ANNALS of THE NEW YORK ACADEMY OF SCIENCES

Challenges and prospects in genome-wide QTL mapping of standing genetic variation in natural populations

Journal:	*Annals of the New York Academy of Sciences*
Manuscript ID:	Draft
Manuscript Type:	Review paper
Date Submitted by the Author:	n/a
Complete List of Authors:	Schielzeth, Holger; Bielefeld University, Evolutionary Biology Husby, Arild; Norwegian University of Science and Technology, Department of Biology
Keywords:	quantitative trait locus, complex trait analysis, genetic architecture, standing genetic variation, genotype-phenotype map

Challenges and prospects in genome-wide QTL mapping of standing genetic variation in natural populations

Holger Schielzeth[1] **& Arild Husby**[2]

[1] Department of Evolutionary Biology, Bielefeld University, Bielefeld, Germany

[2] Centre for Biodiversity Dynamics, Department of Biology, Norwegian University of Science and Technology, Trondheim, Norway

Short title: QTL mapping in natural populations

Word count of abstract: 161

Words count of main text: 6,950

Number of figures and tables: 1 Figure, 4 tables

Number of references: 202

Address for correspondence:

Holger Schielzeth
Department of Evolutionary Biology
Bielefeld University
Morgenbreede 45
33615 Bielefeld, Germany
Phone: +49-521-106 2820
Email: holger.schielzeth@uni-bielefeld.de

Abstract

A considerable challenge in evolutionary genetics is to understand the genetic mechanisms that can facilitate or impede evolutionary adaptation in natural populations. For this we need to understand the genetic loci contributing to trait variation and the selective forces acting them. The decreased costs and increased feasibility of obtaining genotypic data on a large number of individuals have greatly facilitated gene mapping in natural populations. Here we review the methods available to evolutionary biologists interested in dissecting the genetic basis of traits in study populations that are typically outbred. An exciting prospect offered by the technological advance is the possibility to study organisms that have historically been difficult to study in genetic terms, but are now within reach. These new opportunities should open up much needed information on the genetics of complex traits in a wider taxonomic context. We present an overview of the current state of research in the field and draw parallels to studies on crops, livestock and humans.

Keywords: complex trait analysis; genetic architecture; genotype-phenotype map; pedigreed wild populations; quantitative trait locus; standing genetic variation;

1. Mapping standing genetic variation in natural populations

Adaptive evolution is based on selection acting on genetic variants segregating within populations. The statistical description of segregating genetic variation using the tools of quantitative genetics[1] has been largely successful in predicting the response to selection in animal and plant breeding.[2,3] However, quantitative genetics make the simplifying assumption that allelic effects are small and numerous (the infinitesimal model)[1] and while this assumption is in many cases sufficiently robust to make predictions and inferences,[4] it is not fully accurate because loci vary in their effects on trait variation. Moreover, quantitative genetics does not address the identity of the genes involved, the distribution of effect size across loci, the interactions among loci and whether the segregating polymorphisms are regulatory or structural. These are all relevant aspects of the genetic architecture and quantitative trait locus (QTL) mapping offers to shed light on them.

The first applications of QTL mapping date back to the early 20th century when Payne[5] and Sax[6] used monogenic traits as markers for mapping bristle number in *Drosophila* and seed weight size in common bean, respectively. The start of the era of genome-wide QTL scans is marked by a seminal paper by Lander and Botstein in 1989, in which they present the theoretical framework for interval mapping.[7] Since then the possibilities for identifying causal genetic variants have substantially increased, facilitated by the plummeting costs of genotyping and the increased availability of genome sequences. Genetic mapping has been widely used by human geneticists for mapping disease susceptibility and other traits[8-11] as well as by livestock and crop breeders to improve breeding responses.[2,3,12] QTL mapping has also been used extensively in model organisms such as *Drosophila*, mice and *Arabidopsis* but these fields have been reviewed thoroughly.[13-15] Here we here focus on QTL mapping in natural population of non-model organisms, which is a much more recent endeavor.

QTL mapping in natural population contributes to our knowledge about the within-population standing genetic variation (SGV) affecting trait variation (hereafter SGV is used in this sense of being connected to specific phenotypes). A focus on SGV is particularly relevant because a deeper understanding of SGV can lead to a better understanding of microevolutionary dynamics. First, the number and distribution of loci gives an indication of the variants that are readily available for adaptation.[16] Second, knowledge about QTL can be used for studying complex interactions such as pleiotropic effects, genotype-by environment interactions and sex-specific effects.[17] Third, known QTL can be used for studying selection on genetic variants under natural conditions. Selection analyses in different environments allows distinguishing between antagonistic pleiotropy and conditional neutrality of genetic variants in different environments.[18] Finally, QTL mapping can help identify yet unknown developmental, physiological and biochemical pathways and therefore serves as a hypothesis generating enterprise for further analyses.[17]

Gene mapping in outbred non-model organisms have been hampered by the lack of genetic tools (in particular availability of markers) but due to technological advances have seen a growing interest during the last few years. Very few studies date back to the 1990ies[19] and a manageable number till 2003[20], but numbers have rapidly increased since. In contrast to studies on model organisms, there are challenges and constraints specific to QTL mapping in natural, outbred populations, in particular that allele frequencies are unmanipulated.[20,21] This has the unfavorable consequence that power is reduced, because loci under selection are expected to show a U shaped distribution of allele frequencies with large effect variants being uncommon.[22]

QTL mapping is a classical top-down genetic approach that starts with the phenotype and aims to map genetic variants linked to phenotypic variation. We here focus on genome-wide scans for genetic variants, to distinguish QTL analysis (*sensu stricto*) from association mapping in *a priori* selected candidate genes. In some cases, good candidate genes of interest will be known from other (model)

species and these can be used in targeted association studies[23,24]. Such targeted candidate gene analyses have been successfully applied for the identification of causal genetic variants in natural populations (e.g. [25]). Large scale screens of candidate genes have also been used[26] and hold some promise for uncovering loci under selection. Nevertheless, candidate gene approaches are necessarily affected by the amount of prior knowledge and we therefore focus here on an unbiased forward search for trait loci using whole-genome scans.

Any genome-wide QTL scans require accurate and error-free genetic marker information distributed across the genome as well as high-quality phenotypic information on hundreds if not thousands of individuals. The association between genotypes and phenotypes is typically analyzed in a linear (mixed) model framework with phenotypes treated as a response variable and marker genotypes as predictors. There are a large variety of statistical tools that are discussed in detail elsewhere.[22,27,28] We here review genome-wide mapping of trait-specific SGV with a focus on study design. We concentrate on methods for QTL mapping in unmanaged, outbred populations, our focus being motived by our interest in understanding microevolutionary processes and adaptation. The overarching goal is to identify loci of evolution[29] and to study its dynamics in action. Because of this aim, we concentrate on approaches for direct mapping within a single population and only non-exhaustively cover mapping approaches focused on divergent populations, in particular those concerning inter-species cross.

2. Sampling design and mapping strategies

All QTL mapping approaches require that marker alleles are in linkage disequilibrium (LD) with causal genetic variants (trait loci) that influence the trait of interest. There are, however, important differences in how LD enters in the analysis. The most notably distinction can be made between linkage mapping, which exploits LD that runs in pedigrees, and association mapping (sometimes referred to as linkage

disequilibrium mapping), which exploits historical population-wide LD (Figure 1). Linkage and association mapping approaches can both be used for QTL detection in outbred populations and are therefore the focus of our review on mapping SVG.

However, linkage mapping is more frequently used for mapping in experimental crosses, a strategy that is efficient for detecting QTL, but gives only partial information about SGV. In order to highlight the specific challenges within what is commonly subsumed under one term, we discuss linkage mapping separately depending on whether LD is created by crossing divergent lines (which we call "Experimental linkage mapping") or using only naturally occurring LD in outbred pedigrees (which we call "Pedigree linkage mapping", Figure 1).

Experimental linkage mapping using line or population crosses

Linkage mapping in line crosses constitutes the oldest[5,6] and most widely used approach to QTL mapping. In experimental crosses, long-ranging LD between marker and trait loci is experimentally created in the mapping population.[17,22] The key advantage of line crosses is that the allele frequencies at marker and trait loci are equalized, which substantially increases power, because all meiosises are potentially informative for segregating QTL.[21] This is radically different from pedigree linkage mapping in unmanipulated outbred populations where many matings are uninformative because either marker or trait loci are homozygous in both parents.

Experimental linkage mapping requires parents that differ substantially in their allele frequencies, ideally fixed for alternative alleles at trait loci. There are two basic options for selecting parental lines that are suitable for mapping. One requires selective breeding and hence experimentally created lines ("Experimental linkage mapping using line crosses"), while the other uses naturally existing differences among divergent populations ("Experimental linkage mapping using population crosses").

Experimental linkage mapping using line crosses

Classic experimental linkage mapping is based on inbred lines that have been produced such that individuals from each line are (nearly) completely homozygous at both trait and marker loci.[17,22] The development of lines can be done by randomly capturing haplotypes from the base population in inbred lines either through self-fertilization or, somewhat less efficiently, by repeated mating among full-siblings from the same family. Alternatively, selection lines can be used in line crosses and involve targeted local inbreeding at trait loci, a strategy that is most efficient for variants of large effect.

The F1 offspring from a cross between two homozygous parental lines are heterozygous and linkage blocks are only broken up by recombination in the following generation. F1 individuals can be intercrossed to produce a F2 generation or can be backcrossed to one of the parental lines to produce a backcross (B1) generation. One option to increase resolution of the mapping is to use recombinant inbred lines (RILs) that are created by continuously selfing F1 or F2 individuals (typically for six generations).[30] RILs thus consist of (nearly) genetically identical individuals and each RIL captures a different set of recombination from the original cross. Even more complicated lines, such as near-isogenic lines (NIL), co-isogenic lines or chromosome substitution lines can be generated in some study systems.[15,17,22] While in principle such strategies could be used for mapping SGV in outbred populations, this is hampered by practical limitations in most non-model organisms and we refer to the literature on model species such as *Drosophila*[17,31] and *Arabidopsis*[15] for more information.

Each cross and each swarm of RILs captures only two haplotypes from the base population, which does not represent the genetic variance in the base population as a whole.[20,21,32] This can certainly be responsible for to the low reproducibility of QTL peaks in different crosses as in the case of *Arabidopsis*.[33] Nevertheless, line crosses can contribute to our knowledge about SGV. First, they can suggest loci that can be genotyped in the base population in order to study natural allele frequencies

post hoc. More directly, an experimental design targeting SGV may produce many inbred lines that can be crossed among each other, each targeting a subset of the alleles segregating in the population. But while multiple line crosses have been applied in model systems (partly with large joint efforts such as the Collaborative Cross in mice[34,35] or Multiparent Advanced Generation Inter-Crosses, MAGIC, in *Arabidopsis*[36]), we are not aware of any field based study on a non-model organism that has used a large number of inbred lines (but see population crosses below).

Selection lines may be more efficient for mapping large effect variants because they equalize allele frequencies in the cross specifically for variants that have responded to selection. However, the establishment of selection lines is time-consuming, prone to be affected by drift and specific to the trait under selection. As far as we are aware, selection lines generated from natural populations have not been used for QTL mapping in natural populations and are likely to be limited to very specific applications in the future.

The required number of phenotyped individuals in the mapping population is comparatively low, with a few hundred individuals for effect sizes in the order of 5% of the phenotypic variation.[14,32] Conversely, mapping resolution is also low, with typical confidence intervals larger than about 20 Mb.[34] Marker density required are about 100 times lower and the number of individuals in the mapping population about 10 times lower when compared to mapping in an outbred population.[34] While these numbers depend heavily on the specifics of the study system and cannot be taken at face value, they nevertheless give an impression of the difference in power and mapping resolution.

A potential problem when creating experimental lines is differential loss of genetic variants due to selection or stochastic processes (e.g. 77% loss in *Mimulus* RILs[37]). This might impair the possibility to draw conclusions about SGV. Furthermore, relating the QTL variance to the total phenotypic variance (as a standardized effect size) can be problematic because the phenotypic variance in the mapping

population is likely to be reduced due to environmental and genetic homogenization (but is often increased in species crosses[38]). It is therefore useful to relate the QTL variance to the variance in the parental generation.[20]

While inbred line crosses are frequently used in model systems,[13-15] there are relatively few studies that used line crosses for studying natural populations (Table 1), possibly because of the labor-intensive breeding process. Even if the list is non-exhaustive, it becomes clear that crosses between naturally divergent populations are more popular when studying non-model organisms.

Experimental linkage mapping using population crosses

Linkage mapping in population crosses is often the fastest and most efficient way to QTL mapping because population crosses capitalize on naturally existing genetic difference between populations. Local adaptation and drift have done the job of the experimenter. If populations are sufficiently diverged in allele frequencies at trait and marker loci, population crosses allow similar benefits as artificially created lines by raising minor allele frequencies in the mapping population to near 50%.

Population crosses provide insight into loci that have contributed to population divergence and thus only indirectly for SGV. The focus is shifted from contemporary microevolution to past processes of divergence and adaptation. Intraspecies crosses, i.e. crosses among populations of the same species, are closest to the goal of mapping SGV. Interspecies crosses are possible (and the distinction is somewhat arbitrary), but the longer the divergence time among populations, the less they are expected to tell us about contemporary SGV.

Inter-population and ecotype crosses have been conducted in a variety of non-model organisms (Table 1) and a related line of research is the study of loci contributing to domestication in crosses between domesticated organisms and their wild ancestors.[39,40] Most of the studies rely on a mapping population

derived from a single population cross, but more comprehensive studies are beginning to emerge at least in model systems such as *Arabidopsis*.[41-43]

Pedigree linkage mapping in outbred populations

Pedigree-based linkage analyses use segregation within pedigrees for mapping and are nowadays almost exclusively based on interval mapping.[7] Interval mapping offers the distinct advantage that there is no bias towards the most variable marker showing the strongest signal[44] and that QTL effect size and location can be separated.[45] The main statistical and computational tools have been developed in the early 1990ies[28,46-49]. Linkage maps are required for linkage mapping, because the genetic distance (recombination fraction) among markers is needed for estimating IBD probabilities between marker loci. Linkage maps have to be estimated from the segregation patterns of marker loci in a pedigree or line cross and are now available for a number of outbred species.[50,51]

Pedigree linkage mapping in natural populations is based on variance decomposition and involves two-steps.[52-54] First, marker genotypes, pedigree data and linkage information (from a linkage map) are used to estimate IBD probabilities with reference to the base population of founders. This results in a square (N x N, where N is the number of individuals) matrix **Q** of pairwise IBD sharing probabilities at a locus of interest. IBD probabilities can be estimated for arbitrary locations within the genome, provided that they are flanked by at least on marker on either side.

In a second step, the IBD sharing matrix **Q** is used in a linear mixed model to predict phenotypes. The model estimates the amount of variance V_Q explained by the **Q** matrix, which can be scaled by the total phenotypic variance in the population V_P to give the heritability at the putative QTL $h^2_Q = V_Q / V_P$. This ratio provides a naturally standardized effect size with reference to the base population. Because mapping is done in a pedigree, the model includes the additive genetic relatedness matrix **A**, which

describes the pairwise genome-wide IBD sharing probabilities. If the model is fitted without **Q**, the ratio of $h^2 = V_A / V_P$ gives the narrow-sense heritability.[22] **A** and **Q** describe IBD probabilities at different levels, which can be referred to as the global (or genome-wide) and local relatedness matrices, respectively. The two matrices also differ in that **A** is predicted from a pedigree, while **Q** is estimated from genotype data.

Pedigree-based linkage mapping can be conducted in general, multigenerational pedigrees, but also in fragmented pedigrees of multiple core families. Fragmented pedigrees are typical for studies on humans, but similar data structures may also be available in many natural animal populations. Particular statistical tools can be used for analyzing multifamily full-sib data,[55,56] but we here focus on mapping in general pedigrees. The precision of the QTL location estimate is determined by the number of recombination events and deep, well-connected pedigrees contain information on many meiosis per individual and are powerful for mapping, but a larger number of families in shallower pedigrees will be equally suitable.

QTL that are inferred from linkage mapping are characteristics of the founder population. Phenotypes of offspring merely contribute breeding value information for segregating genetic variants present in founders. If pedigree-based linkage mapping is applied in natural populations, it is therefore essential that the population of founders is representative for the base population as a whole. In some systems, for example many plant and fish species, it is possible to generate very large full-sib families and a single full-sib family can sometimes be used for QTL mapping in outbred populations (Table 2). However, the generality of the findings will then be limited to the pair of founders.

Importantly, linkage between marker alleles and trait locus allele can differ between families[32]: M^+T^+/M^-T^- segregating in one family and M^+T^-/M^-T^+ in another (where M^+/M^- are two marker alleles and T^+/T^- are

two trait locus alleles). This also implies that the analysis does not identify particular alleles that are associated with trait variation unlike in an association mapping approach.[57]

The reliance on segregation within a pedigree constitutes the greatest strength, but also the greatest weakness of linkage mapping. Markers are coinherited with trait loci even over large genomic distances, because recombination rates are typically low. Hence, linkage mapping has relatively large power for mapping variants on scales of a few dozen cM[17] and requires comparatively few markers (although higher marker density safeguards against misestimated IBD probabilities in cases of missing genotypes[57]). Unfortunately, these advantages trade off with a lack of resolution, which results in large QTL confidence intervals.[17,49] For example, the largest linkage mapping studies in a well-connected pedigree of c. 1,000 genotyped individuals found QTL peaks that cover 31 ± 16 cM (mean ± SD, range 9-68 cM), 59 ± 49 Mb (range 3-155 Mb) and 602 ± 370 annotated genes (range 53-1,209) in their ΔLOD=1 intervals.[58-60]

Pedigree linkage mapping was first applied to natural populations in 2002 for mapping birth weight in red deer[61] and this was the only study to be included in a review of QTL mapping in natural population from 2005.[20] The situation as substantially change in the last 10 years with a number of studies using this approach, mostly in long-lived species like birds and mammals that are less amenable to experimental linkage mapping (Table 3).

Association mapping

Genome wide association mapping is an extension of early (local) association studies.[62] Instead of mapping the trait in families as in linkage mapping, allele frequencies are compared at candidate loci with respect to the trait of interest. Genome-wide association studies (GWAS) take advantage of LD between a marker and trait loci that exists naturally within populations.[63] The statistical tools for

genome wide analyses were developed in the 1990ies[64,65] in connection with increased marker densities. GWAS have since become the standard tool for gene mapping in human genetics where they have identified mutations for a wide range of traits and diseases.[66]

More recently GWA studies are also starting to be employed in natural populations on Soay sheep[67], great tits[68] and lodgepole pine[69] (Table 3). Due to the increased ease to genotype for a large number of markers, this method is likely to supplement and possibly replace linkage mapping approaches in the near future also in ecological genetic studies. There are two reasons for this: firstly, GWAS bypasses the need to follow many individuals and their relatives over many generations as the analyses requires no information on recombination within a pedigree and second, GWAS offer both increased power and resolution compared to a linkage analysis.[70] The increased resolution is a direct consequence of utilizing historic recombination events accumulated over many generations. Thus linkage blocks are substantially smaller (typically in Kb instead of Mb) with the result that localizing a trait gene or even causative mutation is easier (though it is still by no means easy).[13,17]

Population-wide LD determines the probability that one or more of the markers is in LD with trait alleles. An important consideration is therefore how close we need to be to the causal variant and how many markers we need to have sufficient coverage of the entire genome. Technical improvements will make marker densities less of an issue in the future, but at the current state this is still a real problem. For example, for a typical bird genome of ~1.1 Gb in size[71] one will have on average one marker every 110 kb using 10,000 SNPs and with the largest SNP chip used so far in natural populations of non-model organisms[72] this will increase to one marker every 22 kb using a 50,000 SNP chip. Although a 22kb interval seems still large, comparable LD levels have been found in natural populations,[73] particularly if the effective population size is small.[74] However, many natural populations have large effective population sizes and a long evolutionary history and therefore LD levels are expected to be low[75] with

the consequence that tens of thousands of markers are needed to have sufficient coverage of the genome.

Detecting the effect of the markers on the phenotype can be tested using several different methods and in the simplest case of equally unrelated individuals and no population structure, single locus association scans are based on regressing phenotypes on marker genotypes each locus at a time. The fact that loci are tested one-by-one requires an appropriated type I error control. This level will depend on the effective number of tests carried out and can be estimated in a number of different ways, including stepwise Bonferroni and false-discovery rate control,[76,77] both of which can be overly stringent if some of the markers are in LD with each other.

The absence of cryptic relatedness and population structure is often unrealistic and naïve mapping can therefore lead to increased rates of false positives. Population stratification therefore need to be explicitly modeled to avoid spurious associations.[78] This problem can be especially problematic if both phenotypic and genetic differentiation varies with geographical distance. Several methods have been proposed to control for population stratification but common to them is that they rely on fitting the genomic kinship matrix in a mixed model framework.[79-81] The kinship matrix captures both population structure and cryptic relatedness in the sample and is therefore an efficient way to reduce false positive associations. By being marker-based, the kinship matrix estimates realized relatedness, but it can potentially be replaced by the expected relatedness matrix inferred from a well-connected pedigree.

Admixture mapping

Admixture mapping makes use of natural introgression in hybrid zones and, like association mapping, utilize naturally occurring, population-wide patterns of LD.[82,83] The analysis benefits from the increased LD and increased variation (genetic and phenotypic) in hybridizing populations with different degrees of

backcrossing.[83] An ideal mapping population for admixture mapping therefore harbors recent hybrids with far-ranging LD as well as advanced intercrosses or backcrosses that have accumulated recombinations over many generations. The difference in genetic composition between individuals potentially allows high-resolution QTL mapping with comparatively few markers as compared to association mapping.[83-85]

In admixture mapping marker information is used for estimating a hybrid index that describes the genome-wide degree of mixture among parental genomes for each individual.[83] The analysis contrasts the genome-wide hybrid index with mixture at individual loci. If the locus-specific degree of mixing is larger or smaller than introgression in the remainder of the genome, this is called excess admixture. The basic mapping model fits locus-wise excess admixture as a predictor for the phenotype of interest.[83,86] Similar to association mapping, linkage maps and an annotated genome are not required,, but they greatly assist in interpreting the findings.[82,83]

Admixture mapping is tailored to mapping in natural systems where interbreeding takes place between species, subspecies, ecotypes or any populations that are genetically sufficiently diverged from each other. The rather specific conditions of persistent admixture among sufficiently divergent parental lines make admixture mapping difficult to apply in many natural populations that do not hybridize. Nevertheless, admixture mapping has been successfully used for mapping variants for human diseases,[87-89] and in a few outbred non-human organisms (Table 3, with a few more example of mapping in interspecies hybrid zones[90,91]).

Chromosomal heritabilities

Even relatively well-powered association studies that have used large number of markers and individuals have often only managed to explain a small amount of the heritability. This is perhaps best exemplified

by human height where QTL from GWAS only explained around 5% of the heritability.[92] Yang et al. proposed to fit all markers simultaneously instead of testing the significance of markers individually.[93] This should provide an unbiased estimate of the variance explained by the sum of all trait loci linked to markers. Indeed, this method recovered 45% of the additive genetic variance in human height, with the remaining fraction most likely missing due to incomplete LD between markers and causative sites.[93]

The same idea can also be used to partition the genetic variance across individual chromosomes[94] and has recently been extended for use in ecological studies under high relatedness levels.[95] Under the infinitesimal model one would expect that larger chromosomes harbor more genetic variance than smaller chromosomes and thus that chromosome size should scale positively with proportion of genetic variance. This expectation is indeed fulfilled for many traits,[94] indicative of a polygenic basis, but there is also some variation around a linear relationship that suggests that, for some traits, some chromosomes contribute disproportionately.

Chromosomal heritabilities are a bit departed from mapping at the level of individual loci, but might still allow inferences about the genetic architecture and are therefore included here. Outlier chromosomes could potentially be interpreted as evidence for against a strictly polygenic model.[95,96] However, we would urge caution in using the relationship between chromosome size and proportion of variance explained to infer too much about the number of loci underlying trait variation. A disproportional contribution of a chromosome could be caused by a QTL of large effect, but it could also be due to the clustering of many loci of small effect on a single chromosome. Such clustering is not uncommon[43,97] and therefore even outlier chromosomes could be consistent with a polygenic model.

Combining linkage and association mapping

Association and pedigree linkage mapping are the most targeted approaches for studying SGV (Figure 1). As we have outlined above, the two approaches have different benefits and drawbacks, with a fundamental trade-offs between efficiency (in terms of marker density and sample size) and precision.[57] To take full advantage of the data and increase power, it is therefore desirable to combine linkage mapping and association mapping as they are complementary approaches with different strengths and weaknesses.[40]

Few studies have compared results from linkage and association mapping empirically for the same study population. A recent mapping study on clutch size and egg mass in a population of great tits found no genome wide significant regions were detected in either approach.[68] Moreover, and somewhat surprising, was that nominally significant QTL regions detected in the linkage analysis did not match up with those from the association analysis.[68] Similarly, a joint linkage and association mapping approach of flowering time in *Arabidopsis* found that, while many QTL from the linkage analysis and the GWAS did align, there were also a number of associations from the GWAS that were not present in the linkage analysis.[43]

The discrepancy between results is surprising and requires an explanation. The major difference between the two approaches is the difference in LD structure. First, it is possible that linkage signals are composed of multiple small effect QTL that individually are too weak to be detected by association mapping. Second, it is possible that rare variants of moderate effect are poorly marked in an association study, but show up as segregating with in families. However, if a trait locus is well marked by a marker locus, association mapping is more powerful by combining evidence across families.

A useful approach is therefore to combine linkage and association mapping as two confirmatory approaches (though not independent replication if based on the same dataset). For example, a disagreement might be caused by failure to control for population structure in an association study thereby causing a false positive.[43] When planning follow-up studies, it would be most promising to pursue associations that are identified by both approaches.

3. General challenges

Biased effect size estimation

A notoriously difficult issue is to obtain accurate and unbiased estimation of effect sizes in scans for QTL. Whenever QTL discovery and effect size estimation are conducted on the same dataset, such that the effect size estimation is conditional on significance thresholds, the estimates for the amount of variance explained by a QTL are on average biased upwards. The overestimation of effect sizes is known as the Beavis effect in the context of QTL mapping,[98,99] but applies to conditional effect size estimation in a more general sense.[100]

This overestimation is caused by effects near the detection limit, which makes conditional effect size estimation particularly problematic in underpowered studies. Effects near the detection limit reach statistical significance only if point estimates are comparatively large in the particular dataset at hand. Unfortunately, QTL scans are always working at the detection limit, because most QTL have small to very small effects.[101] The size of the confidence interval is wider in studies with low power, such that only truly large effects and small to moderate effects that are overestimated in the particular sample will yield a point estimated that is large enough so that the CI does not overlap zero. Unfortunately, using a single population sample it is impossible to determine if an estimated effect is truly large or if it was overestimated in the particular sample used.

Empirical data indeed shows a strong negative correlation between estimated effect size and sample size[22,96]. Sample size thresholds, above which the Beavis effect is deemed to be less of an issue, have been suggested (e.g. N > 300[21], N > 500[99]), but this is unlikely to be useful because the problem is continuous and even applies to conditional effect size estimation on a very small-scale.[100] Notably, when mapping in unmanipulated pedigrees, there is also internal heterogeneity in power, because of variation in allele frequencies and/or marker densities across the genome. The only sustainable solution is effect size estimation at a priori defined loci in an independent sample. Unfortunately, replication and accurate effect size estimation is particularly problematic in pedigree linkage analyses because of the difficulty in replicating the sampling design.

The overestimation of effect sizes has an intriguing and often overlooked consequence: Replication studies of similar sample size will tend to result in an inflated number of false-negative finding. The inability to replicate initially significant findings due overoptimistic expectations concerning effect sizes is called the winner's curse.[102,103] Hence, somewhat counter-intuitively, replication studies have to be designed larger than the initial study to avoid the risk of falsely rejecting a QTL.[104,105]

The need of replication

Confirmation of QTL signals is essential for establishing that associations are genuine and is considered the gold standard in human studies.[106,107] Replication is also important, because any fine-mapping is demanding in terms of time, money and labor and replication can therefore avoid wasting resources on spurious signals. Replication of a QTL signal could be done 1) using a different sample from the same population, 2) using a sample from a different population of the same species, 3) in a different species (comparative QTL mapping[108,109]) or 4) ultimately by demonstrating the mechanistic link by functional assays.[107] There are constraints on replication imposed by the study system. For example, a pedigree linkage analyses in natural population cannot be easily replicated in the same population, because

pedigree data often need years to be collected. A useful strategy that also generalizes the results is replication in a different but similar population.[107]

Replication of QTL results has proven difficult in humans.[9] Problems with reproducibility seem to stem mainly from four main issues:[9] Lack of control for population stratification and/or cryptic relatedness, differences in LD between marker and trait loci in different populations,[110] differences in genetic structure between populations and presence of genotype-by-genotype or genotype-by-environment interactions.[9] Since some of the reasons are rooted in differences among populations and are therefore of biological relevance, it is advisable to first replicate the analysis in a population that is similar to the discovery sample and it is important that the phenotype has been measured in a standardized way.[107]

Replication of QTL studies has so far been relatively rare in natural populations of non-model organisms, even though several studies contain internal replication in multiple independent samples.[111,112] An impressive demonstration of replication has been achieved in ecotype crosses of threespined sticklebacks, where a QTL for pelvic spine structures has been identified in a single cross[113], and then been replicated in the multiple populations from the same geographic region[114] and from different continents.[115,116] Another instructive case is the case of a candidate gene approach applied to personality traits in great tits. The *DRD4* gene was found to influence personality,[117] a result that was replicated in the same population but not in others.[118] This lack of replication has subsequently been shown not to be due to inter-population differences in LD between the marker and trait locus.[119]

Strategies for fine-mapping

Confidence regions for QTL signals are typically large, in particular in linkage analyses, and usually cover dozens or hundreds of genes. A better functional understanding of trait-specific SGV requires a more fine-scaled mapping of genetic variants to evaluate if a QTL is caused by a single locus of large effect or if

it constitutes the composite effect of multiple loci with small effect. It is not unusual for a single QTL to decompose into multiple small-effect loci, possibly even spatially offset from the original signal ('ghost QTL').[22,45,120]

Fine-mapping within pedigrees or line crosses requires a very large number of individuals, because LD blocks have to be broken up by recombination. The marginal gain of additional generations for linkage mapping decreases[32] and extending a pedigree is therefore only occasionally a promising option. A follow-up by association mapping can therefore be an attractive choice.[13] This requires a far larger number of markers, but not such a dramatic increase in sample sizes. The statistical power of association mapping can be further increased by large-scale phenotype screens with selective genotyping of extreme phenotypes or sequencing of the QTL region.[21]

A generally promising strategy is to combine QTL mapping with other approaches such as transcriptome profiling,[121] population genomics[51,122] or comparative genomic[123] approaches. Population genomics uses large scale genotyping or resequencing of individuals within populations to identify regions of the genome that are unusually differentiated, but does not focus on particular phenotypes and it can be difficult to separate outlier loci thought to be under selection from demographic effects.[51,124] Similarly, comparative genomics of divergent populations[123] can also help to identify outlier loci of divergence, but is again anonymous to specific phenotypes. QTL mapping is needed to bring in a phenotypic perspective and combining QTL mapping with population and comparative genomics can give evidence that a putative quantitative trait locus is under selection.[125,126]

Most successful studies that have mapped QTL to quantitative trait genes (QTG) have pursued QTL signals by positional searches for candidate genes in the QTL regions (Table 4). The success of such a positional candidate gene strategy[40] depends on the amount of knowledge from other species and is more likely to be successful if an annotated genome assembly is available and if the study species is

closely related to a model organism. Ultimately, the study of post-hoc candidate genes has similar drawbacks as the a priori selection of candidate genes, because even if causal polymorphisms are identified in the candidate, it will remain unclear if this is the only or even the main locus contributing to the initial QTL signal.

Some studies in natural populations have been successful in fine-mapping QTL to the level of a single QTG or even single nucleoid polymorphisms (quantitative trait nucleotide, QTN) (Table 4). Most of these fine-mapping successes have been supported by evidence from other approaches, including population genomics and functional analyses. Admittedly, most of the success stories concern traits with a rather simple genetic architecture. Nevertheless, they nicely demonstrate how QTL mapping can help to elucidate the genotype-phenotype map.

4. What can we learn from model systems?

Some lessons can be learned from the extensive experience with QTL mapping in humans, livestock and crops. A particularly striking and at first glance surprising fact is that, despite substantial efforts and sample sizes in the hundreds of thousands, the QTL that have been identified explain only a small amount of the genetic variance, a phenomena coined the 'missing heritability' mystery.[127] A good example is human height. The trait shows substantial heritable genetic variation that amounts to 80% of the phenotypic variants. Yet, even very large-scale association studies have identified about 180 loci that in sum explain only 10% of the phenotypic variance.[128-131] Such findings have led some authors to have a pessimistic view on the future of QTL mapping.[132,133]

Replication and fine-mapping has been moderately successful in model organisms and humans,[120,134] but the causal functional details of complex traits have remained largely unknown even in humans.[4,134] Furthermore, the results from studies from model organisms are ambiguous with respect to the sharing

of QTL across populations and species with shared QTL among some populations and species,[135] but not in others cases.[33]

The history of QTL studies in model organisms is characterized by widespread reports of large QTL in the initial phase, with smaller effect sizes and a more complex picture of quantitative genetic variation in later studies.[101,120] It seems likely that QTL studies on outbred population are in the process of repeating this history, which is indicated by trends expressed in recent reviews.[50,96] Hence, the field of ecological genomics might ultimately also realize that most quantitative traits are governed by large numbers of loci with small effect, while large effect variants are rare.[101] This observation appears remarkably valid across a wide range of traits and species,[34,120,134,136] even if exceptions do exist.[137] Hence, the infinitesimal model might be surprisingly valid and we cannot expect every QTL mapping effort to discover segregating large effect variants.

5. Outlook

As the costs of sequencing and genotyping continues to decrease,[138] it will become increasingly feasible to use resequencing based methods for QTL mapping. Resequencing will ensure that the causal variants are covered, which will solve the issue with low LD in association studies. However, it will be of little use when mapping in pedigrees, because linkage analyses are not limited by the linkage among markers, but by the lack of recombination. So far resequencing approaches have rarely been used for mapping SGV for fitness traits in natural populations but a notable exception is the detection of candidate genes for adaptation to serpentine soil in *Arabidopsis lyrata*.[139] More resequencing studies are on their way in other organisms and this should yield important insights into the role of other genetic variants such as insertions, deletions, inversions and transposable elements in influencing trait variation in natural populations. Resequencing approaches will also aid QTL identification indirectly, by boosting the

potential for complementary analyses using population genomicc and comparative genomic approaches.

Most QTL mapping studies in natural populations have focused on morphological and life-history traits that are comparatively easy to measure (see Tables 1-3). However, behavioral traits, such as mating and feeding rates, calling activities, aggression and exploration, would be equally interesting (see e.g.[140]), even if it is harder to collect sufficient behavioral data on hundreds or thousands of individuals. As such access to high-quality phenotypes will become highly valuable[57] and long-term studies are therefore likely to continue to play an important role in evolutionary genetics also in the future.[141] Because of the central role of behavioral variation in evolutionary studies of animal populations,[142] we expect to see more attempts of mapping behavioral traits in the future.

The use of linkage mapping and association mapping studies on natural population have successfully allowed the identification of loci important in adaption thereby providing greater insight into the mechanistic underpinnings of evolution. However, identifying the location of a QTL is in itself only the first step towards this goal. What is needed is a mechanistic link between the genotype, phenotype and fitness.[143,144] The paucity of functional knowledge about most loci, even in model organisms, represents a considerable obstacle in genotype-phenotype mapping. Two solutions have been suggested to remedy this situation: the construction of an ecological association ontology database (similar to the gene ontology database available for model organisms) and the use of more functional studies.[143]

It seems likely that the immediate next steps in gene mapping in ecological genomics will be one of scale: more markers and more individuals will be scored to try to find the elusive QTL of quantitative traits. A particularly enticing prospect of this endeavor is measuring selection on the level of the QTL[72,145-147] to better understand how traits can respond to selection and how genetic variance can be maintained in populations. This could be done either experimentally[146] or by measuring fitness of

individuals with known genotype.[145] Fortunately, even if the causal genes remain anonymous, selection analyses can be successfully conducted by studying selection on the closest marker locus.[148-151]

The advance in technology also means that a more diverse range of organisms can be studied, a process that will add important new knowledge about the genetic underpinnings of fitness related traits in natural populations. Hopefully such work will be pursued using a combination of approaches replicated across populations and followed, ultimately, by functional analyses and fitness assays. As more such studies accumulate it should allow for a deeper and more complete understanding of the molecular mechanisms responsible for adaptation.

Acknowledgements

HS was supported by an Emmy Noether fellowship from the German Research Foundation (DFG; SCHI 1188/1-1) and AH by the Norwegian Research council (214553).

References

1 Fisher, R.A. 1918. Correlation between relatives on the supposition of Mendelian inheritance. *Trans. R. Soc. Edinburgh* **52:** 399-433.

2 Collard, B.C.Y. & D.J. Mackill 2008. Marker-assisted selection: an approach for precision plant breeding in the twenty-first century. *Phil. Trans. R. Soc. B* **363:** 557-572.

3 Hospital, F. 2009. Challenges for effective marker-assisted selection in plants. *Genetica* **136:** 303-310.

4 Hill, W.G. 2010. Understanding and using quantitative genetic variation. *Phil. Trans. R. Soc. B* **365:** 73-85.

5 Payne, F. 1918. The effect of artificial selection on bristle number in *Drosophila ampelophila* and its interpretation. *Proc. Natl. Acad. Sci.* **4:** 55-58.

6 Sax, K. 1923. The association of size differences with seed-coat pattern and pigmentation in *Phaseolus vulgaris*. *Genetics* **8:** 552-560.

7 Lander, E.S. & D. Botstein 1989. Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics* **121:** 185-199.

8 Visscher, P.M. *et al.* 2007. Genome partitioning of genetic variation for height from 11,214 sibling pairs. *Am. J. Hum. Genet.* **81:** 1104-1110.

9 Hirschhorn, J.N., K. Lohmueller, E. Byrne & K. Hirschhorn 2002. A comprehensive review of genetic association studies. *Genet. Med.* **4:** 45-61.

10 Hirschhorn, J.N., K.E. Lohmueller, C.L. Pearce, M. Pike & E.S. Lander 2002. Meta-analysis of genetic association studies supports a role for common variants in common disease risk. *Am. J. Hum. Genet.* **71:** 360-360.

11 Lohmueller, K.E., C.L. Pearce, M. Pike, E.S. Lander & J.N. Hirschhorn 2003. Meta-analysis of genetic association studies supports a contribution of common variants to susceptibility to common disease. *Nat. Genet.* **33:** 177-182.

12 Hu, Z.L., C.A. Park, X.L. Wu & J.M. Reecy 2013. Animal QTLdb: an improved database tool for livestock animal QTL/association data dissemination in the post-genome era. *Nucleic Acids Res.* **41:** D871-D879.

13 Mackay, T.F.C. 2001. Quantitative trait loci in *Drosophila*. *Nat. Rev. Genet.* **2:** 11-20.

14 Flint, J. & E. Eskin 2012. Genome-wide association studies in mice. *Nat. Rev. Genet.* **13:** 807-817.

15 Bergelson, J. & F. Roux 2010. Towards identifying genes underlying ecologically relevant traits in *Arabidopsis thaliana*. *Nat. Rev. Genet.* **11:** 867-879.

16 Barrett, R.D.H. & D. Schluter 2008. Adaptation from standing genetic variation. *Trends Ecol. Evol.* **23:** 38-44.

17 Mackay, T.F.C. 2001. The genetic architecture of quantitative traits. *Annu. Rev. Genet.* **35:** 303-339.

18 Colautti, R.I., C.R. Lee & T. Mitchell-Olds 2012. Origin, fate, and architecture of ecologically relevant genetic variation. *Curr Opin Plant Biol* **15:** 199-204.

19 Mitchell-Olds, T. 1995. The molecular basis of quantitative genetic variation in natural populations. *Trends Ecol. Evol.* **10:** 324-328.

20 Slate, J. 2005. Quantitative trait locus mapping in natural populations: progress, caveats and future directions. *Mol. Ecol.* **14:** 363-379.

21 Erickson, D.L., C.B. Fenster, H.K. Stenøien & D. Price 2004. Quantitative trait locus analyses and the study of evolutionary process. *Mol. Ecol.* **13:** 2505-2522.

22 Lynch, M. & B. Walsh. 1998. *Genetics and analysis of quantitative traits*. Sinauer Associates, Inc., Sunderland, MA, USA.

23 Amos, W., E. Driscoll & J.I. Hoffman 2011. Candidate genes versus genome-wide associations: which are better for detecting genetic susceptibility to infectious disease? *Proc. R. Soc. B* **278:** 1183-1188.

24 Fitzpatrick, M.J. *et al.* 2005. Candidate genes for behavioural ecology. *Trends Ecol. Evol.* **20:** 96-104.

25 Hoekstra, H.E., R.J. Hirschmann, R.A. Bundey, P.A. Insel & J.P. Crossland 2006. A single amino acid mutation contributes to adaptive beach mouse color pattern. *Science* **313:** 101-104.

26 Walsh, N., J. Dale, K.J. McGraw, M.A. Pointer & N.I. Mundy 2011. Candidate genes for carotenoid coloration in vertebrates and their expression profiles in the carotenoid-containing plumage and bill of a wild bird. *Proc. R. Soc. B* **279:** 58-66.

27 Liu, B.-H. 1998. Computational tools for study of complex traits. In *Molecular dissection of complex traits* (ed. A. H. Paterson). CRC Press, Boca Raton, pp. 43-79.

28 Blangero, J., J.T. Williams & L. Almasy 2001. Variance component methods for detecting complex trait loci. *Adv. Genet.* **42:** 151-181.

29 Hoekstra, H.E. & J.A. Coyne 2007. The locus of evolution: evo devo and the genetics of adaptation. *Evolution* **61:** 995-1016.

30 Lister, C. & C. Dean 1993. Recombinant inbred lines for mapping RFLP and phenotypic markers in *Arabidopsis thaliana*. *Plant J.* **4:** 745-750.

31 Mackay, T.F. & R.R. Anholt 2006. Of flies and man: *Drosophila* as a model for human complex traits. *Annu. Rev. Genomics Hum. Genet.* **7:** 339-367.

32 Flint, J. & R. Mott 2001. Finding the molecular basis of quantitative traits: successes and pitfalls. *Nat. Rev. Genet.* **2:** 437-445.

33 Mauricio, R. 2001. Mapping quantitative trait loci in plants: uses and caveats for evolutionary biology. *Nat. Rev. Genet.* **2:** 370-381.

34 Valdar, W. *et al.* 2006. Genome-wide genetic association of complex traits in heterogeneous stock mice. *Nat. Genet.* **38:** 879-887.

35 Churchill, G. *et al.* 2004. The Collaborative Cross, a community resource for the genetic analysis of complex traits. *Nat. Genet.* **36:** 1133-1137.

36 Kover, P.X. *et al.* 2009. A multiparent advanced generation inter-cross to fine-map quantitative traits in *Arabidopsis thaliana*. *PLoS Genet.* **5:** e1000551.

37 Hall, M.C., D.B. Lowry & J.H. Willis 2010. Is local adaptation in *Mimulus guttatus* caused by trade-offs at individual loci? *Mol. Ecol.* **19:** 2739-2753.

38 Orr, H.A. 2001. The genetics of species differences. *Trends Ecol. Evol.* **16:** 343-350.

39 Rieseberg, L.H., A. Widmer, A.M. Arntz & J.M. Burke 2003. The genetic architecture necessary for transgressive segregation is common in both natural and domesticated populations. *Phil. Trans. R. Soc. Lond. B* **358:** 1141-1147.

40 Andersson, L. & M. Georges 2004. Domestic-animal genomics: deciphering the genetics of complex traits. *Nat. Rev. Genet.* **5:** 202-212.

41 Fournier-Level, A. *et al.* 2013. Paths to selection on life history loci in different natural environments across the native range of *Arabidopsis thaliana*. *Mol. Ecol.* **22:** 3552-3566.

42 Fournier-Level, A. *et al.* 2011. A map of local adaptation in *Arabidopsis thaliana*. *Science* **334:** 86-89.

43 Brachi, B. *et al.* 2010. Linkage and association mapping of *Arabidopsis thaliana* flowering time in nature. *PLoS Genet.* **6:** e1000940.

44 Knott, S.A. & C.S. Haley 1992. Maximum likelihood mapping of quantitative trait loci using full-sib families. *Genetics* **132:** 1211-1222.

45 Doerge, R.W. 2002. Mapping and analysis of quantitative trait loci in experimental populations. *Nat. Rev. Genet.* **3:** 43-52.

46 Williams, J.T. *et al.* 2001. Distribution of LOD scores in oligogenic linkage analysis. *Genet. Epidemiol.* **21:** S805-S810.

47 Blangero, J., J.T. Williams & L. Almasy 2000. Quantitative trait locus mapping using human pedigrees. *Hum. Biol.* **72:** 35-62.

48 Williams, J.T. & J. Blangero 1999. Power of variance component linkage analysis to detect quantitative trait loci. *Ann. Hum. Genet.* **63:** 545-563.

49 Darvasi, A., A. Weinreb, V. Minke, J.I. Weller & M. Soller 1993. Detecting marker-QTL linkage and estimating QTL gene effect and map location using a saturated genetic map. *Genetics* **134:** 943-951.

50 Slate, J. *et al.* 2010. Genome mapping in intensively studied wild vertebrate populations. *Trends Genet.* **26:** 275-284.

51 Stinchcombe, J.R. & H.E. Hoekstra 2008. Combining population genomics and quantitative genetics: finding the genes underlying ecologically important traits. *Heredity* **100:** 158-170.

52 George, A.W., P.M. Visscher & C.S. Haley 2000. Mapping quantitative trait loci in complex pedigrees: a two-step variance component approach. *Genetics* **156:** 2081-2092.

53 Heath, S.C. 1997. Markov chain Monte Carlo segregation and linkage analysis for oligogenic models. *Am. J. Hum. Genet.* **61:** 748-760.

54 Heath, S.C., G.L. Snow, E.A. Thompson, C. Tseng & E.M. Wijsman 1997. MCMC segregation and linkage analysis. *Genet. Epidemiol.* **14:** 1011-1016.

55 Haseman, J.K. & R.C. Elston 1972. Investigation of linkage between a quantitative trait and a marker locus. *Behav. Genet.* **2:** 3-19.

56 Elston, R.C., S. Buxbaum, K.B. Jacobs & J.M. Olson 2000. Haseman and Elston revisited. *Genet. Epidemiol.* **19:** 1-17.

57 Slate, J. *et al.* 2009. Gene mapping in the wild with SNPs: guidelines and future directions. *Genetica* **136:** 97-107.

58 Knief, U., H. Schielzeth, B. Kempenaers, H. Ellegren & W. Forstmeier 2012. QTL and quantitative genetic analysis of beak morphology reveals patterns of standing genetic variation in an Estrildid finch. *Mol. Ecol.* **21:** 3704-3717.

59 Schielzeth, H., W. Forstmeier, B. Kempenaers & H. Ellegren 2012. QTL linkage mapping of wing length in zebra finch using genome-wide SNP markers. *Mol. Ecol.* **21:** 229-239.

60 Schielzeth, H., B. Kempenaers, H. Ellegren & W. Forstmeier 2012. QTL linkage mapping of zebra finch beak color shows an oligogenic control of a sexually selected trait. *Evolution* **66:** 18-30.

61 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.

62 Aird, I., H.H. Bentall, J.A. Mehigan & J.A.F. Roberts 1954. The blood groups in relation to peptic ulceration and carcinoma of colon, rectum, breast, and bronchus: an association between the ABO groups and peptic ulceration. *Brit. Med. J.* **2:** 320-321.

63 Bush, W.S. & J.H. Moore 2012. Chapter 11: Genome-wide association studies. *PLoS Comput. Biol.* **8:** e1002822.

64 Kruglyak, L. 1999. Prospects for whole-genome linkage disequilibrium mapping of common disease genes. *Nat. Genet.* **22:** 139-144.

65 Donnelly, P.J. & C. Wiuf 1999. SNPs, linkage disequilibrium, shared chromosomal segments, and genome-wide association mapping. *Am. J. Hum. Genet.* **65:** A83-A83.

66 Hindorff, L.A. *et al.* 2009. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc. Natl. Acad. Sci.* **106:** 9362-9367.

67 Johnston, S.E. *et al.* 2011. Genome-wide association mapping identifies the genetic basis of discrete and quantitative variation in sexual weaponry in a wild sheep population. *Mol. Ecol.* **20:** 2555-2566.

68 Santure, A.W. *et al.* 2013. Genomic dissection of variation in clutch size and egg mass in a wild great tit (*Parus major*) population. *Mol. Ecol.* **22:** 3949-3962.

69 Parchman, T.L. *et al.* 2012. Genome-wide association genetics of an adaptive trait in lodgepole pine. *Mol. Ecol.* **21:** 2991-3005.

70 Risch, N. & K. Merikangas 1996. The future of genetic studies of complex human diseases. *Science* **273:** 1516-1517.

71 Ellegren, H. *et al.* 2012. The genomic landscape of species divergence in *Ficedula* flycatchers. *Nature* **491:** 756-760.

72 Johnston, S.E. *et al.* 2013. Life history trade-offs at a single locus maintain sexually selected genetic variation. *Nature***:** doi:10.1038/nature12489.

73 Backström, N., A. Ovarnström, L. Gustafsson & H. Ellegren 2006. Levels of linkage disequilibrium in a wild bird population. *Biol. Lett.* **2:** 435-438.

74 Li, M.H. & J. Merilä 2010. Extensive linkage disequilibrium in a wild bird population. *Heredity* **104:** 600-610.

75 Slatkin, M. 2008. Linkage disequilibrium: understanding the evolutionary past and mapping the medical future. *Nat. Rev. Genet.* **9:** 477-485.

76 Benjamini, Y. & Y. Hochberg 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. B* **57:** 289-300.

77 Storey, J.D. & R. Tibshirani 2003. Statistical significance for genomewide studies. *Proc. Natl. Acad. Sci.* **100:** 9440-9445.

78 Astle, W. & D.J. Balding 2009. Population structure and cryptic relatedness in genetic association studies. *Stat. Sci.* **24:** 451-471.

79 Kang, H.M. *et al.* 2008. Efficient control of population structure in model organism association mapping. *Genetics* **178:** 1709-1723.

80 Korte, A. *et al.* 2012. A mixed-model approach for genome-wide association studies of correlated traits in structured populations. *Nat. Genet.* **44:** 1066-1071.

81 Price, A.L., N.A. Zaitlen, D. Reich & N. Patterson 2010. New approaches to population stratification in genome-wide association studies. *Nat. Rev. Genet.* **11:** 459-463.

82 Rieseberg, L.H. & C.A. Buerkle 2002. Genetic mapping in hybrid zones. *Am. Nat.* **159:** S36-S50.

83 Buerkle, C.A. & C. Lexer 2008. Admixture as the basis for genetic mapping. *Trends Ecol. Evol.* **23:** 686-694.

84 McKeigue, P.M. 2005. Prospects for admixture mapping of complex traits. *Am. J. Hum. Genet.* **76:** 1-7.

85 Smith, M.W. & S.J. O'Brien 2005. Mapping by admixture linkage disequilibrium: advances, limitations and guidelines. *Nat. Rev. Genet.* **6:** 623-U626.

86 Redden, D.T. *et al.* 2006. Regional admixture mapping and structured association testing: Conceptual unification and an extensible general linear model. *PLoS Genet.* **2:** 1254-1264.

87 Reich, D. *et al.* 2005. A whole-genome admixture scan finds a candidate locus for multiple sclerosis susceptibility. *Nat. Genet.* **37:** 1113-1118.

88 Zhu, X.F. *et al.* 2005. Admixture mapping for hypertension loci with genome-scan markers. *Nat. Genet.* **37:** 177-181.

89 Freedman, M.L. *et al.* 2006. Admixture mapping identifies 8q24 as a prostate cancer risk locus in African-American men. *Proc. Natl. Acad. Sci.* **103:** 14068-14073.

90 Rieseberg, L.H., J. Whitton & K. Gardner 1999. Hybrid zones and the genetic architecture of a barrier to gene flow between two sunflower species. *Genetics* **152:** 713-727.

91 Lindtke, D., S.C. González-Martínez, D. Macaya-Sanz & C. Lexer 2013. Admixture mapping of quantitative traits in *Populus* hybrid zones: power and limitations. *Heredity***:** doi:10.1038/hdy.2013.1069.

92 Manolio, T.A. *et al.* 2009. Finding the missing heritability of complex diseases. *Nature* **461:** 747-753.

93 Yang, J.A. *et al.* 2010. Common SNPs explain a large proportion of the heritability for human height. *Nat. Genet.* **42:** 565-569.

94 Yang, J. *et al.* 2011. Genome partitioning of genetic variation for complex traits using common SNPs. *Nat. Genet.* **43:** 519-U544.

95 Robinson, M.R., A.W. Santure, I. DeCauwer, B.C. Sheldon & J. Slate 2013. Partitioning of genetic variation across the genome using multimarker methods in a wild bird population. *Mol. Ecol.* **22:** 3963-3980.

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

97 Yeaman, S. 2013. Genomic rearrangements and the evolution of clusters of locally adaptive loci. *Proc. Natl. Acad. Sci.* **110:** E1743-E1751.

98 Beavis, W.D. 1998. QTL analyses: power, precision, and accuracy. In *Molecular dissection of complex traits* (ed. A. H. Paterson). CRC Press, Boca Raton, pp. 145–162.

99 Beavis, W.D. 1994. The power and deceit of QTL experiments: lessons from comparative QTL studies. In *Proceedings of the 49th Annual Corn and Sorghum Industry Research Conference* (ed. D.B. Wilkinson). American Seed Trade Association, Washington, DC, pp. 250–266.

100 Forstmeier, W. & H. Schielzeth 2011. Cryptic multiple hypotheses testing in linear models: overestimated effect sizes and the winner's curse. *Behav. Ecol. Sociobiol.* **65:** 47-55.

101 Mackay, T.F.C., E.A. Stone & J.F. Ayroles 2009. The genetics of quantitative traits: challenges and prospects. *Nat. Rev. Genet.* **10:** 565-577.

102 Zöllner, S. & J.K. Pritchard 2007. Overcoming the winner's curse: estimating penetrance parameters from case-control data. *Am. J. Hum. Genet.* **80:** 605-615.

103 Ioannidis, J.P.A. 2005. Why most published research findings are false. *PLoS Med.* **2:** 696-701.

104 Yu, K. *et al.* 2007. Flexible design for following up positive findings. *Am. J. Hum. Genet.* **81:** 540-551.

105 Gorroochurn, P., S.E. Hodge, G.A. Heiman, M. Durner & D.A. Greenberg 2007. Non-replication of association studies: "pseudo-failures" to replicate? *Genet. Med.* **9:** 325-331.

106 McCarthy, M.I. *et al.* 2008. Genome-wide association studies for complex traits: consensus, uncertainty and challenges. *Nat. Rev. Genet.* **9:** 356-369.

107 Chanock, S.J. *et al.* 2007. Replicating genotype-phenotype associations. *Nature* **447:** 655-660.

108 Reid, D.P., A. Szanto, B. Glebe, R.G. Danzmann & M.M. Ferguson 2005. QTL for body weight and condition factor in Atlantic salmon (*Salmo salar*): comparative analysis with rainbow trout (*Oncorhynchus mykiss*) and Arctic charr (*Salvelinus alpinus*). *Heredity* **94:** 166-172.

109 Broman, K.W., S. Kim, S. Sen, C. Ane & B.A. Payseur 2012. Mapping quantitative trait loci onto a phylogenetic tree. *Genetics* **192:** 267-U549.

110 Ioannidis, J.P.A., G. Thomas & M.J. Daly 2009. Validating, augmenting and refining genome-wide association signals. *Nat. Rev. Genet.* **10:** 318-329.

111 Protas, M.E. *et al.* 2006. Genetic analysis of cavefish reveals molecular convergence in the evolution of albinism. *Nat. Genet.* **38:** 107-111.

112 Wheeler, N.C. *et al.* 2005. Mapping of quantitative trait loci controlling adaptive traits in coastal Douglas-fir. IV. Cold-hardiness QTL verification and candidate gene mapping. *Mol. Breeding* **15:** 145-156.

113 Shapiro, M.D. *et al.* 2004. Genetic and developmental basis of evolutionary pelvic reduction in threespine sticklebacks. *Nature* **428:** 717-723.

114 Cresko, W.A. *et al.* 2004. Parallel genetic basis for repeated evolution of armor loss in Alaskan threespine stickleback populations. *Proc. Natl. Acad. Sci.* **101:** 6050-6055.

115 Coyle, S.M., F.A. Huntingford & C.L. Peichel 2007. Parallel evolution of *Pitx1* underlies pelvic reduction in Scottish threespine stickleback (*Gasterosteus aculeatus*). *J. Hered.* **98:** 581-586.

116 Chan, Y.F. *et al.* 2010. Adaptive evolution of pelvic reduction in sticklebacks by recurrent deletion of a *Pitx1* enhancer. *Science* **327:** 302-305.

117 Fidler, A.E. *et al.* 2007. *DRD4* gene polymorphisms are associated with personality variation in a passerine bird. *Proc. R. Soc. B* **274:** 1685-1691.

118 Korsten, P. *et al.* 2010. Association between *DRD4* gene polymorphism and personality variation in great tits: a test across four wild populations. *Mol. Ecol.* **19:** 832-843.

119 Mueller, J.C. *et al.* 2013. Haplotype structure, adaptive history and associations with exploratory behaviour of the *DRD4* gene region in four great tit (*Parus major*) populations. *Mol. Ecol.* **22:** 2797-2809.

120 Flint, J. & T.F.C. Mackay 2009. Genetic architecture of quantitative traits in mice, flies, and humans. *Genome Res.* **19:** 723-733.

121 Abzhanov, A. *et al.* 2006. The calmodulin pathway and evolution of elongated beak morphology in Darwin's finches. *Nature* **442:** 563-567.

122 Mitchell-Olds, T. & J. Schmitt 2006. Genetic mechanisms and evolutionary significance of natural variation in *Arabidopsis*. *Nature* **441:** 947-952.

123 Ellegren, H. 2008. Comparative genomics and the study of evolution by natural selection. *Mol. Ecol.* **17:** 4586-4596.

124 Barrett, R.D.H. & H.E. Hoekstra 2011. Molecular spandrels: tests of adaptation at the genetic level. *Nat. Rev. Genet.* **12:** 767-780.

125 Ellegren, H. & B.C. Sheldon 2008. Genetic basis of fitness differences in natural populations. *Nature* **452:** 169-175.

126 Michel, A.P. *et al.* 2010. Widespread genomic divergence during sympatric speciation. *Proc. Natl. Acad. Sci.* **107:** 9724-9729.

127 Maher, B. 2008. Personal genomes: the case of the missing heritability. *Nature* **456:** 18-21.

128 Allen, H.L. *et al.* 2010. Hundreds of variants clustered in genomic loci and biological pathways affect human height. *Nature* **467:** 832-838.

129 Gudbjartsson, D.F. *et al.* 2008. Many sequence variants affecting diversity of adult human height. *Nat. Genet.* **40:** 609-615.

130 Weedon, M.N. *et al.* 2008. Genome-wide association analysis identifies 20 loci that influence adult height. *Nat. Genet.* **40:** 575-583.

131 Lettre, G. *et al.* 2008. Identification of ten loci associated with height highlights new biological pathways in human growth. *Nat. Genet.* **40:** 584-591.

132 Rockman, M.V. 2012. The QTN program and the alleles that matter for evolution: all that's gold does not glitter. *Evolution* **66:** 1-17.

133 Travisano, M. & R.G. Shaw 2013. Lost in the Map. *Evolution* **67:** 305-314.

134 Donnelly, P. 2008. Progress and challenges in genome-wide association studies in humans. *Nature* **456:** 728-731.

135 Nadeau, N.J. & C.D. Jiggins 2010. A golden age for evolutionary genetics? Genomic studies of adaptation in natural populations. *Trends Genet.* **26:** 484-492.

136 Buckler, E.S. *et al.* 2009. The genetic architecture of maize flowering time. *Science* **325:** 714-718.

137 Atwell, S. *et al.* 2010. Genome-wide association study of 107 phenotypes in *Arabidopsis thaliana* inbred lines. *Nature* **465:** 627-631.

138 Metzker, M.L. 2010. Applications of next-generation sequencing sequencing technologies: the next generation. *Nat. Rev. Genet.* **11:** 31-46.

139 Turner, T.L., E.C. Bourne, E.J. Von Wettberg, T.T. Hu & S.V. Nuzhdin 2010. Population resequencing reveals local adaptation of *Arabidopsis lyrata* to serpentine soils. *Nat. Genet.* **42:** 260-263.

140 Poissant, J., D. Reale, J.G.A. Martin, M. Festa-Bianchet & D.W. Coltman 2013. A quantitative trait locus analysis of personality in wild bighorn sheep. *Ecol. Evol.* **3:** 474-481.

141 Clutton-Brock, T. & B.C. Sheldon 2010. Individuals and populations: the role of long-term, individual-based studies of animals in ecology and evolutionary biology. *Trends Ecol. Evol.* **25:** 562-573.

142 Réale, D., S.M. Reader, D. Sol, P.T. McDougall & N.J. Dingemanse 2007. Integrating animal temperament within ecology and evolution. *Biol. Rev.* **82:** 291-318.

143 Pavey, S.A., L. Bernatchez, N. Aubin-Horth & C.R. Landry 2012. What is needed for next-generation ecological and evolutionary genomics? *Trends Ecol. Evol.* **27:** 673-678.

144 Storz, J.F. & C.W. Wheat 2010. Integrating evolutionary and functional approaches to infer adaptation at specific loci. *Evolution* **64:** 2489-2509.

145 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.

146 Lowry, D.B., M.C. Hall, D.E. Salt & J.H. Willis 2009. Genetic and physiological basis of adaptive salt tolerance divergence between coastal and inland *Mimulus guttatus*. *New Phytol.* **183:** 776-788.

147 Barrett, R.D.H., S.M. Rogers & D. Schluter 2008. Natural selection on a major armor gene in threespine stickleback. *Science* **322:** 255-257.

148 Weinig, C. *et al.* 2003. Heterogeneous selection at specific loci in natural environments in *Arabidopsis thaliana*. *Genetics* **165:** 321-329.

149 Lexer, C., M.E. Welch, J.L. Durphy & L.H. Rieseberg 2003. Natural selection for salt tolerance quantitative trait loci (QTLs) in wild sunflower hybrids: implications for the origin of *Helianthus paradoxus*, a diploid hybrid species. *Mol. Ecol.* **12:** 1225-1235.

150 Ungerer, M.C., C.R. Linder & L.H. Rieseberg 2003. Effects of genetic background on response to selection in experimental populations of *Arabidopsis thaliana*. *Genetics* **163:** 277-286.

151 Anderson, J.T., C.R. Lee, C.A. Rushworth, R.I. Colautti & T. Mitchell-Olds 2013. Genetic trade-offs and conditional neutrality contribute to local adaptation. *Mol. Ecol.* **22:** 699-708.

152 Hurme, P., M.J. Sillanpaa, E. Arjas, T. Repo & O. Savolainen 2000. Genetic basis of climatic adaptation in Scots pine by Bayesian quantitative trait locus analysis. *Genetics* **156:** 1309-1322.

153 Byrne, M. *et al.* 1997. Identification and mode of action of quantitative trait loci affecting seedling height and leaf area in *Eucalyptus nitens*. *Theor. Appl. Genet.* **94:** 674-681.

154 Byrne, M., J.C. Murrell, J.V. Owen, E.R. Williams & G.F. Moran 1997. Mapping of quantitative trait loci influencing frost tolerance in *Eucalyptus nitens*. *Theor. Appl. Genet.* **95:** 975-979.

155 Hall, M.C., C.J. Basten & J.H. Willis 2006. Pleiotropic quantitative trait loci contribute to population divergence in traits associated with life-history variation in *Mimulus guttatus*. *Genetics* **172:** 1829-1844.

156 Lowry, D.B. *et al.* 2012. Mapping of ionomic traits in *Mimulus guttatus* reveals Mo and Cd QTLs that colocalize with *MOT1* homologues. *PLoS One* **7**.

157 Friedman, J. & J.H. Willis 2013. Major QTLs for critical photoperiod and vernalization underlie extensive variation in flowering in the *Mimulus guttatus* species complex. *New Phytol.* **199:** 571-583.

158 Linde, M., S. Diel & B. Neuffer 2001. Flowering ecotypes of *Capsella bursa-pastoris* (L.) Medik. (Brassicaceae) analysed by a cosegregation of phenotypic characters (QTL) and molecular markers. *Ann. Bot. (London)* **87:** 91-99.

159 Schranz, M.E., A.J. Manzaneda, A.J. Windsor, M.J. Clauss & T. Mitchell-Olds 2009. Ecological genomics of *Boechera stricta*: identification of a QTL controlling the allocation of methionine- vs branched-chain amino acid-derived glucosinolates and levels of insect herbivory. *Heredity* **102:** 465-474.

160 Anderson, J.T., C.R. Lee & T. Mitchell-Olds 2011. Life-history QTLs and natural selection on flowering time in *Boechera stricta*, a perennial relative of *Arabidopsis*. *Evolution* **65:** 771-787.

161 Moritz, D.M.L. & J.W. Kadereit 2001. The genetics of evolutionary change in *Senecio vulgaris* L.: a QTL mapping approach. *Plant Biol.* **3:** 544-552.

162 Gardner, K.M. & R.G. Latta 2006. Identifying loci under selection across contrasting environments in *Avena barbata* using quantitative trait locus mapping. *Mol. Ecol.* **15:** 1321-1333.

163 Verhoeven, K.J.F., T.K. Vanhala, A. Biere, E. Nevo & J.M.M. van Damme 2004. The genetic basis of adaptive population differentiation: a quantitative trait locus analysis of fitness traits in two wild barley populations from contrasting habitats. *Evolution* **58:** 270-283.

164 Verhoeven, K.J.F., H. Poorter, E. Nevo & A. Biere 2008. Habitat-specific natural selection at a flowering-time QTL is a main driver of local adaptation in two wild barley populations. *Mol. Ecol.* **17:** 3416-3424.

165 Protas, M.E., P. Trontelj & N.H. Patel 2011. Genetic basis of eye and pigment loss in the cave crustacean, Asellus aquaticus. *Proc. Natl. Acad. Sci.* **108:** 5702-5707.

166 Hawthorne, D.J. & S. Via 2001. Genetic linkage of ecological specialization and reproductive isolation in pea aphids. *Nature* **412:** 904-907.

167 Peichel, C.L. *et al.* 2001. The genetic architecture of divergence between threespine stickleback species. *Nature* **414:** 901-905.

168 Colosimo, P.F. *et al.* 2004. The genetic architecture of parallel armor plate reduction in threespine sticklebacks. *PLoS Biol.* **2:** 635-641.

169 Robison, B.D., P.A. Wheeler, K. Sundin, P. Sikka & G.H. Thorgaard 2001. Composite interval mapping reveals a major locus influencing embryonic development rate in rainbow trout (*Oncorhynchus mykiss*). *J. Hered.* **92:** 16-22.

170 Hecht, B.C., F.P. Thrower, M.C. Hale, M.R. Miller & K.M. Nichols 2012. Genetic architecture of migration-related traits in rainbow and steelhead trout, *Oncorhynchus mykiss*. *Genes Genom. Genet.* **2:** 1113-1127.

171 Borowsky, R. & H. Wilkens 2002. Mapping a cave fish genome: polygenic systems and regressive evolution. *J. Hered.* **93:** 19-21.

172 Protas, M., M. Conrad, J.B. Gross, C. Tabin & R. Borowsky 2007. Regressive evolution in the Mexican cave tetra, *Astyanax mexicanus*. *Curr. Biol.* **17:** 452-454.

173 Protas, M. *et al.* 2008. Multi-trait evolution in a cave fish, *Astyanax mexicanus*. *Evol. Dev.* **10:** 196-209.

174 Gross, J.B., R. Borowsky & C.J. Tabin 2009. A Novel Role for Mc1r in the Parallel Evolution of Depigmentation in Independent Populations of the Cavefish Astyanax mexicanus. *PLoS Genet.* **5**.

175 O'Quin, K.E., M. Yoshizawa, P. Doshi & W.R. Jeffery 2013. Quantitative genetic analysis of retinal degeneration in the blind cavefish *Astyanax mexicanus*. *PLoS One* **8**.

176 Jermstad, K.D., D.L. Bassoni, K.S. Jech, N.C. Wheeler & D.B. Neale 2001. Mapping of quantitative trait loci controlling adaptive traits in coastal Douglas-fir. I. Timing of vegetative bud flush. *Theor. Appl. Genet.* **102:** 1142-1151.

177 Jermstad, K.D. *et al.* 2001. Mapping of quantitative trait loci controlling adaptive traits in coastal Douglas-fir. II. Spring and fall cold-hardiness. *Theor. Appl. Genet.* **102:** 1152-1158.

178 Jermstad, K.D. *et al.* 2003. Mapping of quantitative trait loci controlling adaptive traits in coastal Douglas fir. III. Quantitative trait loci-by-environment interactions. *Genetics* **165:** 1489-1506.

179 Höglund, S., A.C. Rönnberg-Wästljung, U. Lagercrantz & S. Larsson 2012. A rare major plant QTL determines non-responsiveness to a gall-forming insect in willow. *Tree Genet. Genomes* **8:** 1051-1060.

180 Freeman, J.S., B.M. Potts & R.E. Vaillancourt 2008. Few Mendelian genes underlie the quantitative response of a forest tree, *Eucalyptus globulus*, to a natural fungal epidemic. *Genetics* **178:** 563-571.

181 Saintagne, C. *et al.* 2004. Distribution of genomic regions differentiating oak species assessed by QTL detection. *Heredity* **92:** 20-30.

182 Scotti-Saintagne, C., E. Bertocchi, T. Barreneche, A. Kremer & C. Plomion 2005. Quantitative trait loci mapping for vegetative propagation in pedunculate oak. *Ann for Sci* **62:** 369-374.

183 Scalfi, M. *et al.* 2004. A RAPD, AFLP and SSR linkage map, and QTL analysis in european beech (*Fagus sylvatica* L.). *Theor. Appl. Genet.* **108:** 433-441.

184 Vasemägi, A., R. Gross, D. Palm, T. Paaver & C.R. Primmer 2010. Discovery and application of insertion-deletion (INDEL) polymorphisms for QTL mapping of early life-history traits in Atlantic salmon. *BMC Genomics* **11:** 156.

185 Moghadam, H.K. *et al.* 2007. Quantitative trait loci for body weight, condition factor and age at sexual maturation in Arctic charr (*Salvelinus alpinus*): comparative analysis with rainbow trout (*Oncorhynchus mykiss*) and Atlantic salmon (*Salmo salar*). *Mol. Genet. Genomics* **277:** 647-661.
186 Malek, T.B., J.W. Boughman, I. Dworkin & C.L. Peichel 2012. Admixture mapping of male nuptial colour and body shape in a recently formed hybrid population of threespine stickleback. *Mol. Ecol.* **21:** 5265-5279.
187 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* **1692:** 2361-2369.
188 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.
189 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.
190 Beraldi, D. *et al.* 2007. Mapping quantitative trait loci underlying fitness-related traits in a free-living sheep population. *Evolution* **61:** 1403-1416.
191 Johnston, S.E., D. Beraldi, A.F. McRae, J.M. Pemberton & J. Slate 2010. Horn type and horn length genes map to the same chromosomal region in Soay sheep. *Heredity* **104:** 196-205.
192 Poissant, J., C.S. Davis, R.M. Malenfant, J.T. Hogg & D.W. Coltman 2012. QTL mapping for sexually dimorphic fitness-related traits in wild bighorn sheep. *Heredity* **108:** 256-263.
193 Morse, A.M. *et al.* 2009. Evolution of genome size and complexity in *Pinus*. *PLoS One* **4:** e4332.
194 Lowry, D.B. & J.H. Willis 2010. A widespread chromosomal inversion polymorphism contributes to a major life-history transition, local adaptation, and reproductive isolation. *PLoS Biol.* **8**.
195 Mojica, J.P., Y.W. Lee, J.H. Willis & J.K. Kelly 2012. Spatially and temporally varying selection on intrapopulation quantitative trait loci for a life history trade-off in *Mimulus guttatus*. *Mol. Ecol.* **21:** 3718-3728.
196 Colosimo, P.F. *et al.* 2005. Widespread parallel evolution in sticklebacks by repeated fixation of *Ectodysplasin* alleles. *Science* **307:** 1928-1933.
197 Barrett, R.D.H., S.M. Rogers & D. Schluter 2009. Environment specific pleiotropy facilitates divergence at the *Ectodysplasin* locus in threespine stickleback. *Evolution* **63:** 2831-2837.
198 Jeffery, W.R. 2009. Regressive evolution in *Astyanax* cavefish. *Annu. Rev. Genet.* **43:** 25-47.
199 Merilä, J. & B. Sheldon 2001. Avian quantitative genetics. *Curr. Ornithol.* **16:** 179-255.
200 van Bers, N.E.M. *et al.* 2012. The design and cross-population application of a genome-wide SNP chip for the great tit *Parus major*. *Mol. Ecol. Res.* **12:** 753-770.
201 Coltman, D. & J. Pemberton. 2004. Inheritance of coat colour and horn type in Soay sheep. In *Soay sheep: dynamics and selection in an island population* (eds. T. H. Clutton-Brock & J.M. Pemberton). Cambridge University Press, Cambridge, UK, pp. 321–327.
202 Pickering, N.K., P.L. Johnson, B. Auvray, K.G. Dodds & J.C. McEwan 2009. Mapping the *Horns* locus in sheep. *Proc. Ass. Adv. Anim. Breed. Genet.* **18:** 88-91.

Figure 1: Conceptual overview of different mapping strategies targeting standing genetic variation. LD = linkage disequilibrium, SGV = Standing genetic variation for trait of interest.

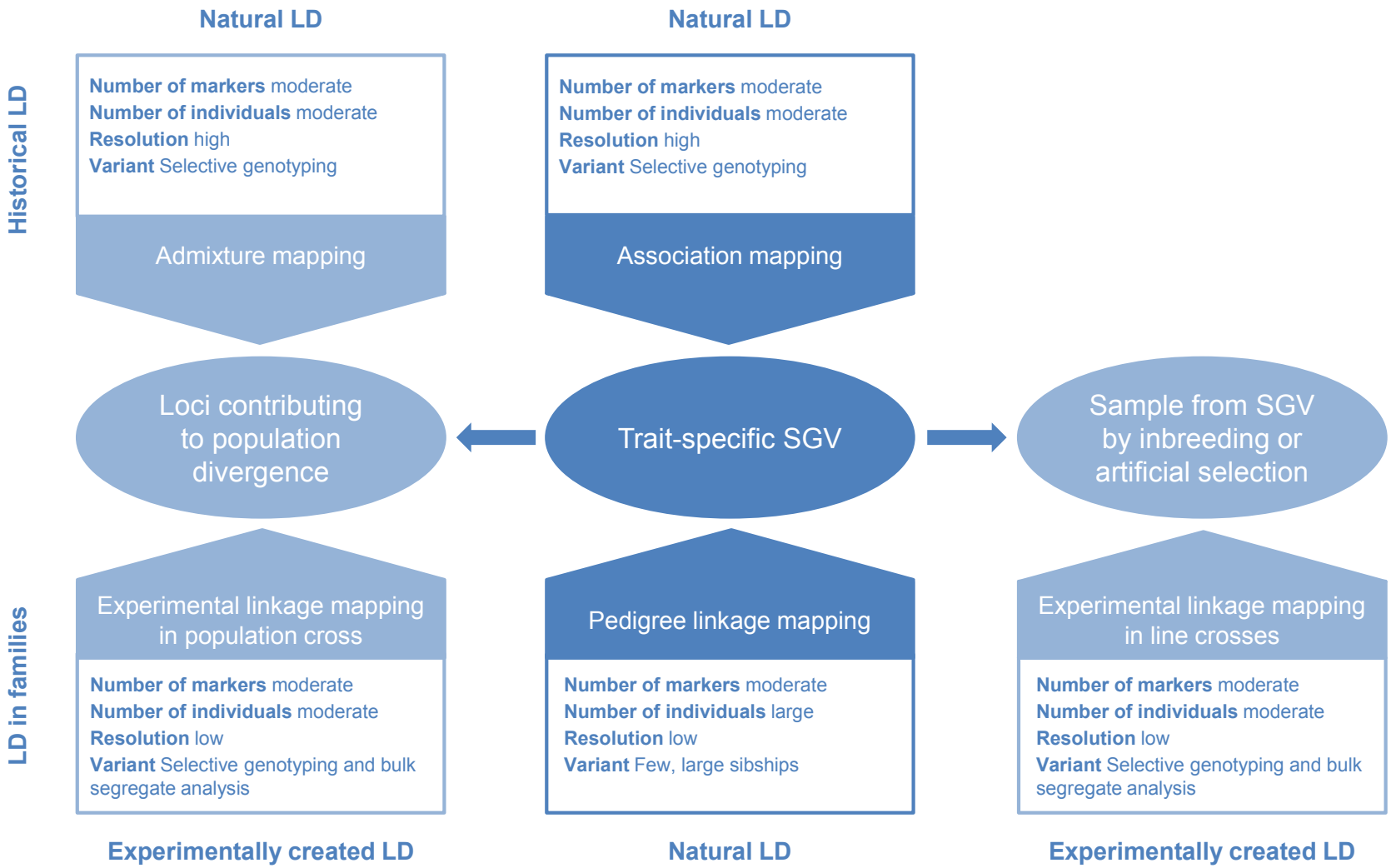

Table 1: Non-exhaustive sample of QTL mapping studies using experimental crosses derived from natural populations of non-model organisms sampled (excluding crops, livestock and crosses of such with their wild ancestors). 'Population cross' refers to geographically separated populations of the same species. 'Ecotype cross' refers to populations of the same species in discretely different habitats. Abbreviations: RFLP = Restriction fragment length polymorphism, AFLP = Amplified fragment length polymorphism, RAPD = Random amplified polymorphic DNA, SSR = Simple sequence repeats (microsatellites), ISSR = Inter inter simple sequence repeat, SNP = Single nucleotide polymorphisms, Iso = Isozymes, Alu = Alu transposable elements, EPIC = Exon-primed intron-crossing markers.

Phylogenetic group	Species	Trait	Method	Sample size: individuals	Sample size: Markers	Ref.
Plants (Pinaceae)	Scots pine (*Pinus sylvestris)*	Timing of bud set and frost hardiness	Population backcross	113	164 RAPD	[152]
Plants (Myrtaceae)	Shining gum (*Eucalyptus nitens*)	Seedling height and leaf area	F2 population cross	178	210 RFLP	[153]
		Frost tolerance	F2 population cross	118	210 RFLP	[154]
Plants (Phrymaceae)	Yellow monkeyflower (*Mimulus guttatus*)	Floral traits, plant morphological traits, age at flowering, pollen viability	F2 ecotype cross	539	112 AFLP+SSR	[155]
		Salt tolerance	RILs from ecotype cross	186 RILs	189 EPIC	[146]
		Corolla and plant morphology and size, flower number, survival, fecundity, timing of flowering	RILs from ecotype cross backcrossed to parental lines (parental inbred lines, PIL)	191 RILs	189 EPIC	[37]
		Accumlation of 17 elemental nutrients and three toxic elemetns	RILs from ecotype cross	186 RILs	189 EPIC	[156]
		Critical photoperiod	F2 population cross, bulk segregate analysis	360	156 EPIC	[157]
		Vernalization	F2 population two crosses, bulk segregate analysis	360 + 360	156 EPIC	[157]

Plants (Brassicaceae)	Shepherd's-purse (*Capsella bursa-pastoris*)	Timing of flowering, plant morphology, number of (sterile) fruits, fecundity, fruit and pedicel length	F2 population cross	113	107 RAPD, 6 Iso	[158]
	Drummond's rockcress (*Boechera stricta*)	Resistence to herbivory	F2 inbred line cross	192	58 SSR	[159]
		Flowering time, leaf number	RILs from population cross	178 RILs	105 SNP, 62 SSR	[160]
Plants (Asteraceae)	Common groundsel (*Senecio vulgaris*)	Phenological, vegetative and reproductive traits	F2 ecotype cross	120	RAPD	[161]
Plants (Poaceae)	Wild oat (*Avena barbata*)	Number of spikelets, plant dry mass	RILs from ecotype cross	188 RIL	129 AFLP	[162]
	Wild barley (*Hordeum spontaneum*)	Viability, fecundity, various seed traits, flower heads per plant and seeds per head	F3 ecotype cross	140	196 AFLP, 6 SSR	[163]
		Flowering time, seed weight, growth rate	F3 ecotype cross	140	196 AFLP, 6 SSR	[164]
Isopoda	Waterlouse (*Asellus aquaticus*)	Body pigmentation and pattern, eye loss	F2 ecotype backcross	194	100 SNP	[165]
Insects (Aphididae)	Pea aphid (*Acyrthosiphon pisum*)	Fecundity, food choice	F2 ecotype cross	194	173 AFLP	[166]
Fish	Theespined stickleback (*Gasterosteus aculeatus*)	Bony armor, feeding morphology	Ecotype backcross	92	227 SSR	[167]
		Bony armor	F2 ecotype cross	360	160 SSR	[168]
		Pelvic spines	Multiple F2 ecotype crosses	33-281	227 SSR	[114]
		Pelvic spines	F2 ecotype cross	375	53 SSR	[113]
	Rainbow trout (*Oncorhynchus mykiss*)	Embryonic development rate	F2 inbred line cross	170	219 AFLP, 2 SSR, 1 Alu	[169]
		Body size, condition, growth, morphology, skin reflectance, and osmoregulatory ability	F2 ecotype cross	235	164 SSR, 414 SNP	[170]
	Mexican tetra (*Astyanax mexicanus*)	Eye size, melanophore number, condition factor, albinism	Ecotype backcross	111	81 RAPD	[171]
		Albinism	Ecotype backcross	111	267 SSR	[111]
		Eye size (jaw size, number of teeth, tast	F2 ecotype cross	539	178 SSR	[172]

buds and melanophores)		(117-227)		
Eye size, body length, body condition (melanophore number, chemical sensitivity, body and jaw morphology, body length,body condition)	F2 ecotype cross	533-539 (113-361)	177 or 294 SSR	[173]
Brown phenotype	F2 ecotype cross	488	262 SSR	[174]
Retina thickness	F2 ecotype cross	115	463 SNP, 235 SSR	[175]

Table 2: Examples of linkage mapping studies for in outbred large fullsib families with parents or recent ancestors collected from natural population. Abbreviations: RFLP = Restriction fragment length polymorphism, AFLP = Amplified fragment length polymorphism, RAPD = Random amplified polymorphic DNA, SSR = Simple sequence repeats (microsatellites), SNP = Single nucleotide polymorphisms, SCAR = Sequenced characterized amplified regions, INDEL = Insertion/deletion polymorphisms.

				Sample size		
Phylogenetic group	**Species**	**Trait(s)**	**Mapping approach**	**Individuals**	**Markers**	**Ref**
Plants (Pinaceae)	Douglas-fir (*Pseudotsuga menziesii*)	Timing of spring bud flush	One fullsib family (parentally selected extreme phenotypes)	190	74 RFLP	[176]
		Cold-hardiness	One fullsib family	186	74 RFLP	[177]
			One fullsib family	383	74 RFLP	[112]
		Timing of seasonal growth initiation, cessation and bud flush	One fullsib family	357-429	72 RFLP	[178]
Plants (Salicaceae)	Common osier (*Salix viminalis*)	Parasite resistence	One fullsib family	282	214 SNP, 41 SSR	[179]
Plants (Myrtaceae)	Southern blue gum (*Eucalyptus globulus*)	Parasite resistance	One outbred families (parentally selected phenotypes)	112	132 AFLP, 33 SSR	[180]
			Two outbred families (selected genotyping of extremes)	50 + 40	132 AFLP, 33 SSR	[180]
Plants (Fagaceae)	Pedunculate oak (*Quercus robur*).	Leaf morphology	One outbred full-sib family	390	34 SSR, 84 AFLP, 1 SCAR, 9 RAPD	[181]
		Vegetative propagation	One outbred full-sib family	232	34 SSR, 84 AFLP, 1 SCAR, 9 RAPD	[182]
	European beech (*Fagus sylvatica*)	Leaf number, leaf area and shape, tree height	On full sib family	143	28 RAPD, 274 AFLP, 10 SSR	[183]
Fish	Atlantic salmon (*Salmo salar*)	Body weight, body condition	Three outbred full-sib families	3 x 46	91 SSR	[108]
		Time of emergence, tail fork	Two outbred full-sib families	370 + 279	50 INDEL, 77	[184]

	length			SSR	
Arctic charr (*Salvelinus alpinus*)	Body weight, body condition, age at maturation	Two outbred full-sib families	2 x 94	100 SSR	[185]

963

Table 3: Overview of QTL mapping studies in outbred populations of non-model organisms. The overview covers pedigree linkage mapping, genome-wide association mapping and admixture mapping approaches. In the second part of the table, we also include examples of studies that analyze a small number of full-sib families when parents were sampled from natural populations. Abbreviations: AFLP = Amplified fragment length polymorphism, SSR = Simple sequence repeats (microsatellites), SNP = Single nucleotide polymorphisms, Iso = Isozymes.

Phylogenetic group	Species	Trait(s)	Mapping approach	Sample size: Individuals	Sample size: Markers	Ref
Mapping in large, diverse mapping populations						
Plants (Pinaceae)	Lodgepole pine (*Pinus contorta*)	Cone serotiny	Association mapping based on selection of extreme phenotypes	98	97,616 SNP	69
Fish	Threespined stickleback (*Gasterosteus aculeatus*)	Nuptial coloration	Admixture mapping in ecotype hybrid zone (used for QTL confirmation)	508	576 SSR	186
Birds	Great tit (*Parus major*)	Clutch size, egg mass	Chromosome partitioning, pedigree linkage mapping, association mapping	902-969	7,203 SNP	68
		Wing length	Chromosome partitioning	2,644	7,203 SNP	95
	Great reed warbler (*Acrocephalus arundinaceus*)	Wing length	Pedigree linkage mapping	333	57 SSR, 36 AFLP	187
	Zebra finch (*Taeniopygia guttata*)	Wing length	Pedigree linkage mapping	1,066	1,404 SNP	59
		Beak color	Pedigree linkage mapping	1,019	1,404 SNP	60
		Beak morphology	Pedigree linkage mapping	992	1,404 SNP	58
Mammals	Soay sheep (*Ovis aries*)	Horn type, coat color, coat pattern	Pedigree linkage mapping	560	247 SSR, 4 Iso	188
		Pathogen resistance	Pedigree linkage mapping	588	247 SSR, 4 Iso	189
		Birth date, birth weight, leg length, body weight, jaw and	Pedigree linkage mapping	588	247 SSR, 4 Iso	190

	metacarpal length				
	Horn type, horn size	Linkage mapping (local only)	588	21 SSR	[191]
	Horn type	Association mapping	445	35,831 SNP	[67]
	Horn size	Association mapping	160	35,831 SNP	[67]
Bighorn sheep (*Ovis canadensis*)	Horn size, body mass	Pedigree linkage mapping	310	247 SSR	[192]
	Docility, boldness	Pedigree linkage mapping	310	238 SSR	[140]
Red deer (*Cervus elaphus*)	Birth weight	Pedigree linkage mapping	295	90 SSR	[61]

Table 4: A selection of QTL mapping case studies in natural populations illustrating how a variety of approaches can lead to the identification, replication and fine-mapping of trait loci. The studies also give examples for how knowledge about QTL can be used for studying selection under natural conditions.

Serotiny in lodgepole pines (*Pinus contorta*) In many species of conifers the ability to release the seeds inside cones in response to an environmental trigger, such as wildfires, is an important adaptive trait but the genetic basis to this has been unknown. Recently, Parchman and colleagues[69] used high throughput sequencing and a GWA mapping approach to remedy this situation. They sampled three populations of lodgepole pines from the Rocky Mountains and obtained a reference assembly from which they called more than 97,000 SNPs to be used in a GWAS on 98 individuals that were selected for unambiguous serotinous or non-serotinous phenotypes. Rather surprisingly given the low number of markers (compared to the genome size) and individuals, the authors were able to detect eleven loci that were associated with serotiny, although the function of these loci was unknown.[69] This study illustrates the possibilities offered by high throughput sequencing and a GWA approach in a species with huge genome (18-40,000 Mbp[193]) to detect genetic polymorphisms affecting fitness traits in natural populations.
Local adaptation and life-history evolution in yellow monkeyflowers (*Mimulus guttatus*) Yellow monkeyflowers are distributed throughout western North America and show two distinct ecotypes that are locally adapted to coastal and inland habitats. Ecotypes differ in many characteristics including whether they are annual or perennial, time of flowering, plant height and other morphological traits.[155] Experimental linkage mapping in population crosses were used to uncover the genetic basis of traits contributing to local adaptation, including mapping of a suite of 20 morphological and life-history traits in a 539 F2 indiviudals.[155] However, most of the mapping was

done using the possibility of constructing RILs in monkeyflowers. For example, Lowry et al.[146] mapped salt tolerance in RILs and performed reciprocal transplant experiment to demonstrate the fitness benefit of the salt tolerance QTL. Furthermore, population crosses based on sampling from a larger geographical range helped to pinpoint an inversion polymorphism affecting flowering time and reciprocal transplants demonstrated its contribution to local adpatiation.[194] Other work from the same group also shows the potential for studying selection on QTL under natural conditions. For example, key life-history traits have been found to be under spatially and temporally variable balancing selection.[195]

Adaptation to freshwater habitats in Threespined stickleback (*Gasterosteus aculeatus*)

Threespined sticklebacks occur globally widespread in marine habitats, but have colonized freshwater habitats on multiple independent occasions. Adaptations to freshwater habitats involve striking changes in morphology, most prominently the loss of pelvic spines and armor plates. Both of these traits have been mapped in genome-wide linkage scans based on F2 population crosses between marine and benthic populations sampled from native habitats.[113,168] The identification of the *Eda* locus as a QTG for armor plates was based a positional candidate genes approach and validated by positional cloning and high-resolution association mapping.[196] The initial linkage mapping of pelvic spines revealed one major and 4 minor QTL.[113] The leading QTL signal was confirmed in multiple independent crosses, including some from independently derived populations.[114,115] Fine-mapping to a very small genomic region upstream of the *Pitx1* gene was done by combining positional cloning, comparative genomics, expression analysis and artificial breeding.[113,116] Knowledge about the *Eda* QTG was used for studying pleiotropic effects under laboraroty and selection under field conditions.[147,197].

Regressive evolution in Mexican tetra (*Astyanax mexicanus*)

Mexican tetra is a central American fish species that has colonized cave habitats at least three times independently.[198] Cave-dwelling populations are characterized by several regressive characters, most notably the loss of pigmentation and eye reduction. The species readily reproduces in the laboratory and crosses between cave-dwelling populations and their surface-dwelling conspecifics have been used for mapping cave-specific traits. For example, Protas and collegues[111] mapped albinism in a backcross family and found one strong QTL signal. The QTL was confirmed in an independent F2 cross that involved a different cave population. Lack of complementation in a cross between the two cave populations further suggested that the very same locus was involved in loss of pigmentation in both populations. A positional candidate genes search resulted in only one gene (*Oca2*) that matched the linkage peak. The functional role of *Oca2* was validated by genetic transfection in mouse cells. Further analyses suggest at least three independent mutations in the *Oca2* gene that have led to a albinism in cave populations, including two different exon deletions in two different cave populations.[111] This study nicely demonstrates the general stepwise procedure of linkage mapping, replication and fine-mapping with careful choice of good candidate genes in QTL regions.

Life history traits in great tits (*Parus major*)

Clutch size in birds is a classical avian life-history trait and numerous studies have demonstrated that clutch size is under selection and has a genetic basis,[199] yet so far no genes influencing this trait is known from natural populations. To address this Santure et al.[68] genotyped 650 females using 5500 polymorphic great tit SNPs [200] to map QTL for clutch size and egg mass using a combination of three approaches: chromosome partitioning, linkage analysis and genome wide association mapping. Neither the linkage mapping approach nor the GWAS were able to detect any genome wide significant QTL, probably because power was too low to detect loci with the small effect sizes

expected from a polygenic trait. This latter conclusion is supported by the fact that the amount of genetic variance on each chromosome and the size of the chromosome was strongly positively correlated, which is suggestive of a largely polygenic basis to these traits.[68] The study illustrates that even with a relatively large sample size it may be problematic to detect loci for ecologically important quantitative traits in natural populations. For the great tits the search for clutch size QTL continue.

Genetic basis of sexual ornamentation in Soay sheep (*Ovies aries*)

One of the first large-scale QTL study in a natural population aimed to map the genetic basis of horn morphology in island population of Soay sheep.[67] The discrete horn type polymorphism observed suggests a largely Mendelian basis to this trait and previous research has indicated that it may be controlled by a single locus with three alleles.[201] Using linkage mapping Johnston et al. were able to map the location to a QTL on chromosome 10 covering 7.4 cM region.[191] With the availability of a commercial Ovine 50k SNP chip, it was possible to follow this up with a genome wide association scan, which confirmed the linkage mapping signal on chromosome 10 and narrowed it down to three markers located close to the *RXFP2* gene[67] that has previously been found to associate with horn type in domestic sheep.[202] The result was further strengthened by a smaller scale SNP array for genotyping 17 SNPs within and around the *RXFP2*. Johnston et al. then used the QTL mapping results for further study of the selective processes that maintain variation and found that the two alleles had opposing effects on reproductive success and survival, with heterozygotes being the most successful genotype overall, a pattern that could contribute to maintenance of genetic variance at this locus.[72]

971

972 .

973