about the putative function of the locus in the absence of direct evidence. I used this criterion to classify the markers as presumably functional or neutral, and found differences (in the number of (rare) alleles, heterozygosity and distance to genes) that would be expected to occur between markers with genuine differences in functionality (Chapter 2). The results obtained in the analyses of HFC and identity disequilibrium (ID) (Chapters 2 and 3) further supported the classification of the loci and justified to perform separate analyses with the two groups of markers.

The study design is important when investigating HFC. High variance in the level of inbreeding is required to detect general effects of heterozygosity (Balloux *et al.* 2004; Szulkin *et al.* 2010), whereas controlling for the inbreeding coefficient is useful to detect localized (local or direct) effects (Hansson *et al.* 2004; Hansson and Westerberg 2008). Therefore, I used both approaches to investigate HFC in the blue tit population. In Chapters 2 and 3 I used reproductive and survival data from breeding birds without controlling for the level of individual inbreeding and found significant positive effects of multilocus heterozygosity of neutral markers on different fitness-related traits. In contrast, using survival data from full-siblings, I found significant negative effects of multilocus heterozygosity of functional markers (Chapter 3). The characteristics of the study design suggest that the first HFC are caused by negative effects of inbreeding, whereas the second HFC are caused by local or direct effects of few loci. Yet, additional analyses are necessary to validate this interpretation.

Attributing significant correlations between heterozygosity and fitness-related traits to inbreeding, or disregarding the effects of inbreeding on such

correlations, is not straightforward and should be done with caution. This is especially true when HFC are studied in populations where the occurrence of inbreeding is difficult to document and, therefore, the fitness and level of heterozygosity of inbred and outbred individuals cannot be directly compared. In this thesis I attempted to document the two main causes of inbreeding (within- and between-population, Keller and Waller 2002) in the studied blue tit population by reconstructing the pedigrees with molecular data and by investigating genetic population structure (Chapter 2). The great majority (80.5%) of the breeding individuals in the population were immigrants (with an assumed inbreeding coefficient of 0) and most pedigrees were shallow. Nevertheless, I could identify that in 2% of the pairs that produced young, the male and the female were related. I also detected a fine-scale population structure of the breeding individuals in the study area. The subpopulations found, however, did not seem to represent genetically isolated groups where genetic drift occurred and different sets of deleterious alleles became fixed. Instead, the structure was detected only in males and could be attributed to their low rates and distances of dispersal (Greenwood 1980). Although this type of genetic structuring does not correspond to the definition of between-population inbreeding (Keller and Waller 2002), it may still contribute to the occurrence of inbreeding events. Since not all young females disperse away of their natal area and adult females are relatively faithful to their breeding territories (Valcu and Kempenaers 2008), the detected structuring of males in the population can promote the coupling of related individuals. Consanguineous matings and the existence of a population structure that can contribute to inbreeding indicate that, in the studied blue tit population, variance in the individual inbreeding coefficient could generate HFC. Under these circumstances, identity disequilibrium (ID) occurs and marker heterozygosity should reflect genome-wide heterozygosity. Therefore, I tested to which extent all the markers used for this study, as well as the subsets of functional and neutral loci, reflect a general state of the genome (Balloux *et al.* 2004; David *et al.* 2007; Szulkin *et al.* 2010). I found a significant degree of ID when heterozygosity was measured with all markers or with neutral loci only, but not when measured with functional loci. Altogether, the above results indicate that the significant positive HFC found between neutral markers and reproductive success traits and survival of breeding birds could be caused by the general effects of inbreeding (Chapters 2 and 3). On the other hand, the negative effects of heterozygosity at functional markers on

early survival and local recruitment of full-siblings may be better explained by localized effects than by outbreeding depression (Chapter 3).

Positive effects of multilocus heterozygosity on reproductive traits in the blue tit have been reported previously in the same population that was studied for this thesis (Foerster *et al.* 2003) and in a Spanish population (Garcia-Navas *et al.* 2009). In both of these studies, a smaller set of neutral microsatellites was used. It has been suggested that less than 10 markers are sufficient to reveal inbreeding depression if there is variance in the level of inbreeding and large numbers of individuals are sampled (Bierne *et al.* 2000; Lesbarreres *et al.* 2007; Townsend *et al.* 2009). Therefore, population processes similar to the ones described in this thesis, causing variation in the individual level of inbreeding, may be widespread in the blue tit. This is somewhat surprising, as inbreeding is not anticipated in large, open populations of common species. It would be interesting to investigate if similar processes occur in species where significant correlations between heterozygosity at few markers and fitness-related traits have been described.

Negative effects of multilocus heterozygosity on early survival, as the ones found in Chapter 3, have not been described before for blue tits or other bird species. In contrast, positive effects of genetic diversity on early survival have been reported in at least three blue tit studies (Kempenaers *et al.* 1996; Krokene and Lifjeld 2000; Foerster *et al.* 2003) and in several other bird species (Hansson *et al.* 2001; Hansson *et al.* 2004; Markert *et al.* 2004; Jensen *et al.* 2007; Blomqvist *et al.* 2010; Taylor *et al.* 2010). I have argued that the study design (i.e. full-sibling approach) and the type of markers used (i.e. functional) can explain the negative effects described in Chapter 3 and the contrast with the results found in other bird studies. More investigations where the level of inbreeding is controlled for and heterozygosity is measured with EST-derived microsatellites are needed to validate the importance of these factors for finding negative HFC. On the other hand, if the negative effects of parental genetic similarity on hatching success (Kempenaers *et al.* 1996; Krokene and Lifjeld 2000) or the positive effects of individual heterozygosity on recruitment (Foerster *et al.* 2003) that were reported for blue tits are indeed caused by inbreeding, analyzing hatching and recruitment success in individuals that show sufficient variance in the level of inbreeding (e.g. extra- vs. within-pair young, young from different nests) should reveal positive HFC with the set of neutral markers.

There is an increasing tendency in HFC studies to test for localized effects of heterozygosity at single loci separately (e.g. Lieutenant-Gosselin and Bernatchez 2006; Luikart *et al.* 2008; Da Silva *et al.* 2009; Hoffman *et al.* 2010a). When there are no clear *a-priori* expectations about which markers should have strong HFC and when multiple testing is not corrected for, this approach has an exploratory character and significant results at single loci should be interpreted cautiously. Knowledge about the type of selection affecting each locus could also be helpful to formulate predictions about the strength and direction of HFC at single markers (Mueller *et al.* 2010). In Chapters 2 and 3 of my thesis I have tested for the effect of heterozygosity at single loci on reproduction and survival of blue tits. Contrary to what might be expected, my results did not confirm that markers classified as functional, or markers located within a gene, show stronger effects than neutral or out-of-gene markers. Moreover, the number of significant tests (per trait and in total) did not exceed the expected number of false positives. Additionally, when heterozygosity of a single marker showed significant correlations with more than one fitness-related trait, the direction of the effect differed between traits. This illustrates the problematic of generalizing the relevance and direction of significant single-locus effects for fitness measured as a whole. As pointed out by Szulkin *et al.* (2010), using single-locus HFC to identify the genes that contribute to variation in a given phenotype is not very effective. Even if multilocus HFC reveal the combined action of genes causing inbreeding depression, the low coverage of the genome in HFC studies limits the likelihood of identifying the multiplicity of loci affecting a complex phenotype. For these reasons, I did not propose any of the loci used for this study as a strong candidate for causing variation in fitness-related traits in the blue tit population.

Findings and conclusions of HFC studies can have an important impact on the field of mate choice (Kempenaers 2007). The positive effects of neutral heterozygosity on fitness-related traits found here (i.e. in clutch size, number of sired eggs, offspring recruitment success, survival of yearlings; Chapters 2 and 3) imply that a bird would gain non-additive genetic benefits for its offspring by choosing a non-related mate. In previous studies on the same population, however, inbreeding avoidance by females choosing unrelated social or extra-pair partners was not detected (Foerster *et al.* 2003; Foerster *et al.* 2006). The costs of inbreeding avoidance may exceed the costs of inbreeding or, alternatively, these findings could reflect a lack of mate choice opportunities at the moment of pair formation (Foerster *et al.* 2003). Interestingly, in broods of mixed paternity, a

fertilization bias towards less related copulation partners was described (Foerster *et al.* 2006), suggesting that blue tits in this population could reduce the costs of inbreeding through a postcopulatory mechanism. The positive correlation between neutral heterozygosity and fitness traits that I found also implies that more heterozygous (i.e. less inbred) individuals are of higher quality. Males could benefit from the higher fecundity of heterozygous females (Foerster *et al.* 2003; Garcia-Navas *et al.* 2009), and females, in turn, could benefit from the higher feeding rates provided by heterozygous males (Garcia-Navas *et al.* 2009). Hence, both males and females would increase their probability of raising recruits by choosing a more heterozygous individual as a social partner (Chapter 3). In a study on a Spanish blue tit population, Garcia-Navas *et al.* (2009) found that heterozygosity was positively correlated between social partners and suggested that such a pattern of assortative mating may be maintained by direct (quality of the social partner) as well as indirect (by increasing heterozygosity of the offspring) benefits. Unfortunately, the benefits of individual heterozygosity on the offspring themselves was not directly investigated (Garcia-Navas *et al.* 2009). In contrast, in the blue tit population used for this thesis, heterozygosity between social partners was not correlated (r MLH_{male}-MLH_{female} < 0.001, $p = 0.982$, unpublished data). The expected offspring heterozygosity was highly correlated with both mother and father heterozygosity (r MLH_{father}-$MLH_{offspring} = 0.25$, $p < 0.001$; r MLH_{mother}-$MLH_{offspring} = 0.22$, $p < 0.001$, unpublished data). However, the negative effects of individual heterozygosity on early survival of females (Chapter 3) reveal a complex situation in which an increase in offspring heterozygosity is not necessarily beneficial, at least during early life stages. Therefore, heterozygous birds may experience a trade-off between the direct benefits of choosing a heterozygous partner and the disadvantages of producing heterozygous offspring that will have lower probabilities of survival early in life. I found that clutches without extra-pair young tend to have lower offspring heterozygosity than expected (unpublished data), but further analyses are needed to determine if this trade-off exists and how it can be resolved.

Conclusions and further directions

One of the main contributions of my thesis to the field of HFC is that it demonstrated the importance of considering the putative functionality of the

microsatellites used to measure individual heterozygosity. Markers classified as functional were different from neutral markers not only in locus-specific characteristcs (e.g. diversity), but also in the power to detect correlations with fitness traits and identity disequilibrium. I showed that classifying markers can improve the interpretation of significant HFC and reveal correlations that would otherwise remain hidden. Therefore, for future HFC studies I strongly recommend to consider the functionality of the markers employed.

I have demonstrated the difficulties involved in identifying loci that are responsible for variation in fitness by performing HFC. I found that the effects of heterozygosity at single loci are not predicted by the putative function of the loci and that the strength and direction of the effects vary according to the sex of the individuals and to the fitness trait considered. Therefore, HFC should not be seen as the "easy" way to map the loci responsible for a given phenotype.

The HFC literature is full of evidence that demonstrates the existence of correlations between heterozygosity and phenotypic traits (Coltman and Slate 2003; Chapman *et al.* 2009). However, much work needs to be done until we understand the population characteristics that can promote such correlations. In this thesis I have shed light on the population processes that could contribute to the occurrence of HFC in an open, natural population of a common and highly mobile species. Contrary to what may be expected for such a population, I found that certain conditions can generate more-than-expected variance in the individual level of inbreeding. In order to push forward this field of study, researchers should carefully investigate the phenomena causing HFC in their study species and populations.

Summary

The fitness consequences of genetic diversity are a fundamental area of study in evolutionary and conservation biology and constitute a topic of primary interest for animal and plant breeders. Correlations between individual heterozygosity at multiple loci and fitness-related traits, known as heterozygosity-fitness correlations or HFC, have been used to study the relationship between selection and genetic variation and to explore the consequences of inbreeding. For this thesis I studied HFC in a natural population of blue tits (*Cyanistes caeruleus*) that was carefully monitored during nine years. The blue tit, a common European passerine bird, is a popular model species for behavioral and ecological studies in the wild.

The number, characteristics and function of the molecular markers are important factors to consider when interpreting the results of HFC studies and establishing the causes of significant correlations. Therefore, in a first step, I characterized a set of at least 95 polymorphic microsatellite loci with utility for population genetic studies in the blue tit (Chapter 1). I demonstrated the value of Expressed Sequence Tag (EST) libraries of the model genetic passerine, the zebra finch (*Taeniopygia guttata)*, for developing molecular markers for other passerines. In addition, I predicted the location of the microsatellites in the avian genome using whole-genome sequence data of the chicken (*Gallus gallus*) and the zebra finch. I used 79 microsatellites to genotype all the breeding individuals (N = 794) and a selected group of chicks (N = 1005) in the studied blue tit population. I classified the markers as presumably functional or neutral by considering whether the genomic region where the markers are located is transcribed to RNA and therefore expressed. Interestingly, this classification captured differences in the characteristics of the loci (e.g. number of alleles, expected heterozygosity, distance to genes) that would be expected to occur between markers with genuine differences in functionality (Chapter 2).

I considered two different approaches for investigating HFC in the blue tit population. First, I used the sample of breeding birds, without controlling for the individual level of inbreeding, in order to capture the existent variance in inbreeding in the population and establish its role in HFC. With this approach, I found positive correlations between multilocus heterozygsity at neutral markers

and three reproductive and one survival traits (Chapters 2 and 3). Supporting a role of inbreeding on the detected HFC, I documented some consanguineous matings and a particular type of population structure that can contribute to the occurrence of inbreeding events (Chapter 2). Moreover, I detected identity disequilibrium (i.e. the correlation between heterozygosity or homozygosity across loci within individuals, caused by inbreeding) with all markers and with the group of neutral loci. For the second approach I controlled for the level of inbreeding by using groups of full-siblings (Chapter 3). In this case, the correlations between multilocus heterozygosity and two estimates of early survival were detected only with the group of functional markers and were negative. Taking into account the study design employed for the two approaches and the population structure and identity disequilibrium analyses, the positive HFC found with neutral markers may be attributed to the negative effects of inbreeding, whereas localized effects at single loci explain better the negative HFC found with functional loci. Since I did not find markers with disproportionately large and consistent effects on the phenotypic traits considered, I did not propose any of the loci as a strong candidate for causing variation in fitness in this population of blue tits.

The results of my thesis highlight the importance of considering the putative functionality of the markers when interpreting the results of HFC studies. Although HFC have limitations for identifying the specific loci causing variation in complex traits, I demonstrated that multilocus HFC are useful for revealing broad patterns of genotype – phenotype associations. I have illustrated the value of a thorough consideration of the population processes that can contribute to the genesis of HFC, especially for species or populations where strong variation in the level of inbreeding is not expected.

Zusammenfassung

Fitnesskonsequenzen der genetischen Diversität sind von fundamentaler Bedeutung in der Evolutions- und Naturschutzbiologie, wie auch in der Tier- und Pflanzenzucht. Eine Methode, die oft verwendet wird, um die Wechselwirkung zwischen Selektion und genetischer Variation und die Konsequenzen von Inzucht zu untersuchen, sind Korrelationen zwischen dem individuellen Heterozygotiegrad an mehreren Loci (MLH) und Fitnessmerkmalen, bekannt als Heterozygotie-Fitness-Korrelationen (HFC). In der vorliegenden Dissertation habe ich HFC in einer natürlichen Blaumeisenpopulation, die über neun Jahre sorgfältig überwacht wurde, studiert.

Zum Interpretieren der Resultate in HFC-Studien ist es wichtig, die Anzahl, Eigenschaften und Funktionen der verwendeten molekularen Marker zu kennen. In einem ersten Schritt habe ich deshalb ein Set von mindestens 95 polymorphen Mikrosatelliten beschrieben, welche in populationsgenetischen Studien über Blaumeisen Verwendung finden können (Kapitel 1). Ich konnte zeigen, dass „Expressed Sequence Tag" Bibliotheken von Zebrafinken (*Taeniopygia guttata*), der Modellart für genetische Studien bei Singvögeln, sehr nützlich sind, um molekulare Marker für andere Vogelarten zu entwickeln. Darüber hinaus habe ich mit Hilfe des publizierten Huhn- (*Gallus gallus*) und Zebrafinkengenoms die Lage meiner Mikrosatelliten im Vogelgenom definiert. Alle Brutvögel (N = 794) und eine ausgewählte Anzahl von Jungvögeln (N = 1005) wurden an 79 Mikrosatelliten genotypisiert. Je nachdem, ob die Genomregion, zu der ein Marker gemappt wurde, als transkribiert annotiert wurde oder nicht, habe ich meine Marker als potentiell funktional oder neutral klassifiziert. Die Charakteristiken der Marker (z.Bsp. Anzahl verschiedener Allele, erwarteter Heterozygotiegrad, Distanz zu Genen) decken sich mit dem, was bei funktionalen Unterschieden zwischen den Markern erwartet werden würde (Kapitel 2).

Um HFC in meiner Blaumeisenpopulation zu untersuchen, habe ich zwei verschiedene Ansätze verwendet. Im ersten Ansatz habe ich alle Brutvögel berücksichtigt, um die Varianz der Inzucht in der Population abzuschätzen (ohne für den individuellen Inzuchtgrad zu kontrollieren) und um Einsicht in deren Rolle bei HFC zu erhalten. Dabei fand ich positive Assoziationen zwischen der

MLH von neutralen Markern und mehreren Fitnessmerkmalen (Kapitel 2 und 3). Ich konnte einige Paarungen zwischen nah verwandten Individuen dokumentieren und eine ungewöhnliche Populationsstruktur aufdecken (Kapitel 2). Es erscheint deshalb plausibel, dass Inzuchtvariation die gefundenen HFC zumindest teilweise erklären kann. Darüber hinaus zeigten sowohl die Gruppe aller Marker als auch die Untergruppe der neutralen Marker ein Identitätsungleichgewicht (identity disequilibrium), d.h. eine durch Inzucht bedingte Korrelation zwischen heterozygoten und homozygoten Genotypen innerhalb eines Individuums. Der zweite Ansatz beinhaltete eine Kontrolle für den Inzuchtgrad, indem für die Analyse nur Geschwister berücksichtigt wurden. In diesem Falle waren die Korrelationen zwischen MLH und zwei Fitnessmerkmalen negativ und konnten nur innerhalb der funktionalen Marker nachgewiesen werden. Unter Berücksichtigung von Studiendesign, Populationsstruktur und Identitätsungleichgewichtsanalysen lässt sich schliessen, dass die positiven HFC in den neutralen Markern nachteiligen Effekten von Inzucht zugeschrieben werden können, während die negativen HFC in den funktionalen Loci durch lokale Effekte an einzelnen Genorten zustande kommen dürften. Da ich keine Marker mit überproportional grossen und konsistenten Effekten auf die untersuchten phänotypischen Merkmale entdecken konnte, scheint keiner der verwendeten Loci einen substantiellen Teil der Fitnessvariation in meiner Blaumeisenpopulation erklären zu können.

Die Resultate der vorliegenden Dissertation unterstreichen die Notwendigkeit, bei der Interpretation von Resultaten in HFC-Studien die Funktionalität der verwendeten Marker zu berücksichtigen. Obwohl HFC an einer begrenzten Anzahl Marker wenig geeignet sind, um spezifische Loci, die die Variation von komplexen Merkmalen erklären, zu identifizieren, konnte ich aufzeigen, dass Multilokus-HFC Inzuchtvariation als Spezialfall einer Genotyp-Phänotyp-Beziehung aufdecken können. Weiter konnte ich zeigen, dass es wichtig ist, Populationsprozesse zu berücksichtigen, welche zur Entstehung von HFC beitragen können, insbesondere auch bei Arten oder Populationen, wo eine starke Inzuchtvariation nicht unbedingt erwartet werden würde.

Acknowledgements

I would like to thank Bart Kempenaers for giving me the opportunity to pursue my doctoral studies at his group in the MPIO. I will always be grateful for the trust he put in me from our first conversation and for his support during my time in Seewiesen. Special thanks go to Jakob Müller for providing the scientific basis for this project and supervising my work, for having his door always open for questions and discussions and for his speed reading and commenting manuscript drafts. Thank you both for all the guidance throughout these years, I certainly learnt a lot from you!

Producing all the genotypic data for this thesis would not have been possible without the expert hands of Alexander Girg in the lab. It was great to work with you, Alex, we proved to be a good "Sherlock Holmes team"! I would like to thank Melanie Schneider for teaching me how to make PA gels and all the lab people who were directly or indirectly involved with the "hunt for microsats" project. Many thanks to all the members of the blue tit team between 1998 and 2006 for collecting the phenotypic data used for this thesis.

Many thanks go to Mihai Valcu for his invaluable guidance through the world of statistics, R, and the Kolbeterberg database. Mihai, thanks for sharing your time and knowledge, this definitely made my doctoral studies easier! I would like to thank Aurèlie Cohas for introducing me to the HFC analyses in R and for providing scripts for analyzing genotype data. Thanks also to Holger Schielzeth, Wolfgang Forstmeier and Niels Dingemanse for advise on statistical issues, and to Kathi Foerster for an early discussion about this project. Thanks to Hendrik Reers for help in formatting this thesis.

I am grateful to Bengt Hansson for participating in my advisory committee and for including me in the blue tit linkage map project. Thanks to Deborah A. Dawson and Marcus Ljungqvist for designing some of the microsatellites used to genotype the blue tits. Thanks to the three of them for their contributions on Chapter 1 of this thesis.

Thanks to the many people in the MPIO, and particularly to those in the Behavioral Ecology and Evolutionary Genetics group, who supported me in the several scientific and non-scientific aspects of pursuing a doctoral thesis. Special

thanks to the fellow doctoral students and friends for sharing the joys and worries, to the blue tit team 2008 - 2010 for the fun field seasons, to Carmen Dobus, Daniel Piechowski and Martin Wikelski for all the help getting enrolled at the University, and to Susann Rössel for the German lessons.

Thanks to Mathias, for being my best discovery in Seewiesen. Your support, patience, company and practical help were fundamental for the completion of this thesis. To Bina, Thomas, Claudia, Monika and Flurina, thanks for making me feel in family so far away from home.

Very special thanks to all my family and friends for being a great source of support, even from the distance. Mami y papi, ustedes siempre serán uno de los ejes principales de mis logros, gracias por todo! Cata, gracias por venir a acompañar a su hermanita en tierras tan lejanas. Clau, gracias por haberme organizado la vida otra vez hace cuatro años!

Funding for my thesis was provided by the German Research Foundation (Project: MU 1479/2-1) and the Max Planck Society.

Seewiesen

May 2011

References

Acevedo-Whitehouse K, Spraker TR, Lyons E, Melin SR, Gulland F, Delong RL, Amos W (2006) Contrasting effects of heterozygosity on survival and hookworm resistance in California sea lion pups. Molecular Ecology 15:1973-1982

Amos W, Wilmer JW, Fullard K, Burg TM, Croxall JP, Bloch D, Coulson T (2001) The influence of parental relatedness on reproductive success. Proceedings of the Royal Society of London Series B - Biological Sciences 268:2021-2027

Aparicio JM, Ortego J, Cordero PJ (2006) What should we weigh to estimate heterozygosity, alleles or loci? Molecular Ecology 15:4659-4665

Ashley MV, Dow BD (1994) The use of microsatellite analysis in population biology: background, methods and potential applications. In: Schierwater B, Streit B, Wagner GP, DeSalle R (eds) Molecular Ecology and Evolution: Approaches and Applications. Birkhauser Verlag, Boston, pp 185-202

Backstrom N, Karaiskou N, Leder EH, Gustafsson L, Primmer CR, Qvarnstrom A, Ellegren H (2008) A gene-based genetic linkage map of the collared flycatcher (*Ficedula albicollis*) reveals extensive synteny and gene-order conservation during 100 million years of avian evolution. Genetics 179:1479-1495

Balloux F, Amos W, Coulson T (2004) Does heterozygosity estimate inbreeding in real populations? Molecular Ecology 13:3021-3031

Balloux F, Lugon-Moulin N (2002) The estimation of population differentiation with microsatellite markers. Molecular Ecology 11:155-165

Bates D, Maechler M (2010) lme4: Linear mixed-effects models using S4 classes. R package version 0.999375-34

Bed'Hom B, Coullin P, Guillier-Gencik Z, Moulin S, Bernheim A, Volobouev V (2003) Characterization of the atypical karyotype of the black-winged kite *Elanus caeruleus* (Falconiformes: Accipitridae) by means of classical and molecular cytogenetic techniques. Chromosome Research 11:335-343

Bell G (2010) Fluctuating selection: the perpetual renewal of adaptation in variable environments. Philosophical Transactions of the Royal Society B - Biological Sciences 365:87-97

Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate - a practical and powerful approach to multiple testing. Journal of the Royal Statistical Society Series B-Methodological 57:289-300

Bensch S, Price T, Kohn J (1997) Isolation and characterization of microsatellite loci in a *Phylloscopus* warbler. Molecular Ecology 6:91-92

Bierne N, Bezuart I, Vonau V, Bonhomme F, Bedier E, Aquacop (2000) Microsatellite-associated heterosis in hatchery-propagated stocks of the shrimp *Penaeus stylirostris*. Aquaculture 184:203-219

Blackmore CJ, Adcock GJ, Ebert D, Heinsohn R (2006) Microsatellite loci for population and behavioural studies of grey-crowned babblers (*Pomatostomus temporalis*: Aves). Molecular Ecology Notes 6:412-414

Blanchet S, Rey O, Berthier P, Lek S, Loot G (2009) Evidence of parasite-mediated disruptive selection on genetic diversity in a wild fish population. Molecular Ecology 18:1112-1123

Blomqvist D, Pauliny A, Larsson M, Flodin LA (2010) Trapped in the extinction vortex? Strong genetic effects in a declining vertebrate population. BMC Evolutionary Biology 10:33

Bouck A, Vision T (2007) The molecular ecologist's guide to expressed sequence tags. Molecular Ecology 16:907-924

Britten HB (1996) Meta-analyses of the association between multilocus heterozygosity and fitness. Evolution 50:2158-2164

Brouwer L, Komdeur J, Richardson DS (2007) Heterozygosity-fitness correlations in a bottlenecked island species: a case study on the Seychelles warbler. Molecular Ecology 16:3134-3144

Bulgin NL, Gibbs HL, Vickery P, Baker AJ (2003) Ancestral polymorphisms in genetic markers obscure detection of evolutionarily distinct populations in the endangered Florida grasshopper sparrow (*Ammodramus savannarum floridanus*). Molecular Ecology 12:831-844

Burnham KP, Anderson DR (2002) Model selection and multimodel inference. A practical information-theoretic approach, 2nd edn. Springer, USA

Chakraborty R (1981) The distribution of the number of heterozygous loci in an individual in natural populations. Genetics 98:461-466

Chapman JR, Nakagawa S, Coltman DW, Slate J, Sheldon BC (2009) A quantitative review of heterozygosity-fitness correlations in animal populations. Molecular Ecology 18:2746-2765

Charlesworth D, Charlesworth B (1987) Inbreeding depression and its evolutionary consequences. Annual Review of Ecology and Systematics 18:237-268

Charlesworth D, Willis JH (2009) The genetics of inbreeding depression. Nature Reviews Genetics 10:783-796

Charmantier A, Blondel J, Perret P, Lambrechts MM (2004) Do extra-pair paternities provide genetic benefits for female blue tits *Parus caeruleus*? Journal of Avian Biology 35:524-532

Charpentier M, Setchell JM, Prugnolle F, Wickings EJ, Peignot P, Balloux F, Hossaert-Mickey M (2006) Life history correlates of inbreeding depression in mandrills (*Mandrillus sphinx*). Molecular Ecology 15:21-28

Charpentier MJE, Williams CV, Drea CM (2008) Inbreeding depression in ring-tailed lemurs (*Lemur catta*): genetic diversity predicts parasitism, immunocompetence, and survivorship. Conservation Genetics 9:1605-1615

Coltman DW, Pilkington JG, Smith JA, Pemberton JM (1999) Parasite-mediated selection against inbred Soay sheep in a free-living, island population. Evolution 53:1259-1267

Coltman DW, Slate J (2003) Microsatellite measures of inbreeding: A meta-analysis. Evolution 57:971-983

Cordeiro GM, Casu R, McIntyre CL, Manners JM, Henry RJ (2001) Microsatellite markers from sugarcane (*Saccharum spp*.) ESTs cross transferable to erianthus and sorghum. Plant Science 160:1115-1123

Coster A (2009) pedigree: Package to deal with pedigree data. R package version 1.2

Coulon A (2010) GENHET: an easy-to-use R function to estimate individual heterozygosity. Molecular Ecology Resources 10:167-169

Coulson T, Albon S, Slate J, Pemberton J (1999) Microsatellite loci reveal sex-dependent responses to inbreeding and outbreeding in red deer calves. Evolution 53:1951-1960

Coulson TN, Pemberton JM, Albon SD, Beaumont M, Marshall TC, Slate J, Guinness FE, Clutton-Brock TH (1998) Microsatellites reveal heterosis in red deer. Proceedings of the Royal Society of London Series B - Biological Sciences 265:489-495

Crnokrak P, Roff DA (1995) Dominance variance: associations with selection and fitness. Heredity 75:530-540

Crnokrak P, Roff DA (1999) Inbreeding depression in the wild. Heredity 83:260-270

Da Silva A, Gaillard JM, Yoccoz NG, Hewison AJM, Galan M, Coulson T, Allaine D, Vial L, Delorme D, Van Laere G, Klein F, Luikart G (2009) Heterozygosity-fitness correlations revealed by neutral and candidate gene markers in roe deer from a long-term study. Evolution 63:403-417

Da Silva A, Luikart G, Yoccoz NG, Cohas A, Allaine D (2006) Genetic diversity-fitness correlation revealed by microsatellite analyses in European alpine marmots (*Marmota marmota*). Conservation Genetics 7:371-382

Dauwe T, Janssens E, Pinxten R, Eens M (2005) The reproductive success and quality of blue tits (*Parus caeruleus*) in a heavy metal pollution gradient. Environmental Pollution 136:243-251

David P (1998) Heterozygosity-fitness correlations: new perspectives on old problems. Heredity 80:531-537

David P, Delay B, Jarne P (1997) Heterozygosity and growth in the marine bivalve *Spisula ovalis*: testing alternative hypotheses. Genetical Research 70:215-223

David P, Pujol B, Viard F, Castella V, Goudet J (2007) Reliable selfing rate estimates from imperfect population genetic data. Molecular Ecology 16:2474-2487

Dawson DA (2007) Genomic analysis of passerine birds using conserved microsatellite loci. PhD Thesis. University of Sheffield, UK

Dawson DA, Akesson M, Burke T, Pemberton JM, Slate J, Hansson B (2007) Gene order and recombination rate in homologous chromosome regions

of the chicken and a passerine bird. Molecular Biology and Evolution 24:1537-1552

Dawson DA, Burke T, Hansson B, Pandhal J, Hale MC, Hinten GN, Slate J (2006) A predicted microsatellite map of the passerine genome based on chicken-passerine sequence similarity. Molecular Ecology 15:1299-1320

Dawson DA, Hanotte O, Greig C, Stewart IRK, Burke T (2000) Polymorphic microsatellites in the blue tit *Parus caeruleus* and their cross-species utility in 20 songbird families. Molecular Ecology 9:1941-1944

Dawson DA, Horsburgh GJ, Kupper C, Stewart IRK, Ball AD, Durrant KL, Hansson B, Bacon I, Bird S, Klein A, Krupa AP, Lee JW, Martin-Galvez D, Simeoni M, Smith G, Spurgin LG, Burke T (2010) New methods to identify conserved microsatellite loci and develop primer sets of high cross-species utility - as demonstrated for birds. Molecular Ecology Resources 10:475-494

de Oliveira EHC, Habermann FA, Lacerda O, Sbalqueiro IJ, Wienberg J, Muller S (2005) Chromosome reshuffling in birds of prey: the karyotype of the world's largest eagle (Harpy eagle, *Harpia harpyja*) compared to that of the chicken (*Gallus gallus*). Chromosoma 114:338-343

Delhey K, Johnsen A, Peters A, Andersson S, Kempenaers B (2003) Paternity analysis reveals opposing selection pressures on crown coloration in the blue tit (*Parus caeruleus*). Proceedings of the Royal Society of London Series B - Biological Sciences 270:2057-2063

Dempster AP, Laird NM, Rubin DB (1977) Maximum likelihood from incomplete data via the EM algorithm. Journal of the Royal Statistical Society Series B-Methodological 39:1-38

Derjusheva S, Kurganova A, Habermann F, Gaginskaya E (2004) High chromosome conservation detected by comparative chromosome painting in chicken, pigeon and passerine birds. Chromosome Research 12:715-723

DeRose MA, Roff DA (1999) A comparison of inbreeding depression in life-history and morphological traits in animals. Evolution 53:1288-1292

DeWoody YD, DeWoody JA (2005) On the estimation of genome-wide heterozygosity using molecular markers. Journal of Heredity 96:85-88

Dhondt AA, Kempenaers B, Adriaensen F (1992) Density-dependent clutch size caused by habitat heterogeneity. Journal of Animal Ecology 61:643-648

Dibattista JD, Feldheim KA, Gruber SH, Hendry AP (2008) Are indirect genetic benefits associated with polyandry? Testing predictions in a natural population of lemon sharks. Molecular Ecology 17:783-795

Double MC, Dawson D, Burke T, Cockburn A (1997) Finding the fathers in the least faithful bird: A microsatellite-based genotyping system for the superb fairy-wren *Malurus cyaneus*. Molecular Ecology 6:691-693

Double MC, Peakall R, Beck NR, Cockburn A (2005) Dispersal, philopatry, and infidelity: Dissecting local genetic structure in superb fairy-wrens (*Malurus cyaneus*). Evolution 59:625-635

Doutrelant C, Gregoire A, Grnac N, Gomez D, Lambrechts MM, Perret P (2008) Female coloration indicates female reproductive capacity in blue tits. Journal of Evolutionary Biology 21:226-233

Doutrelant C, Leitao A, Giorgi H, Lambrechts MM (1999) Geographical variation in blue tit song, the result of an adjustment to vegetation type? Behaviour 136:481-493

Dranchak PK, Chaves LD, Rowe JA, Reed KM (2003) Turkey microsatellite loci from an embryonic cDNA library. Poultry Science 82:526-531

Earl DA (2010) Structure Harvester v0.56.4. Available from http://taylor0.biology.ucla.edu/struct_harvest/

Ellegren H, Sheldon BC (2008) Genetic basis of fitness differences in natural populations. Nature 452:169-175

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Molecular Ecology 14:2611-2620

Falush D, Stephens M, Pritchard JK (2003) Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. Genetics 164:1567-87

Fitzpatrick JL, Evans JP (2009) Reduced heterozygosity impairs sperm quality in endangered mammals. Biology Letters 5:320-323

Fitzpatrick MJ, Ben-Shahar Y, Smid HM, Vet LEM, Robinson GE, Sokolowski MB (2005) Candidate genes for behavioural ecology. Trends in Ecology & Evolution 20:96-104

Foerster K, Delhey K, Johnsen A, Lifjeld JT, Kempenaers B (2003) Females increase offspring heterozygosity and fitness through extra-pair matings. Nature 425:714-717

Foerster K, Valcu M, Johnsen A, Kempenaers B (2006) A spatial genetic structure and effects of relatedness on mate choice in a wild bird population. Molecular Ecology 15:4555-4567

Forstmeier W, Kempenaers B, Meyer A, Leisler B (2002) A novel song parameter correlates with extra-pair paternity and reflects male longevity. Proceedings of the Royal Society of London Series B - Biological Sciences 269:1479-1485

Fossoy F, Johnsen A, Lifjeld JT (2009) Cell-mediated immunity and multi-locus heterozygosity in bluethroat nestlings. Journal of Evolutionary Biology 22:1954-1960

Fox CW, Scheibly KL, Wallin WG, Hitchcock LJ, Stillwell RC, Smith BP (2006) The genetic architecture of life span and mortality rates: Gender and species differences in inbreeding load of two seed-feeding beetles. Genetics 174:763-773

Fridolfsson AK, Gyllensten UB, Jakobsson S (1997) Microsatellite markers for paternity testing in the willow warbler *Phylloscopus trochilus*: high frequency of extra-pair young in an island population. Hereditas 126:127-132

Galbusera P, Van Dongen S, Matthysen E (2000) Cross-species amplification of microsatellite primers in passerine birds. Conservation Genetics 1:163-168

Garant D, Dodson JD, Bernatchez L (2005) Offspring genetic diversity increases fitness of female Atlantic salmon (*Salmo salar*). Behavioral Ecology and Sociobiology 57:240-244

Garcia-Navas V, Ortego J, Sanz JJ (2009) Heterozygosity-based assortative mating in blue tits (*Cyanistes caeruleus*): implications for the evolution of mate choice. Proceedings of the Royal Society B - Biological Sciences 276:2931-2940

Garnier-Gere P, Dillmann C (1992) A computer program for testing pairwise linkage disequilibria in subdivided populations. Journal of Heredity 83:239-239

Gibbs HL, Tabak LM, Hobson K (1999) Characterization of microsatellite DNA loci for a neotropical migrant songbird, the Swainson's thrush (*Catharus ustulatus*). Molecular Ecology 8:1551-1552

Gibbs M, Dawson DA, McCamley C, Wardle AF, Armour JAL, Burke T (1997) Chicken microsatellite markers isolated from libraries enriched for simple tandem repeats. Animal Genetics 28:401-417

Glutz von Blotzheim UN, Bauer KM (1993) Handbuch der Vögel Mitteleuropas Vol. 13. Aula-Verlag, Wiesbaden

Greenwood PJ (1980) Mating systems, philopatry and dispersal in birds and mammals. Animal Behaviour 28:1140-1162

Griffin DK, Robertson LBW, Tempest HG, Skinner BM (2007) The evolution of the avian genome as revealed by comparative molecular cytogenetics. Cytogenetic and Genome Research 117:64-77

Griffith SC, Stewart IRK, Dawson DA, Owens IPF, Burke T (1999) Contrasting levels of extra-pair paternity in mainland and island populations of the house sparrow (*Passer domesticus*): is there an 'island effect'? Biological Journal of the Linnean Society 68:303-316

Griffiths R, Double MC, Orr K, Dawson RJG (1998) A DNA test to sex most birds. Molecular Ecology 7:1071-1075

Guillot G (2008) Inference of structure in subdivided populations at low levels of genetic differentiation - the correlated allele frequencies model revisited. Bioinformatics 24:2222-2228

Guillot G, Estoup A, Mortier F, Cosson JF (2005) A spatial statistical model for landscape genetics. Genetics 170:1261-1280

Haas F, Hansson B (2008) Identification of 20 polymorphic microsatellite loci in European crow (*Corvus corone*) from existing passerine loci. Molecular Ecology Resources 8:846-850

Haldane JBS (1954) The biochemistry of genetics. George Allen and Unwin, London

Hale MC, Jensen H, Birkhead TR, Burke T, Slate J (2008) A comparison of synteny and gene order on the homologue of chicken chromosome 7 between two passerine species and between passerines and chicken. Cytogenetic and Genome Research 121:120-129

Hansson B, Akesson M, Ljungqvist M (2010a) Which molecular diversity index should we use to estimate inbreeding in natural populations? In: M. Ljungqvist. Have a fling with your neighbour, you might be less related. The use of molecular markers to estimate genetic diversity and population structure in conservation biology. PhD thesis. Lund University, Lund

Hansson B, Akesson M, Slate J, Pemberton JM (2005) Linkage mapping reveals sex-dimorphic map distances in a passerine bird. Proceedings of the Royal Society B - Biological Sciences 272:2289-2298

Hansson B, Bensch S, Hasselquist D, Akesson M (2001) Microsatellite diversity predicts recruitment of sibling great reed warblers. Proceedings of the Royal Society of London Series B - Biological Sciences 268:1287-1291

Hansson B, Ljungqvist M, Dawson DA, Mueller JC, Olano-Marin J, Ellegren H, Nilsson JA (2010b) Avian genome evolution: insights from a linkage map of the blue tit (*Cyanistes caeruleus*). Heredity 104:67-78

Hansson B, Westerberg L (2002) On the correlation between heterozygosity and fitness in natural populations. Molecular Ecology 11:2467-2474

Hansson B, Westerberg L (2008) Heterozygosity-fitness correlations within inbreeding classes: local or genome-wide effects? Conservation Genetics 9:73-83

Hansson B, Westerdahl H, Hasselquist D, Akesson M, Bensch S (2004) Does linkage disequilibrium generate heterozygosity-fitness correlations in great reed warblers? Evolution 58:870-879

Hedrick PW, Kalinowski ST (2000) Inbreeding depression in conservation biology. Annual Review of Ecology and Systematics 31:139-162

Hoffman JI, Boyd IL, Amos W (2004) Exploring the relationship between parental relatedness and male reproductive success in the antarctic fur seal *Arctocephalus gazella*. Evolution 58:2087-2099

Hoffman JI, Forcada J, Amos W (2010a) Exploring the mechanisms underlying a heterozygosity-fitness correlation for canine size in the antarctic fur seal *Arctocephalus gazella*. Journal of Heredity 101:539-552

Hoffman JI, Hanson N, Forcada J, Trathan PN, Amos W (2010b) Getting long in the tooth: a strong positive correlation between canine size and heterozygosity in antarctic fur seals *Arctocephalus gazella*. Journal of Heredity 101:527-538

Hoglund J, Piertney SB, Alatalo RV, Lindell J, Lundberg A, Rintamaki PT (2002) Inbreeding depression and male fitness in black grouse. Proceedings of the Royal Society of London Series B - Biological Sciences 269:711-715

Hothorn T, Bretz F, Westfall P (2008) Simultaneous inference in general parametric models. Biometrical Journal 50:346-363

Houle D (1989) Allozyme-associated heterosis in *Drosophila melanogaster*. Genetics 123:789-801

Hubby JL, Lewontin RC (1966) A molecular approach to the study of genic heterozygosity in natural populations .I. Number of alleles at different loci in *Drosophila pseudoobscura*. Genetics 54:577-594

International Chicken Genome Sequencing Consortium (2004) Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. Nature 432:695-716

Ishibashi Y, Mikami O, Abe S (2000) Isolation and characterization of microsatellite loci in the Japanese marsh warbler *Locustella pryeri*. Molecular Ecology 9:373-375

Jarne P, Lagoda PJL (1996) Microsatellites, from molecules to populations and back. Trends in Ecology & Evolution 11:424-429

Jensen H, Bremset EM, Ringsby TH, Saether BE (2007) Multilocus heterozygosity and inbreeding depression in an insular house sparrow metapopulation. Molecular Ecology 16:4066-4078

Johannessen LE, Slagsvold T, Hansen BT, Lifjeld JT (2005) Manipulation of male quality in wild tits: effects on paternity loss. Behavioral Ecology 16:747-754

Johnsen A, Fidler AE, Kuhn S, Carter KL, Hoffmann A, Barr IR, Biard C, Charmantier A, Eens M, Korsten P, Siitari H, Tomiuk J, Kempenaers B (2007) Avian Clock gene polymorphism: evidence for a latitudinal cline in allele frequencies. Molecular Ecology 16:4867-4880

Johnson AM, Chappell G, Price AC, Rodd FH, Olendorf R, Hughes KA (2010) Inbreeding depression and inbreeding avoidance in a natural population of guppies (*Poecilia reticulata*). Ethology 116:448-457

Kantety RV, La Rota M, Matthews DE, Sorrells ME (2002) Data mining for simple sequence repeats in expressed sequence tags from barley, maize, rice, sorghum and wheat. Plant Molecular Biology 48:501-510

Karaiskou N, Buggiotti L, Leder E, Primmer CR (2008) High degree of transferability of 86 newly developed zebra finch EST-linked microsatellite markers in 8 bird species. Journal of Heredity 99:688-693

Kawano KM (2003) Isolation of polymorphic microsatellite markers in the great tit (*Parus major minor*). Molecular Ecology Notes 3:314-315

Keller LF, Waller DM (2002) Inbreeding effects in wild populations. Trends in Ecology & Evolution 17:230-241

Kempenaers B (2007) Mate choice and genetic quality: A review of the heterozygosity theory. Advances in the Study of Behavior 37:189-278

Kempenaers B, Adriaensen F, Van Noordwijk AJ, Dhondt AA (1996) Genetic similarity, inbreeding and hatching failure in blue tits: Are unhatched eggs infertile? Proceedings of the Royal Society of London Series B - Biological Sciences 263:179-185

Kempenaers B, Verheyen GR, Vandenbroeck M, Burke T, Vanbroeckhoven C, Dhondt AA (1992) Extra-pair paternity results from female preference for high-quality males in the blue tit. Nature 357:494-496

Kempenaers B, Verheyren GR, Dhondt AA (1997) Extrapair paternity in the blue tit (*Parus caeruleus*): female choice, male characteristics, and offspring quality. Behavioral Ecology 8:481-492

Kim KS, Ratcliffe ST, French BW, Liu L, Sappington TW (2008) Utility of EST-Derived SSRs as population genetics markers in a beetle. Journal of Heredity 99:112-124

Kimura M (1983) Rare variant alleles in the light of the neutral theory. Molecular Biology and Evolution 1:84-93

King TL, Eackles MS, Henderson AP, Bocetti CI, Currie D, Wunderle JM (2005) Microsatellite DNA markers for delineating population structure and kinship among the endangered Kirtland's warbler (*Dendroica kirtlandii*). Molecular Ecology Notes 5:569-571

Kinghorn BP, Kinghorn S (2006) Pedigree Viewer version 5.5. Available from http://www-personal.une.edu.au/~bkinghor/

Klein A, Horsburgh GJ, Küpper C, Major A, Lee PLM, Hoffmann G, Mátics R, Dawson DA (2009) Microsatellite markers characterized in the barn owl (*Tyto alba*) and of high utility in other owls (Strigiformes: AVES). Molecular Ecology Resources 9:1513-1519

Kong Q, Xiang C, Yu Z, Zhang C, Liu F, Peng C, Peng X (2007) Mining and charactering microsatellites in *Cucumis melo* expressed sequence tags from sequence database. Molecular Ecology Notes 7:281-283

Krokene C, Lifjeld JT (2000) Variation in the frequency of extra-pair paternity in birds: A comparison of an island and a mainland population of blue tits. Behaviour 137:1317-1330

Kupper C, Horsburgh GJ, Dawson DA, Ffrench-Constant R, Szekely T, Burke T (2007) Characterization of 36 polymorphic microsatellite loci in the Kentish plover (*Charadrius alexandrinus*) including two sex-linked loci and their amplification in four other *Charadrius* species. Molecular Ecology Notes 7:35-39

Kupper C, Kosztolanyi A, Augustin J, Dawson DA, Burke T, Szekely T (2010) Heterozygosity-fitness correlations of conserved microsatellite markers in Kentish plovers *Charadrius alexandrinus*. Molecular Ecology 19:5172-5185

Latch EK, Dharmarajan G, Glaubitz JC, Rhodes OE (2006) Relative performance of Bayesian clustering software for inferring population substructure and individual assignment at low levels of population differentiation. Conservation Genetics 7:295-302

Ledig FT, Guries RP, Bonefeld BA (1983) The relation of growth to heterozygosity in pitch pine. Evolution 37:1227-1238

Lee JW, Jang BS, Dawson DA, Burke T, Hatchwell BJ (2009) Fine-scale genetic structure and its consequence in breeding aggregations of a passerine bird. Molecular Ecology 18:2728-2739

Lesbarrères D, Primmer CR, Laurila A, Merilä J (2005) Environmental and population dependency of genetic variability-fitness correlations in *Rana temporaria*. Molecular Ecology 14:311-323

Lesbarrères D, Schmeller DS, Primmer CR, Merilä J (2007) Genetic variability predicts common frog (*Rana temporaria*) size at metamorphosis in the wild. Heredity 99:41-46

Lewontin RC, Hubby JL (1966) A molecular approach to the study of genic heterozygosity in natural populations. II. Amount of variation and degree of heterozygosity in natural populations of *Drosophila pseudoobscura*. Genetics 54:595-609

Li SH, Huang YJ, Brown JL (1997) Isolation of tetranucleotide microsatellites from the Mexican jay *Aphelocoma ultramarina*. Molecular Ecology 6:499-501

Li YC, Korol AB, Fahima T, Nevo E (2004) Microsatellites within genes: structure, function, and evolution. Molecular Biology and Evolution 21:991-1007

Lieutenant-Gosselin M, Bernatchez L (2006) Local heterozygosity-fitness correlations with global positive effects on fitness in threespine stickleback. Evolution 60:1658-1668

Ljungqvist M, Akesson M, Hansson B (2010) Do microsatellites reflect genome-wide genetic diversity in natural populations? A comment on Vali et al. (2008). Molecular Ecology 19:851-855

Luikart G, Pilgrim K, Visty J, Ezenwa VO, Schwartz MK (2008) Candidate gene microsatellite variation is associated with parasitism in wild bighorn sheep. Biology Letters 4:228-231

Lynch M (1991) The genetic interpretation of inbreeding depression and outbreeding depression. Evolution 45:622-629

MacColl ADC, Hatchwell BJ (2003) Heritability of parental effort in a passerine bird. Evolution 57:2191-2195

Mainguy J, Cote SD, Coltman DW (2009) Multilocus heterozygosity, parental relatedness and individual fitness components in a wild mountain goat, *Oreamnos americanus* population. Molecular Ecology 18:2297-2306

Mannen H, Murata K, Kikuchi S, Fujima D, Sasazaki S, Fujiwara A, Tsuji S (2005) Development and mapping of microsatellite markers derived from cDNA in Japanese quail (*Coturnix japonica*). Journal of Poultry Science 42:263-271

Markert JA, Grant PR, Grant BR, Keller LF, Coombs JL, Petren K (2004) Neutral locus heterozygosity, inbreeding, and survival in Darwin's ground finches (*Geospiza fortis* and *G. scandens*). Heredity 92:306-315

Marshall TC, Coltman DW, Pemberton JM, Slate J, Spalton JA, Guinness FE, Smith JA, Pilkington JG, Clutton-Brock TH (2002) Estimating the prevalence of inbreeding from incomplete pedigrees. Proceedings of the Royal Society of London Series B - Biological Sciences 269:1533-1539

Marshall TC, Spalton JA (2000) Simultaneous inbreeding and outbreeding depression in reintroduced Arabian oryx. Animal Conservation 3:241-248

Martin-Galvez D, Dawson DA, Horburgh GJ, Burke T (2009) Isolation, characterization and chromosome locations of polymorphic black-billed magpie *Pica pica* (Corvidae, AVES) microsatellite loci. Molecular Ecology Resources 9:1506-1512

Matthysen E, Adriaensen F, Dhondt AA (2001) Local recruitment of great and blue tits (*Parus major*, *P. caeruleus*) in relation to study plot size and degree of isolation. Ecography 24:33-42

Merilä J, Kruuk LEB, Sheldon BC (2001) Cryptic evolution in a wild bird population. Nature 412:76-79

Mitton JB (1997) Selection in natural populations. Oxford University Press, New York

Mitton JB, Grant MC (1984) Associations among protein heterozygosity, growth-rate, and developmental homeostasis. Annual Review of Ecology and Systematics 15:479-499

Moore SS, Sargeant LL, King TJ, Mattick JS, Georges M, Hetzel DJS (1991) The conservation of dinucleotide microsatellites among mammalian genomes allows the use of heterologous PCR primer pairs in closely related species. Genomics 10:654-660

Mueller JC, Hermisson J, Olano-Marin J, Hansson B, Kempenaers B (2010) Linking genetic mechanisms of heterozygosity-fitness correlations to footprints of selection at single loci. Evolutionary Ecology 25:1-11

Neff BD (2004) Stabilizing selection on genomic divergence in a wild fish population. Proceedings of the National Academy of Sciences of the United States of America 101:2381-2385

Norris K (1993) Seasonal variation in the reproductive success of blue tits: an experimental study. Journal of Animal Ecology 62:287-294

Ohta T (1971) Associative overdominance caused by linked detrimental mutations. Genetical Research 18:277-286

Ohta T, Kimura M (1969) Linkage disequilibrium at steady state determined by random genetic drift and recurrent mutation. Genetics 63:229-238

Ohta T, Kimura M (1970) Development of associative overdominance through linkage disequilibrium in finite populations. Genetical Research 16:165-177

Olano-Marin J, Dawson DA, Girg A, Hansson B, Ljungqvist M, Kempenaers B, Mueller JC (2010) A genome-wide set of 106 microsatellite markers for the blue tit (*Cyanistes caeruleus*). Molecular Ecology Resources 10:516-532

Oliveira KM, Pinto LR, Marconi TG, Mollinari M, Ulian EC, Chabregas SM, Falco MC, Burnquist W, Garcia AAF, Souza AP (2009) Characterization of new polymorphic functional markers for sugarcane. Genome 52:191-209

Ortego J, Calabuig G, Bonal R, Munoz A, Aparicio JM, Cordero PJ (2009) Temporal variation of heterozygosity-based assortative mating and related benefits in a lesser kestrel population. Journal of Evolutionary Biology 22:2488-2495

Otsuka R, Nishiumi I, Wada M (2003) Characterization of 12 polymorphic microsatellite loci in the Japanese bush warbler *Cettia diphone*. Molecular Ecology Notes 3:44-46

Otter K, Ratcliffe L, Michaud D, Boag PT (1998) Do female black-capped chickadees prefer high-ranking males as extra-pair partners? Behavioral Ecology and Sociobiology 43:25-36

Pemberton J (2004) Measuring inbreeding depression in the wild: the old ways are the best. Trends in Ecology & Evolution 19:613-615

Perez F, Ortiz J, Zhinaula M, Gonzabay C, Calderon J, Volckaert FAMJ (2005) Development of EST-SSR markers by data mining in three species of shrimp: *Litopenaeus vannamei*, *Litopenaeus stylirostris*, and *Trachypenaeus birdy*. Marine Biotechnology 7:554-569

Petren K (1998) Microsatellite primers from *Geospiza fortis* and cross-species amplification in Darwin's finches. Molecular Ecology 7:1782-1784

Piertney SB, Marquiss M, Summers R (1998) Characterization of tetranucleotide microsatellite markers in the Scottish crossbill (*Loxia scotica*). Molecular Ecology 7:1261-1263

Pinheiro CJ, Bates MD (2000) Mixed effects models in S and S-Plus. Springer, New York

Pinheiro J, Bates D, DebRoy S, Sarkar D, the R Core team (2009) nlme: linear and nonlinear mixed effects models. R package version 3.1-93

Poesel A, Kunc HP, Foerster K, Johnsen A, Kempenaers B (2006) Early birds are sexy: male age, dawn song and extrapair paternity in blue tits, *Cyanistes* (formerly *Parus*) *caeruleus*. Animal Behaviour 72:531-538

Postma E, van Noordwijk AJ (2005) Gene flow maintains a large genetic difference in clutch size at a small spatial scale. Nature 433:65-68

Primmer CR, Møller AP, Ellegren H (1996) A wide-range survey of cross-species microsatellite amplification in birds. Molecular Ecology 5:365-378

Primmer CR, Painter JN, Koskinen MT, Palo JU, Merilä J (2005) Factors affecting avian cross-species microsatellite amplification. Journal of Avian Biology 36:348-360

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. Genetics 155:945-959

Pritchard JK, Wen X, Falush D (2010) Documentation for structure software: Version 2.3. Available from http://pritch.bsd.uchicago.edu/structure.html

Przybylo R, Wiggins DA, Merilä J (2001) Breeding success in blue tits: good territories or good parents? Journal of Avian Biology 32:214-218

Pulido FJ, Diaz M (2000) Foraging behaviour of blue tits *Parus caeruleus* in a patchy environment under contrasting levels of natural food supply. Journal of Avian Biology 31:81-86

Pusey A, Wolf M (1996) Inbreeding avoidance in animals. Trends in Ecology & Evolution 11:201-206

R Development Core Team (2008) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria

Raberg L, Stjernman M, Hasselquist D (2003) Immune responsiveness in adult blue tits: heritability and effects of nutritional status during ontogeny. Oecologia 136:360-364

Reed DH, Frankham R (2003) Correlation between fitness and genetic diversity. Conservation Biology 17:230-237

Reid JM, Arcese P, Keller LF, Elliott KH, Sampson L, Hasselquist D (2007) Inbreeding effects on immune response in free-living song sparrows (*Melospiza melodia*). Proceedings of the Royal Society B - Biological Sciences 274:697-706

Rexroad CE, Rodriguez MF, Coulibaly I, Gharbi K, Danzmann RG, DeKoning J, Phillips R, Palti Y (2005) Comparative mapping of expressed sequence tags containing microsatellites in rainbow trout (*Oncorhynchus mykiss*). BMC Genomics 6:54

Richardson DS, Jury FL, Dawson DA, Salgueiro P, Komdeur J, Burke T (2000) Fifty Seychelles warbler (*Acrocephalus sechellensis*) microsatellite loci polymorphic in Sylviidae species and their cross-species amplification in other passerine birds. Molecular Ecology 9:2226-2231

Rioux-Paquette E, Festa-Bianchet M, Coltman DW (2011) Sex-differential effects of inbreeding on overwinter survival, birth date and mass of bighorn lambs. Journal of Evolutionary Biology 24:121-31

Rossiter SJ, Jones G, Ransome RD, Barratt EM (2001) Outbreeding increases offspring survival in wild greater horseshoe bats (*Rhinolophus ferrumequinum*). Proceedings of the Royal Society of London Series B - Biological Sciences 268:1055-1061

Rousset F (2008) GENEPOP'007: a complete re-implementation of the GENEPOP software for Windows and Linux. Molecular Ecology Resources 8:103-106

Ruyter-Spira CP, de Koning DJ, van der Poel JJ, Crooijmans RPMA, Dijkhof RJM, Groenen MAM (1998) Developing microsatellite markers from cDNA: a tool for adding expressed sequence tags to the genetic linkage map of the chicken. Animal Genetics 29:85-90

Saccheri IJ, Lloyd HD, Helyar SJ, Brakefield PM (2005) Inbreeding uncovers fundamental differences in the genetic load affecting male and female fertility in a butterfly. Proceedings of the Royal Society of London Series B - Biological Sciences 272:39-46

Saito DS, Saitoh T, Nishiumi I (2005) Isolation and characterization of microsatellite markers in Ijima's leaf warbler, *Phylloscopus ijimae* (Aves : Sylviidae). Molecular Ecology Notes 5:666-668

Saladin V, Bonfils D, Binz T, Richner H (2003) Isolation and characterization of 16 microsatellite loci in the great Tit *Parus major*. Molecular Ecology Notes 3:520-522

Scribner KT, Pearce JM (2000) Microsatellites: evolutionary and methodological background and empirical applications at individual, population, and phylogenetic levels. In: Baker A (ed) Molecular Methods in Ecology. Blackwell Science Limited, London, pp 235-271

Seddon N, Amos W, Mulder RA, Tobias JA (2004) Male heterozygosity predicts territory size, song structure and reproductive success in a cooperatively breeding bird. Proceedings of the Royal Society of London Series B - Biological Sciences 271:1823-1829

Shields GF (1982) Comparative avian cytogenetics: a review. Condor 84:45-58

Simeoni M, Dawson DA, Gentle L, Coiffait L, Wolff K, Evans KL, Gaston K, Hatchwell BJ (2009) Characterization of 38 microsatellite loci in the blackbird Turdus merula (Turdidae, AVES). Molecular Ecology Resources 9:1520-1526

Slate J, David P, Dodds KG, Veenvliet BA, Glass BC, Broad TE, McEwan JC (2004) Understanding the relationship between the inbreeding coefficient and multilocus heterozygosity: theoretical expectations and empirical data. Heredity 93:255-265

Slate J, Hale MC, Birkhead TR (2007) Simple sequence repeats in zebra finch (*Taeniopygia guttata*) expressed sequence tags: a new resource for evolutionary genetic studies of passerines. BMC Genomics 8:52

Slate J, Pemberton JM (2002) Comparing molecular measures for detecting inbreeding depression. Journal of Evolutionary Biology 15:20-31

Smallegange IM, Fiedler W, Koppen U, Geiter O, Bairlein F (2010) Tits on the move: exploring the impact of environmental change on blue tit and great tit migration distance. Journal of Animal Ecology 79:350-357

Stapley J, Birkhead TR, Burke T, Slate J (2008) A linkage map of the zebra finch *Taeniopygia guttata* provides new insights into avian genome evolution. Genetics 179:651-667

Steinmeyer C, Mueller JC, Kempenaers B (2009) Search for informative polymorphisms in candidate genes: clock genes and circadian behaviour in blue tits. Genetica 136:109-117

Steinmeyer C, Schielzeth H, Mueller JC, Kempenaers B (2010) Variation in sleep behaviour in free-living blue tits, *Cyanistes caeruleus*: effects of sex, age and environment. Animal Behaviour 80:853-864

Stenzler LM, Fitzpatrick JW (2002) Isolation of microsatellite loci in the Florida scrub-jay *Aphelocoma coerulescens*. Molecular Ecology Notes 2:547-550

Stenzler LM, Fraser R, Lovette IJ (2004) Isolation and characterization of 12 microsatellite loci from golden-winged warblers (*Vermivora chrysoptera*)

with broad cross-taxon utility in emberizine songbirds. Molecular Ecology Notes 4:602-604

Svensson L (1992) Identification guide to european passerine birds, 4th edn, Stockholm

Szulkin M, Bierne N, David P (2010) Heterozygosity-fitness correlations: a time for reappraisal. Evolution 64:1202-1217

Tang JF, Baldwin SJ, Jacobs JME, van der Linden CG, Voorrips RE, Leunissen JAM, van Eck H, Vosman B (2008) Large-scale identification of polymorphic microsatellites using an in silico approach. BMC Bioinformatics 9:374

Taylor SS, Sardell RJ, Reid JM, Bucher T, Taylor NG, Arcese P, Keller LF (2010) Inbreeding coefficient and heterozygosity-fitness correlations in unhatched and hatched song sparrow nestmates. Molecular Ecology 19:4454-4461

Temple HJ, Hoffman JI, Amos W (2006) Dispersal, philopatry and intergroup relatedness: fine-scale genetic structure in the white-breasted thrasher, *Ramphocinclus brachyurus*. Molecular Ecology 15:3449-3458

Tomiuk J, Stauss M, Segelbacher G, Fietz J, Kompf J, Bachmann L (2007) Female genetic heterogeneity affects the reproduction of great tits (*Parus major* L., 1758) in low-quality woodlands. Journal of Zoological Systematics and Evolutionary Research 45:144-150

Townsend AK, Clark AB, McGowan KJ, Buckles EL, Miller AD, Lovette IJ (2009) Disease-mediated inbreeding depression in a large, open population of cooperative crows. Proceedings of the Royal Society B - Biological Sciences 276:2057-2064

Tremblay I, Thomas DW, Lambrechts MM, Blondel J, Perret P (2003) Variation in blue tit breeding performance across gradients in habitat richness. Ecology 84:3033-3043

Trifonov EN (2004) Tuning function of tandemly repeating sequences: a molecular device for fast adaptation. In: Wasser SP (ed) Evolutionary theory and processes: modern horizons, papers in honor of Eviatar Nevo. Kluwer Academic Publishers, The Netherlands, pp 115-138

Valcu M, Kempenaers B (2008) Causes and consequences of breeding dispersal and divorce in a blue tit, *Cyanistes caeruleus*, population. Animal Behaviour 75:1949-1963

van de Pol MV, Wright J (2009) A simple method for distinguishing within- versus between-subject effects using mixed models. Animal Behaviour 77:753-758

Vasemagi A, Nilsson J, Primmer CR (2005) Expressed sequence tag-linked microsatellites as a source of gene-associated polymorphisms for detecting signatures of divergent selection in Atlantic salmon (*Salmo salar* L.). Molecular Biology and Evolution 22:1067-1076

von Hardenberg A, Bassano B, Festa-Bianchet M, Luikart G, Lanfranchi P, Coltman D (2007) Age-dependent genetic effects on a secondary sexual trait in male Alpine ibex, *Capra ibex*. Molecular Ecology 16:1969-1980

Wang MT, Hsu YC, Yao CT, Li SH (2005) Isolation and characterization of 12 tetranucleotide repeat microsatellite loci from the green-backed tit (*Parus monticolus*). Molecular Ecology Notes 5:439-442

Waples RS, Gaggiotti O (2006) What is a population? An empirical evaluation of some genetic methods for identifying the number of gene pools and their degree of connectivity. Molecular Ecology 15:1419-39

Warren WC, Clayton DF, Ellegren H, Arnold AP, Hillier LW, Kunstner A, Searle S, White S, Vilella AJ, Fairley S, Heger A, Kong LS, Ponting CP, Jarvis ED, Mello CV, Minx P, Lovell P, Velho TAF, Ferris M, Balakrishnan CN, Sinha S, Blatti C, London SE, Li Y, Lin YC, George J, Sweedler J, Southey B, Gunaratne P, Watson M, Nam K, Backstrom N, Smeds L, Nabholz B, Itoh Y, Whitney O, Pfenning AR, Howard J, Voelker M, Skinner BM, Griffin DK, Ye L, McLaren WM, Flicek P, Quesada V, Velasco G, Lopez-Otin C, Puente XS, Olender T, Lancet D, Smit AFA, Hubley R, Konkel MK, Walker JA, Batzer MA, Gu WJ, Pollock DD, Chen L, Cheng Z, Eichler EE, Stapley J, Slate J, Ekblom R, Birkhead T, Burke T, Burt D, Scharff C, Adam I, Richard H, Sultan M, Soldatov A, Lehrach H, Edwards SV, Yang SP, Li XC, Graves T, Fulton L, Nelson J, Chinwalla A, Hou SF, Mardis ER, Wilson RK (2010) The genome of a songbird. Nature 464:757-762

Weir BS, Cockerham CC (1973) Mixed self and random mating at two loci. Genetical Research 21:247-262

Whitlock M (1993) Lack of correlation between heterozygosity and fitness in forked fungus beetles. Heredity 70:574-581

Wright S (1922) Coefficients of inbreeding and relationship. American Naturalist 56:330-338

Wright S (1965) The interpretation of population structure by F-statistics with special regard to systems of mating. Evolution 19:395-420

Yodogawa Y, Nishiumi I, Saito D, Okanoya K (2003) Characterization of eight polymorphic microsatellite loci from the Bengalese finch (*Lonchura striata* var. *domestica*). Molecular Ecology Notes 3:183-185

Zane L, Bargelloni L, Patarnello T (2002) Strategies for microsatellite isolation: a review. Molecular Ecology 11:1-16

Zouros E, Foltz DW (1987) The use of allelic isozyme variation for the study of heterosis. Isozymes: Current Topics in Biological and Medical Research 13:1-59

Author contributions

Chapter 1

B.K., J.M. and J.O.M. contributed with concept and study design. D.A.D, B.H., M.L., J.C.M and J.O.M with marker design. A.G. and J.O.M. with multiplex optimization and genotyping of the Austrian population. B.H. and M.L. with genotyping and population genetic analysis of the Swedish population. J.O.M. with *in-silico* mapping, population genetic analyses from the Austrian population and writing of the chapter. D.A.D., B.H., B.K. and J.C.M with discussion and comments on the chapter.

Chapters 2 and 3

J.O.M. contributed with concept, study design, genotyping, data analysis and writing of the chapters. J.M. contributed with concept, study design, guidance on data analysis, discussion and comments on the chapters. B.K. provided the phenotypic data from the study population and contributed with concept, study design, discussion and comments on the chapters.

Addresses of co-authors

Deborah A. Dawson
University of Sheffield
Department of Animal and Plant Sciences
Sheffield, S10 2TN, UK
Email: D.A.Dawson@sheffield.ac.uk

Alexander Girg
Max Planck Institute for Ornithology
Department of Behavioral Ecology and
Evolutionary Genetics
Eberhard-Gwinner-Strasse 7/8
D-82319 Seewiesen, Germany
Email: girg@orn.mpg.de

Bengt Hansson
Lund University
Department of Animal Ecology,
Ecology Building
SE-223 62 Lund, Sweden
Email: Bengt.Hansson@biol.lu.se

Bart Kempenaers
Max Planck Institute for Ornithology
Department of Behavioral Ecology and
Evolutionary Genetics
Eberhard-Gwinner-Strasse 7/8
D-82319 Seewiesen, Germany
Email: b.kempenaers@orn.mpg.de

Marcus Ljungqvist
Lund University
Department of Animal Ecology,
Ecology Building
SE-223 62 Lund, Sweden
Email: marcus.ljungqvist@zooekol.lu.se

Jakob C. Mueller
Max Planck Institute for Ornithology
Department of Behavioral Ecology and
Evolutionary Genetics
Eberhard-Gwinner-Strasse 7/8
D-82319 Seewiesen, Germany
Email: mueller@orn.mpg.de

Curriculum vitae

Juanita Olano Marín

Max Planck Institute for Ornithology
Department of Behavioral Ecology and Evolutionary Genetics
Eberhard-Gwinner-Str. 7/8
82319 Seewiesen, Germany
Email: olano@orn.mpg.de
Phone: ++49 (0)8157 932-328
Fax: ++49 (0)8157 932-400

Date and place of birth

17 October 1982, Bogotá, Colombia

Education

2007 - 2011	Doctoral student, Max Planck Institute for Ornithology, Seewiesen Disssertation "Exploring the causes of heterozygosity-fitness correlations in the blue tit". Supervisors: Dr. Jakob C. Mueller, Prof. Dr. Bart Kempenaers
1999 - 2005	BSc in biology and microbiology, Universidad de los Andes, Bogotá Thesis "Correlation between mutations in DNA repair genes and antibiotics resistance in *Mycobacterium tuberculosis*". Supervisor: Dr. Maria Mercedes Zambrano

List of publications

Peer-reviewed publications

Mueller J. C., Hermisson J., **Olano-Marin J.**, Hansson B., Kempenaers B. (2011) Linking genetic mechanisms of heterozygosity-fitness correlations to footprints of selection at single loci. Evolutionary Ecology 25:1-11.

Olano-Marin J., Dawson D. A., Girg A., Hansson B., Ljungqvist M., Kempenaers B., Mueller J. C. (2010). A genome-wide set of 106 microsatellite markers for the blue tit (*Cyanistes caeruleus*). Molecular Ecology Resources 10: 516–532.

Hansson B., Ljungqvist M., Dawson D. A., Mueller J. C., **Olano-Marin J.**, Ellegren H., Nilsson J.-A. (2010). Avian genome evolution: insights from a linkage map of the blue tit (*Cyanistes caeruleus*). Heredity 104:67–78.

Farji-Brener A. G., Carvajal D., Gei M. G., **Olano J.**, Sanchez J. D. (2008). Direct and indirect effects of soil structure on the density of an antlion larva in a tropical dry forest. Ecological Entomology 33:183–188.

Olano J., Lemos M. P., López B., Reyes A., Correa N., Del Portillo P., Robledo J., Ritacco V., Zambrano M. M. (2007). Mutations in DNA repair genes are associated with the Haarlem lineage of *Mycobacterium tuberculosis* independently of their antibiotic resistance. Tuberculosis 87:502-508.

Amat E., **Olano J**., Forero F., Botero C. (2007). Notas sobre *Philornis vulgaris* (Diptera: Muscidae) en nidos del Sinsonte Tropical (*Mimus gilvus*) en el centro de Colombia (Notes about *Philornis vulgaris* (Diptera: Muscidae) in Tropical Mockingbird (*Mimus gilvus*) nests in central Colombia). Acta Zoológica Mexicana (nueva serie) 23: 205-207.

Manuscripts accepted or in press

Olano-Marin J., Mueller J. C., Kempenaers B. Correlations between heterozygosity and reproductive success in the blue tit (*Cyanistes caeruleus*): an analysis of inbreeding and single locus effects. *Evolution*, in press.

Olano-Marin J., Mueller J. C., Kempenaers B. Heterozygosity and survival in blue tits (*Cyanistes caeruleus*): contrasting effects of presumably functional and

neutral loci. Accepted for publication in *Molecular Ecology*, pending minor revisions.

Non-peer-reviewed publications

Murillo A., Burbano C., **Olano J**. (2003). Cangrejo Ermitaño busca casa móvil (Hermit crab searches for mobile house). Hipótesis 2:22–29.

Other publications

Olano J. *et al.* (2007). Ecología Tropical y Conservación 2007–02 (Tropical Ecology and Conservation 2007-02). Course memories. Organización para Estudios Tropicales (OET) and Universidad de Costa Rica. In: http://www.ots.duke.edu/en/education/pdfs/graduate/coursebooks/07-2.pdf.

Olano J. (2005). Correlación entre mutaciones en genes involucrados en reparación del ADN y resistencia a antibióticos en *Mycobacterium tuberculosis* (Correlation between mutations in DNA repair genes and antibiotics resistance in *Mycobacterium tuberculosis*). BSc. Biology and BSc. Microbiology thesis. Universidad de los Andes.

Erklärung

Ich erkläre hiermit, dass ich die vorliegende Arbeit ohne unzulässige Hilfe Dritter und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe. Die aus anderen Quellen direkt oder indirekt übernommenen Daten und Konzepte sind unter Angabe der Quelle gekennzeichnet. Weitere Personen, insbesondere Promotionsberater, waren an der inhaltlichen materiellen Erstellung dieser Arbeit nicht beteiligt. Die Arbeit wurde weder im In- noch im Ausland in gleicher oder ähnlicher Form einer Prüfungsbehörde vorgelegt.

Juanita Olano Marín