Review

Genome mapping in intensively studied wild vertebrate populations

Jon Slate, Anna W. Santure, Philine G.D. Feulner, Emily A. Brown, Alex D. Ball, Susan E. Johnston and Jake Gratten

Department of Animal & Plant Sciences, University of Sheffield, Sheffield, S10 2TN, UK

Over the past decade, long-term studies of vertebrate populations have been the focus of many quantitative genetic studies. As a result, we have a clearer understanding of why some fitness-related traits are heritable and under selection, but are apparently not evolving. An exciting extension of this work is to identify the genes underlying phenotypic variation in natural populations. The advent of next-generation sequencing and high-throughput single nucleotide polymorphism (SNP) genotyping platforms means that mapping studies are set to become widespread in those wild populations for whom appropriate phenotypic data and DNA samples are available. Here, we highlight the progress made in this area and define evolutionary genetic questions that have become tractable with the arrival of these new genomics technologies.

Evolutionary genetics in pedigreed wild populations (PWPs)

Globally, there is a modest, yet important number of long-term studies of wild vertebrate populations in which individuals have been intensively monitored in the field, and detailed records of breeding, lifespan, reproductive success and morphological characters have been collected (Figure 1). These datasets typically span many generations and contain records on thousands, or even tens of thousands, of individuals. For the purposes of this review, we refer to this type of population as a PWP. Over the past decade, PWPs have been the focus of a series of quantitative genetic studies that have aimed to explain the evolutionary dynamics of fitness-related traits in natural populations [1–3]. Most of this research uses the quantitative geneticist's 'animal model' because this can accommodate the complex and uncontrolled pedigree structures typically encountered in the wild [4], and the researchers sometimes describe themselves as 'wild animal modellers' (http://www.wildanimalmodels.org). A key aim of researchers working in this field is to attempt to describe the evolutionary dynamics of heritable traits under selection, and in particular to explain why anticipated evolutionary responses expected under the single trait 'breeders' equation' are not always observed in nature [2,5].

Although much progress in understanding microevolution has been made with quantitative genetics, a limitation of such an approach is that it cannot identify the individual genes responsible for trait variation. Therefore, there is

Glossary

Animal model: A form of statistical mixed effects model in which pedigree information is used to separate phenotypic variation into genetic and non-genetic sources. First used by animal breeders but widely adopted by ecological geneticists in the past decade.

Association mapping: A method of detecting QTL that relies on linkage disequilibrium between a marker and the QTL. This usually means only markers that are very close to the QTL can detect it. Therefore, QTL can be mapped with great precision, but the power to detect QTL is limited when marker density is low. Although association mapping does not require pedigree information, it can be very prone to false positives if population structure or relatedness between individuals is not controlled for.

Breeders' equation: $R = h^2S$. The evolutionary response (R) to selection is a product of the selection differential S (i.e. the difference in mean phenotype between individuals that successfully reproduce and the mean phenotype of the overall population) and the heritability h^2 of a trait. The univariate (single trait) version of the breeders' equation works well in artificial selection experiments, but is less reliable in wild populations, in part because natural selection acts on many traits simultaneously.

Evolutionary stasis: The commonly observed situation in the wild when a trait is heritable and under selection but does not seem to be evolving in a way predicted by the breeders' equation.

Fitness trait: A phenotypic trait that is a component of, or contributes to, the number of offspring an individual leaves in the next generation.

Fixed effect: One of two forms of explanatory variable in a statistical model. Fixed effects generally affect the mean of a response variable. For the purposes of mapping experiments, the different alleles at QTL might be regarded as fixed effects.

Genetic architecture: A description of the number, effect size and allele frequencies of loci that contribute to quantitative genetic variation, as well as their mode of action (additive or dominant), the way they interact with one another (epistasis) and the way their effect depends on environmental conditions (gene by environment interaction).

Identity by descent (IBD): If two individuals share a particular allele that they both inherited from a common ancestor then that allele is IBD. Pedigree information can be used to estimate the overall proportion of the genome that two individuals share IBD, whereas marker information can be used to estimate IBD coefficients at specific points of the genome. It is the former that is required to estimate heritabilities with the animal model, whereas the latter is required to search for QTL.

Linkage disequilibrium: The non-random association between alleles at two different loci. Note that linkage disequilibrium can occur between linked or unlinked loci, but because it is broken down by recombination it will tend to persist longer when the loci are in tight linkage.

Microevolution: Relatively small evolutionary changes measurable within a population e.g. as a function of a change in gene frequencies.

Pleiotropy: When variation at a gene can affect more than one character.

Quantitative trait loci (QTL): A region(s) of the genome that explains some (often a small part) of the genetic variation in a continuously varying trait.

Random effect: The other type (see fixed effect) of explanatory variable in a model. Random effects explain a proportion of variation in the response variable, but it is often not known how many levels of a given random effect are present in the population. Examples of random effects include the polygenic (heritable) component of a phenotypic trait and a QTL effect when detected in an animal model/variance components framework. In this setting, the number of alleles at the QTL is unknown.

Variance components: The different sources of phenotypic variation that can be measured as different random effects e.g. using the animal model to partition phenotypic variation into the components caused by polygenic variation, a QTL, environmental effects and other unknown, or residual, sources of variation.

Corresponding author: Slate, J. (j.slate@shef.ac.uk)

0168-9525/$ – see front matter doi:10.1016/j.tig.2010.03.005 Trends in Genetics 26 (2010) 275–284

Figure 1. Examples of PWPs that are the focus of gene mapping studies. Clockwise from top left. Red deer (*C. elaphus*) on Rum, Scotland; photo: J Slate, Soay sheep (*O. aries*) on St Kilda, Scotland; photo: A Ozgul, great tits (*P. major*) in de Hoge Veluwe, The Netherlands; photo: K van Oers, house sparrows (*Passer domesticus*) on various Norwegian islands; photo: H Jensen, song sparrows (*Melospiza melodia*) on Mandarte Island, Canada; photo: L Keller.

interest in developing genetic maps of PWPs to identify quantitative trait loci (QTL) for fitness-related traits [6,7]. Until recently, mapping was only feasible for PWPs that are closely related to model organisms or important domestic species, and efforts to integrate quantitative genetic and QTL studies were rare, relying heavily on genetic markers developed for domestic cattle and sheep [8,9]. However, the advent of next-generation sequencing technologies [10–12] has made it possible to develop and type a large number of genetic markers in any organism [13,14]. The aim of this review is to describe the kinds of questions that can (and cannot) be addressed by mapping in wild populations and outline how the advent of next-generation genomics tools is set to dramatically increase the rate of progress in this area.

How gene mapping in PWPs is done

As a first step, it is necessary to generate a genetic map of the study species. Briefly, this is done by typing a pedigree of individuals, spanning two or more generations, at a suite of genetic markers. The markers are assigned map positions by studying their cosegregation in the mapping pedigree. There are now genetic maps for at least six PWPs with more on the way (Table 1), but all of these maps have relied on genomics resources from related organisms, e.g. domestic cattle and sheep microsatellites to build maps of red deer (*Cervus elaphus*) and Soay sheep (*Ovis aries*) and chicken genome sequences to develop SNPs in collared flycatchers (*Ficedula albicollis*). They all have a modest marker density of ~1 marker per 10–15 centiMorgans (cM), which is likely to impact on the accuracy of the location of QTL. Typical mapping pedigrees contain 300–500 individuals, most of which will also be measured at the phenotype(s) of interest. Notably, though, there are several PWPs that have no genetic maps to date (e.g. bighorn sheep *Ovis canadensis*, great tits *Parus major*), despite being the focus of a large number of quantitative genetic analyses and being, in all other aspects, ideal systems for studying evolutionary genetics in the natural environment (Table 1). Once a map is constructed, QTL are identified by variance components linkage mapping (Box 1).

Table 1. PWPs that have been the focus of mapping studies

Species	Population	Investigators	Genetic map	Linkage disequilibrium description	Trait mapping	Next-generation approach	References
Soay sheep (*O. aries*)	St Kilda, Scotland	J Pemberton, J Slate	Y	Y	Y	Solexa, 454, 50 K SNP chip	[8,17–19,48,49,54]
Red deer (*C. elaphus*)	Rum, Scotland	J Pemberton	Y	Y	Y	–	[9,27]
Great tit (*P. major*)	Wytham Woods, Oxford	B Sheldon, J Slate	N*	N*	N*	454, 10 K SNP chip	
	De Hoge Veluwe, The Netherlands	M Groenen, M Visser, K Van Oers	N*	N*	N*	Solexa, 10 K SNP chip	
Collared flycatcher (*F. albicollis*)	Gotland, Sweden	H Ellegren, L Gustafsson	Y	Y	N		[55–57]
House sparrow (*P. domesticus*)	Norway	H Jensen, B-E Sæther	N*	N	N*	454	
	Lundy, UK	T Burke	N*	N	N	454	
Siberian jay (*Perisoreus infaustus*)	Finland	J Merilä	Y	Y	N	–	[58,59]
Blue tit (*P. caeruleus*)	Sweden	B Hansson	Y	N	N	–	[60]
Great reed warbler (*Acrocephalus arundinaceus*)	Sweden	B Hansson	Y		Y	–	[61,62]
Bighorn sheep (*O. canadensis*)	Canada	D Coltman, J Poissant	N*	N	N*	–	
Song sparrow (*M. melodia*)	Mandarte Island, Canada	L Keller, E Postma	N*	N	N*	454	
Californian condor (*Gymnogyps californianus*)	California, Arizona, Baja California	M Romanov, O Ryder	N*	N	N*	454	[63]

*Unpublished work, known to be in progress.

Box 1. Variance components QTL mapping in wild populations

Once a map is constructed, QTL detection in PWPs requires a relatively complex statistical framework compared with studies in laboratory crosses. This is because matings are not controlled and so the cosegregation of markers and phenotypes occurs in a complicated pedigree with overlapping generations and many different kinds of relatives (parent–offspring, full-sibs, half-sibs, grandparent–grandoffspring, cousins etc.). The solution to this problem is to carry out the analysis using a variance components approach. Essentially, this is an animal model whereby the presence of QTL is tested by fitting, as a random effect, an IBD relationship matrix for that specific part of the genome in addition to the relationship (i.e. the 'genome-wide average' IBD) matrix used in 'traditional' animal models to measure the polygenic additive genetic variance of the trait [7,71]. Variance component mapping is flexible in the sense it can handle complex pedigrees, but it comes at the disadvantage that the effect of a QTL is reported in terms of the total amount of variance explained, rather than as a mean effect size of each QTL allele. Indeed, the number of alleles at a given QTL is not even estimated. This is an important limitation because many of the most interesting evolutionary questions that can be addressed by mapping require knowledge of the alleles at the QTL. Association mapping (Box 3) offers a solution to this problem.

How new genomic resources have changed the way it is done

For those species where maps have been developed, map construction was a long and painstaking process. With the advent of next-generation sequencing the situation has changed, and it is now possible to rapidly identify large numbers (e.g. many thousands) of SNPs in virtually any species. This has been exemplified by a recent sequencing study of an intensively studied PWP of great tits (*P. major*) in The Netherlands [15] (Box 2). Furthermore, the availability of a much larger number of markers, alongside medium- (100s–1000 s) and high- (3000–1 000 000) throughput SNP typing platforms [16], means that mapping studies in PWPs are now likely to become increasingly powerful and sophisticated. Genotyping costs have fallen sufficiently to allow much larger numbers of individuals to be typed. For example, typing ~1400 individuals at 384 SNPs currently costs ~£0.04 per genotype, whereas a larger experiment with ~2800 individuals and 10 000 SNPs will cost ~£0.01 per genotype. Earlier mapping studies, where ~400 individuals were typed at ~100–200 microsatellites cost ~£0.50–£1.00 per

Box 2. SNP discovery by Solexa sequencing in a wild great tit population

Making the transition from a variance components mapping approach with a few hundred markers to an association mapping approach with many thousands of markers was, until recently, impossible because no PWP had a sufficiently large number of markers. However, a recent study in great tits (*P. major*) has outlined how next-generation sequencing approaches can be used to efficiently identify many thousands of markers.

Great tits are one of the most intensively studied vertebrate species in ecological genetics, with several PWPs being the focus of decades-long field research, resulting in some landmark quantitative genetic studies. Therefore, it is surprising that no genetic map and very few polymorphic markers have been described for this species. A recent study [15] used short read Illumina Solexa sequencing of birds from the well-known de Hoge Veluwe population in The Netherlands to mine 20 000 SNPs, which has laid the foundations for gene mapping in this species (Figure I). An alternative, and currently more widespread, approach to SNP discovery is to sequence cDNA (the transcriptome) using Roche-454 sequencing and identify SNPs in the coding regions and untranslated regions of known genes (for example, [13]). Regardless of how the SNPs are discovered, the important point is that several companies will manufacture bespoke SNP chips for non-model organisms, which means experiments can be designed where every individual is typed at tens of thousands of markers. This paves the way for association mapping studies (Box 3), which in turn opens up a new set of evolutionary questions that can be tackled in wild populations.

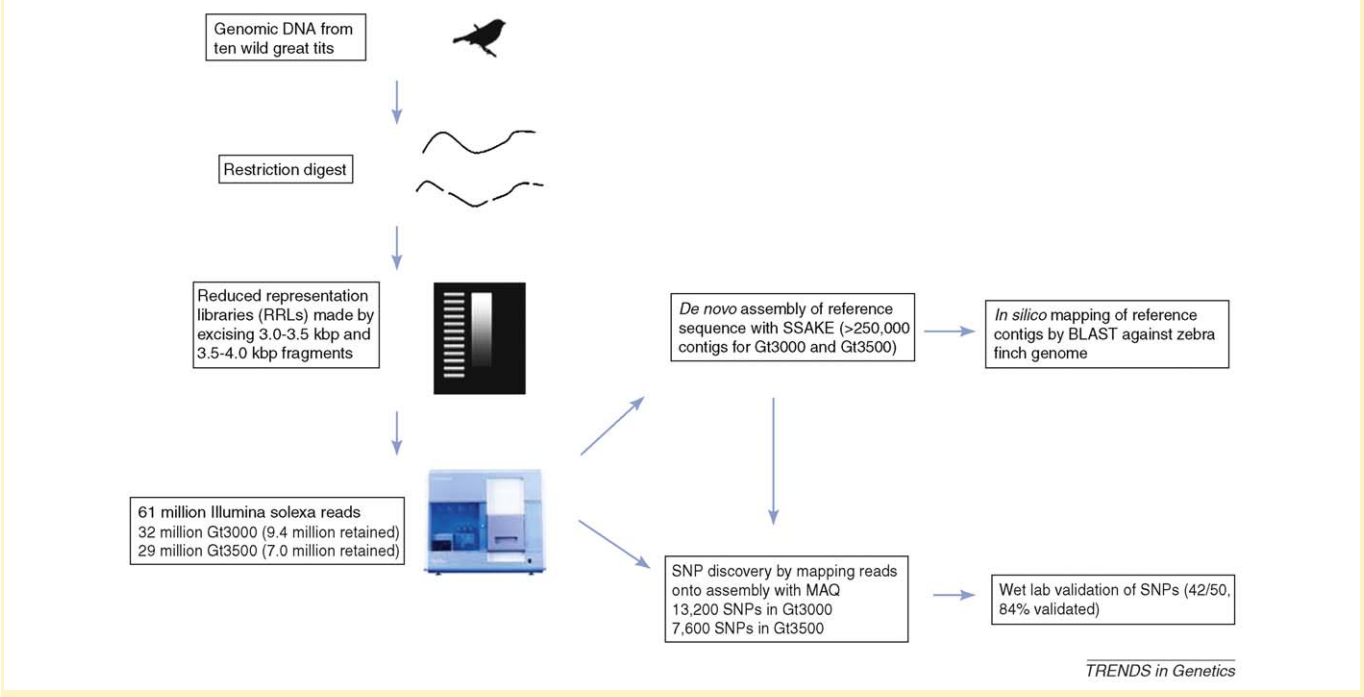

Figure I. SNP discovery in the great tit by Illumina Solexa sequencing.

Box 3. GWAS in wild populations

The creation of SNP chips with 10 000 or more markers has made GWAS possible in PWPs. In association mapping, QTL are detected by fitting each marker as a fixed effect and pedigree information is not required (although fitting pairwise relatedness, by either estimating it directly from the pedigree or indirectly with markers, will result in more robust results). This means that QTL are only detected by SNPs in strong linkage disequilibrium with the causative locus, and so the resolution of position should be improved. One consequence of association mapping is that QTL effects can be described in population genetic terms (whereby the frequencies and mean effect sizes of different alleles can be measured), whereas variance components mapping is a quantitative genetic approach that reports the proportion of trait variance explained by the QTL. The former approach opens up new opportunities because it means QTL allele frequencies can be tracked temporally and responses to selection can be measured Table I .

Table I. Comparisons between variance components linkage mapping and genome-wide association mapping

	Variance components mapping	Association mapping
Required marker interval	Sparse – a 15 cM interval is common	Dense – markers must be in linkage disequilibrium with the QTL, so usually less than 1 cM
Pedigree required	Yes	No, although it can help prevent type 1 errors
Risk of type 1 error	Not if appropriate corrections for multiple testing are made	Yes, if population structure not accounted for. Additionally, there are usually more tests conducted than in variance components mapping
Resolution of QTL position	Generally poor – intervals are typically tens of cM	Much better than linkage mapping, determined by the extent of linkage disequilibrium around the QTL
Are different QTL alleles/genotypes identified?	No, QTL are only estimated as a proportion of trait variance	Yes, provided marker is in linkage disequilibrium with the QTL. The phenotypic mean of each allele can be estimated.
Statistical model	$y = X\beta + Za + Zq + e$ y is the phenotype of interest, β is a vector of fixed effects, a is a vector of additive genetic effects, q is a vector of QTL effects and e is a vector of residual effects. X and Z are design matrices relating records to the appropriate fixed or random effects. Note that the QTL is a random effect	$y = X\beta + Za + e$ y is a vector of the phenotype of all individuals, β is a vector of fixed effects, including the SNP, fitted as a fixed effect to test for QTL, a is a vector of additive genetic effects and e is a vector of residual effects. X and Z are design matrices relating records to the appropriate fixed and random effects
Questions that can be addressed		
The proportion of trait variation explained by a QTL	✓	✓
Estimating the phenotypic mean of each QTL allele	x	✓
The QTL mode of action (are alleles additive or dominant?)	x	✓
Whether genetic correlations are caused by colocalised QTL	✓	✓
Whether QTL have sexually antagonistic effects	✓	✓
Whether the same QTL are important in different populations	✓	✓
Whether QTL alleles have the same effect in different populations	x	✓
Are there QTL by environment interactions?	✓	✓
Has the QTL responded to selection?	x	✓

genotype. Therefore, the power to detect QTL will be improved. More important than the issue of power, however, is the fact that higher marker densities and more typed individuals means that QTL detection can be performed in a fundamentally different way; for the first time it is possible to carry out genome-wide association studies (GWAS) (Box 3) in wild populations. The big advantage of this is that QTL effects can now be fitted in models as fixed effects, which should make evolutionary analyses that were hitherto difficult far more tractable. Association mapping studies in PWPs will be greatly aided by several recently developed bioinformatics and statistical tools that will aid SNP discovery, the management of large datasets, experimental design and the detection of QTL (Table 2).

Evolutionary questions that gene mapping in PWPs can address

The remainder of this review discusses the types of evolutionary questions that can be addressed now that new genomics tools have opened up the opportunity to map QTL in more populations and at a greater resolution. A possible criticism of the first mapping studies in PWPs is that they were unable to reveal how traits responded to selection in the wild. The main reason for this is that they were unable to resolve QTL locations to a very precise

Table 2. Recent resources to help gene mapping in PWPs.

Main use	Name	Description	URL/Reference
Assembly of sequence data and SNP discovery	SEQanswers	Forum featuring a very comprehensive discussion of all things next-generation sequencing-related. Bioinformatics section contains detailed information on most of the available programs.	http://seqanswers.com/forums/index.php
QTL power analysis and simulation	QMSim	QTL and marker simulator. Excellent tool for generating dummy datasets under very realistic scenarios. Can simulate QTL, polygenic variation and selection.	http://www.aps.uoguelph.ca/~msargol/qmsim/ [64]
	Pedantix	R package for power analyses of quantitative genetic studies of wild populations. Includes the option to simulate marker data.	http://wildevolution.biology.ed.ac.uk/awilson/pedantix.html [65,66]
	powQ	Software for the power analysis of QTL mapping by variance components mapping. Allows for complex pedigrees, although assumes marker information content is perfect.	http://www.twin-research.ac.uk/WebPowQ/PowQ.htm [67]
	Quantinemo	Program to simulate individual-based marker, QTL and trait data under predefined genetic architectures.	http://www2.unil.ch/popgen/softwares/quantinemo/ [68]
	Genetic Power Calculator	Website that provides simple interface for estimating power to detect QTL by variance components linkage mapping and association study.	http://pngu.mgh.harvard.edu/~purcell/gpc/ [69]
SNP data management and genome-wide association mapping	PLINK	Whole genome association analysis toolset. Very useful for marker data manipulation and exploration, as well as performing association studies.	http://pngu.mgh.harvard.edu/~purcell/plink/ [70]
Linkage map construction	CriMap	Old linkage mapping software, but still the best program for building maps in complex pedigrees.	http://compgen.rutgers.edu/Crimap/Default.aspx
QTL mapping and animal model construction	GridQTL	Grid-based QTL mapping software that includes options for general pedigrees, as well as combined linkage disequilibrium and linkage mapping. Standard error of QTL effect not currently reported, though.	http://cleopatra.cap.ed.ac.uk/gridsphere/gridsphere [52]
	WamWiki	Help pages, including example scripts for running animal models in various packages (As-REML, WOMBAT, MCMCglmm). These packages can be used to map QTL if IBD matrices are obtained via other programs.	http://wildanimalmodels.org/tiki-index.php [25]

location, and so they failed to identify marker alleles or haplotypes that were in linkage disequilibrium with the genes explaining quantitative trait variation. As a result, it was impossible to measure the fitness of different alleles at QTL or see whether selection was causing a change in frequency of these alleles. Despite this valid criticism, the early studies can be regarded as pioneering in the sense that they showed gene mapping in PWPs was achievable.

Single genes of major effect – is there really any evidence?

The first mapping studies of PWPs examined the number and magnitude of QTL underlying continuous trait variation. To date, all studies have found evidence that much genetic variation is explained by a very small number of loci of large effect [9,17–20]. This finding is both surprising and important. It is surprising because although studies of model organisms have also provided some evidence of reasonably large effect QTL, their effects are smaller than in PWPs even though QTL are expected to explain more phenotypic variance in inbred lines than in outbred populations [21]. Furthermore, some of the apparently medium-to-large effect QTL found in model organisms were subsequently shown to be several tightly linked smaller effect QTL [21]. In human genetics, very large GWAS of morphological traits such as height have typically found QTL of small effect [22], and in fact have failed to find the loci underlying much of the genetic variation in complex traits. These unknown loci are sometimes referred to as the 'missing heritability' of a trait. It is likely that high-powered studies in humans have failed to find these QTL because they have rare alleles, small effect sizes or both [23]. If the opposite situation (i.e. genes of large effect) really is true in wild vertebrate populations, then it would seem that morphological traits in PWPs have a profoundly different genetic architecture than model organisms or humans, and microevolutionary responses to natural selection can be rapid.

How strong is the evidence for a few genes/large effects model of continuous variation in PWPs? The 'Beavis Effect' [24] is a well-established phenomenon in QTL mapping studies that describes how the magnitude of significant QTL tends to be overestimated, particularly when sample sizes are modest (n ~300–500), as is the case for PWP studies. There has been no formal evaluation of whether the observed data from wild populations are inconsistent with a polygenic model, although it is acknowledged in all studies to date that the Beavis Effect is likely to be an

Box 4. Inflated QTL and the Beavis Effect

Some insight into the accuracy of QTL estimates in PWPs can be obtained with simulated datasets, where approximately 400 individuals in a randomly mating population are typed at either 10 microsatellites or 100 SNPs on a 100 cM chromosome with a QTL at 50 cM. The simulated trait has a heritability of 0.4, of which the QTL effect is varied to explain between zero and all of the overall heritability. Data were generated using QMSim, and QTL detected by variance components mapping with GridQTL (Table 2). The purpose of the simulations is not to formally or rigorously compare the SNP and microsatellite datasets, but to demonstrate that estimated QTL effect size estimates are not particularly robust and can be dramatically inflated (Figure I).

For a QTL of major effect (defined here as one explaining 0.2 of the total variance, with the remaining polygenic variance also explaining 0.2 of the overall variance), the power to detect the QTL at a $P < 0.0024$ (equivalent to a LOD score of 2.0, which roughly corresponds to 'suggestive linkage' thresholds used in PWP mapping studies) is modest. QTL estimates are upwardly biased, especially if estimated from significant results only. If published QTL effects in PWPs are *always* highly inflated, then even meta-analyses of the published data are likely to lead to wrong conclusions. The bias seems to be worse in the SNP dataset, although this might be an artefact of a modest number of replicates.

The arrival of high-throughput SNP typing is likely to improve the situation. The dramatic decrease in genotyping costs means that the next generation of mapping studies will be conducted on datasets 5–10 times as large as those published to date. If the dataset is ~4000 rather than ~400 individuals the problem of bias is largely eliminated (Table I), indicating that it is a result of the Beavis Effect, rather than any innate problem with the variance components method of QTL mapping. To formally analyse a greater amount of parameter space (e.g. different sized QTL, pedigrees, genomes, marker densities etc.) more simulations would be required. However, the parameters used here are typical of the study populations described to date.

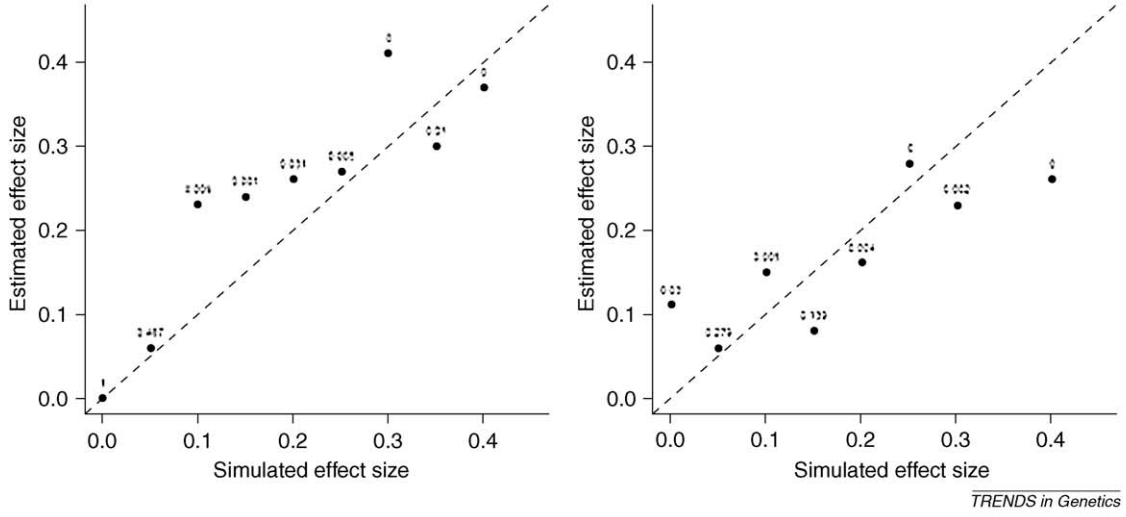

Figure I. Simulated and estimated QTL effect sizes for microsatellites (left) and SNPs (right). Numbers above each data point are *P* values.

Table I. Precision of estimates of QTL of effect size 0.2

Dataset	Mean LRT	Replicates significant at $P < 0.05$	Replicates significant at $P < 0.0024$	All replicates		Significant (at $P < 0.0024$) replicates only	
				Mean estimated QTL effect	QTL RMSE	Mean estimated QTL effect	QTL RMSE
Microsatellites	9.76	10/10	4/10	0.22	0.04	0.26	0.07
SNPs, n = 400	13.53	7/10	4/10	0.23	0.13	0.37	0.18
SNPs, n = 4000	85.27	10/10	10/10	0.19	0.03	0.19	0.03

RMSE = root mean square error of QTL effect, LRT = likelihood ratio test statistic, 10 replicates for each scenario.

issue, and that QTL estimates might be overestimated. That said, at least one study that found evidence of a major QTL for a morphological trait in a PWP [19] was in a candidate region that was *a priori* expected to be important, and in this situation bias due to the Beavis Effect is expected to be less of a problem. Fortunately, the scale on which PWP mapping studies are conducted is set to increase (Box 3), and the overestimation of QTL effects should become less of a problem (Box 4). At that stage, the issue of whether major effect QTL really are segregating in wild populations will be resolved.

Although a description of the genetic architecture of traits studied in wild vertebrate populations is desirable, that is not the sole reason for carrying out mapping studies. Below we outline other important questions, which are less sensitive to the Beavis Effect, that, if addressed, will help explain how evolution by natural selection operates.

Genetic correlations between traits

A major motivation for studying PWPs is to understand the causes of apparent evolutionary stasis [2,5,25]. One explanation is that the breeders' equation, which predicts the response to selection given the strength of selection and a trait's heritability, considers the trait in isolation, whereas in reality different traits are genetically correlated. This means that an evolutionary response to selection on a focal trait can be accelerated, dampened, stalled or even reversed depending on the sign of the correlation and the direction of selection on other non-independent traits. These correlations are typically regarded as either being

caused by pleiotropy (in which case they are fixed) or linkage disequilibrium between linked genes (in which case their magnitude and sign can be changed by recombination). However, there are alternative explanations. For example, in wild populations ephemeral linkage disequilibrium between unlinked loci can cause short-term genetic correlations that are broken down by independent assortment within a few generations. This scenario could arise either if a small number of individuals achieve a high proportion of reproductive success, or if immigrants bring novel alleles at multiple loci into the population [26,27]. Either scenario is plausible, and possibly common, in PWPs. Gene mapping has the potential to unravel the causes of genetic correlations. If QTL for two correlated traits are colocalised, then pleiotropy or close linkage explains the correlation. Potentially, if sample sizes are large enough, it might be possible to disentangle pleiotropy and tight linkage. Alternatively, if QTL for the two traits are in different locations then genetic correlation is probably an ephemeral one. To date, the only genetically correlated traits in any PWP that have also been mapped are parasite load (measured as strongyle worm faecal egg count) and morphological traits (body weight and hind leg length) in Soay sheep. These traits have a negative genetic correlation of about –0.3 [28], but there is no evidence that faecal egg count QTL [17] and morphological QTL [18] overlap.

Gene by environment interactions

When the effect of a given genotype on the resulting phenotype varies with environmental conditions, phenotypic plasticity is said to occur [29]. Phenotypic plasticity has become increasingly well studied in PWPs because it can provide a mechanism by which populations can adapt to a changing environment [30] and can even be a potential explanation for the maintenance of genetic variation in the face of directional selection [2]. Although genetic variation in plasticity has been demonstrated in PWPs [31,32], there have been no attempts to examine plasticity at the level of individual QTL. In principle, QTL–environment interactions could be investigated for traits that are expressed and measured multiple times within an individual's lifetime (e.g. antler size in red deer, clutch size or laying date in passerine birds). By treating phenotypes recorded in separate environments (e.g. warm versus cold springs, high versus low population density) as discrete traits, or by mapping the QTL effect in an individual as a function of an environmental variable (i.e. in a random regression animal model [2]), it should be possible to determine which, if any, QTL have effects that are sensitive to environmental conditions. If the sign of a QTL effect varies between environments then variation at that locus is more likely to be maintained.

Genetic correlations between sexes

Gene mapping studies also have the potential to provide insight into sexual selection and sexual conflict. For example, there is considerable interest in whether male ornaments are genetically correlated with female preference [33], a prediction of several hypotheses of sexual selection [34]. If so, genetic correlations are most likely to occur (and be maintained) when QTL are linked, and in the case of male ornamentation and female preference it has been suggested that sex chromosomes might contribute disproportionately to the phenotypic variance [34], although the evidence supporting this is weak [35]. Similarly, theory predicts that sexual conflict, which has been described in PWPs [36], has several 'resolutions' [37] that might be evaluated by mapping. These include testing whether QTL alleles for fitness traits have opposite effects in males and females (there is some evidence they have [38]) and testing whether they show evidence for either sex linkage or genomic imprinting.

To test whether QTL have opposite fitness effects in males and females one needs to treat male and female fitness as separate traits and then perform a bivariate trait variance components QTL analysis. If there is a negative covariance between the male and female QTL variance components then there is evidence that a QTL affecting fitness acts antagonistically between the sexes. Testing whether a QTL is sex-linked is relatively straightforward provided the chromosome of interest is the one found in both sexes (X chromosome in mammals, Z chromosome in birds). The mammalian Y chromosome and avian W chromosome are not easy to investigate in mapping studies because only one copy is ever found in an individual (in one sex) and, therefore, it is not possible to follow cosegregation between a marker and a trait.

Testing for imprinting is possible by examining parent-of-origin effects. For example, a half-sib Haley-Knott linear regression approach [39], which examines QTL segregation independently in maternal and paternal sibships, has provided some evidence that birth weight QTL in red deer are only maternally expressed [9]. Modelling the parent of origin of each allele is not trivial in a variance components mapping framework, although methods are now available to do so [40]. Initially, imprinting was thought to be a feature of viviparous (live young bearing) rather than oviparous (egg-laying) animals, because it was thought most likely to evolve in organisms with polyandrous mating systems and *in utero* maternal provisioning of the offspring. However, some theories predict that intralocus sexual conflict could lead to the evolution of imprinting in a broader set of taxa [37,41]; wild bird populations with reasonably large maternal and paternal sibships and measures of fitness would make an interesting place to look for imprinted QTL.

Are the same loci relevant in different populations?

To date, no mapping study of a PWP has been replicated in a second population of the same study species. It would be useful to replicate mapping studies in independent populations for two reasons. First, the discovery of QTL in a second population is often regarded as the only way to confirm that the QTL is real [42], although the absence of the QTL in the second population does not necessarily mean that it was a false positive in the first. Second, although there are well-known examples of adaptive traits evolving via the same genes (parallel evolution; for examples see [43]), we currently have little idea whether a common set of genes can explain segregating variation in fitness-related traits within geographically isolated popu-

lations. Fortunately, some species have been the focus of multiple PWP studies (e.g. great tits), and so there is an opportunity to replicate mapping studies (Box 1). For example, a recent study [44] followed up an association between the dopamine receptor D4 gene (*Drd4)* and exploratory personality in a Dutch population of great tits [45] by examining the relationship in three additional populations (in Belgium, the UK and a second Dutch population). The original association, explaining around 5% of phenotypic variation, although never confirmed by linkage mapping, was replicated in a second sample from that population, but was absent in the other populations. In this instance, it seems that *Drd4* does not explain variation in multiple populations.

Locus of evolution

There has been considerable debate in the evolutionary genetics literature as to whether adaptive evolution is caused predominantly by replacement substitutions in coding regions of genes or whether regulatory mutations are more important [46,47]. In wild vertebrates, most of the identified mutations to date are coding region substitutions. However, these are likely to be much easier to discover than regulatory mutations because candidate genes for many traits are known from model organisms, and the effect on gene function of mutations in coding regions can be readily predicted (i.e. replacement substitutions are more likely to affect phenotype than silent ones). Regarding coding region and regulatory mutations as mutually exclusive alternatives is unhelpful because it is likely that both are important sources of natural variation, and that other types of mutation such as indels, copy number variants and chromosomal inversions are also likely to have significant effects. Indeed, proponents of the 'coding region' hypothesis have taken this more holistic viewpoint from the outset [46].

In PWPs, gene mapping studies of single locus traits are now beginning to emerge and, perhaps unsurprisingly, there is evidence for both types of mutation. In Soay sheep, a polymorphism in coat colour is explained by a single base replacement substitution in the coding region of the tyrosinase-related protein 1 gene (*TYRP1*) on sheep chromosome 2 [48]. A second single locus coat polymorphism in Soay sheep has also recently been mapped; the coat pattern polymorphism is independent to coat colour and is caused by mutations in the agouti signalling protein *ASIP* on sheep chromosome 13. However, sequencing *ASIP* revealed a surprisingly complex genetic basis of the coat pattern polymorphism. At *ASIP* there are three mutations in and around this gene that can all result in the same derived phenotype, a uniform coat pattern termed 'self' [49]. One of these mutations is a replacement substitution in the coding region, one is a deletion in the coding region that results in a frameshift and a premature stop codon (and therefore a non-functional protein) and the third (unknown) mutation regulates *ASIP* expression. All three mutations are recessive and in strong linkage disequilibrium, which means they effectively segregate as a single locus. Therefore, differences in both amino acid sequence and gene expression can generate the self phenotype in this population. However, it is still too early to say whether we can generalise about the type of mutation influencing traits under selection in PWPs because no single locus traits have been mapped in other populations and individual genes/mutations underlying variation in quantitative traits have not yet been identified in any PWP, including Soay sheep.

Using association mapping to detect and track fitness-associated haplotypes

A major motivation for mapping QTL in PWPs is to measure the fitness of different alleles at a locus and track whether they have responded to selection e.g. by observing whether their frequencies have changed over the duration of a long-term project. This has been achieved in Soay sheep where the mutation for a coat colour polymorphism [48] was typed in several thousand individuals, and associations with fitness-related traits were examined [50]. Of particular interest was the observation that the frequency of the light coat colour had increased over 25 years when it was expected to decline due to an association with smaller body size, which is disadvantageous. This apparently counterintuitive trend was resolved when it was shown that linkage disequilibrium between *TYRP1* and adjacent genes influences variation in both body size and fitness, with the result that both light sheep and heterozygous dark sheep were fitter than homozygous dark sheep [50]. The differences in fitness were undetectable by studying the phenotype alone because homozygous and heterozygous dark sheep are phenotypically indistinguishable. Therefore, for the first time, gene mapping was able to explain an evolutionary trend that was unexplainable by quantitative genetic approaches. Interestingly, the body size QTL was not originally detected in a variance components mapping study [18] but was later detected by an approach that combines linkage and association mapping [51]. The subsequent use of the LDLA module of the GridQTL software [52], which combines linkage mapping, association and historical population demography parameters, has further resolved the QTL location.

Concluding remarks

Next-generation sequencing and high-throughput SNP typing, along with the development of new association mapping frameworks, are set to revolutionise genetic studies of pedigreed wild vertebrate populations. For the first time, it should be possible to identify genomic regions associated with fitness variation and track the origin, dynamics and fate of loci under selection. We should not lose sight of the fact that long-term field data are the foundation of many evolutionary studies, and that keeping data collection going over decades is extremely difficult [53], making datasets such as the ones described here immensely valuable. Nor should we overlook the challenges that we will face with bioinformatics, genotyping and statistical genetic analyses. However, there is now a tremendous opportunity to build on the ecological genetics legacy of EB Ford, Theodosius Dobzhansky and others by integrating quantitative and population genetic analyses to better understand the evolution of traits under selection in some of the richest ecological datasets.

Acknowledgements

We thank our 'gene mapping in the wild' collaborators, especially Josephine Pemberton, for stimulating discussions over many years on these and related topics. Kees van Oers, Henrik Jensen, Lukas Keller and Arpat Ozgul provided photos of organisms. Three anonymous reviewers and the editor provided insightful comments on an earlier draft of the manuscript. Funding was provided by the European Research Council, the Biotechnology and Biological Sciences Research Council, the Natural Environment Research Council and the European Commission Transfer of Knowledge program.

References

1 Kruuk, L.E.B. and Hill, W.G. (2008) Introduction. Evolutionary dynamics of wild populations: the use of long-term pedigree data. *Proc. R. Soc. B* 275, 593–596
2 Kruuk, L.E.B. *et al.* (2008) New answers for old questions: the evolutionary quantitative genetics of wild animal populations. *Annu. Rev. Ecol. Evol. Syst.* 39, 525–548
3 Merilä, J. and Sheldon, B.C. (2001) Avian quantitative genetics. *Curr. Ornithol.* 16, 179–255
4 Kruuk, L.E.B. (2004) Estimating genetic parameters in natural populations using the 'animal model'. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 359, 873–890
5 Merilä, J. *et al.* (2001) Explaining stasis: microevolutionary studies in natural populations. *Genetica* 112, 199–222
6 Ellegren, H. and Sheldon, B.C. (2008) Genetic basis of fitness differences in natural populations. *Nature* 452, 169–175
7 Slate, J. (2005) QTL mapping in natural populations: progress, caveats and future directions. *Molec. Ecol.* 14, 363–379
8 Beraldi, D. *et al.* (2006) Development of a linkage map and mapping of phenotypic polymorphisms in a free-living population of Soay sheep (*Ovis aries*). *Genetics* 173, 1521–1537
9 Slate, J. *et al.* (2002) A genome scan for quantitative trait loci in a wild population of red deer (*Cervus elaphus*). *Genetics* 162, 1863–1873
10 Mardis, E.R. (2008) The impact of next-generation sequencing technology on genetics. *Trends Genet.* 24, 133
11 Hudson, M.E. (2008) Sequencing breakthroughs for genomic ecology and evolutionary biology. *Mol. Ecol. Res.* 8, 3–17
12 Metzker, M.L. (2010) Sequencing technologies – the next generation. *Nat. Rev. Genet.* 11, 31–46
13 Vera, J.C. *et al.* (2008) Rapid transcriptome characterization for a nonmodel organism using 454 pyrosequencing. *Molec. Ecol.* 17, 1636–1647
14 Baird, N.A. *et al.* (2008) Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS One* 3, 7
15 van Bers, N. *et al.* (2010) Genome-wide SNP detection in the great tit *Parus major* using high throughput sequencing. *Molec. Ecol.* 19, 89–99
16 Ragoussis, J. (2009) Genotyping technologies for genetic research. *Annu. Rev. Genomics Hum. Genet.* 10, 117–133
17 Beraldi, D. *et al.* (2007) Quantitative trait loci (QTL) mapping of resistance to strongyles and coccidia in the free-living Soay sheep (*Ovis aries*). *Int. J. Parasitol.* 37, 121–129
18 Beraldi, D. *et al.* (2007) Mapping QTL underlying fitness-related traits in a free-living sheep population. *Evolution* 61, 1403–1416
19 Johnston, S.E. *et al.* (2010) Horn type and horn length genes map to the same chromosomal region in Soay sheep. *Heredity* 104, 196–205
20 Tarka, M. *et al.* (2010) A strong quantitative trait locus for wing length on chromosome 2 in a wild population of great reed warblers. *Proc. R. Soc. B* (doi:10.1098/rspb.2010.0033)
21 Flint, J. and Mackay, T.F.C. (2009) Genetic architecture of quantitative traits in mice, flies, and humans. *Genome Research* 19, 723–733
22 Visscher, P.M. (2008) Sizing up human height variation. *Nat. Genet.* 40, 489–490
23 Manolio, T.A. *et al.* (2009) Finding the missing heritability of complex diseases. *Nature* 461, 747–753
24 Beavis, W.D. (1994) The power and deceit of QTL experiments: lessons from comparative QTL studies, In *Proceedings of the forty-ninth annual corn and sorghum industry research conference*, American Seed Trade Association. pp. 250–266
25 Wilson, A.J. *et al.* (2010) An ecologist's guide to the animal model. *Journal Of Animal Ecology* 79, 13–26
26 Hansson, B. *et al.* (2004) Does linkage disequilibrium generate heterozygosity-fitness correlations in great reed warblers? *Evolution* 58, 870–879
27 Slate, J. and Pemberton, J.M. (2007) Admixture and patterns of linkage disequilibrium in a free-living vertebrate population. *J. Evol. Biol.* 20, 1415–1427
28 Coltman, D.W. *et al.* (2001) Positive genetic correlation between parasite resistance and body size in a free-living ungulate population. *Evolution* 55, 2116–2125
29 Nussey, D.H. *et al.* (2007) The evolutionary ecology of individual phenotypic plasticity in wild populations. *J. Evol. Biol.* 20, 831–844
30 Charmantier, A. *et al.* (2008) Adaptive phenotypic plasticity in response to climate change in a wild bird population. *Science* 320, 800–803
31 Nussey, D.H. *et al.* (2005) Selection on heritable phenotypic plasticity in a wild bird population. *Science* 310, 304–306
32 Reale, D. *et al.* (2003) Genetic and plastic responses of a northern mammal to climate change. *Proc. R. Soc. B.* 270, 591–596
33 Qvarnström, A. *et al.* (2006) Testing the genetics underlying the coevolution of mate choice and ornament in the wild. *Nature* 441, 84–86
34 Kirkpatrick, M. and Hall, D.W. (2004) Sexual selection and sex linkage. *Evolution* 58, 683–691
35 Mank, J.E. *et al.* (2006) Sex chromosomes and male ornaments: a comparative evaluation in ray-finned fishes. *Proc. R. Soc. Lond. B. Biol. Sci.* 273, 233–236
36 Foerster, K. *et al.* (2007) Sexually antagonistic genetic variation for fitness in red deer. *Nature* 447, 1107–1109
37 Bonduriansky, R. and Chenoweth, S.F. (2009) Intralocus sexual conflict. *Trends Ecol. Evol.* 24, 280–288
38 Poissant, J. *et al.* (2010) Sex-specifc genetic variance and the evolution of sexual dimorphism: a systematic review of cross-sex genetic correlations. *Evolution* 64, 97–107
39 Knott, S.A. *et al.* (1996) Methods for multiple-marker mapping of quantitative trait loci in half-sib populations. *Theor. Appl. Genet.* 93, 71–80
40 Rowe, S.J. *et al.* (2009) Detecting parent of origin and dominant QTL in a two-generation commercial poultry pedigree using variance component methodology. *Genet. Sel. Evol.* 41, 11
41 Day, T. and Bonduriansky, R. (2004) Intralocus sexual conflict can drive the evolution of genomic imprinting. *Genetics* 167, 1537–1546
42 Lander, E. and Kruglyak, L. (1995) Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. *Nat. Genet.* 11, 241–247
43 Dalziel, A. *et al.* (2009) Linking genotypes to phenotypes and fitness: how mechanistic biology can inform molecular ecology. *Molec. Ecol.* 18, 4997–5017
44 Korsten, P. *et al.* (2010) Association between DRD4 gene polymorphism and personality variation in great tits: a test across four wild populations. *Molec. Ecol.* 19, 832–843
45 Fidler, A.E. *et al.* (2007) Drd4 gene polymorphisms are associated with personality variation in a passerine bird. *Proc. R. Soc. Lond. B. Biol. Sci.* 274, 1685–1691
46 Hoekstra, H.E. and Coyne, J.A. (2007) The locus of evolution: evo devo and the genetics of adaptation. *Evolution* 61, 995–1016
47 Carroll, S.B. (2008) Evo-devo and an expanding evolutionary synthesis: a genetic theory of morphological evolution. *Cell* 134, 25–36
48 Gratten, J. *et al.* (2007) Compelling evidence that a single nucleotide substitution in TYRP1 is responsible for coat-colour polymorphism in a free-living population of Soay sheep. *Proc. R. Soc. B.* 274, 619–626
49 Gratten, J. *et al.* (2010) The genetic basis of recessive self-colour pattern in a wild sheep population. *Heredity* 104, 206–214
50 Gratten, J. *et al.* (2008) A localized negative genetic correlation constrains microevolution of coat color in wild sheep. *Science* 319, 318–320
51 Hernández-Sánchez, J. *et al.* (2003) Candidate gene analysis for quantitative traits using the transmission disequilibrium test: the example of the melanocortin 4-receptor in pigs. *Genetics* 164, 637–644
52 Hernández-Sánchez, J. *et al.* (2009) A web application to perform linkage disequilibrium and linkage analyses on a computational grid. *Bioinformatics* 25, 1377–1383
53 Clutton-Brock, T. and Sheldon, B.C. (2010) The seven ages of *Pan*. *Science* 327, 1207–1208
54 McRae, A.F. *et al.* (2005) Modeling linkage disequilibrium in natural populations: the example of the Soay sheep population of St. Kilda, Scotland. *Genetics* 171, 251–258

55 Backström, N. *et al.* (2006) Genetic mapping in a natural population of collared flycatchers (*Ficedula albicollis*): conserved synteny but gene order rearrangements on the avian Z chromosome. *Genetics* 174, 377–386
56 Backström, N. *et al.* (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
57 Backström, N. *et al.* (2006) Levels of linkage disequilibrium in a wild bird population. *Biol. Lett.* 2, 435–438
58 Li, M.H. and Merilä, J. (2009) Extensive linkage disequilibrium in a wild bird population. *Heredity* (doi:10.1038/hdy.2009.150)
59 Jaari, S. *et al.* (2009) A first-generation microsatellite-based genetic linkage map of the Siberian jay (*Perisoreus infaustus*): insights into avian genome evolution. *BMC Genomics* 10, 1
60 Hansson, B. *et al.* (2009) Avian genome evolution: insights from a linkage map of the blue tit (*Cyanistes caeruleus*). *Heredity* 104, 67–78
61 Åkesson, M. *et al.* (2007) Linkage mapping of AFLP markers in a wild population of great reed warblers: importance of heterozygosity and number of genotyped individuals. *Molec. Ecol.* 16, 2189–2202
62 Hansson, B. *et al.* (2005) Linkage mapping reveals sex-dimorphic map distances in a passerine bird. *Proc. R. Soc. B.* 272, 2289–2298
63 Romanov, M. *et al.* (2009) The value of avian genomics to the conservation of wildlife. *BMC Genomics* 10, S10
64 Sargolzaei, M. and Schenkel, F.S. (2009) QMSim: a large-scale genome simulator for livestock. *Bioinformatics* 25, 680–681
65 Morrissey, M.B. and Wilson, A. (2009) pedantics: an r package for pedigree-based genetic simulation and pedigree manipulation, characterization and viewing. *Mol. Ecol. Res.*
66 Morrissey, M.B. *et al.* (2007) A framework for power and sensitivity analyses for quantitative genetic studies of natural populations, and case studies in Soay sheep (*Ovis aries*). *J. Evol. Biol.* 20, 2309–2321
67 Falchi, M and Borlino, C. (2006) PowQ: a user-friendly package for the design of variance component multipoint linkage analysis studies. *Bioinformatics* 22, 1404–1405
68 Neuenschwander, S. *et al.* (2008) quantiNemo: an individual-based program to simulate quantitative traits with explicit genetic architecture in a dynamic metapopulation. *Bioinformatics* 24, 1552–1553
69 Purcell, S. *et al.* (2003) Genetic power calculator: design of linkage and association genetic mapping studies of complex traits. *Bioinformatics* 19, 149–150
70 Purcell, S. *et al.* (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* 81, 559–575
71 George, A.W. *et al.* (2000) Mapping quantitative trait loci in complex pedigrees: a two-step variance component approach. *Genetics* 156, 2081–2092