MOLECULAR ECOLOGY

Molecular Ecology (2013) **22**, 3963–3980 doi: 10.1111/mec.12375

Partitioning of genetic variation across the genome using multimarker methods in a wild bird population

MATTHEW R. ROBINSON,*† ANNA W. SANTURE,* ISABELLE DeCAUWER,*‡
Ben C. Sheldon§ and JON SLATE*

Department of Animal and Plant Science, University of Sheffield, Western Bank, Sheffield, S10 2TN, UK, †Queensland Statistical Genetics Laboratory, Queensland Institute of Medical Research, Royal Brisbane Hospital, 300 Herston Road, Brisbane 4029, Australia, ‡Laboratoire de Génétique et Evolution des Populations Végétales, UMR CNRS 8198, Bâtiment SN2, Université des Sciences et Technologies de Lille – Lille 1, F-59655 Villeneuve d'Ascq Cedex, France, §Department of Zoology, Edward Grey Institute, University of Oxford, Oxford, OX1 3PS, UK

Abstract

The underlying basis of genetic variation in quantitative traits, in terms of the number of causal variants and the size of their effects, is largely unknown in natural populations. The expectation is that complex quantitative trait variation is attributable to many, possibly interacting, causal variants, whose effects may depend upon the sex, age and the environment in which they are expressed. A recently developed methodology in animal breeding derives a value of relatedness among individuals from high-density genomic marker data, to estimate additive genetic variance within livestock populations. Here, we adapt and test the effectiveness of these methods to partition genetic variation for complex traits across genomic regions within ecological study populations where individuals have varying degrees of relatedness. We then apply this approach for the first time to a natural population and demonstrate that genetic variation in wing length in the great tit (*Parus major*) reflects contributions from multiple genomic regions. We show that a polygenic additive mode of gene action best describes the patterns observed, and we find no evidence of dosage compensation for the sex chromosome. Our results suggest that most of the genomic regions that influence wing length have the same effects in both sexes. We found a limited amount of genetic variance in males that is attributed to regions that have no effects in females, which could facilitate the sexual dimorphism observed for this trait. Although this exploratory work focuses on one complex trait, the methodology is generally applicable to any trait for any laboratory or wild population, paving the way for investigating sex-, age- and environment-specific genetic effects and thus the underlying genetic architecture of phenotype in biological study systems.

Keywords: chromosome partitioning, genetic architecture, genomic relatedness, heritability, molecular quantitative genetics, partitioning genetic variance, sex-specific genetic variance

Received 21 January 2013; revision received 11 April 2013; accepted 11 April 2013

Long-term, individual-based studies of wild populations experiencing naturally occurring environmental conditions are fundamental to our understanding of evolution (Both *et al.* 2006; Charmantier *et al.* 2008; Robinson *et al.* 2008; Ozgul *et al.* 2009). However, studying the diversity of morphology, behaviour and physiology within and among natural populations is limited by a lack of knowledge of the underlying genetic architecture of continuously varying quantitative traits (Mackay *et al.* 2009; Hill 2012). Genetic variation in these complex traits may be attributable to many, possibly interacting, genes whose expression may be context dependent (Falconer

Correspondence and Present address: Matthew R. Robinson, The University of Queensland, Queensland Brain Institute (QBI), QBI Building (#79) St Lucia, QLD 4072, Australia
Fax: 0044 (0)114 222 0002;
E-mail: matthew.r.robinson@sheffield.ac.uk

& Mackay 1996; Mackay 2001; Hill *et al.* 2008). Dissection of this genetic variation into underlying causative factors is thus likely to be difficult, requiring an approach that is able to accurately describe the overall genetic architecture of complex quantitative traits in populations that contain close relatives.

The primary goal of most genetic mapping experiments is to identify the locations of genes that affect among-individual trait variation and to estimate the genetic effects of these quantitative trait loci (QTL). QTL mapping by linkage analysis has been used to great effect to address these goals (Johnson *et al.* 2010; Slate *et al.* 2010), but there are two main limitations of this approach. First, it does not always accurately estimate the variance attributed to QTL, and second, it does not accurately detect the number of QTL. True undetected QTL have effects that contribute to the estimates of the regions that are detected as having QTL and sampling thus occurs from an upwardly biased set of estimated effects (Beavis 1994; Göring *et al.* 2001). Significant associations are reported only when test statistics exceed a predetermined critical threshold, which results in declared QTL regions having effects that appear much larger than they really are, particularly in studies with small sample size (Slate 2013). Therefore, QTL mapping by linkage analysis is useful for locating genomic regions of large effect, but this approach is unable to accurately describe the overall genetic architecture of complex quantitative traits in terms of the number and effect size of underlying causal variants.

Genome-wide association studies (GWAS) have led to the discovery of hundreds of marker loci that are associated with complex traits, including disease and quantitative phenotypes in human, livestock and crop populations (Buckler *et al.* 2009; Goddard & Hayes 2009; International Schizophrenia Consortium 2009; Allen *et al.* 2010). For most traits, however, the associated variants cumulatively explain only a small fraction of total heritability, as loci whose effect sizes are too small to reach genome-wide statistical significance will not be detected. This results in a 'missing heritability' (Manolio *et al.* 2009), meaning that GWAS is also unable to accurately describe the overall genetic architecture of complex quantitative traits in terms of the number and effect size of underlying causal variants.

To overcome the limitations of these two approaches, a recently developed multimarker method has been used to partition additive genetic variance across the genome. The actual genome-wide relatedness between pairs of individuals, defined as the proportion of the genome that two relatives share identity by descent (IBD), varies around its expectation because of Mendelian segregation, with the exception of monozygotic twins and parent–offspring pairs (Visscher 2009; Hill & Weir 2011). Dense single nucleotide polymorphism (SNP) genotyping has now made it possible to estimate the actual genome-wide relatedness among individuals using genetic markers (Visscher 2009). Estimating identity by state from high-density SNP markers was first developed in animal breeding to accurately predict breeding values in livestock populations that contain close relatives ('genomic prediction'). This revolutionized the field, enabling increased accuracy of breeding value prediction, and allowed candidate individuals to be selected without requiring the measurement of their phenotype (Van Raden *et al.* 2009; Hayes *et al.* 2010; Dekkers 2012).

These multimarker methods have recently been applied to estimate additive genetic variation in human height and other quantitative phenotypes in unrelated human study populations (Yang *et al.* 2010, 2011) and among sib pairs (Visscher *et al.* 2006, 2007). Human medical genetic studies typically use 'unrelated' individuals, and thus, the approach is conceptually different than in animal breeding. Applied in this setting, multimarker methods capture the contribution from all causal variants that are in linkage disequilibrium (LD) with genotyped markers, in the same way as for single markers in GWAS, meaning that the total heritability explained by the additive effects of all causal variants in linkage with the SNP markers is estimated. These developments have lead to revolutionary insights into the underlying genetic architecture of human disease and other complex phenotypes (e.g. Yang *et al.* 2010, 2011).

For ecological study populations, multimarker approaches could provide accurate estimates of heritability and allow genetic variance to be accurately partitioned across the genome. Importantly, this approach can also be adopted irrespective of whether a pedigree is available or not, enabling additive genetic variance to be calculated without a pedigree or breeding design. This would enable a description of the overall genetic architecture of complex quantitative traits in terms of the number and effect size of underlying causal variants for any population. Previous studies provide evidence that sharing of haplotype blocks within families can be estimated with a high degree of accuracy with sparsely distributed genome-wide markers (Habier *et al.* 2009), and that genetic variance could be estimated based on few full-sib families (Ødergård & Meuwissen 2012). However, in more complex pedigrees and study designs typical of ecological data sets, where more distant relatives are common, tracing haplotype blocks over multiple generations may be more challenging. This is because DNA blocks will be shortened due to more recombination, and actual genomic relationships are expected to deviate more from their expected values in more distantly related individuals (Hill & Weir 2011). While the accuracy of 'genomic prediction' in large half-sib livestock

populations has been tested, where marker density and sample sizes are far greater than ecological study populations, there has been no testing of the bias and the accuracy of these approaches in recovering estimated effects across different regions of the genome. Therefore, before these approaches can be applied for the first time in molecular ecology, the accuracy and bias of multimarker methods to estimate additive genetic variance and partition this variance across the genome requires testing.

Here, we adapt these multimarker methods, developed in animal breeding, to partition additive genetic variance across chromosomal regions within a wild population for the first time, and to test for sex-specific genetic effects across the genome. We also conduct the first simulation study to test for bias and the accuracy of these approaches in recovering estimated effects, using data that are similar to that available from studies of naturally occurring populations. Having established that the approach is appropriate for ecological data, we then analyse the complex quantitative trait of wing length, measured within a comprehensively studied wild population of great tits, *Parus major*, breeding at Wytham Woods, near Oxford, UK (McCleery *et al.* 2004; Charmantier *et al.* 2008). We use wing length as a model trait because it allowed us to demonstrate multimarker methods in full, due to the fact that wing length is highly heritable, and because it is the trait for which a large number of phenotypic measures were available for a large number of genotyped individuals (2098 individuals and 4575 measures).

Wing length variation is partly caused by differences among individuals in the length of primary feathers in the wing, partly by differences among individuals in forelimb length, and also reflects differences in body size, sex, age and general condition (Jensen *et al.* 2003; Gienapp & Merilä 2010). This trait is therefore a classic example of a complex quantitative trait, which is highly heritable and shows sexual dimorphism across many populations. Wing length is associated with migratory behaviour and thus shows striking adaptive diversity within and among avian species (Lockwood *et al.* 1998; Dawideit *et al.* 2009; Rolshausen *et al.* 2009), with abundant within-population genetic variation that has been shown to respond rapidly to environmental changes (Rolshausen *et al.* 2009). In general, the analyses of within-species genetic variation will yield a deeper understanding of microevolution of morphological traits in natural populations (Jensen *et al.* 2003; Gienapp & Merilä 2010; Tarka *et al.* 2010). Therefore, in addition to revealing the underlying genetic architecture of this trait within this population, our aim is to adapt multimarker approaches to pave the way for investigating sex-, age- and environment-specific genetic effects across the genome, which we believe will provide great insight into the underlying genetic architecture of phenotype for any ecological study population.

Materials and methods

Observational data used in this study

The population. Great tits have been studied at Wytham Woods, near Oxford, United Kingdom, since the 1940s, with nest boxes first erected in 1947. As described elsewhere (McCleery *et al.* 2004; Charmantier *et al.* 2008), a wide range of morphological and life history phenotypes have been recorded for individuals since the early 1960s. Blood samples have been collected for most birds since 2005, and from nestlings for a smaller subset of individuals between 1985 until 2005, allowing DNA to be extracted from nucleated blood cells (Santure *et al.* 2011).

Phenotypic data. There were 4293 measures on 1949 genotyped individuals born between 1985 and 2009 for which the sex and age were known. Age was classified as 0 for individuals measured during the year of their birth (305 measures); 1 for those caught in the year after their birth (2293 measures); and 2 for those of any other age (1695 measures).

Genomic data. A panel of almost 10 000 SNPs have previously been selected for inclusion on an Illumina iSelect BeadChip, 'SNP chip' (van Bers *et al.* 2012). A total of 2644 individuals were successfully genotyped, with 7203 of the 9193 SNPs included on the SNP chip displaying population-level variation (van Bers *et al.* 2012).

Pedigree and identity checking. A social pedigree was available based on field observations, and this was corrected using the genotype information by (i) checking the consistency between field-recorded and genetic sex based on the observed heterozygosity of Z-linked markers, (ii) identifying and removing parent–offspring pairs raising large numbers of Mendelian inheritance errors at autosomal markers and (iii) using categorical paternity analysis to assign a number of genetic fathers to individuals resulting from extra-pair paternities (van Bers *et al.* 2012). A total of 2497 individuals of confirmed identity were included in further analyses.

Construction of a genetic linkage map. A great tit genetic linkage map was constructed for 32 chromosomes (1–15, 17–24, 26–28, 1A, 4A, 25A, 25B, Z and an additional linkage group LGE22) containing 4878 markers. An additional 714 markers whose chromosomal location could be inferred by comparative mapping with the assembled zebra finch genome were added to the map, giving a set of 5312 'chromosome-assigned' markers (van Bers *et al.* 2012). Given the small number of markers on some chromosomes, a total of 23 chromosomes or chromosome sets were constructed: chromosomes 1–15, 17–20, 1A, 4A, a

pooled set of markers from microchromosomes 21–28 plus the linkage group LGE422, and the Z chromosome. These chromosomes contained a predicted total of 15 448 genes based on homology with the zebra finch genome (van Bers *et al*. 2012).

Simulated data used in this study

Data were simulated with the exact same marker structure and number of individuals as for the Wytham great tit data. Three different trait architectures were simulated: (i) a polygenic architecture where a heritability of 0.4 is explained by 200 QTL spaced every 10 centimorgan (cM) throughout the genome; (ii) an architecture where a heritability of 0.4 is explained 22 QTL, with each QTL on a different linkage group; and (iii) an architecture where a heritability of 0.4 is explained by five QTL on separate linkage groups, with the remaining linkage groups containing no QTL. For each type of architecture, we analysed 10 replicate data sets. QTL did not have a uniform effect size, either between replicates or between chromosomes within a replicate (see below for an explanation and justification for this).

To create the simulated data, we used the software QMSim (Sargolzaei & Schenkel 2009), which simulates populations based on a forward in time process (Sargolzaei & Schenkel 2009). The QMSim software has been used previously to simulate genetic architectures and test multimarker methods for genomic prediction (e.g. Brito *et al*. 2011).

In the first simulation step, 100 generations with a constant population size of 500 were simulated. The number of individuals of each sex was equal, and the mating system was based on the random mating of gametes, sampled from both the male and female gamete pools. Therefore, only two evolutionary forces were considered: mutation and drift. We assumed a recurrent mutation model where mutation alters the allelic state with equal transition probabilities from one allele to another. The mutation rates for both markers and QTL were specified at a rate of 2.5 e−5 per gene per generation, with the number of mutations sampled from a Poisson distribution, and the assumption that mutation rates are equal for all loci within markers and QTL. Recombination across these generations was sampled from a Poisson distribution with a mean derived from the length of the chromosomes and the location of crossover assigned at random across the chromosomes. We simulated the same number and length of chromosomes, similar LD structure and similar number and spacing per chromosome of SNP markers as in the Wytham great tit data described above.

In the second step, recent generations are then simulated by randomly selecting 200 males and 200 females, representing those first sampled at the start of the study (founding individuals). From these individuals, we simulated five overlapping generations, giving a total of 1800 individuals. Mating was random where an average of two offspring are produced from each dam, with gametes randomly selected from the male and female gametic pools. Fifty percent of the sires and dams were replaced in all generations, and the selection design was random with respect to the simulated phenotype, but the probability of culling was based on age (so that the oldest 50% were replaced). Phenotype and genotype data were outputted for all individuals in the recent population, along with the pedigree. LD was calculated in the final generation as the pooled square correlation between adjacent loci (Hill & Robertson 1968).

This entire simulation process was repeated 30 times, to give 10 independently generated replicate populations for each of three simulated trait architectures. All traits had a simulated heritability of 0.4 and a total phenotypic variance of 1. We simulated the following architectures:

1. for the first architecture, we simulated a scenario where a heritability of 0.4 is divided across 200 QTL that were evenly spaced every 10 cM across the genome. The individual QTL effects sampled from a gamma distribution with shape parameter 0.1 and the heritability of all the QTL summed to 0.4. This gave a trait with many QTL, each of small average effect (average effect size of each QTL: $h^2 = 0.002$), spread evenly across the genome so that the effect size of a linkage group should on average equal its length (hereafter termed 'polygenic').
2. for the second architecture, we simulated a scenario where a heritability of 0.4 is divided across 22 QTL, with each QTL on a different linkage group. The QTL effects were sampled from a gamma distribution with shape parameter 0.5. This gives a trait with some moderately large QTL effects, where on average across replicates all chromosomes should contribute towards the additive genetic variance (average effect size of each QTL: $h^2 = 0.018$), but not in proportion to their length (trait with moderate effect QTL).
3. for the third architecture, we simulated a scenario where a heritability of 0.4 is divided across 5 QTL located on linkage groups 1, 5, 6, 9 and 12 in the same position across all replicate populations. QTL effects were sampled from a gamma distribution with shape parameter 1. This gave a trait with 5 large effect QTL (average effect size of each QTL: $h^2 = 0.08$), where some chromosomes contribute towards the heritability and others will not (trait with large effect QTL).

Although there are the same numbers of QTL across replicates, and all QTL are in the same locations across

replicates, the effect size of an individual QTL and thus a particular chromosome will vary across replicates. This is because QTL effects were sampled from a gamma distribution within each replicate and were not fixed, as we felt this was more likely to be realistic given theory predicts that effect sizes in natural populations will follow a distribution (Otto & Jones 2000), and we wanted to assess the ability of our approach to detect a range of effect sizes.

Applying multimarker methods to ecological study populations

The concept of IBD is used to indicate two homologous alleles that have descended from a common ancestor, with the probability of IBD defined with respect to a base population (Lynch & Walsh 1998). Estimating IBD coefficients has been used for some linkage mapping studies (typical termed variance component QTL mapping studies) in many populations where relatedness is known, where individuals with no known ancestors form a natural base population of 'unrelated' founding individuals (Lynch & Walsh 1998). However, if the aim is to gain an unbiased estimate of the genome-wide relatedness between all individuals, then estimating IBD may not be the most appropriate approach. All individuals are related if traced back further enough; and consequently, estimating IBD loses much of the information contained in the data. A recent review has argued that IBD probabilities are mostly used to predict the probability that individuals carry alleles that are identical by state (IBS) at unobserved loci and demonstrated that relatedness calculated by IBS more accurately represents true relatedness than IBD coefficients (Powell *et al.* 2010). Therefore, in this study, we present an association-based variance partitioning approach that uses 'multimarker' techniques, rather than using single markers and estimating their effects individually as in GWAS studies.

In the next section, we provide an overview of using multimarker methods to estimate heritability and to partition additive genetic variation across the genome. A workflow is provided in Box 1, which provides an overview of each step. Each step is then explained in detail below. All statistical testing was conducted in ASReml (Gilmour *et al.* 2006), and all matrix algebra was conducted in R version 2.15 (R Foundation for Statistical Computing 2013).

Estimating relatedness from SNP markers

Marker-derived IBS relatedness estimates can be obtained using numerous methods, and below we used three of the main methods that have been presented to date following Van Raden (2007). Let **M** be a matrix of autosomal marker allele scores, with rows corresponding to the number of individuals (n) and columns corresponding to the number of loci (m). Elements of **M** are −1, 0 or 1 depending upon whether the individual is a homozygote, heterozygote or the other homozygote, respectively, for each marker. The diagonals of an $n \times n$ matrix of **MM′** counts the number of homozygous loci for each individual, and off-diagonals measure the number of alleles shared by relatives. If the frequency of the second allele at locus i is p_i and **P** contains these allele frequencies expressed as a difference from 0.5 and multiplied by 2, then column i of **P** is $2(p_i - 0.5)$. If we subtract **P** from **M** to get **Z**, then the mean values of the allele effects are set to 0.

The first approach weights markers by reciprocals of their expected variance using the formula $\mathbf{G}_1 = \mathbf{ZDZ}'$, where **D** is diagonal with $D_{ii} = 1/m[2p_i(1-p_i)]$. This scales $\mathbf{G}_1$ to be analogous to the relationship matrix **A**, and we used the observed allele frequencies of the SNPs in the current population to do this. This approach is identical to that proposed and utilized in human medical genetic studies (Yang *et al.* 2010, 2011) and in the 'genomic selection' methods used to predict breeding values in livestock (Hayes *et al.* 2010). A second approach does not require allele frequencies but regresses **MM′** onto a kinship matrix calculated from the pedigree, **A**, to obtain $\mathbf{G}_2$ using the model $\mathbf{MM}' = g_0\mathbf{11}' + g_1\mathbf{A} + \mathbf{E}$. The regression is fit with **MM′** as the dependent and **A** as independent, with $\mathbf{G}_2 = \mathbf{MM}' - g_0(\mathbf{11}')/g_1$. Finally, a third approach was used, analogous to the first, where a normalized matrix is obtained using $\mathbf{G}_3 = \mathbf{ZZ}'/[\text{trace}(\mathbf{ZZ}')/m]$. This scales $\mathbf{G}_3$ by its actual variance rather than the expected variance.

Estimating relatedness at the sex chromosomes

For the sex chromosomes, we adopt the first approach of weighting marker relatedness by reciprocals of their expected variance (Yang *et al.* 2010, 2011). Because the heterogametic sex only has one copy of the sex chromosome, SNP alleles on the sex chromosomes are scored as 0 or 1 for the heterogametic sex and 0, 1 or 2 for the homogametic sex referring to the number of copies of a given SNP allele. Relatedness estimators are then modified to:

$\mathbf{G}^{\text{het}} = (\mathbf{M}-\mathbf{P})\mathbf{D}(\mathbf{M}-\mathbf{P})'$ for the heterogametic pair, where **D** is diagonal with $D_{ii} = 1/m[p_i(1-p_i)]$ and **P** is $(p_i - 0.5)$.

$\mathbf{G}^{\text{hom}} = (\mathbf{M}-\mathbf{P})\mathbf{D}(\mathbf{M}-\mathbf{P})'$ for the homogametic pair, where **D** is diagonal with $D_{ii} = 1/m[2p_i(1-p_i)]$ and **P** is $2(p_i - 0.5)$.

$\mathbf{G}^{\text{het,hom}} = (\mathbf{M}-\mathbf{P})\mathbf{D}(\mathbf{M}-\mathbf{P})'$ for a heterogametic and homogametic pair, where **D** is diagonal with $D_{ii} = 1/m[\sqrt{2}p_i(1-p_i)]$ and **P** is $2(p_i - 0.5)$ for the homogametic individual and $(p_i - 0.5)$ for the heterogametic individual.

Box 1: Applying multi-marker methods to ecological study populations

Step 1—Estimating relatedness from markers

Markers →
Individuals ↓

0	1	-1	0	1
1	-1	0	1	1
-1	0	1	-1	0
0	1	1	-1	-1

M is a matrix of SNP marker scores where 0 is a heterozygote, 1 is a homozygote, and -1 is the other homozygote

P is a matrix of the allele frequencies *p* of each SNP *i* expressed as $2(p_i-0.5)$

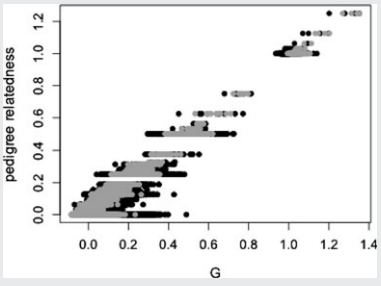
Subtracting **P** from **M** gives **Z**, a matrix of allele effects that are set to 0 mean value

Three different approaches can then be used to provide a matrix of relatedness **G** containing point estimates of the pair-wise relatedness among individuals from the matrix **Z** (see Materials and Methods).

Step 2—Weighting marker-derived relatedness estimates

Point estimates of marker-derived relatedness may not accurately reflect actual relatedness. This is because markers may not be in LD with all of the unobserved loci that we take them to represent. Thus, there may be sampling error in estimating coefficients of relatedness.

The figure shows marker-derived relatedness matrix **G** plotted against an identity-by-descent kinship coefficient as estimated from the pedigree **A**. Black shows the un-weighted estimates of **G** and grey gives the weighted estimates of **G**. Weighting **G** towards **A**, reduces the sampling error around the expected values.



If the population has a pedigree—**G** can be weighted to the expected relationships estimated from the pedigree relatedness matrix **A**, using a proxy measure of LD.

If there is no pedigree—**G** can be weighted directly using a proxy measure of LD if the population contains'unrelated' individuals.

Step 3—Estimating heritability from marker-based estimates of relatedness

Weighted estimates of marker based relatedness can then be used to estimate additive genetic variance within an 'animal model'

$y = \mathbf{X}\beta + \mathbf{Z}v + e$ where **Z** is a design matrix relating individuals to additive genetic effects *a* with $V(v) = \mathbf{G}\sigma_a^2$ where **G** is the relationship matrix and σ_a^2 is the polygenic variance.

Step 4—Partitioning genetic effects across the genome using marker-based relatedness

Additive genetic variance can then be partitioned across different regions of the genome. To do this multiple **G** matrices from different regions of the genome are included within one model:

$y = \mathbf{X}\beta + \mathbf{Z}_x v_x + \mathbf{Z}_j v_j + e$ where $\mathbf{Z}_x$ and $\mathbf{Z}_y$ are design matrices relating individuals to additive genetic effects *a* at two different regions of the genome *x* and *j*

If we assume no dosage compensation, where each allele has a similar effect on the trait irrespective of the sex, then the genetic variance on the sex chromosome for the homogametic sex is twice that of the heterogametic sex. For this scenario, we can redefine the relationship matrix for the sex chromosome as $\mathbf{G} = 1/2\mathbf{G}^{het}$ for heterogametic pairs, $\mathbf{G} = \mathbf{G}^{hom}$ for homogametic pairs and $\mathbf{G} = 1/\sqrt{2}\mathbf{G}^{het,hom}$ for heterogametic–homogametic pairs.

If we assume full dosage compensation, where each allele in the homogametic sex has only half the effect of an allele in the heterogametic sex, we can redefine the relationship matrix for the sex chromosome as $\mathbf{G} = 2\mathbf{G}^{het}$ for heterogametic pairs, $\mathbf{G} = \mathbf{G}^{hom}$ for homogametic pairs and $\mathbf{G} = \sqrt{2}\mathbf{G}^{het,hom}$ for heterogametic–homogametic pairs. Finally, we can assume equal variance on the sex chromosome for either sex and not adjust the sex chromosome relatedness in any way.

Weighting marker-derived relatedness estimates

In practice, we know little about the actual pairwise relatedness between individuals; rather we use marker-derived estimates to approximate this relatedness. There may be sampling error in estimating coefficients of relatedness, as point estimates of marker-derived relatedness may not accurately reflect actual relatedness, because markers may not be in LD with all of the unobserved loci that we take them to represent.

However, we can weight the marker-based relatedness estimates using a proxy measure of LD, which reduces the sampling variance around the estimates. Weighting the marker-derived relationship matrix can be performed for the whole genome, for individual chromosomes or for individual genomic segments. The way in which this is done will depend upon whether there are close relatives within the data or not.

If the data contain close relatives. If a pedigree is available, then we can follow a recently advocated approach in animal breeding (Goddard *et al.* 2011) and weight the marker-derived relatedness values to their expected pedigree-based values using a proxy measure of the LD between the markers and the underlying unobserved loci. This removes sampling error in our estimates of **G** that occur due to a finite number of markers. In brief, we (i) create two **M** matrices by randomly selecting SNPs to give 0.5 *m* SNPs per group (where *m* is the number of SNP loci); (ii) calculate the marker-derived relatedness matrix **G** for each set, giving **G1** and **G2**; (iii) calculate the expected relatedness from the pedigree **A** and subtract **A** from **G** within each set (**G1a** = **G1**−**A**; **G2a** = **G2**−**A**); (iv) estimate the covariance between the estimates of **G1a** and **G2a**; and (v) assume that **G1a** represents our marker-derived estimates or relatedness and **G2a** represents the actual relatedness, and thus that this covariance represents a proxy measure of the LD between the markers and the underlying unobserved loci, which is used as a weighting factor w, to weight **G** calculated from all markers towards the expected relationships **A** using $\mathbf{G}_n^{\mathrm{A}} = w\mathbf{G}_n + (1-w)\mathbf{A}$, where $\mathbf{G}_n^{\mathrm{A}}$ is a weighted marker-derived relatedness matrix estimated using either of the three approaches of estimating marker-derived relatedness (n) outlined above.

Box 1 plots un-weighted and weighted estimates of **G** against the expected pedigree relatedness values **A**. This shows the shrinkage of the variance in relatedness that occurs around the expected pedigree values due to the weighting process, removing sampling error created by low LD of SNP markers. It is important to note that accounting for relatedness before estimating the proxy measure of LD is required when relatedness is high. This is because in population of high relatedness both **G1a** and **G2a** will contain close relatives, and thus, **A** must be subtracted before estimating the proxy measure of LD, otherwise an artificially high estimate of LD will be obtained that will result in a poor estimate of additive genetic variance. As a result, programs such as GCTA (Yang *et al.* 2010, 2011) are unlikely to be appropriate for estimating heritability from marker-derived relatedness in naturally occurring study populations containing close relatives.

If a pedigree is not available but the relatedness estimates from the markers are greater than 0.125, then one can exclude closely related individuals and follow the protocol outlined below for unrelated individuals. If this is not possible/desired, then steps (1) through (4) above can be conducted using only the diagonal of **G** to calculate the proxy measure of LD w, and then, this can be used to weight **G** directly as $w\mathbf{G}_n$ for the off-diagonal and $1 + w\mathbf{G}_n$ for the diagonal. This is because estimating the covariance between the estimates of **G1a** and **G2a** in step (4) without first accounting for high levels of relatedness will give an artificially high measure of LD because both **G1a** and **G2a** will contain close relatives biasing the regression. Therefore, it is convenient just to use the diagonals to calculate w in this situation.

If individuals are 'unrelated'. If all individuals are related by <0.125, then the weighting methods used by Yang *et al.* (2010) that are implemented in GCTA (Yang *et al.* 2010, 2011) can be used. This is steps 1 through 4 as outlined above for where pedigree data are available where the off-diagonal of **G** is weighted by multiplying it by w, and the diagonal of **G** is weighted by $1 + w\mathbf{G}_n$.

Estimating heritability from marker-based estimates of relatedness

Under the null hypothesis of a polygenic additive model, all locations in the genome contribute a small amount to the overall variance. Total additive genetic variance can therefore be estimated as the sum of the effects of many markers across the genome in a hierarchical mixed effects model framework:

$$\boldsymbol{y} = \mathbf{X}\beta + \mathbf{Z}v + \boldsymbol{e} \quad (1)$$

where $\boldsymbol{y}$ is a vector of phenotypic measures for a polygenic trait. $\mathbf{X}$ is a design matrix relating trait records to vectors of fixed effects β. $\boldsymbol{e}$ is a vector of random residual deviates, where σ_e^2 is the error variance. $\mathbf{Z}$ is a design matrix relating individuals to additive genetic effects a. $V(v) = \mathbf{A}\sigma_a^2$ where $\mathbf{A}$ is the average relationship matrix and σ_a^2 is the polygenic variance. The matrix of genome-wide relatedness estimates, $\mathbf{G}_a$, calculated by the methods outlined above using all SNP markers typed across the entire genome, can be used in eqn 1 instead of the $\mathbf{A}$ matrix estimated from a breeding design or pedigree. For wing length, additional random effects of year of measurement to control for temporal variation, area of the study site to control for spatial variation, and a permanent environment effect to account for nonadditive differences among individuals were used (Table 1). Fixed effects included age to control for age-specific patterns and sex to control for mean sexual dimorphism.

Partitioning genetic effects using marker-based relatedness

Equation 1 can then be extended to partition additive genetic variance across different regions of the genome. To do this, relatedness can be estimated for different genomic regions separately (e.g. different chromosomes), and these multiple relatedness estimates can all be included as separate parameters within one model:

$$\boldsymbol{y} = \mathbf{X}\beta + \mathbf{Z}_x v_x + \mathbf{Z}_j v_j + \boldsymbol{e} \quad (2)$$

where in this example $\mathbf{Z}_x$ and $\mathbf{Z}_j$ are design matrices relating individuals to additive genetic effects v where $V_x(v_x) = \mathbf{G}_a^x\sigma_a^2$ and $V_j(v_j) = \mathbf{G}_a^j\sigma_a^2$ which estimates genetic variance attributable to two separate regions of the genome x and genome j. This model can be extended to include multiple relationship matrices, thus partitioning additive genetic variance into effects attributable to multiple genomic regions. We used this approach to obtain estimates of the additive genetic variance and heritability of each genomic region.

We can also conduct statistical testing using likelihood ratio testing by estimating a series of models and using log-likelihood ratio test statistics to test among them. Under the null hypothesis of a polygenic model, it is predicted that the total additive genetic variance is distributed across the chromosomes in the genome according to the gene content of each chromosome. To test the alternative hypothesis that a small number of genes account for most of the genetic variance (i.e. an oligogenic model), four model sets can be constructed and contrasts made for every chromosome.

For a given chromosome, the models are given as follows:

1 A mixed model with only one **G** constructed excluding markers on that chromosome.
2 A mixed model with one **G** constructed excluding markers on that chromosome, plus another **G** constructed with only markers on that chromosome.
3 A mixed model with **G** constructed with all markers.
4 Mixed model with **G** constructed with all markers, plus **G** constructed with only markers on that chromosome.

The contrasts are then:

1 Contrast 1: for each autosome, compare the variance explained by model (ii) vs. model (i) to test whether the chromosome explains any variation in the trait. Across all chromosomes, the expectation is that there is a positive linear relationship between the number of genes and the amount of variance explained per chromosome if the trait is polygenic.
2 contrast 2: for each autosomal chromosome and the Z chromosome, compare the variance explained from model (iv) to model (iii) to test whether there is evidence that the variance explained by the chromosome is greater than the amount expected given the size (chromosome length/gene content) of the chromosome.

The contribution of each chromosome to the overall phenotypic variance is tested by comparing the log likelihood of the genome-wide model (model i or iii) with the log likelihood of the genome-wide plus chromosome model (model ii or iv), with a likelihood ratio test (LRT).

Using the simulated data to test bias and accuracy of our approach

Before analysing the Wytham Woods great tit data, we first assessed the accuracy and bias in the analytical approach described below in a number of ways:

1 We estimated heritability within our simulated data using the pedigree and both the un-weighted and weighted estimates of marker-based relatedness from the three different methods. We compared the estimates gained and the model fit to the data.
2 We partitioned additive genetic variance across 22 simulated autosomal linkage groups within our simulated data. For each of the three simulated trait

Table 1 Accuracy of heritability estimates made from estimates of genomic marker-based relatedness across simulated traits of different genetic architecture

	V_A	ΔV_A	Δ logLK	V_A	ΔV_A	Δ logLK	V_A	ΔV_A	Δ logLK
	Polygenic trait			Trait with moderate effect QTL			Trait with large effect QTL		
A	0.374	−0.026 (−0.047, 0.004)	—	0.398	−0.002 (−0.024, 0.022)	—	0.426	0.026 (−0.004, 0.040)	—
$\mathbf{G}_1$	0.311	−0.089 (−0.106, −0.072)	5.283 (1.171, 9.375)	0.328	−0.072 (−0.098, −0.046)	1.886 (−2.226, 5.998)	0.345	−0.055 (−0.065, −0.045)	4.759 (0.179, 9.697)
$\mathbf{G}_2$	0.312	−0.088 (−0.107, −0.073)	5.303 (1.188, 9.418)	0.328	−0.072 (−0.098, −0.046)	1.913 (−2.102, 6.028)	0.346	−0.054 (−0.066, −0.046)	4.757 (0.184, 9.698)
$\mathbf{G}_3$	0.312	−0.084 (−0.105, −0.063)	5.305 (1.189, 9.421)	0.327	−0.073 (−0.099, −0.047)	1.914 (−2.102, 6.030)	0.345	−0.055 (−0.065, −0.045)	4.758 (0.182, 9.698)
$\mathbf{G}_1^A$	0.374	−0.026 (−0.049, 0.006)	9.938 (7.272, 12.704)	0.388	−0.012 (−0.033, 0.009)	7.034 (4.368, 9.700)	0.415	0.015 (−0.003, 0.033)	9.943 (7.310, 12.576)
$\mathbf{G}_2^A$	0.374	−0.026 (−0.049, 0.006)	9.944 (7.279, 12.709)	0.388	−0.013 (−0.033, 0.009)	7.051 (4.386, 9.716)	0.416	0.014 (−0.003, 0.031)	9.945 (7.311, 12.579)
$\mathbf{G}_3^A$	0.374	−0.026 (−0.049, 0.006)	9.949 (7.270, 12.619)	0.389	−0.011 (−0.033, 0.099)	7.059 (4.380, 9.729)	0.416	0.014 (−0.003, 0.031)	9.947 (7.308, 12.586)

QTL, quantitative trait loci.

Data were simulated with the exact same marker structure and number of individuals as for the Wytham Woods great tit data. Three different trait architectures were simulated: (i) a polygenic architecture where the heritability is explained by 200 QTL spaced every 10 centimorgan (cM) throughout the genome; (ii) an architecture where all of the heritability is explained by 22 QTL, with one QTL per linkage group; and (iii) an architecture where all of the heritability is explained by five QTL on separate linkage groups, with the remaining linkage groups containing no QTL. For each type of architecture, we analysed 10 replicate data sets. V_A gives the average estimates of additive genetic variance across replicates; ΔV_A gives the average difference of the estimated genetic variance from the simulated value of 0.4 across replicates, with the values in brackets representing the 95% confidence intervals (95% CI) of the differences across replicates assuming a normal distribution of differences between simulated and estimated values (if the intervals overlap zero, then the estimates overlap with the simulated value); Δ logLK gives the average difference in log-likelihood values as compared to the pedigree estimates, with values in brackets representing the 95% CI of the difference in log likelihood across replicates (positive value gives a better model fit). $\mathbf{G}_1$–$\mathbf{G}_3$ refer to the three different methods of estimating genomic relatedness and $\mathbf{G}_1^A$ to refer $\mathbf{G}_3^A$ to the weighted relationship matrices of each of these methods (see Materials and Methods).

architectures, estimates of the heritability of each of the 22 linkage groups across 10 replicates were made. We used qqplots to compare estimated and simulated effect sizes; we tested for a difference between the simulated values and the point estimates using generalized linear models with a gamma error distribution and log link; we also used a root-mean-square test statistic of the observed from the expected and estimated a *P*-value using Monte Carlo simulation in the R package rms.gof.

3 We tested for an association between chromosome length/gene content and effect size across the three simulated architectures to describe whether a chromosome contributes more to the heritability than expected under a polygenic model.

4 We used the chromosomes that contained no QTL in the large effect QTL architecture simulations to test whether a chi-square distribution with one degree of freedom captured all of the potential for estimation error.

Testing for sex-specific genetic effects across the genome

Finally, for the Wytham Woods empirical data, we then extended eqn 2 to test for sex-specific effects at the chromosome level, using log-likelihood ratio testing to examine whether fitting separate effects for each sex provides a better fit to the data. To do this, we adopted a statistical testing approach where we fit a series of models:

$$\boldsymbol{y} = \mathbf{X}\beta + \mathbf{Z}_{x,s}v_{x,s} + \mathbf{Z}_{j,s}v_{j,s} + \boldsymbol{e} \quad (3)$$

where $\mathbf{Z}_{x,s}$ and $\mathbf{Z}_{j,s}$ are design matrices relating individuals to additive genetic effects v for each sex s. x is a chromosome and j is the rest of the genome, and thus, eqn 3 fits a marker-derived relatedness matrix for markers on chromosome x, and a marker-derived relatedness matrix for markers on the rest of the genome j. Thus, this partitions genetic variance to a chromosome x and to the rest of the genome j.

We first fit a model containing an unstructured covariance matrix for j, which estimates genetic variance for each sex and the across-sex genetic covariance, alongside a single variance component for chromosome x. In this model, fitting a single variance component for chromosome x assumes equal variance across the sexes and a cross-sex correlation of unity.

Second, we then extend this model to estimate separate variance for each sex at chromosome x with a covariance across the sexes that is fixed to unity, alongside the unstructured covariance matrix for j. This was done using a CORGH variance structure in ASReml for chromosome x. This structure estimates a correlation (here fixed to one) and then two variances, thus testing for sex-limited variance at a particular chromosome, while simultaneously estimating sex-limited variance and covariance at the rest of the genome. We compared the fit of this model to the data using likelihood ratio test statistics (termed LRT_1 Table 3) assuming a chi-square distribution and one degree of freedom.

If this second model provided a better fit to the data and then the first model assumes constant variance across the sexes, then this is evidence for sex-specific effects at chromosome x. We did not extend our model comparisons any further because there was no scenario where a chromosome has an effect that was significantly different to zero in both sexes, and thus, we could not test for differences in additive genetic covariance between the sexes across chromosomes.

Results

Estimating heritability from markers

We first estimated heritability within our simulated data using (i) the pedigree, (ii) un-weighted marker-based estimates of relatedness calculated using the three different approaches described in the methods and (iii) weighted marker-based estimates of relatedness calculated using the three different approaches. The simulated heritability was 0.4, and we examined the accuracy of the different approaches using 95% confidence intervals of the estimates across replicates and by the difference in log likelihood across models. The estimated weighting factor was 0.460.

Using a pedigree to estimate heritability generally returned the simulated values across the three different simulated trait architectures (Table 1). Using weighted, marker-based relatedness to estimate heritability also provides accurate estimates and also gave a better likelihood than the un-weighted or pedigree-based methods, suggesting that it provides a significantly better fit to the data (LRT, Table 1). Heritability estimates made using un-weighted genomic relatedness matrices were lower than the heritability values estimated by the pedigree (Table 1) and generally provided a poorer model fit as compared to weighted marker-based relatedness (Table 1).We found no differences between the three different methods of calculating marker-based relatedness in either the estimates returned or the model fit (Table 1). Therefore, any of the three methods of estimating genomic relatedness provide reliable estimates and a better fit to the data, provided that the point estimates of relatedness are scaled by an estimate of the LD between markers and the underlying causal variants. Furthermore, in our simulations, the average r^2

among adjacent markers (a measure of LD) was 0.012, which compares to the Wytham Woods great tit population where the value is 0.015. Therefore, we provide evidence that in populations containing a range of relatives, with relatively low LD among markers and low marker density, three different methods of estimating genomic relatedness are able to accurately return the simulated heritability.

Using markers to partition genetic variance across the genome

Having established that weighting the genomic relationship matrix provided an accurate estimate of the simulated heritability and a better fit to the data, we used this approach to partition additive genetic variance across 22 linkage groups within our simulated data. For each of the three simulated trait architectures, estimates of the heritability of each of the 22 linkage groups across 10 replicates were made (Fig. 1). We compared the estimated and simulated effect sizes in qqplots and tested for differences in distribution between estimated and simulated effects.

We compare estimated and simulated effect sizes using qqplots and found that the 95% confidence intervals of all of the estimates overlapped with their expected value (Fig. 1a). We also present the root square difference of the estimates from the simulated value (root square error: Fig. 1b), showing that the difference between the point estimates and the simulated values are never greater than a heritability of 8%, and that this difference occurred in less than one in a hundred models.

Second, we then tested for a difference between the simulated values and the point estimates using generalized linear models with a gamma error distribution and log link and found no evidence for a difference between the estimated and the simulated effects for the polygenic architecture (mean observed effect: 0.024, 95% CI 0.019: 0.029; mean simulated effect: 0.023, 95% CI: 0.019: 0.027; $P = 0.684$), for the trait with moderate effect QTL (mean observed effect: 0.024, 95% CI 0.019: 0.030; mean simulated effect: 0.019, 95% CI: 0.014: 0.024; $P = 0.180$) or for a trait with large QTL effects (mean observed effect: 0.026, 95% CI 0.017: 0.035; mean simulated effect: 0.019, 95% CI: 0.010: 0.028; $P = 0.684$). We also used a root-mean-square test statistic of the observed from the expected and estimated a *P*-value using Monte Carlo simulation and gained a *P*-value of 1 for each architecture, indicating no difference between the observed and simulated values (root-mean-square deviation for polygenic architecture: 0.0012; for the trait with moderate effect QTL: 0.0014; for the trait with large effect QTL: 0.0013). Therefore, we have shown that marker-based estimates of relatedness can be used to partition additive genetic variation across regions of the genome with accuracy, and no systematic bias, in populations containing a range of relatives, with relatively low LD among markers and low marker density.

Third, we also provide evidence that log-LRT statistics provide an accurate method of assessing a chromosome's contribution towards the additive genetic variance. We used the chromosomes that contained no QTL in the large effect QTL architecture simulations to test whether a chi-square distribution with one degree of freedom captured all of the potential for estimation error (Fig. 1c). We found that LRT of the effects of chromosomes containing no QTL was never significant (Fig. 1c), supporting the use of LRT to assess the significance of a chromosomes contribution to the overall heritability of a trait.

Finally, we tested for an association between chromosome length/gene content and effect size to describe whether a chromosome contributes more to the heritability than expected under a polygenic model (Fig. 2). As expected, the polygenic trait had average effect sizes across replicates that scaled with the chromosome length (Fig. 2), and in a linear mixed effects model testing for an association with chromosome length, the slope of the regression was significant ($P < 0.001$) and equivalent to the simulated effect size of 1.496 e−04 (95% CI: 9.886 e−05, 2.004 e−04; Fig. 2). We found no evidence for an interaction across replicates (5.773 e−06; 95% CI: −1.062 e−05, 2.558 e−05; $P = 0.532$), demonstrating that this association was consistent across replicates. We found no association between chromosome length and effect size for either the trait of moderate effect QTL (1.233 e−05; 95% CI: −7.078 e−05, 9.170 e−05; $P = 0.798$) or large effect QTL (−8.740 e−05; 95% CI: −3.590 e−04, 1.718 e−04; $P = 0.528$) across any replicates (no association was expected for these architectures), demonstrating that a significant regression of chromosome effect size on chromosome length/gene content provides robust evidence of whether effect sizes are indicative of a polygenic basis of inheritance.

The genetic architecture of wing length in a wild bird population

For wing length, when marker-derived relatedness estimates were weighted (by 0.407) to correct for incomplete LD with the underlying genome, all methods of estimating marker-based relatedness gave similar heritability values to the pedigree value (Table 2). In contrast, un-weighted estimates returned lower heritability values, a poorer model fit, and an overestimation of the permanent environment effect (Table 2).

Second, we then estimated genetic variance associated with individual chromosomes and found that four of the 22 estimates of the chromosomal heritability were

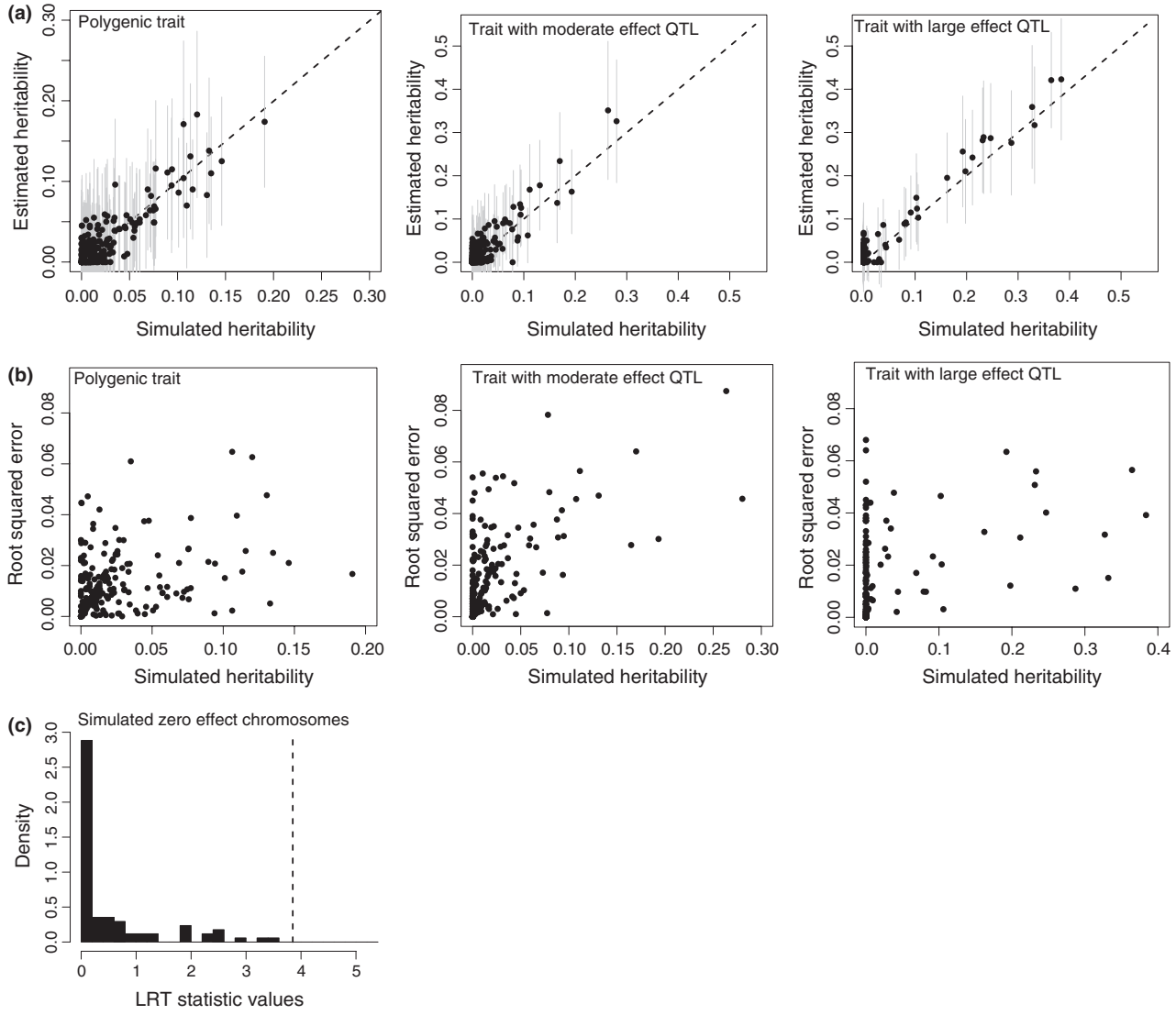


Fig. 1 Data were simulated with the exact same marker structure and number of individuals as for the Wytham Woods great tit data. Three different trait architectures were simulated: (i) a polygenic architecture where the heritability is explained by 200 quantitative trait loci (QTL) spaced every 10 centimorgan (cM) throughout the genome; (ii) an architecture where all of the heritability is explained 22 QTL, with one QTL per linkage group (trait with moderate effect QTL); and (iii) an architecture where all of the heritability is explained by five QTL each on separate linkage groups, with the remaining linkage groups containing no QTL (trait with large effect QTL). For each type of architecture, we analysed 10 replicate data sets. (a) For all replicates of each simulated architecture, a qqplot is shown of the estimated heritability of each linkage group plotted against the simulated value, with grey lines giving the 95% confidence intervals (95% CI) of the estimates gained from the models (note that the scale is different across the architectures). (b) We examined the accuracy of the point estimates across different simulated values by estimating the root squared difference of each estimate from the simulated value (root squared error). (c) Histogram of the log-likelihood ratio test statistics for linkage groups (n = 170) that were simulated to have no effects for trait architecture (iii), with dashed line indicating a significant test statistic at $P < 0.05$.

significantly different from zero, on chromosomes 1A, 5, 7 and 11 (Fig. 3a; Table 3: LRT_1). We then estimated variance on the Z chromosome separately from the autosomes and tested different models of dosage compensation and found that the Z chromosome effects were not significantly different to zero, with the model of best fit reflecting no dosage compensation for the Z chromosome for this trait (Table 3). The sum of the chromosome-specific heritability was 0.568 (Table 3), which was almost identical to the estimates made from the pedigree (Table 2).

Finally, we then tested whether the contribution of different genomic regions to adult wing length was associated with the expected gene content of the chro-

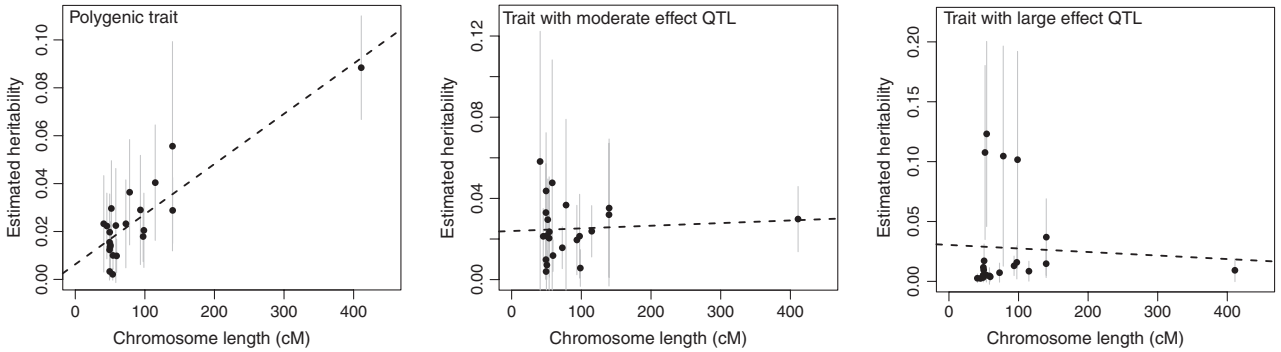


Fig. 2 For each simulated architecture, the average heritability of each linkage group across replicates is plotted against the chromosome length in centimorgan (cM), with grey lines giving the 95% CI of the point estimates across replicates. Note that in the simulated data, the recombination rate is the same across chromosomes, and thus, the linkage map distances are directly comparable with physical distance and with gene content. The regression is also not dependent upon any single data point.

Table 2 Estimates of additive genetic variance for wing length in the Wytham Woods great tit population using relationship estimates derived from a pedigree and from markers using three different methods

	V_{PE}	V_A	V_{YR}	V_S	V_R	logLK	h^2
A	0.094 (0.071)	1.243 (0.103)	0.045 (0.024)	0.059 (0.025)	0.753 (0.022)	−2934.26	0.567 (0.038)
$\mathbf{G}_1$	0.552 (0.061)	0.787 (0.083)	0.045 (0.024)	0.051 (0.023)	0.758 (0.022)	−2949.44	0.359 (0.032)
$\mathbf{G}_2$	0.582 (0.062)	0.757 (0.083)	0.045 (0.024)	0.052 (0.023)	0.758 (0.022)	−2953.82	0.345 (0.032)
$\mathbf{G}_3$	0.578 (0.062)	0.761 (0.082)	0.045 (0.024)	0.052 (0.023)	0.758 (0.022)	−2953.75	0.347 (0.032)
$\mathbf{G}_1^A$	0.141 (0.067)	1.193 (0.099)	0.045 (0.024)	0.054 (0.024)	0.754 (0.022)	−2920.68	0.545 (0.036)
$\mathbf{G}_2^A$	0.158 (0.067)	1.178 (0.099)	0.045 (0.024)	0.054 (0.024)	0.754 (0.022)	−2923.98	0.538 (0.036)
$\mathbf{G}_3^A$	0.153 (0.067)	1.183 (0.099)	0.045 (0.024)	0.054 (0.024)	0.754 (0.022)	−2923.37	0.546 (0.036)

A represents the estimates made using a relationship matrix from a pedigree. $\mathbf{G}_1$–$\mathbf{G}_3$ represent the estimates made using un-weighted marker-derived estimates of relatedness from the three approaches outlined in Materials and Methods. $\mathbf{G}_1^A$–$\mathbf{G}_3^A$ represent estimates made using weighted marker-derived estimates of relatedness from the three approaches. V_{PE} gives the variance estimated for the permanent environment effect; V_A gives the additive genetic variance; V_{YR} gives the yearly temporal variance in measurement; V_S gives the variance estimate for the spatial measurement effects; V_R gives the residual variance; logLK is the log likelihood of the model; and h^2 gives the heritability.

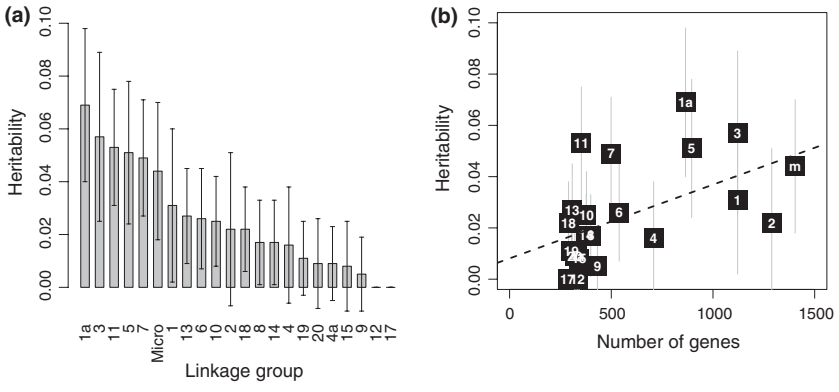


Fig. 3 (a) Heritability of wing length in a wild bird population attributed to each linkage group is (b) correlated with the number of genes predicted through homology with the zebra finch genome. The linkage group is displayed on the *x*-axis of (a) and as a data point in (b) with micro and m both representing the microchromosomes. Error bars give the standard error of the estimates.

mosome (Table 3) thus further testing whether the effects fit that expected from a polygenic model. In general, the greater the expected gene content of a chromosome, the higher the heritability estimate for that chromosome ($F_{1,20} = 8.26$, $P = 0.009$), and the slope of the regression line of 2.975×10^{-5} heritability per gene is consistent with a genome-wide heritability value of 0.383 if the heritability scales with the number of expected genes identified from the zebra finch genome (Fig. 3b). In contrast, a multiple regression including the number of markers per chromosome and chromosome size alongside the estimated gene content found no evidence for any effects of marker density ($F_{1,20} = 0.24$, $P = 0.618$) or chromosome size

Table 3 Estimates of additive genetic variance of wing length for each chromosome in the Wytham Woods great tit population

LG	Markers	Genes	Length (cM)	V_A	PV_A	h^2	LRT_1	LRT_2
1	579	1254	139.88	0.068 (0.065)	0.057 (0.054)	0.031 (0.029)	0.02	2.56
1A	415	972	93.58	**0.153 (0.064)**	**0.129 (0.053)**	**0.069 (0.029)**	**5.94**	0.56
2	700	1450	139.69	0.048 (0.065)	0.041 (0.055)	0.022 (0.029)	0.02	1.04
3	596	1290	114.97	0.125 (0.069)	0.106 (0.058)	0.057 (0.032)	1.84	0.16
4	356	811	97.56	0.035 (0.048)	0.029 (0.041)	0.016 (0.022)	1.44	0.06
4A	103	39	59.38	0.019 (0.030)	0.016 (0.025)	0.009 (0.014)	0.64	0.02
5	346	998	98.57	**0.113 (0.051)**	**0.095 (0.045)**	**0.051 (0.025)**	**4.18**	0.84
6	177	596	78.02	0.058 (0.041)	0.049 (0.035)	0.026 (0.019)	3.58	1.42
7	176	562	72.64	**0.107 (0.049)**	**0.090 (0.041)**	**0.049 (0.022)**	**8.58**	3.38
8	134	575	53.83	0.038 (0.036)	0.032 (0.030)	0.017 (0.016)	0.64	0.00
9	130	497	54.20	0.010 (0.031)	0.008 (0.026)	0.005 (0.014)	0.50	0.00
10	148	444	50.47	0.055 (0.039)	0.046 (0.032)	0.025 (0.017)	1.20	0.08
11	135	397	58.23	**0.116 (0.048)**	**0.098 (0.040)**	**0.053 (0.022)**	**9.58**	**5.30**
12	152	369	51.92	0.000	0.000	0.000	0.00	2.26
13	117	379	40.95	0.061 (0.039)	0.051 (0.033)	0.027 (0.018)	1.82	0.40
14	126	426	49.20	0.038 (0.035)	0.032 (0.029)	0.017 (0.016)	1.60	0.20
15	173	381	49.15	0.018 (0.037)	0.016 (0.031)	0.008 (0.017)	0.06	0.44
17	96	336	45.36	0.000	0.000	0.000	0.00	0.62
18	93	334	49.91	0.048 (0.035)	0.040 (0.030)	0.022 (0.016)	0.86	0.04
19	97	348	49.43	0.023 (0.031)	0.020 (0.026)	0.011 (0.014)	0.72	0.02
20	155	356	49.44	0.022 (0.038)	0.018 (0.032)	0.009 (0.017)	0.48	0.02
Micro	308	1808	411.04	0.097 (0.057)	0.082 (0.048)	0.044 (0.026)	2.34	0.14
Z_{ND}	279	826	51.21	0.032 (0.033)	0.018 (0.018)		—	0.80
Z_{FD}	279	826	51.21	0.016 (0.018)	0.009 (0.008)		—	0.62
Z_{EV}	279	826	51.21	0.008 (0.022)	0.004 (0.005)		—	0.40

LG gives the linkage group, with the number of markers, number of expected genes and length in cM. Different models of dosage compensation for the Z chromosome were tested with Z_{ND} assuming no dosage compensation; Z_{FD} assuming full dosage compensation; and Z_{EV} assuming equal variance. V_A gives the additive genetic variance; PV_A gives the proportion of genome-wide additive genetic variance; and h^2 gives the heritability. Estimates are made from a model where the relationship matrices of each chromosomes are fit simultaneously within a single model, and the sum of the heritability of each chromosome is 0.568. LRT (log-likelihood test) statistics are gained from models containing two relationship matrices: one for the chromosome and another for the rest of the genome. LRT_1 is a log-likelihood ratio test statistics for whether an effect size is significantly greater than zero, and LRT_2 tests whether the contribution of a chromosome is greater or less than expected from a polygenic model (contrast 1 and contrast 2 in Materials and Methods). Significance at $P < 0.05$ is given in bold. Linkage groups with effects smaller than 1.0×10^{-4} are shown as zero because their effects and the error are bound to zero within the model.

($F_{1,20} = 0.26$, $P = 0.616$). Therefore, as the slope of the regression line estimated in Fig. 3b, returns an estimate that is lower than directly estimated heritability, it suggests that there may be additional gene effects that are not captured by simply the number of predicted genes. However, the hypothesis that all chromosomes contribute towards variation in proportion to their gene content, in line with an additive mode of gene action cannot be rejected from these data as it is clear that most chromosomes contribute modestly to the overall additive genetic variance.

Testing for sex-specific genetic effects across the genome

As plumage characteristics and body size are sexually dimorphic in this population (wing length: $F_{1,4575} = 284.58$, $P < 0.001$; male mean: 76.225, 1.588 SD; female mean: 73.212, 1.519), we examined whether there was any evidence for sex-specific genetic effects across different regions of the genome, by testing for the effects of sex on the estimated genetic variance associated with individual chromosomes (model 3, Table 4).

Overall, the cross-sex genetic correlation was close to one ($r_G = 0.885 \pm 0.217$), and testing for sex-specific effects across chromosomes revealed that across most of the genome, there was no evidence for sex differences in the additive genetic variance attributed to chromosomes (Table 4). We found evidence for sex-specific effects at chromosome 8, which contributed 12% to the additive genetic variance in males, but only 0.1% in females (Table 4). As suggested by the increased standard error of the estimated effects in Table 3, as compared to the standard errors of Table 2, we appear to be restricted to detecting a single region of large sex-specific effect, because of our limited sample size for

Table 4 Estimates of additive genetic variance of wing length for each chromosome in males and females in the Wytham Woods great tit population

	Males			Females			
LG	V_A	PV_A	h^2	V_A	PV_A	h^2	LRT_1
1	0.000	0.000	0.000	0.090 (0.117)	0.073 (0.095)	0.043 (0.050)	0.74
1a	0.253 (0.124)	0.189 (0.092)	0.114 (0.055)	0.166 (0.108)	0.138 (0.088)	0.079 (0.051)	1.26
2	0.048 (0.138)	0.037 (0.105)	0.022 (0.062)	0.127 (0.105)	0.103 (0.097)	0.059 (0.059)	3.18
3	0.094 (0.122)	0.069 (0.089)	0.042 (0.054)	0.177 (0.119)	0.144 (0.096)	0.084 (0.056)	1.98
4	0.125 (0.098)	0.092 (0.072)	0.056 (0.043)	0.038 (0.078)	0.031 (0.063)	0.018 (0.036)	3.44
4a	0.000	0.000	0.000	0.059 (0.064)	0.047 (0.051)	0.028 (0.031)	0.26
5	0.050 (0.096)	0.036 (0.071)	0.022 (0.043)	0.000	0.000	0.000	2.46
6	0.093 (0.076)	0.069 (0.056)	0.041 (0.034)	0.038 (0.064)	0.031 (0.053)	0.018 (0.030)	1.60
7	0.120 (0.092)	0.089 (0.057)	0.054 (0.041)	0.186 (0.084)	0.153 (0.066)	0.089 (0.034)	0.14
8	0.164 (0.072)	0.115 (0.059)	0.076 (0.036)	0.001 (0.053)	0.001 (0.043)	0.001 (0.025)	**6.20**
9	0.024 (0.065)	0.018 (0.047)	0.011 (0.029)	0.003 (0.053)	0.002 (0.043)	0.001 (0.025)	1.24
10	0.010 (0.062)	0.007 (0.046)	0.004 (0.028)	0.052 (0.063)	0.042 (0.051)	0.025 (0.030)	0.30
11	0.086 (0.077)	0.071 (0.057)	0.034 (0.034)	0.080 (0.066)	0.067 (0.055)	0.038 (0.031)	0.22
12	0.000	0.000	0.000	0.080 (0.076)	0.064 (0.061)	0.038 (0.036)	0.66
13	0.000	0.000	0.000	0.069 (0.068)	0.056 (0.055)	0.033 (0.033)	0.04
14	0.000	0.000	0.000	0.053 (0.064)	0.043 (0.052)	0.025 (0.030)	0.26
15	0.088 (0.088)	0.064 (0.063)	0.039 (0.039)	0.000	0.000	0.000	3.40
17	0.000	0.000	0.000	0.002 (0.051)	0.001 (0.042)	0.001 (0.025)	0.32
18	0.000	0.000	0.000	0.042 (0.057)	0.034 (0.046)	0.020 (0.027)	0.10
19	0.000	0.000	0.000	0.060 (0.062)	0.049 (0.049)	0.029 (0.029)	0.36
20	0.151 (0.094)	0.109 (0.067)	0.067 (0.041)	0.000	0.000	0.000	3.54
Micro	0.036 (0.108)	0.026 (0.077)	0.016 (0.048)	0.000	0.000	0.000	2.24
Z_{ND}	0.000	0.000	0.000	0.027 (0.042)	0.022 (0.034)	0.013 (0.020)	0.40

LG gives the linkage group. V_A gives the additive genetic variance; PV_A gives the proportion of genome-wide additive genetic variance; and h^2 gives the heritability. Z_{ND} is a model assuming no dosage compensation for the Z chromosome. LRT_1 is a log-likelihood ratio test statistics for whether the genetic variance of a chromosome differs between the sexes as compared to a model that estimates a single effect for both sexes at that chromosome. Bold indicates a significant difference in variance across the sexes at $P < 0.05$. In all models, we allowed for sex-specific variance across the rest of the genome before testing sex-limited effects as a specific chromosome.

each sex (Table 4). Therefore, there may be additional genomic regions where the effect sizes differ more subtly, but we do not have the power in this study to differentiate this. However, it is clear that the combined-sex estimates of chromosome 8 in Table 3 do not reflect the effect sizes estimated for males in Table 4, and thus, by testing for sex-specific effects, we have potentially identified a sex-specific region that would otherwise have been deemed to have little effect in both sexes.

Discussion

In this study, we adapt, test and apply multimarker methods that enable genetic variation for complex phenotypic traits, measured on individuals from ecological study populations, to be partitioned across regions of the genome. Our application of this approach demonstrates for the first time that genetic variation for a complex quantitative trait within a wild population reflects contributions from multiple regions of the genome, whose effects scale with their expected gene content. We also provide one of the first tests of the distribution of sex-limited genetic effects across chromosomes. Our results suggest that most of the genomic regions that influence wing length have the same effects in both sexes. However, some of the genetic variance in males can be attributed to regions that have no effects in females, which could facilitate the sexual dimorphism observed for this trait.

Studying individuals within their natural habitat provides the opportunity to examine patterns of genetic variance and covariance of, and the strength and direction of selection acting upon, complex phenotypic traits in the wild. However, this often requires many decades of effort before sample sizes become large enough to support the analyses. The multimarker approach presented here estimates the actual genome-wide relatedness among individuals using genetic markers and is unbiased of the sampling error. Importantly, this approach can be adopted irrespective of whether a pedigree is available or not, enabling additive genetic variance to be calculated without a pedigree or breeding design, and within this study, we describe how this can be done. Here, we show

that marker-based relatedness estimates are capable of returning accurate additive genetic variance estimates, provided that they are weighted by a proxy measure of LD among markers and the genome. The way in which the weighting is conducted depends upon the relatedness structure of the data and whether a pedigree is available, and we outline different approaches that can be taken within the methods and show that the weighting factor reduces the estimation error of the pairwise relatedness estimates (see Box 1). We also demonstrate that this approach is capable of revealing novel insights into the underlying genetic architecture of phenotype in ecological study populations, as general patterns of genetic architecture can be tested for, and genetic variance can be accurately estimated and partitioned across the genome, even under relatively modest marker density and sample size compared with crop, livestock and human medical genetic studies.

Given the significance of the regression of chromosome effect size and expected gene content, we cannot reject the hypothesis for a genetic architecture with a polygenic additive mode of gene action for wing length within this population. There is the potential for moderate effects to exist, given the fact that the slope of the chromosome estimates on their expected gene content returns a heritability lower than the directly estimated value, that chromosome 11 may contribute more than expected under a polygenic model and that some chromosomes have not been mapped. However, we do not find a pattern of moderate or large QTL effects that we see in the simulated data, and thus, we simply conclude that the majority of the heritability of wing length within this natural population is likely to come from many genes spread throughout the genome.

It has recently been argued that rare alleles of large effect may explain much of the genetic variance for traits under selection (Orr 2005; Eyre-Walker 2010), which is supported by linkage mapping QTL studies conducted to date in wild pedigreed populations (e.g. the wing length studies of Tarka *et al.* 2010; Schielzeth *et al.* 2012a,b). Our results are in contrast to these studies, as we find most of the heritability of consistent with a polygenic basis, where all chromosomes contribute towards variation in proportion to their gene content. Our findings are consistent with recent studies of quantitative traits in both human medical genetic and livestock populations, which despite the power to detect rare variants, typically find little evidence for QTL of large effect (Mackay 2001; Hill *et al.* 2008; Allen *et al.* 2010; Yang *et al.* 2010). While it is possible that different architectures exist across different populations, differences between our findings and previous studies may simply reflect the limited sample size and power of evolutionary biology studies to conduct QTL mapping, which result in inflated estimates of effect sizes (Slate 2013). Therefore, it will now be important to repeat these studies with the approach we present here.

Consistent directional and stabilizing selection acting upon quantitative traits is common in wild populations, and thus, examining the underlying genetic architecture of a range of sexually selected, morphological and life history traits, for many different species and habitats, is likely to contribute significantly towards our understanding of the architecture of traits under selection. The next step in this study would be to fine map the trait in question by partitioning variance across regions of the genome smaller than a chromosome (i.e. 50SNP blocks), which will yield a greater understanding of whether the chromosome effect stems from an isolated region or whether they are distributed across the chromosome.

One way in which we can better understand the architecture of traits under selection is to examine whether the effects of genomic regions depend upon the context in which they are expressed. Our findings reveal that through partitioning genetic variance across chromosomes and testing for sex-specific effects, regions of the genome can be detected, which may contain causal variants whose effects are context specific. Estimating genetic covariance using relatedness estimated either from a pedigree or from a genome-wide marker sharing revealed a highly positive cross-sex genetic correlation, and this is supported by no evidence for sex differences in the additive genetic variance attributable to the majority of chromosomes. Many studies have interpreted highly positive cross-sex correlations in the between-family genetic variance as a constraint to an independent evolutionary trajectory for each sex (Lande 1980; Poissant *et al.* 2009). Our results suggest that certain genomic regions may have effects that act independently in each sex, despite positively correlated genetic effects across the rest of the genome. Theoretically, differential selection acting upon these regions in males and females could facilitate sexual dimorphism for this trait. Therefore, these findings, if replicated across other wild populations, will have implications for our understanding of sexual selection, population dynamics and speciation (Lande 1980; Butler *et al.* 2007; Poissant *et al.* 2009), because they suggest that failing to model sex-specific architecture could hamper our ability to detect regions underlying complex traits.

In summary, this is the first time that genetic variance for a complex trait in a naturally occurring population has been partitioned across the genome and localized sex-specific effects assessed. It remains to be seen both in this and across other populations, whether similar effects can be found or whether QTL of large effects are segregating as expected under theories of adaptation (Otto & Jones 2000; Slate *et al.* 2010). Our application

was to a single trait; however, the method is entirely general and could be applied to any quantitative trait to examine sex-, age- and environment-specific effects, as well as to multiple traits to partition genetic covariance across the genome in any ecological study population.

Acknowledgements

We would like to thank three reviewers and the editor for their comments, which improved the manuscript. Genotyping was performed by Christine Blancher and her team at The Wellcome Trust Centre for Human Genetics at Oxford. MRR is supported by a Natural Environment Council (NERC) fellowship (Linking classical and molecular genetics—NE/G013535/1) and JS is supported by a European Research Council (ERC) grant (Avian EGG—grant number 202487).

References

Allen HL, Estrada K, Lettre G *et al.* (2010) Hundreds of variants clustered in genomic loci and biological pathways affect human height. *Nature*, **467**, 832–838.

Beavis WD (1994) *The power and deceit of QTL experiments: lessons from comparative QTL studies*. pp. 250–266. Proceedings of the forty-ninth annual corn and sorghum industry research conference. American Seed Trade Association, Washington, DC.

van Bers NE, Santure AW, van Oers K *et al.* (2012) The design and cross-population application of a genome-wide SNP chip for the great tit *Parus major*. *Molecular Ecology Resources*, **12**, 753–770.

Both C, Bouwhuis S, Lessels CM, Visser ME (2006) Climate change and population declines in a long-distance migratory bird. *Nature*, **441**, 81–83.

Brito FV, Neto JB, Sargolzaei M, Cobuci JA, Schenkel FS (2011) Accuracy of genomic selection in simulated populations mimicking the extent of linkage disequilibrium in beef cattle. *BMC Genetics*, **12**, 80.

Buckler ES, Holland JB, Bradbury PJ *et al.* (2009) The genetic architecture of maize flowering time. *Science*, **325**, 714–718.

Butler MA, Sawyer SA, Losos JB (2007) Sexual dimorphism and adaptive radiation in *Anolis* lizards. *Nature*, **447**, 202–205.

Charmantier A, McCleery RH, Cole LR, Perrins C, Kruuk LEB, Sheldon BC (2008) Adaptive phenotypic plasticity in response to climate change in a wild bird population. *Science*, **320**, 800–803.

Dawideit BA, Phillimore AB, Laube I, Leisler B, Böhning-Gaese K (2009) Ecomorphological predictors of natal dispersal distances in birds. *Journal of Animal Ecology*, **78**, 388–395.

Dekkers JCM (2012) Application of genomics tools to animal breeding. *Current Genomics*, **13**, 207–212.

Eyre-Walker A (2010) Genetic architecture of a complex trait and its implications for fitness and genome-wide association studies. *Proceedings of the National Academy of Sciences of the United States of America*, **107**, 1752–1756.

Falconer DS, Mackay TFC (1996) *Introduction to Quantitative Genetics*. Longman Group Ltd, Essex, UK.

Gienapp P, Merilä J (2010) Genetic and environmental effects on a condition-dependent trait: feather growth in Siberian jays. *Journal of Evolutionary Biology*, **23**, 715–723.

Gilmour AR, Gogel BJ, Cullis BR, Thompson R (2006) *ASReml User Guide Release 2.0*. VSN International Ltd, Hemel Hempstead, UK.

Goddard ME, Hayes BJ (2009) Mapping genes for complex traits in domestic animals and their use in breeding programs. *Nature Reviews Genetics*, **10**, 381–391.

Goddard ME, Hayes BJ, Meuwissen THE (2011) Using the genomic relationship matrix to predict accuracy of genomic selection. *Journal of Animal Breeding and Genetics*, **128**, 409–421.

Göring HHH, Terwilliger JC, Blangero J (2001) Large upward bias in estimation of locus-specific effects from genome-wide scans. *American Journal of Human Genetics*, **69**, 1357–1369.

Habier D, Fernando RL, Dekkers JC (2009) Genomic selection using low-density marker panels. *Genetics*, **182**, 343–353.

Hayes BJ, Pryce J, Chamberlain AJ, Bowman PJ, Goddard ME (2010) Genetic architecture of complex traits and accuracy of genomic prediction: coat color, milk-fat percentage, and type in Holstein cattle as contrasting model traits. *PLoS Genetics*, **6**, e1001139.

Hill WG (2012) Quantitative genetics in the genomics era. *Current Genomics*, **13**, 196–206.

Hill WG, Robertson A (1968) Linkage disequilibrium in finite populations. *Theoretical and Applied Genetics*, **38**, 226–231.

Hill WG, Weir BS (2011) Variation in actual relationship as a consequence of Mendelian sampling and linkage. *Genetical Research*, **93**, 47–64.

Hill WG, Goddard ME, Visscher PM (2008) Data and theory point to mainly additive genetic variance for complex traits. *PLoS Genetics*, **4**, e1000008.

Jensen H, Sæther BE, Ringsby TH *et al.* (2003) Sexual variation in heritability and genetic correlations of morphological traits in house sparrow (*Passer domesticus*). *Journal of Evolutionary Biology*, **16**, 1296–1307.

Johnson SE, Beraldi D, McRae AF, Pemberton JM, Slate J (2010) Horn type and horn length genes map to the same chromosomal region in Soay sheep. *Heredity*, **104**, 196–205.

Lande R (1980) Sexual dimorphism, sexual selection and adaptation in polygenic characters. *Evolution*, **34**, 292–305.

Lockwood R, Swaddle JP, Rayner JMV (1998) Avian wingtip shape reconsidered: wingtip shape indices and morphological adaptations to migration. *Journal of Avian Biology*, **29**, 273–292.

Lynch M, Walsh JB (1998) *Genetics and Analysis of Quantitative Traits*. 980 pp. Sinauer Assocs., Inc., Sunderland, MA.

Mackay TFC (2001) The genetic architecture of quantitative traits. *Annual Review of Genetics*, **35**, 202–339.

Mackay TFC, Stone EA, Ayroles JF (2009) The genetics of quantitative traits: challenges and prospects. *Nature Reviews Genetics*, **19**, 565–577.

Manolio TA, Collins FS, Cox NJ *et al.* (2009) Finding the missing heritability of complex diseases. *Nature*, **461**, 747–753.

McCleery RH, Pettifor RA, Armbruster P *et al.* (2004) Components of variance underlying fitness in a natural population of the great tit *Parus major*. *American Naturalist*, **164**, E62–E72.

Ødergård J, Meuwissen THE (2012) Estimation of heritability from limited family data using genome-wide identity-by-descent sharing. *Génétique, Sélection, Évolution*, **44**, 16–18.

Orr HA (2005) The genetic theory of adaptation: a brief history. *Nature Reviews Genetics*, **6**, 119–127.

Otto SP, Jones CD (2000) Detecting the undetected: estimating the total number of loci underlying a quantitative trait. *Genetics*, **156**, 2093–2107.

Ozgul A, Tuljapurkar S, Benton TG, Pemberton JM, Clutton-Brock TH, Coulson T (2009) The dynamics of phenotypic change and the shrinking sheep of St. Kilda. *Science*, **325**, 464–467.

Poissant J, Wilson AJ, Coltman DW (2009) Sex-specific genetic variance and the evolution of sexual dimorphism: a systematic review of cross-sex genetic correlations. *Evolution*, **64**, 97–107.

Powell JE, Visscher PM, Goddard ME (2010) Reconciling the analysis of IBD and IBS in complex trait studies. *Nature Reviews Genetics*, **11**, 800–805.

Robinson MR, Pilkington JG, Clutton-Brock TH, Pemberton JM, Kruuk LEB (2008) Environmental heterogeneity generates fluctuating selection on a secondary sexual trait. *Current Biology*, **18**, 751–757.

Rolshausen G, Segelbacher G, Hobson KA, Schaefer HM (2009) Contemporary evolution of reproductive isolation and phenotypic divergence in sympatry along a migratory divide. *Current Biology*, **19**, 2097–2101.

Santure AW, Gratten J, Mossman JA, Sheldon BC, Slate J (2011) Characterization of the transcriptome of a wild great tit *Parus major* population by next generation sequencing. *BMC Genomics*, **12**, 753–770.

Sargolzaei M, Schenkel FS (2009) QMSim: a large-scale genome simulator for livestock. *Bioinformatics*, **25**, 680–681.

Schielzeth H, Forstmeir W, Kempenaers B, Ellegren H (2012a) QTL linkage mapping of wing length in zebra finch using genome-wide single nucleotide polymorphisms markers. *Molecular Ecology*, **21**, 329–339.

Schielzeth H, Kempenaers B, Ellegren H, Forstmeir W (2012b) QTL linkage mapping of zebra finch beak color shows an oligogenic control of a sexually selected trait. *Evolution*, **66**, 18–30.

Slate J (2013) From beavis to beak colour: a simulation study to examine how much QTL mapping can reveal about the genetic architecture of quantitative traits. *Evolution*, **67**, 1251–1262.

Slate J, Santure AW, Feulner PGD *et al.* (2010) Genome mapping in intensively studied wild vertebrate populations. *Trends in Genetics*, **26**, 275–284.

Tarka M, Akesson M, Beraldi D *et al.* (2010) A strong quantitative trait locus for wing length on chromosome 2 in a wild population of reed warblers. *Proceedings of the Royal Society of London, Series B: Biological Sciences*, **1692**, 2361–2369.

The International Schizophrenia Consortium (2009) Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature*, **460**, 748–752.

Van Raden PM (2007) Efficient estimation of breeding value from dense genomic data. *Journal of Dairy Science*, **90**(Suppl), 1374–1375.

Van Raden PM, Van Tassel CP, Wiggans GR *et al.* (2009) Invited review: reliability of genomic predictions for North American Holstein bulls. *Journal of Dairy Science*, **92**, 16–24.

Visscher PM (2009) Whole genome approaches to quantitative genetics. *Genetica*, **136**, 351–358.

Visscher PM, Medland SE, Ferreira MAR *et al.* (2006) Assumption-free estimation of heritability from genome-wide identity-by-descent sharing between full siblings. *PLoS Genetics*, **2**, 316–325.

Visscher PM, MacGregor S, Beben B *et al.* (2007) Genome partitioning of genetic variation for height from 11,214 sibling pairs. *American Journal of Human Genetics*, **81**, 1104–1110.

Yang J, Benyamin B, McEvoy BP *et al.* (2010) Common SNPs explain a large proportion of the heritability for human height. *Nature Genetics*, **42**, 565–567.

Yang J, Manolio TA, Pasquale LR *et al.* (2011) Genome partitioning of genetic variation for complex traits using common SNPs. *Nature Genetics*, **43**, 519–544.

B.C.S. managed the population study and designed the collection of phenotypic data. J.S. and A.W.S. conceived and designed the collection of genomic data. M.R.R., B.C.S., J.S., A.W.S. and I.D.C. all contributed data. M.R.R. conceived the experiment alongside J.S. M.R.R. designed and performed the experiments, analysed the data and wrote the article. All authors commented on versions of the manuscript.

Data accessibility

Single nucleotide polymorphism marker names, positions and genotypes, pedigree information, phenotype files including fixed and random effects and simulated data: Dryad doi:10.5061/dryad.t12v8.