Gene mapping in the wild with SNPs: guidelines and future directions

Jon Slate · Jake Gratten · Dario Beraldi · Jessica Stapley · Matt Hale · Josephine M. Pemberton

Abstract One of the biggest challenges facing evolutionary biologists is to identify and understand loci that explain fitness variation in natural populations. This review describes how genetic (linkage) mapping with single nucleotide polymorphism (SNP) markers can lead to great progress in this area. Strategies for SNP discovery and SNP genotyping are described and an overview of how to model SNP genotype information in mapping studies is presented. Finally, the opportunity afforded by new generation sequencing and typing technologies to map fitness genes by genome-wide association studies is discussed.

Keywords Gene discovery · QTL · Linkage · Mapping · SNP · Wild population

Introduction

Longitudinal studies of wild animal populations have proven invaluable for studying selection, genetic architecture and microevolution of fitness-related traits, especially since the widespread uptake of the 'animal model' approach to quantitative genetic parameter estimation (Kruuk 2004;

J. Slate (✉) · J. Gratten · J. Stapley · M. Hale
Department of Animal & Plant Sciences, University of Sheffield, Sheffield S10 2TN, UK
e-mail: j.slate@shef.ac.uk

D. Beraldi · J. M. Pemberton
Institute of Evolutionary Biology, University of Edinburgh, Edinburgh EH9 3JT, Scotland, UK

Present Address:
M. Hale
Department of Forestry and Natural Resources, Purdue University, West Lafayette, Indiana 47907-1159, USA

Kruuk and Hill 2008; Merilä et al. 2001). Unfortunately, identifying the actual genes responsible for variation in fitness has proven more difficult, even though the statistical framework and some appropriate study populations have been available for some time (Slate 2005). To date, there are relatively few examples of quantitative trait locus (QTL) studies being conducted in unmanipulated, wild populations (Beraldi et al. 2007a, b; Slate et al. 2002) and these studies have only identified approximate locations of a limited number of QTL. However, there is now a great opportunity to synthesise gene discovery with quantitative genetic studies of wild populations due to the increasing ease (and decreasing cost) with which genomics studies can now be conducted in non-model organisms (Ellegren 2008; Ellegren and Sheldon 2008).

To date, QTL mapping studies in the wild have all been conducted by typing a suite of microsatellite markers, originally identified in closely related model organisms. Although microsatellites have a number of well-documented properties that make them excellent for molecular ecology research (Jarne and Lagoda 1996), they are not ideal for gene mapping. The main limitation of microsatellites is that typing methods are not highly automated; it is difficult to type more than about ten loci in a single reaction. Relative to other markers they are not as abundant in the genome, and marker discovery has traditionally been time-consuming, especially when large numbers of loci are required. There is a suspicion that previous mapping studies in wild populations have approached the limits of what can be realistically achieved with microsatellites and a pedigree of several hundred individuals; i.e. low marker density genome scans that yield crude estimates of QTL location and magnitude.

Single nucleotide polymorphisms (SNPs) are the most abundant type of genetic polymorphism in most, if not all, genomes. In recent years SNPs have attracted growing

interest from researchers who have recognised their potential for addressing a number of outstanding questions in evolutionary biology and ecology (Luikart et al. 2003; Morin et al. 2004; Nielsen 2005). The advantages (and disadvantages) of SNPs relative to other types of molecular marker have been reviewed elsewhere (Morin et al. 2004), and it is not the aim of this article to duplicate that material. However, a relatively new application of SNPs is as a tool for carrying out gene mapping experiments in wild vertebrate populations. The main reasons for using SNPs over (or in addition to) microsatellites in mapping studies are that (i) they can be typed on a much larger scale and (ii) they are much more abundant, meaning that any genomic location can be analysed.

Our research groups have been using SNPs for mapping experiments for approximately 5 years, and we have witnessed a dramatic change in the ways in which SNPs can be identified, genotyped and analysed. The aim of this article is to provide an overview of some of the methods, problems and pitfalls we have encountered during this period, which we hope will act as a guide to others wishing to carry out similar projects. We are mostly interested in using SNPs to map genes relevant underlying traits under selection in unmanipulated, pedigreed vertebrate populations and refer the reader to other reviews for the underlying rationale behind this work (Slate 2005; Ellegren and Sheldon 2008; Kruuk et al. 2008). The main areas we discuss are: methods of SNP discovery, SNP typing, and analyses of SNP data in mapping studies. We also discuss the feasibility of performing genome-wide association studies in wild populations using many thousands of markers. Examples from our own research laboratories are used to compare alternative methods and approaches, but the points we address are generally applicable to other laboratories, other taxonomic groups and other evolutionary questions. In particular the described method are relevant to the QTL or population genomics approaches to the identification of loci involved in population divergence, reproductive isolation and speciation (sensu Rogers and Bernatchez 2005), a topic that is addressed in another paper in this volume (Butlin 2008).

Methods for SNP discovery

There are many different methods of SNP detection available to molecular ecologists studying non-model organisms. Broadly these can be divided into two categories; (i) sequencing of targeted individual genomic regions and (ii) random sequencing of genomic regions, followed by identification of segregating SNPs. The two strategies are complementary, rather than competing and we have used both approaches in the course of our research.

EPIC and related approaches

The acronym EPIC refers to Exon-Priming, Intron-Crossing primers which are used to PCR amplify intronic regions of genes (Fig. 1). The idea behind the approach is that, in the absence of sequence data for the focal organism, PCR primers can still be designed by performing sequence alignment of exonic sequences from other species in the same or related taxa. If the two primers are designed in adjacent exons, then the intron they flank can be amplified and sequenced (usually bi-directionally by capillary sequencing). This approach to SNP discovery in non-model organisms is a reliable method that can be employed in most taxa (Aitken et al. 2004; Lyons et al 1997; Palumbi 1996). The EPIC method works best when DNA sequence data are available for protein-coding regions of genes from organisms closely related to the focal species (for recent examples see Cappuccio et al. 2006; Elfstrom et al. 2006; Morin et al. 2007; for a bioinformatics pipeline applicable to plants see Fredslund et al 2006). There are several variants on the EPIC approach, including amplification of exonic rather than intronic sequence (Elfstrom et al. 2007; Ryynanen and Primmer 2004). The main reason to sequence exons is that nonsynonymous substitutions can be identified, and these are potentially functionally important. The main disadvantage of sequencing exons is that they tend to have lower levels of nucleotide diversity than intronic sequence, as they are typically under greater functional constraint. Thus, the decision on whether to sequence exons or introns may be
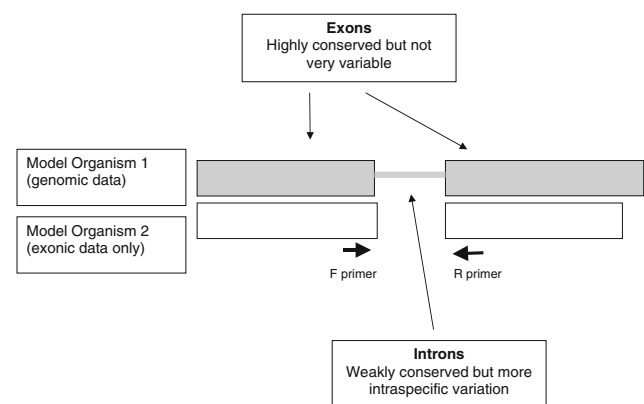

Fig. 1 SNP detection using EPIC primers. *Note*: DNA sequence data from a specified gene are aligned for two (or more) organisms related to the focal species. Here, intronic and exonic sequence has been obtained from organism 1, but only exonic sequence from organism 2. Exons that flank an intron of sequencable length (e.g. 500-800 bp) are identified, and regions of high sequence conservation are chosen for primer design. The rationale is that primers in these regions have the maximal chance of amplification success in the focal species. The intron (which is assumed to be of similar size in the focal species) is amplified by PCR and, following PCR product purification, sequenced in both directions in a small number of individuals. Putative SNPs are identified in the amplified intronic sequence and validated in a larger number of individuals

determined by the initial questions that are being addressed (exonic SNPs may be preferred if hypotheses specific to that gene are being investigated; intronic SNPs may be preferred if a suite of neutral markers are required for e.g., constructing a linkage map). One interesting variant on the EPIC strategy was employed in salmon by Ryynanen and Primmer (2006) who designed primers in intronic regions to amplify exons i.e. IPEC primers. The reason for adopting this 'reverse strategy' is that salmonid genomes contain large numbers of duplicated genes, and by designing primers in less conserved introns the problem of non-specific amplification of target genes was reduced.

The EPIC approach is now beginning to be employed specifically for gene mapping projects, both for genome scans, and studies that focus on specific genes. For genome or chromosome wide scans, genomic resources in closely related model species can be used to design primers that are approximately evenly spaced (based on their predicted locations in the related model organism). Thus, a panel of SNPs can be identified that are suitable for linkage map construction; an approach that has most notably been employed in studies of wild passerine bird populations, where the sequenced chicken genome can be used as a comparative genomics reference (Backström et al. 2006a, 2008; Hale et al. 2008).

EPIC has also been employed for designing SNPs in genes that are the focus of a candidate gene study. For example, Gratten et al. (2007) identified SNPs in five genes regarded as candidates for a coat colour polymorphism in a free-living population of Soay sheep. SNPs were initially identified in intronic regions, because it was reasoned that they would be most prevalent in introns of each candidate, and would likely be in linkage disequilibrium (LD) with causative mutations whether they were coding or regulatory. An intronic SNP in the gene Tyrosinase related protein 1 (*Tyrp1*) was found to be associated with coat colour, and linkage mapping of the gene confirmed it co-localised to the coat colour locus, which had been mapped to a small region of sheep chromosome 2 with a genome-wide panel of microsatellites (Beraldi et al. 2006). Subsequent sequencing of the *Tyrp1* coding region identified a non-synonymous substitution in a highly conserved site, which is probably causative for the polymorphism (Gratten et al. 2007).

In summary, SNP discovery by EPIC sequencing is a reliable method that can be used to target specific gene regions. The main disadvantage of this approach is that it is a relatively laborious method as each locus has to be investigated individually.
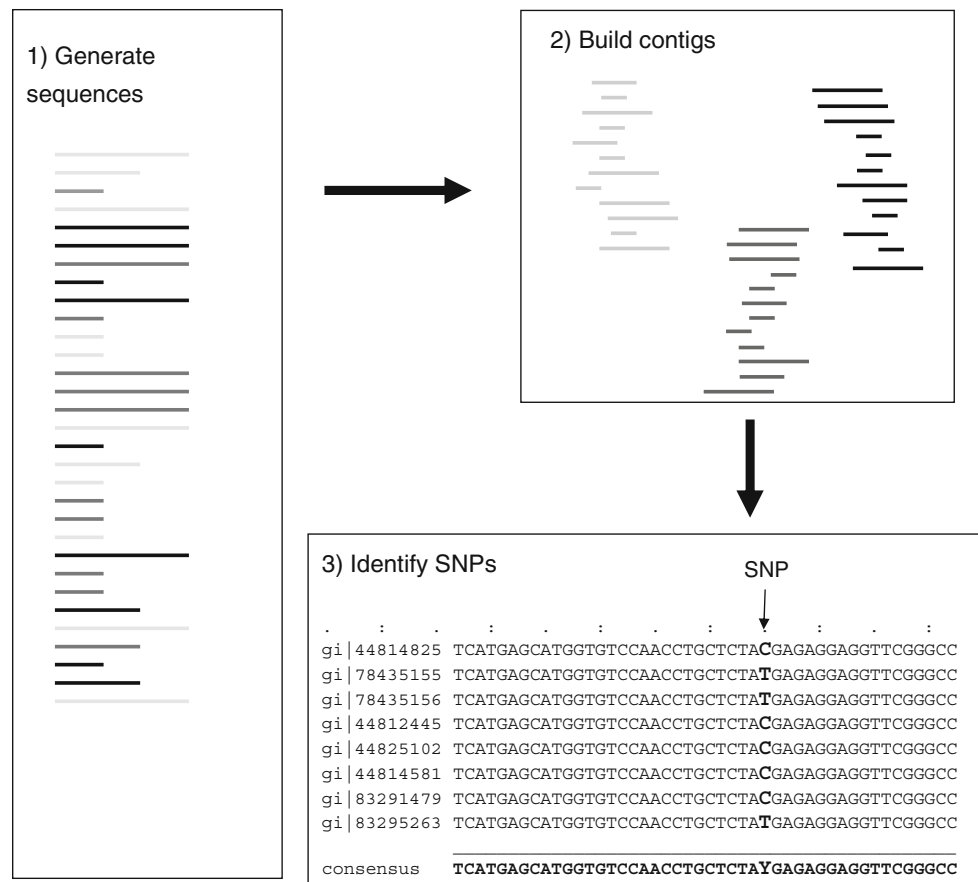
Random sequencing

The second main approach to SNP discovery involves sequencing of random genomic fragments in a limited number of individuals, followed by SNP discovery and validation. Some success has been achieved by capillary sequencing of random clones from genomic DNA libraries. For example, Rosenblum et al. (2007) identified 158 SNPs in a lizard species while Lin et al. (2007) used a similar approach to identify >40 SNPs in a bird species. A related approach was adopted by Adams et al. (2006) who identified SNPs in clones that were originally sequenced as part of a microsatellite library construction.

One method with the potential to rapidly generate large numbers of SNPs is to examine existing expressed sequence tag (EST) databases for putative SNPs. Provided sufficient numbers of EST are available for redundant sequences to be aligned, it is possible to identify SNPs in silico using a number of different computer programs such as PolyBayes (Marth et al. 1999), AutoSNP (Barker et al. 2003), SNPDetector (Zhang et al. 2005), PolyScan Chen et al. (2007) and QualitySNP (Tang et al. 2006). This approach to SNP discovery has been used in humans (Irizarry et al. 2000), model organisms (Fahrenkrug et al. 2002; Schmid et al. 2003; Stone et al. 2002) and more recently in non-model systems such as polar leaf rust (Feau et al. 2007) and *Bicyclus* butterflies Beldade et al. (2006).

EST libraries are currently unavailable for most of the species that are the focus of pedigree-based longitudinal population studies. However, the advent of ultrahigh-throughput sequencing technologies, such as 454 pyrosequencing (Hudson 2008; Mardis 2008; Margulies et al. 2005) makes SNP discovery feasible in any species. Here, the idea is that many genomic regions can be sequenced at very high coverage, usually by outsourcing to a service provider. Following contig assembly, it is then possible to identify SNPs from overlapping sequences (Fig. 2). The high-throughput sequencing method can be carried out on complementary DNA (cDNA) synthesised from messenger RNA by reverse transcriptase, in which case SNPs within the transcriptome will be identified (Vera et al. 2008), or from genomic DNA such that SNPs from all over the genome will be reported. The key to identifying SNPs is that several individuals must be sequenced with high sequence redundancy to detect segregating sites. For example, if we assume that a single run of a 454 sequencer can generate 400,000 sequences of 250 bp, and that an organism has a genome of 2 Gbp of which 100 Mbp is transcribed, then 10 runs would produce 10-fold coverage of the transcriptome, but 200 runs would be required to achieve similar coverage of the genome. At current costs of ∼€12 k per run, 10 runs is within the budget of a medium-large research grant, but 200 runs is probably not. That said, sequencing costs continue to fall very rapidly, and other technologies have even greater throughput than the 454 GS-FLX system, and offer great potential for SNP discovery if reference genomes are available (e.g. Van Tassell et al. 2008). We have used 454

Fig. 2 SNP discovery process with 454 sequence data. *Note*: Many thousands of short sequence reads are obtained from a small number of individuals. Overlapping sequences are identified and aligned computationally, to build contigs. Redundant contigs are then compared and aligned, and segregating SNPs are identified


```
1) Generate
   sequences
```

```
2) Build contigs
```

```
3) Identify SNPs                          SNP
                                           ↓
       .    :    .    :    .    :    .    :    .    :    .    :
gi|44814825  TCATGAGCATGGTGTCCAACCTGCTCTACGAGAGGAGGTTCGGGCC
gi|78435155  TCATGAGCATGGTGTCCAACCTGCTCTATGAGAGGAGGTTCGGGCC
gi|78435156  TCATGAGCATGGTGTCCAACCTGCTCTATGAGAGGAGGTTCGGGCC
gi|44812445  TCATGAGCATGGTGTCCAACCTGCTCTACGAGAGGAGGTTCGGGCC
gi|44825102  TCATGAGCATGGTGTCCAACCTGCTCTACGAGAGGAGGTTCGGGCC
gi|44814581  TCATGAGCATGGTGTCCAACCTGCTCTACGAGAGGAGGTTCGGGCC
gi|83291479  TCATGAGCATGGTGTCCAACCTGCTCTACGAGAGGAGGTTCGGGCC
gi|83295263  TCATGAGCATGGTGTCCAACCTGCTCTATGAGAGGAGGTTCGGGCC

consensus    TCATGAGCATGGTGTCCAACCTGCTCTAYGAGAGGAGGTTCGGGCC
```

sequencing of cDNA and the QualitySNP pipeline (Tang et al. 2006) to identify several thousand SNPs in the zebra finch transcriptome (Stapley et al. 2008). Assays were designed for 1536 SNPs that were detected in silico, of which 1298 (84.5%) were confirmed, indicating that high conversion rates of putative SNP to scoreable segregating SNPs can be achieved. More importantly, this approach yields large numbers of SNPs rapidly and (per SNP) cheaply. Methods for SNP detection from 454 sequence data are still being refined as the short reads present a challenge to software designed for SNP detection from EST databases, especially if no reference genome assembly is available. However, software for detecting SNPs from short sequence reads is now appearing (e.g. Quinlan et al. 2008).

A comparison of the advantages and disadvantages of the alternative approaches to SNP detection are outlined in Table 1 We provide qualitative rather than actual estimates as the relative costs of EPIC-based methods tend to vary between laboratories, depending on the infrastructure. Furthermore, prices of high-throughput sequencing are changing so rapidly that any figure reported here will be redundant in the near future. Importantly, although high-throughput methods offer many advantages, there are some scenarios for which EPIC type approaches are preferable, most obviously when a small number of specific genes are

under investigation. Therefore, it remains useful to retain the capacity to identify SNPs even if the growing trend of outsourcing large-scale laboratory work to service providers continues to gain momentum.

Methods for SNP typing

In much the same way that SNP detection methods vary depending on the scale of the project, there are a large number of alternative SNP typing strategies, the relative suitability of which depends on the number of loci and individuals that require typing.

Methods used in our group

We have used three different methods of typing SNPs. Their relative merits are summarised in Table 2 We are likely to continue with two of these into the medium term future, but one we have abandoned on grounds of assay complexity and relative cost. The main considerations when planning a SNP typing experiment are the number of loci and individuals that need to be typed. One then needs to compare the relative merits of outsourcing the work or performing it in-house.

Table 1 Relative comparison of SNP discovery methods

	EPIC	Existing EST databases	454
Scale	Small	Variable	Large
Cost per SNP	High	Low (zero)	Low
Initial outlay	Low	Low	High
Outsourceable	No	No	Yes
Intronic or exonic	Intronic (usually)	Exonic	Exonic (usually)
Targeted regions	Yes	No	Not usually[a]

Cost, scale and outlay are relative and assume that SNP discovery will be carried out in-house. In silico SNP detection from EST databases is included, although this approach assumes an EST database for the focal organism already exists. In reality, this is unlikely for most species with pedigreed wild populations

[a] It is possible to pool PCR products from many targeted EPIC regions and sequence them on a single 454 run

Table 2 Comparison of three SNP typing methods

	SNP-SCALE	SNPlex	Illumina
Min loci	1	24	384
Max loci	$\sim$30	48	1536
Expertise required	Modest	Considerable	None
Equipment required	Sequencer	Sequencer, robotics, two laboratories	Beadstation, unless outsourced
In-house or Outsourced[a]	In-house	In-house	Outsourced
Total cost	Low	High	High
Cost per locus	Medium	Medium	Low
Comments	Very flexible	Technically tricky	Very rapid, excellent service

[a] It is possible to outsource SNPlex or to perform Illumina GoldenGate typing in-house, but these two platforms are most often regarded as in-house and outsourced methods respectively

For in-house SNP genotyping we have developed an allele-specific PCR-based method termed SNP-SCALE (Hinten et al. 2007) that uses locked nucleic acids (LNAs) at the 3'-SNP positions of primers to enhance allele specificity. This method does not require specialist equipment (we use the same capillary sequencer set-up for screening microsatellites, AFLPs, SNPs and DNA sequencing) and it is flexible in terms of the number of loci and individuals that can be typed. The SNP-SCALE method has recently been refined and extended such that multiplexing of 25–30 loci is now possible (Kenta et al. 2008).

An alternative medium-throughput typing technology is the Applied Biosystems SNPlex system (Tobler et al. 2005). SNPlex reactions involve two steps—allele specific oligonucleotide ligation (OLA), followed by PCR. Genotypes are resolved by electrophoresis, and upto 48 loci can be multiplexed. We attempted to type up to 64 putative SNPs in Soay sheep with SNPlex (Table 3) It was possible to design assays for 59 (92%) loci, of which 47 were assayed and 31 (66%) could be reliably scored. Thus, around 60% of loci identified in silico could be converted to useful genotype data, although the manufacturers claim conversion rates in excess of 80%. Of course, one should expect conversion rates to be lower in wild populations than in model organisms, although the considerable

genomic resources for closely related domestic sheep and cattle mean that conversion rates for Soay sheep may be higher than for most non-model organisms. We also found SNPlex to be a technically difficult method. It is recommended that the OLA and PCR steps are performed in different laboratories to avoid typing error/contamination, and the method is manually demanding without liquid-handling robotics. We also found that SNPlex performed poorly on our more degraded or low concentration samples, which may be a consideration to other researchers collecting samples in the field.

The third approach we have taken for SNP genotyping is to outsource genotyping to a service provider (in our case the GoldenGate platform provided by Illumina). This system, which requires specialist equipment (a 'beadstation'), is cost-effective and rapid provided large numbers of SNPs (384 or more) are typed. Goldengate uses allele-specific extension followed by PCR to assay SNPs. PCR products are bound to beads on a Sentrix® microarray which is then read by the beadstation. We have used this approach to type 1536 zebra finch SNPs identified in silico, of which 1298 (85%) could be typed. There was a 97% call rate among genotyped loci, 100% reproducibility and 99.8% Mendelian inheritance consistency (Stapley et al. 2008). These figures are slightly lower than the manufacturer's advertised benchmarks (93%

Table 3 Effect on IBD coefficient estimation of adding SNPs to a microsatellite map

Position[a]	Dataset	IBD mean[b]	IBD variance	Increase in variance relative to msats (%)	Proportion of theoretical maximum variance (%)[c]
1 (224.5 cM)	Microsats only	0.254	0.027		
	All loci	0.255	0.039	42.6	62.6
2 (249.5 cM)	Microsats only	0.253	0.036		
	All loci	0.257	0.052	45.4	83.7
3 (279.5 cM)	Microsats only	0.257	0.034		
	All loci	0.256	0.047	37.3	74.9

[a] Data are provided from three positions on Soay sheep chromosome 3 (see Fig. 3)

[b] Mean and variance of IBD coefficients between half-sibs in the Soay sheep mapping panel are reported

[c] The theoretical maximum variance is 0.0625. Adding SNPs to the map improves IBD estimation, and therefore power to detect QTL

conversion rate and 99.9% call rate), although some DNA samples in our mapping panel came from material known to be of low quality and/or quantity. Generally, one should expect that DNA obtained from natural populations will often be of lower quality than is typically used in studies from model organisms or human subjects, because sampling may have taken place in difficult conditions, there may have been a delay between sampling and extraction, DNA may have been archived in freezers for considerable periods and small amounts of material may have been sampled. Although it would be facile to make a direct comparison between our SNPlex and GoldenGate data, because different samples and loci were compared, we believe the data obtained with GoldenGate were of higher quality.

There are alternative methods for typing the numbers of loci that might be required for mapping studies. We do not have experience with these alternatives, so do not discuss them further. However, popular medium-throughput methods, many of which are offered by service providers, include the Beckman SNPStream (12 or 48-plex) platform Bell et al. (2002) and the Sequenom iPLex (up to 40-plex) assay, performed on the MassArray platform Buetow et al. (2001).

In summary, we find SNP-SCALE to be an excellent method when small-medium numbers of SNPs need to be typed (often in a large number of individuals), while outsourcing to providers of the GoldenGate platform works better for larger numbers of SNPs. Typically, large numbers (100 s) of SNPs might be typed when performing an initial linkage scan, while more modest numbers might be typed when performing association studies on a more limited number of genomic regions; see for example Gratten et al. (2007).

Analytical issues

Prior to the collection of SNP genotype data there are a number of analytical questions that need to be addressed.

How many SNPs are required to build genetic linkage maps (given the size of a mapping panel, the size of a genome and the marker density required)? How does one use SNP data to detect QTL by linkage mapping? How does one use SNP data to examine whether variation at candidate genes explains trait variation? Can adding SNPs to a map improve the power and resolution of QTL detection? In this last section we consider these questions, using empirical examples wherever possible. However, it must be remembered that SNPs have only been applied to mapping projects in a handful of natural populations to date, and further data are required before general conclusions can be reached.

Building linkage maps with SNPs

To date most linkage maps of wild populations have been constructed using microsatellites (Beraldi et al. 2006; Hansson et al. 2005; Slate et al. 2002), as their high levels of genetic variability mean they are informative about whether recombination has (or has not) occurred between markers during meiosis. Of course, SNPs are less variable and therefore a larger number of loci are required to construct linkage maps. However, simulation studies show that SNPs at 2 cM (and probably larger) intervals are able to produce robust and accurate linkage maps in typical pedigrees of wild populations (Slate 2008). Maps built entirely from SNPs have been used to map the Z chromosome of the collared flycatcher *Ficedula albicollis* (Backström et al. 2006a), while maps combining microsatellites and SNPs have been used to study the homologue of chicken chromosome 7 in various passerine birds (Hale et al. 2008). Although a higher density of SNPs than microsatellites is required to map genomes, this constraint is unlikely to be a problem as high-throughput typing becomes the norm. Furthermore, newer SNP typing technologies have error rates that are considerably lower than those of microsatellites (indeed, some SNP platforms have error rates lower

than the *mutation* rate of some microsatellites), and so map error or map inflation due to typing error are likely to be less problematic than is the case for microsatellites (Slate 2008). Resolution of map errors caused by genotyping error can be a frustrating and time-consuming process. Prior to performing SNP-based map construction it is certainly worth performing simulation studies to ensure that marker density will be sufficiently high to detect linkage between syntenic markers. For example, marker data segregating in a mapping panel can be simulated with predetermined variability and chromosomal positions, with software such as SimPed (Leal et al. 2005).

Does the addition of SNPs to microsatellite maps help detect QTL?

QTL mapping studies conducted in natural populations to date (Beraldi et al. 2007a, b; Slate et al. 2002) have used evenly spaced microsatellite markers, typically at low density (e.g. 15 cM intervals), and then employed a two-step variance component approach to QTL detection (George et al. 2000; Slate 2005). The first stage in this process is to estimate the proportion of alleles that are identical-by-descent (IBD) *at every genomic location* that is to be tested for a QTL e.g. at 2 cM intervals. This means that IBD coefficients are estimated from markers that may be some distance (5–10 cM) from the test location. The second step is to fit the IBD matrix as a random effect in an 'animal model' (a form of linear mixed model widely used in quantitative genetics). Studies to date have reported QTL of marginal genome-wide significance (Beraldi et al. 2007b; Slate et al. 2002). In principle, typing additional markers in a region can improve IBD estimates at putative QTL locations. These improved estimates should enhance the power to detect (or disprove) QTL, as well as providing more accurate estimates of QTL position and magnitude.

We have examined whether the typing of additional SNPs improves the accuracy of IBD estimation in general pedigrees. A linkage mapping study in the St Kilda population of Soay sheep has been conducted using a panel of 250 microsatellites (Beraldi et al. 2007a, b). Here, we examine the mean and variance of IBD coefficients at three genomic regions of sheep chromosome 3, with and without the addition of SNP markers (Fig. 3). Within the Soay sheep mapping panel most of the power to detect QTL comes from half-sibs as full-sibs are rare in this population. At any given location half-sibs are expected to have an IBD coefficient of either 0 or 0.5 (expected mean = 0.25), with a variance of 0.0625 (Almasy and Blangero 1998). However, when marker information is imperfect, IBD estimates will not be as low as 0 or as high as 0.5, but instead will be closer to the mean value of 0.25, expected when there is no
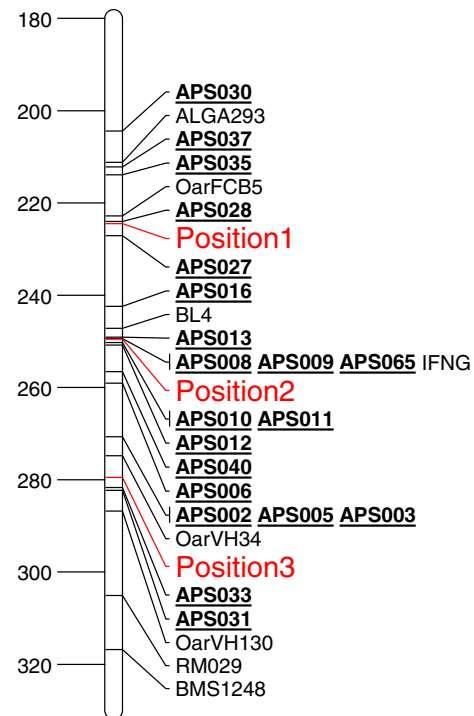

Fig. 3 Markers on Soay Sheep chromosome 3. *Note*: Microsatellite markers on the original (Beraldi et al. 2006) map are in plain font. SNPs subsequently added are in bold, underlined font. The mean and variance of IBD coefficient estimates were estimated at three positions (see Table 3)

marker information (Visscher and Hopper 2001). Therefore, the variance will be reduced relative to the theoretical maximum. By comparing the mean and variance of IBD coefficients between half-sibs with and without the addition of SNPs it is possible to measure the extent to which additional markers enhance IBD coefficient estimation (Table 3) Adding a modest number of SNPs increased the IBD coefficient variance by 37–45%, with the three locations yielding variances between 63 and 83% of the theoretical maximum. In this population the addition of SNPs in targeted locations, once a low marker density genome scan has identified putative QTL, appears to be a useful strategy. Note that this is in contrast to other QTL mapping strategies such as interval mapping in backcross or F2 populations created from divergent lines, where marker spacing less than 10 cM makes little difference to power (Darvasi et al. 1993; Piepho 2000).

How to model SNPs in mapping studies?

Linkage mapping in natural populations by the two-step variance components method outlined above involves fitting the estimated IBD matrix as a random effect in a mixed effects linear model (George et al. 2000; Slate 2005). The variance component associated with the IBD

matrix gives an estimate of QTL magnitude, and its statistical significance is assessed by likelihood ratio tests (by making a comparison to a model with the QTL random effect excluded). Several points are perhaps not immediately obvious until this type of analysis has been performed. First, QTL effects are only reported as a proportion of trait variation explained; mean trait values can not be assigned to individual alleles or genotypes. Second, only additive genetic effects at the QTL are estimated; this is in contrast to least squares linear regression or maximum likelihood approaches used in F2 crosses, where additive and dominance effects can be estimated (Haley and Knott 1992; Haley et al. 1994).

An alternative approach to detecting QTL with SNPs is to fit a SNP genotype as a fixed effect in a linear model (or in a mixed effects 'animal model' where polygenic effects are accounted for as a random effect). Although this approach is intuitively appealing (as the mean value of each genotype can be evaluated) it is highly prone to Type 1 error as population stratification can yield false associations between genotype and phenotype. One scenario where this approach may be justified is when a handful of candidate genes are being evaluated for linkage to a single locus trait, and associations can be tested by Fisher's Exact Test or other contingency table type tests. However, it is still preferable to confirm putative associations by linkage analysis (e.g. Gratten et al. 2007).

One feature of linkage analysis is that test locations need not be particularly close (>1 cM) to a causative mutation, yet they are still able to detect an association to a trait of interest. This is because linkage analysis is sensitive to recombination events between the marker and the causative locus within the mapping pedigree members only, while association studies are sensitive to historical recombination events that pre-date the pedigree. Although this means that low density marker coverage can detect linkage, it also means the confidence interval surrounding a QTL is wide. This problem can be remedied by typing additional SNPs around a candidate region and then performing association studies. This approach has rarely been taken in natural populations, although one exception is reported by Gratten et al. (2008). By performing transmission disequilibrium tests or TDTs (Hernandez-Sanchez et al. 2003), Gratten et al. simultaneously tested for linkage and linkage disequilibrium between a SNP and a locus (or loci) affecting both body size and lifetime fitness in Soay sheep. By testing for linkage as well as LD the problem of false positives in association mapping studies is remedied. Current methods for performing TDTs in general pedigrees are computationally quite demanding, although methods suited to this type of analysis continue to be refined (Chen and Abecasis 2006, 2007).

The future—whole genome association analyses in wild populations

Future methods

The advent of ultra-high throughput sequencing means that it will be possible to discover tens of thousands of SNPs in ecological organisms. In principle, and depending on the extent of linkage disequilibrium in the genome, it would then be possible to perform typing of many thousands of SNPs in a large enough number of individuals to perform whole-genome association mapping without first conducting linkage mapping. Several platforms now exist for typing thousands of SNPs on a chip (Gunderson et al. 2005; Hardenbol et al. 2005; Syvanen 2001). Although these have mostly been developed for humans (e.g. the Affymetrix GeneChip® 500 k array set, the Illumina Human1 M BeadChip) or model organisms (e.g. Illumina's canine SNP20 and bovine SNP50 Beadchips with >20,000 and >50,000 SNPs respectively), the technology can be used in any organism. Excitingly, both of the main providers of SNP chips offer the opportunity to develop customised panels (up to 60,000 SNPs on the Illumina Infinium iSelect platform) and up to 10,000 SNPs per kit, with the opportunity for construction of multiple kits, on the Affymetrix GeneChip® system. At present, the idea of typing 10 s or even 100 s of thousands of SNPs in wild populations may seem fanciful, but studies of this kind will shortly be upon us. For example, a 60 k domestic sheep SNP chip will be available in 2008, and preliminary data suggest that two thirds of the SNPs will be segregating in a wild Soay sheep population, which will likely be typed on this platform shortly.

How many SNPs for genome-wide association mapping?

When studies of linkage disequilibrium were first carried out in humans it soon became apparent that regions of high linkage disequilibrium (haploblocks) were prevalent throughout the genome (Goldstein 2001; Reich et al. 2001; Stephens et al. 2001; Weiss and Clark 2002); these blocks are usually separated by recombination hotspots. Typing many SNPs from the same haploblocks is redundant for genome-wide association scans, and so a better strategy is to type the minimal number of SNPs that describe the main haplotypes within each block (so-called tagSNPs). Considerable efforts have been taken to optimise strategies for tagSNP selection in humans (Carlson et al. 2004; Zhang et al. 2002), and the larger (250–500 K) SNP chip arrays have sufficient power to identify disease-causing variants by association mapping (Docherty et al. 2007). Work is underway to estimate LD in other organisms (Aerts et al.

2007; Heifetz et al. 2005; Morrell et al. 2005; Nordborg et al. 2002; Nsengimana et al. 2004; Remington et al. 2001; Sutter et al. 2004), which can then be used to estimate how many tagSNPs are required for genome-wide association mapping. For example, two estimates from dairy cattle suggest that just 30–100 k SNPs will suffice (Khatkar et al. 2007; McKay et al. 2007, mainly because LD extends long distances in cattle (Farnir et al. 2000).

If researchers studying wild populations are to conduct whole genome association studies using SNP chips then a first step is to measure the extent of linkage disequilibrium in the genomes of wild populations. Studies of this type are in their infancy (Backström et al. 2006b; Slate and Pemberton 2007), but are essential to evaluate how many tagSNPs are required to perform genome-wide association scans. The cost of studies of this type can be substantially reduced if pooling of individuals from extremes of a trait distribution can be performed and SNP allele frequencies estimated from the pools (Macgregor et al. 2008). One attractive feature of association studies is that pedigrees are not necessary, so potentially a larger number of wild populations will be amenable to this type of analysis.

Concluding remarks

SNPs are now being used for a number of different applications in molecular ecology research, including gene mapping. Advances in DNA sequencing and typing technologies mean that mapping studies are now feasible in any non-model organism for which adequate phenotypic or life history data are available. Indeed, the biggest challenge in gene mapping studies in the wild is the painstaking collection of field data, as there are no technology-driven shortcuts to this component of the work. In the next 5 years we expect to see more mapping projects being carried out in pedigreed wild populations, although we caution that moving from QTL detection to identification of the actual underlying gene or mutation will be very difficult. Therefore, researchers should carefully consider what they want to get from a mapping project before embarking on one. Simple detection of a QTL and reporting of its magnitude may not reveal much about fitness variation or microevolution in the wild. However, mapping does have the potential to build on the quantitative genetic studies conducted to date, including yielding a greater understanding of the architecture of genetic correlations and gene by environment interaction. Furthermore, if causative SNPs (or SNPs in near-perfect LD with a causative SNP) can be found, it will be possible to combine population genetic and quantitative genetic approaches to studying fitness variation, such that selection on underlying genotypes can be identified sensu Gratten et al (2008). These are exciting times for researchers studying the genetics of wild populations, and we eagerly await the findings of further mapping projects.

References

Adams RI, Hallen HE, Pringle A (2006) PRIMER NOTE. Using the incomplete genome of the ectomycorrhizal fungus Amanita bisporigera to identify molecular polymorphisms in the related Amanita phalloides. Mol Ecol Notes 6:218–220

Aerts J, Megens HJ, Veenendaal T, Ovcharenko I, Crooijmans R et al (2007) Extent of linkage disequilibrium in chicken. Cytogenet Genome Res 117:338–345

Aitken N, Smith S, Schwarz C, Morin PA (2004) Single nucleotide polymorphism (SNP) discovery in mammals: a targeted-gene approach. Mol Ecol 13:1423–1431

Almasy L, Blangero J (1998) Multipoint quantitative-trait linkage analysis in general pedigrees. Am J Hum Genet 62:1198–1211

Backström N, Brandström M, Gustafsson L, Qvarnström A, Cheng H et al (2006a) 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

Backström N, Qvarnström A, Gustafsson L, Ellegren H (2006b) Levels of linkage disequilibrium in a wild bird population. Biol Lett 2:435–438

Backstrom N, Fagerberg S, Ellegren H (2008) Genomics of natural bird populations: a gene-based set of reference markers evenly spread across the avian genome. Mol Ecol 17:964–980

Barker G, Batley J, O'Sullivan H, Edwards KJ, Edwards D (2003) Redundancy based detection of sequence polymorphisms in expressed sequence tag data using autoSNP. Bioinformatics 19:421–422

Beldade P, Rudd S, Gruber JD, Long AD (2006) A wing expressed sequence tag resource for Bicyclus anynana butterflies, an evo-devo model. BMC Genomics 7:130

Bell PA, Chaturvedi S, Gelfand CA, Huang CY, Kochersperger M et al (2002) SNPstream (R) UHT: ultra-high throughput SNP genotyping for pharmacogenomics and drug discovery. Biotechniques 70

Beraldi D, McRae AF, Gratten J, Slate J, Visscher PM 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

Beraldi D, McRae AF, Gratten J, Pilkington JG, Slate J et al (2007a) 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

Beraldi D, McRae AF, Gratten J, Slate J, Visscher P et al (2007b) Mapping QTL underlying fitness-related traits in a free-living sheep population. Evolution 61:1403–1416

Buetow KH, Edmonson M, MacDonald R, Clifford R, Yip P et al (2001) High-throughput development and characterization of a genomewide collection of gene-based single nucleotide polymorphism markers by chip-based matrix-assisted laser

desorption/ionization time-of-flight mass spectrometry. Proc Natl Acad Sci USA 98:581–584

Butlin RK (2008) Population genomics and speciation. Genetica (this issue). doi:10.1007/s10709-008-9321-3

Cappuccio I, Pariset L, Ajmone-Marsan P, Dunner S, Cortes O et al (2006) Allele frequencies and diversity parameters of 27 single nucleotide polymorphisms within and across goat breeds. Mol Ecol Notes 6:992–997

Carlson CS, Eberle MA, Rieder MJ, Yi Q, Kruglyak L et al (2004) Selecting a maximally informative set of single-nucleotide polymorphisms for association analyses using linkage disequilibrium. American Journal Of Human Genetics 74:106–120

Chen WM, Abecasis GR (2006) Estimating the power of variance component linkage analysis in large pedigrees. Genet Epidemiol 30:471–484

Chen WM, Abecasis GR (2007) Family-based association tests for genomewide association scans. Am J Hum Genet 81:913–926

Chen K, McLellan MD, Ding L, Wendl MC, Kasai Y et al (2007) PolyScan: an automatic indel and SNP detection approach to the analysis of human resequencing data. Genome Res 17:659–666

Darvasi A, Weinreb A, Minke V, Weller JI, Soller M (1993) Detecting marker-Qtl linkage and estimating Qtl gene effect and map location using a saturated genetic-map. Genetics 134:943–951

Docherty S, Butcher L, Schalkwyk L, Plomin R (2007) Applicability of DNA pools on 500 K SNP microarrays for cost-effective initial screens in genomewide association studies. BMC Genomics 8:214

Elfstrom CM, Smith CT, Seeb JE (2006) Thirty-two single nucleotide polymorphism markers for high-throughput genotyping of sockeye salmon. Mol Ecol Notes 6:1255–1259

Elfstrom CM, Smith CT, Seeb LW (2007) Thirty-eight single nucleotide polymorphism markers for high-throughput genotyping of chum salmon. Mol Ecol Notes 7:1211–1215

Ellegren H (2008) Sequencing goes 454 and takes large-scale genomics into the wild. Mol Ecol 17:1629–1631

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

Fahrenkrug SC, Freking BA, Smith TPL, Rohrer GA, Keele JW (2002) Single nucleotide polymorphism (SNP) discovery in porcine expressed genes. Anim Genet 33:186–195

Farnir F, Coppieters W, Arranz J-J, Berzi P, Cambisano N et al (2000) Extensive genome-wide linkage disequilibrium in cattle. Genome Res 10:220–227

Feau N, Bergeron M-J, Joly DL, Roussel F, Hamelin RC (2007) Detection and validation of EST-derived SNPs for poplar leaf rust Melampsora medusae f. sp. deltoidae. Mol Ecol Notes 7:1222–1228

Fredslund J, Madsen LH, Hougaard BK, Nielsen AM, Bertioli D, Sandal N, Stougaard J, Schauser L (2006) A general pipeline for the development of anchor markers for comparative genomics in plants. BMC genomics 7:207

George AW, Visscher PM, Haley CS (2000) Mapping quantitative trait loci in complex pedigrees: a two-step variance component approach. Genetics 156:2081–2092

Goldstein DB (2001) Islands of linkage disequilibrium. Nat Genet 29:109–111

Gratten J, Beraldi D, Lowder BV, McRae AF, Visscher PM 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 Biol Sci 274:619–626

Gratten J, Wilson AJ, McRae AF, Beraldi D, Visscher PM et al (2008) A localized negative genetic correlation constrains microevolution of coat color in wild sheep. Science 319:318–320

Gunderson KL, Steemers FJ, Lee G, Mendoza LG, Chee MS (2005) A genome-wide scalable SNP genotyping assay using microarray technology. Nat Genet 37:549

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

Haley CS, Knott SA (1992) A simple regression method for mapping quantitative trait loci in line crosses using flanking markers. Heredity 69:315–324

Haley CS, Knott SA, Elsen JM (1994) Mapping quantitative trait loci in crosses between outbred lines using least squares. Genetics 136:1195–1207

Hansson B, Åkesson M, Slate J, Pemberton JM (2005) Linkage mapping reveals sex-dimorphic map distances in a passerine bird. Proc R Soc B Biol Sci 272:2289–2298

Hardenbol P, Yu F, Belmont J, MacKenzie J, Bruckner C et al (2005) Highly multiplexed molecular inversion probe genotyping: over 10, 000 targeted SNPs genotyped in a single tube assay. Genome Res 15:269–275

Heifetz EM, Fulton JE, O'Sullivan N, Zhao H, Dekkers JCM et al (2005) Extent and consistency across generations of linkage disequilibrium in commercial layer chicken breeding populations. Genetics 171:1173–1181

Hernandez-Sanchez J, Visscher P, Plastow G, Haley C (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

Hinten GN, Hale MC, Gratten J, Mossman JA, Lowder BV et al (2007) SNP-SCALE: SNP scoring by colour and length exclusion. Mol Ecol Notes 7:377–388

Hudson ME (2008) Sequencing breakthroughs for genomic ecology and evolutionary biology. Mol Ecol Resour 8:3–17

Irizarry K, Kustanovich V, Li C, Brown N, Nelson S et al (2000) Genome-wide analysis of single-nucleotide polymorphisms in human expressed sequences. Nat Genet 26:233–236

Jarne P, Lagoda PJL (1996) Microsatellites, from molecules to populations and back. Trends Ecol Evol 11:424–429

Kenta T, Gratten J, Hinten GN, Slate J, Butlin RK et al (2008) Multiplex SNP_SCALE: a cost:effective medium-throughput SNP genotyping method. Mol Ecol Resour (in press)

Khatkar MS, Zenger KR, Hobbs M, Hawken RJ, Cavanagh JAL et al (2007) A primary assembly of a bovine haplotype block map based on a 15, 036-single-nucleotide polymorphism panel genotyped in Holstein-Friesian cattle. Genetics 176:763–772

Kruuk LEB (2004) Estimating genetic parameters in natural populations using the 'animal model'. Philos Trans R Soc Lond B Biol Sci 359:873–890

Kruuk LEB, Hill WG (2008) Introduction. Evolutionary dynamics of wild populations: the use of long-term pedigree data. Proc R Soc B Biol Sci 275:593–596

Kruuk LEB, Slate J, Wilson AJ (2008) New answers for old questions: the evolutionary quantitative genetics of wild animal populations. Annu Rev Ecol Evol Syst. doi:10.1146/annurev.ecolsys.39.110707.173542

Leal SM, Yan K, Müller-Myhsok B (2005) SimPed: a simulation program to generate haplotype and genotype data for pedigree structures. Hum Hered 60:119

Lin R-C, Yao C-T, Lo W-S, Li S-H (2007) Characterization and the broad cross-species applicability of 20 anonymous nuclear loci isolated from the Taiwan Hwamei (Garrulax taewanus). Mol Ecol Notes 7:156–159

Luikart G, England PR, Tallmon D, Jordan S, Taberlet P (2003) The power and promise of population genomics: from genotyping to genome typing. Nat Rev Genet 4:981–994

Lyons L, Laughlin T, Copeland NG, Jenkins NA, Womack JE et al (1997) Comparative anchor tagged sequences (CATS) for integrative mapping of mammalian genomes. Nat Genet 15:47–56

Macgregor S, Zhao ZZ, Henders A, Nicholas MG, Montgomery GW et al (2008) Highly cost-efficient genome-wide association studies using DNA pools and dense SNP arrays. Nucl Acids Res 36(6):e35

Mardis ER (2008) The impact of next-generation sequencing technology on genetics. Trends Genet 24:133

Margulies M, Egholm M, Altman WE, Attiya S, Bader JS et al (2005) Genome sequencing in microfabricated high-density picolitre reactors. Nature 437:376–380

Marth GT, Korf I, Yandell MD, Yeh RT, Gu Z et al (1999) A general approach to single-nucleotide polymorphism discovery. Nat Genet 23:452

McKay SD, Schnabel RD, Murdoch BM, Matukumalli LK, Aerts J et al (2007) Whole genome linkage disequilibrium maps in cattle. BMC Genetics 8:74

Merilä J, Sheldon BC, Kruuk LEB (2001) Explaining stasis: microevolutionary studies in natural populations. Genetica 112:199–222

Morin PA, Luikart G, Wayne RK (2004) SNPs in ecology, evolution and conservation. Trends Ecol Evol 19:208–216

Morin PA, Aitken NC, Rubio-Cisneros N, Dizon AE, Mesnick S (2007) Characterization of 18 SNP markers for sperm whale (Physeter macrocephalus). Mol Ecol Notes 7:626–630

Morrell PL, Toleno DM, Lundy KE, Clegg MT (2005) Low levels of linkage disequilibrium in wild barley (Hordeum vulgare ssp spontaneum) despite high rates of self-fertilization. Proc Natl Acad Sci USA 102:2442–2447

Nielsen R (2005) Molecular signatures of natural selection. Annu Rev Genet 39:197–218

Nordborg M, Borevitz JO, Bergelson J, Berry CC, Chory J et al (2002) The extent of linkage disequilibrium in Arabidopsis thaliana. Nat Genet 30:190–193

Nsengimana J, Baret P, Haley CS, Visscher PM (2004) Linkage disequilibrium in the domesticated pig. Genetics 166:1395–1404

Palumbi S (1996) Nucleic acids II: the polymerase chain reaction. In: Hillis D, Moritz C, Mable B (eds) Molecular systematics. Sinauer, Sunderland, Massachusetts, pp 205–247

Piepho HP (2000) Optimal marker density for interval mapping in a backcross population. Heredity 84:437–440

Quinlan AR, Stewart DA, Stromberg MP, Marth GT (2008) Pyrobayes: an improved base caller for SNP discovery in pyrosequences. Nat Methods 5:179

Reich DE, Cargill M, Bolk S, Ireland J, Sabeti PC et al (2001) Linkage disequilibrium in the human genome. Nature 411:199–204

Remington DL, Thornsberry JM, Matsuoka Y, Wilson LM, Whitt SR et al (2001) Structure of linkage disequilibrium and phenotypic associations in the maize genome. Proc Natl Acad Sci USA 98:11479–11484

Rogers SM, Bernatchez L (2005) Integrating QTL mapping and genome scans towards the characterization of candidate loci under parallel selection in the lake whitefish (Coregonus clupeaformis). Mol Ecol 14:351–361

Rosenblum EB, Belfiore NM, Moritz C (2007) Anonymous nuclear markers for the eastern fence lizard, Sceloporus undulatus. Mol Ecol Notes 7:113–116

Ryynanen HJ, Primmer CR (2004) Primers for sequence characterization and polymorphism detection in the Atlantic salmon (Salmo salar) growth hormone 1 (GH1) gene. Mol Ecol Notes 4:664–667

Ryynanen HJ, Primmer CR (2006) Single nucleotide polymorphism (SNP) discovery in duplicated genomes: intron-primed exon-crossing (IPEC) as a strategy for avoiding amplification of duplicated loci in Atlantic salmon (Salmo salar) and other salmonid fishes. BMC Genomics 7:192

Schmid KJ, Sorensen TR, Stracke R, Torjek O, Altmann T et al (2003) Large-scale identification and analysis of genome-wide single-nucleotide polymorphisms for mapping in Arabidopsis thaliana. Genome Res 13:1250–1257

Slate J (2005) QTL mapping in natural populations: progress, caveats and future directions. Mol Ecol 14:363–379

Slate J (2008) Robustness of linkage maps in natural populations: a simulation study. Proc R Soc B Biol Sci 275:695–702

Slate J, Pemberton JM (2007) Admixture and patterns of linkage disequilibrium in a free-living vertebrate population. J Evol Biol 20:1415–1427

Slate J, Visscher PM, MacGregor S, Stevens D, Tate ML et al (2002) A genome scan for quantitative trait loci in a wild population of red deer (Cervus elaphus). Genetics 162:1863–1873

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

Stephens JC, Schneider JA, Tanguay DA, Choi J, Acharya T et al (2001) Haplotype variation and linkage disequilibrium in 313 human genes. Science 293:489–493

Stone RT, Grosse WM, Casas E, Smith TPL, Keele JW et al (2002) Use of bovine EST data and human genomic sequences to map 100 gene-specific bovine markers. Mamm Genome 13:211–215

Sutter NB, Eberle MA, Parker HG, Pullar BJ, Kirkness EF et al (2004) Extensive and breed-specific linkage disequilibrium in Canis familiaris. Genome Res 14:2388–2396

Syvanen AC (2001) Accessing genetic variation: genotyping single nucleotide polymorphisms. Nat Rev Genet 2:930–942

Tang J, Vosman B, Voorrips R, van der Linden CG, Leunissen J (2006) QualitySNP: a pipeline for detecting single nucleotide polymorphisms and insertions/deletions in EST data from diploid and polyploid species. BMC Bioinformatics 7:438

Tobler AR, Short S, Andersen MR, Paner TM, Briggs JC et al (2005) The SNPlex genotyping system: a flexible and scalable platform for SNP genotyping. J Biomol Tech 16:398–406

Van Tassell CP, Smith TPL, Matukumalli LK, Taylor JF, Schnabel RD et al (2008) SNP discovery and allele frequency estimation by deep sequencing of reduced representation libraries. Nat Methods 5:247

Vera JC, Wheat CW, Fescemyer HW, Frilander MJ, Crawford DL et al (2008) Rapid transcriptome characterization for a nonmodel organism using 454 pyrosequencing. Mol Ecol 17:1636–1647

Visscher PM, Hopper JL (2001) Power of regression and maximum likelihood methods to map QTL from sib-pair and DZ twin data. Ann Hum Genet 65:583–601

Weiss KM, Clark AG (2002) Linkage disequilibrium and the mapping of complex human traits. Trends Genet 18:19–24

Zhang K, Calabrese P, Nordborg M, Sun FZ (2002) Haplotype block structure and its applications to association studies: power and study designs. Am J Hum Genet 71:1386–1394

Zhang J, Wheeler DA, Yakub I, Wei S, Sood R et al (2005) SNPdetector: a software tool for sensitive and accurate SNP detection. PLoS Comput Biol 1:e53