

Review

## Applications and Implications of Neutral *versus* Non-neutral Markers in Molecular Ecology

Heather Kirk and Joanna R. Freeland \*

Department of Biology, Trent University, Peterborough, Ontario K9J 7B8, Canada;

E-Mail: heatherkirk@trentu.ca

\* Author to whom correspondence should be addressed; E-Mail: joannafreeland@trentu.ca;

Tel.: +001-705-748-1011 ext. 7180; Fax: +001-705-748-1003.

Received: 28 April 2011; in revised form: 6 June 2011 / Accepted: 7 June 2011 /

Published: 14 June 2011

---

**Abstract:** The field of molecular ecology has expanded enormously in the past two decades, largely because of the growing ease with which neutral molecular genetic data can be obtained from virtually any taxonomic group. However, there is also a growing awareness that neutral molecular data can provide only partial insight into parameters such as genetic diversity, local adaptation, evolutionary potential, effective population size, and taxonomic designations. Here we review some of the applications of neutral *versus* adaptive markers in molecular ecology, discuss some of the advantages that can be obtained by supplementing studies of molecular ecology with data from non-neutral molecular markers, and summarize new methods that are enabling researchers to generate data from genes that are under selection.

**Keywords:** molecular markers; natural selection; genetic diversity; genetic drift; genetic differentiation; gene expression; next-generation sequencing

---

### 1. Introduction

The contributions that molecular biology has made to ecological research over the past two decades are phenomenal, and have created the relatively new field that is known as molecular ecology. During that time, methods for genetically characterizing individuals, populations, and species have become almost routine, and have provided us with fascinating new insights into the ecology and evolution of virtually all taxonomic groups [1]. Molecular markers allow us, among other things, to quantify

genetic diversity [2,3], track the movements of individuals [4,5], measure inbreeding [6,7], identify species from mixed samples (for example soil samples or gut contents) [8,9], characterize new species [10,11] and retrace historical patterns of dispersal [12,13]. Building on these accomplishments, the field of molecular ecology continues to evolve, and among the more recent developments is a growing awareness that neutral molecular data—on which the majority of published studies in molecular ecology are based—can provide only partial insight into parameters such as genetic diversity, local adaptation, evolutionary potential, effective population size, and taxonomic designations [14–16] (but see [17,18]).

Biologists constantly strive to better understand evolution, and this quest is an important reason why we increasingly seek the information that can be obtained from adaptive genes (*i.e.*, genes that directly influence fitness). The relatively recent focus on non-neutral (adaptive) markers in molecular ecology can be further attributed to the potential practical applications of this approach, for example the identification of disease-causing genes or genes that can improve crop yields. In addition, there is growing concern over the rate at which environmental change is now occurring around the world. Species have three options that may allow them to survive rapidly changing environments: dispersal, phenotypic plasticity, or adaptation. If a species is unable to disperse from its native range to other suitable habitats, and is incapable of a plastic response, its survival will require rapid adaptive change which is possible only if an adequate level of adaptive genetic variation has been maintained [19,20]; therefore neutral and adaptive genetic diversity will likely have different impacts on long-term survival because only one (adaptive diversity) will allow a population to adapt to changing environmental conditions [21,22].

Another reason for the growing interest in adaptive variation is more practical: we are increasingly able to develop and utilize molecular markers that allow us to characterize non-neutral genomic regions. In recent years researchers have not only been able to identify those gene regions that are most likely to be under selection in natural populations, but in some cases have then been able to identify the function of adaptive genes and, ultimately, to link phenotype to genotype across a range of environmental conditions (Table 1). Recent advances in our technological capabilities to capture markers at hundreds or thousands of loci, combined with ongoing improvements in the abilities of statistical tools and software to tease apart expectations based on neutral *versus* non-neutral models of evolution, have led to an explosion in the number of studies that incorporate or target non-neutral markers for questions in the fields of population genetics, molecular ecology, and evolutionary biology. This relatively recent ability to identify DNA regions and even genes under the influence of selection is rapidly closing the gap between molecular biologists who study mechanisms of gene transcription, translation, and regulation, and those biologists who are interested in addressing the role of selection in shaping biodiversity.

**Table 1.** Some of the candidate genes and the phenotypic traits that they influence in natural populations of non-model species. Adapted from [23].

Candidate Gene	Species	Phenotypic Trait	Reference
Calmodulin (CALM1)	Darwin's finches ( <i>Geospiza</i> spp.)	Beak morphology	[24]
cGMP-dependent protein kinase (KGPI)	Honey bee ( <i>Apis mellifera</i> )	Foraging behavior division of labour	[25]
Pantophysin (PAN1)	Cod ( <i>Gadus morhua</i> )	Growth	[26]
Phosphoglucose isomerase (PGI)	Glanville fritillary butterfly ( <i>Melitaea cinxia</i> )	Dispersal	[27]
Protein tyrosine phosphatase (PTEN)	Nasonia wasps	Longevity/incompatibility	[28]
Thyroid hormone receptor alpha	Ambystomatid salamanders	Timing of metamorphosis	[29]
Wingless	Heliconius butterflies	Wing patterning	[30]
Tyrosine related protein kinase I (TRYPI)	Soay sheep ( <i>Ovis aries</i> )	Coat colour polymorphism	[31]
Gp-9-odorant binding protein precursor	Fire ants	Social organization behaviour	[32]

## 2. Adaptive Genes and Genetic Diversity

Genetic diversity is a critical measure in population genetics because it can tell us a great deal about the current and likely future health of a population: low levels of genetic diversity can lead to inbreeding depression in the short-term, and to reduced evolutionary potential in the longer term. To date, the vast majority of genetic diversity estimates have been based on neutral markers. Although these data continue to provide us with invaluable insights into the overall levels of genetic variation within populations, in recent years they have been increasingly supplemented with data from adaptive genetic variation. Below, we shall discuss some of the ways in which these more recently acquired data have improved our understanding of inbreeding and evolutionary potential.

### 2.1. Inbreeding

Inbreeding occurs when individuals mate with their relatives. Depending on how closely related the parents are, the resulting inbred offspring will have a moderate to large proportion of alleles that are identical by descent, in other words they will exhibit a genome-wide increase in homozygosity relative to outbred individuals. This often leads to a reduction in fitness through a phenomenon that is known as inbreeding depression. Two processes can lead to inbreeding depression: dominance and overdominance. Dominance refers to the unmasking of deleterious recessive alleles that accompanies the overall increase in homozygosity; this occurs when unfavourable alleles that formerly occurred primarily in heterozygous individuals become more prevalent in a homozygous state, and therefore their deleterious effects are manifested. Overdominance, also known as heterozygote advantage, means that individuals that are heterozygous at a particular locus have higher fitness than individuals that are homozygous for either allele; the general increase in homozygosity that accompanies inbreeding means that beneficial heterozygotes become less common, once again reducing fitness.

Quantifying genome-wide heterozygosity is impractical, and is therefore typically inferred from a subset of loci such as microsatellites (e.g., [33]). Similarly, there are logistical constraints to quantifying fitness based on lifetime reproductive success, and therefore one or more surrogate measures such as clutch size, sperm count, or seed production is most commonly used (e.g., [34]). Multilocus genotype data and fitness estimates can be combined to test for heterozygosity fitness correlations (HFCs), which occur when there is a correlation between overall heterozygosity and a measure of fitness; a positive HFC suggests that low heterozygosity is reducing fitness within a population. Although a correlation between heterozygosity and fitness is widely accepted as evidence of inbreeding depression (reviewed in [35]), others have argued that HFCs that are based on only a small number of neutral markers may not reflect inbreeding depression because they are unlikely to represent genome-wide changes in homozygosity [36]. This was recently illustrated by a study of a free-ranging pedigreed population of the endangered takahe (*Porphyrio hochstetteri*) in which even relatively large numbers (>20) of microsatellite loci provided imprecise estimates of individual genome-wide heterozygosity [37]. The shortcomings of inbreeding estimates based on HFCs may therefore be twofold: first, it may be inappropriate to extrapolate genome-wide estimates of heterozygosity from small numbers of loci, and second, such extrapolation may be further weakened by the fact that heterozygosity is typically calculated on a subset of alleles that are neutral, and that have no functional significance in terms of adaptation and fitness (but see [38]).

An alternative approach to studying inbreeding depression is to seek specific information about its underlying molecular basis [39]. The first whole-genome study on the relationship between inbreeding and gene expression was done on *Drosophila melanogaster* [40]. The authors of that study compared gene expression in inbred and outbred lines of *D. melanogaster*, and determined that inbreeding changes transcription levels for a number of genes. The genes that showed differential expression in inbred lines were disproportionately involved in metabolism and stress responses, for example heat shock protein genes, which are involved in stress response, were upregulated more (*i.e.*, expressed in greater amounts) in inbred flies. This suggests that inbreeding acts like an environmental stressor that confers metabolic costs, and therefore leaves less energy for reproduction; in other words, inbreeding reduces fitness because stress responses are using energy that would otherwise be allocated to reproduction. This effect was even more pronounced when flies were placed in a high temperature environment, which conferred even greater stress and had the effect of further increasing the differential expression of heat-shock protein and metabolism genes in inbred *versus* outbred flies [41]. This latter study supports the idea that inbred organisms will be particularly challenged in stressful environments, and is consistent with an earlier study which found that inbreeding depression is on average 6.9 times higher for mammals in the wild compared to mammals that are kept in the relatively stress-free confines of captivity [42].

Demontis *et al.* [43] extended the study of gene expression in inbred *Drosophila* by investigating 40 SNPs in coding regions of genes that were identified in the earlier studies as being differentially expressed in inbred and outbred lines. They compared fast inbred lines, which took one generation to reach a predefined level of inbreeding, with slow inbred lines, which took 19 generations to reach the same level of inbreeding. Specifically, they wished to test the hypothesis that slow inbreeding leads to lower levels of inbreeding depression compared to fast inbreeding, because the former may allow more efficient purging of deleterious alleles and/or more efficient selection for heterozygotes. They found

a significantly higher level of genetic variation in the slow inbred lines compared to the fast inbred lines, including fewer homozygotes, and concluded that higher genetic diversity in slow inbred lines is a result of more efficient selection for heterozygotes (balancing selection) compared to the fast inbred lines. This indicates that, at least in this case, overdominance is likely the primary mechanism of inbreeding depression. Studies such as these strongly suggest that the use of “omic” approaches (e.g., genomics, proteomics) to unravel some of the cellular mechanisms behind inbreeding depression will feature more prominently in the near future [44].

## 2.2. Evolutionary Potential

Populations with low levels of genetic diversity should be less able to adapt to novel selection pressures, because a limited gene pool should decrease the likelihood that adaptive alleles will be present within a population. This expectation has been upheld by a growing number of studies. For example, populations of *Mercurialis annua* with reduced genetic diversity following range expansion had a reduced ability to respond to natural selection on a key life history trait [45]. In another example, laboratory populations of an estuarine crustacean (*Americamysis bahia*) with low genetic diversity had reduced fitness compared to populations with high genetic diversity; under stressful conditions the majority of low diversity populations went extinct, whereas populations with high genetic diversity were able to survive, albeit with reduced population sizes and less frequent reproduction [46].

The genetic diversity within populations is influenced by a range of factors, the most important of which is effective population size ( $N_e$ ), a measure that was introduced in the 1930s by Sewall Wright [47,48] who defined it as “the number of breeding individuals in an idealized population that would show the same amount of dispersion of allele frequencies under random genetic drift or the same amount of inbreeding as the population under consideration” [48]. In other words, the  $N_e$  of a population reflects the rate at which genetic diversity will be lost following genetic drift: only in an ideal population (sensu Wright) will the loss of genetic diversity as a result of drift occur at a rate that is commensurate with its actual population size. Understanding  $N_e$  is relevant to predictions about the viability of populations, because populations with low  $N_e$  are expected to have little evolutionary potential, and hence may be unable to respond to changing environmental conditions. However, this leaves us with a conundrum: estimates of  $N_e$  that are derived from molecular genetic data must be based on neutral markers (most commonly microsatellites) because  $N_e$  reflects the rate at which genetic drift—not selection—is altering allele frequencies from one generation to the next. As a result,  $N_e$  may tell us little about adaptive potential. This was recently illustrated by a study of the evolution of pesticide resistance in populations of the fruitfly *Drosophila melanogaster*, which concluded that resistance alleles have evolved quickly and repeatedly within multiple populations [49]. The authors of that study argue that such extensive evolutionary change would require a substantially larger (>100-fold) effective population size than had previously been identified. They further suggest that this discrepancy arises from the fact that estimates of  $N_e$  are usually derived from levels of standing variation which in turn is influenced by long-term population dynamics, whereas short-term effective population sizes are more relevant for rapid adaptation, and these may be much closer to  $N_c$ .

To date, most studies that have managed to quantify adaptively important genetic diversity have been based on three gene families whose diversity is maintained by balancing selection: major histocompatibility complex (MHC) loci in vertebrates [50], self-incompatibility loci in plants [51], and

sex loci in Hymenoptera [52]. However, these families collectively represent only a modest proportion of all adaptive genetic variation. Furthermore, it is not entirely clear what impacts the loss of diversity at these loci may have on the survival of populations, because the interactions between selection and drift have often resulted in correlations between levels of MHC variation and variation at neutral loci [53]. An understanding of the link between neutral and adaptive diversity, and their collective influence on long-term survival, must therefore be based on a larger number of adaptive genes. There are a number of different methods that can be used to identify these genes, some of which are outlined in Box 1. Depending on the methods used, researchers may be able to identify genes that appear to be under selection (candidate genes) on the basis of allele frequency distributions. An example of this was reported in a study of threespine stickleback (*Gasterosteus aculeatus*), in which the authors used next generation sequencing to genotype 100 fish from each of three freshwater and two oceanic populations at 45,000 single nucleotide polymorphisms (SNPs) [54]. The population genetic signal from neutral markers indicated that a panmictic oceanic population gave rise to freshwater populations multiple independent times, while outlier loci provided evidence that balancing and divergent selection occurred in parallel genomic regions in different freshwater populations with independent origins. A number of candidate genes involved in differentiation were identified, providing the basis for further studies of adaptation at these loci.

The search for a link between adaptive genes and evolutionary potential has been complicated in recent years by the growing awareness that gene expression can play a role in the adaptive divergence of populations. Gene expression is influenced by both genetic and environmental factors, with the relevant genetic factors being changes in either regulatory genes or *cis*-regulatory regions (as opposed to protein-coding regions) of functional genes. Examples in the literature which show how gene expression can influence the adaptive divergence of populations are growing. In one study, at least 4% of the compared transcriptome significantly differed between two sympatric ecotypes of the marine snail *Littorina saxatilis*. One of the identified transcripts was cytochrome c oxidase subunit I (COI), a mitochondrial gene involved in energy metabolism. This gene was overexpressed in the lower shore ecotype which is subject to the strongest wave action, and which therefore may need a particularly effective energy supply [55]. In another study, this time on the model species *Drosophila melanogaster*, population differentiation of gene expression (measure as  $Q_{ST}$ , or quantitative trait variation; see [56]) was not correlated with  $G_{ST}$  (an analogue of  $F_{ST}$  [57]) when based on all nucleotide polymorphisms; however, a correlation between  $Q_{ST}$  and  $G_{ST}$  was found when based on a more specific comparison in which  $G_{ST}$  was based solely on nucleotide differences in the 5' coding regions of genes, in other words the regions that contain regulatory sequences [58].

Overall, neutral molecular markers have some clear advantages when used to estimate the genetic diversity of populations: they are relatively easy to characterize, and they can provide unbiased estimates of random processes such as genetic drift [59,60]. However, microsatellites, which are currently the most widely used markers for inferring genetic diversity [1], may not accurately reflect the genome-wide genetic diversity of natural populations [61,62], in part because a relatively small number of microsatellite loci are usually characterized. Although neutral markers will undoubtedly continue to play an important role in at least initial estimates of heterozygosity, we will likely see in the future a greater emphasis on whole genome scans, patterns of gene expression, and the functional analyses of genes [44,63]. (See also Box 1).

**Box 1. Genomics techniques to generate sequence, genotype and gene expression data.**

Recent and ongoing developments in both analytical and statistical tools have advanced the capabilities of molecular ecologist and evolutionary biologists to address complex questions regarding population genetic structure and processes of adaptation. We do not intend here to thoroughly review new methodologies that can be used to identify, characterize, and analyse genomic information, because comprehensive reviews have been published elsewhere (e.g., [64–66]). However, it is fitting to briefly summarize a few of the relatively recent techniques that are facilitating the large scale analysis of both neutral and non-neutral markers, including next generation sequencing (NGS), novel genotyping strategies, and strategies for studying gene expression.

NGS methods permit the rapid sequencing of genomic DNA, mRNA, or cDNA at relatively low (and rapidly falling) costs. Using various platforms (the best known of which include those manufactured by Illumina, Roche and ABI), it is possible to generate hundreds of thousands to millions of reads from a single lane, with read lengths between approximately 30 and several hundred base pairs. Each lane may contain a single sample, or it may contain pooled samples, each of which may be labeled with a unique nucleotide tag. This permits the development of large databases of genomic information from model- and non-model organisms alike, and is also driving the demand for the expansion of the bioinformatics field. Third generation sequencing technologies, set to be released over the next several years, promise to increase read length to approximately 10,000 bp at much greater speed (e.g., [67]), which will greatly increase the ease and accuracy of *de novo* assembly.

In many studies of non-model organisms, NGS is currently used for marker discovery because the comparison of whole genomes or whole transcriptomes (or expressed sequence tags; ESTs) remains expensive, time consuming, and analytically daunting. Recent advances in genotyping technologies have also improved the economy of including a large number of loci in various types of studies (reviewed by [68]). Both single nucleotide polymorphisms (SNPs) and microsatellites (or simple sequence repeats; SSRs) remain the most commonly used types of markers, particularly for population genomics studies, or gene association studies. SNPs are useful because they are ubiquitous in most genomes (and can therefore yield excellent coverage of the genome), and are relatively cost-effective and easy to genotype because most are biallelic (only two alternative nucleotides at a single SNP). A great variety of different commercial genotyping methods are offered; these include commercially available SNP microchips for model organisms and common agricultural species, and commercial genotyping services, such as the GoldenGate Assay offered by Illumina [69]. SNPs can also be genotyped at a small scale in-house using commercially available kits, but the costs of doing so are generally higher than outsourced options. On the other hand, microsatellites can yield a much greater amount of information per locus because they often exhibit a large number of alleles. However they are generally more time consuming and expensive than SNPs to genotype (with a smaller variety of commercial options), and this tends to be reflected in less extensive genome coverage in studies incorporating microatellites.

Although the genome coverage for microsatellites is lower than that of SNPs, they continue to be widely used because of their tremendous utility in molecular ecology [70]. One limitation of microsatellites, however, is the time and expense required for *de novo* development [71,72]; this is particularly problematic in some taxonomic groups such as Lepidoptera [73]. In addition, the PCR primers that are

used to amplify microsatellite loci are often species-specific and therefore cannot be used on multiple taxa (Ellis and Burke, 2007). However, a more recently developed approach for characterizing microsatellites uses publicly available expressed sequence tag (EST) databases. An EST represents a single sequencing run starting from one end of a cDNA, and yields a sequence that is a small portion of the expressed gene. The growing use of NGS means that EST databases such as the National Center for Biotechnology Information (NCBI) EST database (dbEST; [74]) can be increasingly used for efficiently developing so-called EST-SSRs for a wide variety of taxa (reviewed in [75]). The evidence so far suggests that EST-SSRs are more likely to be transferrable between taxa than the more traditionally-developed SSRs which are isolated from a species' genome in an anonymous manner [76–78]. EST-SSRs may also facilitate the generation of molecular markers that are directly associated with a trait of interest, and are therefore increasingly common in studies of molecular ecology (reviewed in [79]).

Advances also continue to be made in the area of gene expression studies, which can be helpful for the identification of important functional genes that may be under selection. NGS now permits direct transcriptome sequencing (RNA-seq), which can provide quantitative information on gene expression in different tissues, individuals, or populations. The number of reads generated for any particular transcript is expected to be proportional to the level of transcription, so that the so-called “read depth” can be used to generate information on relative transcription levels from different samples. Custom microarrays can also be commercially constructed for non-model organisms at reasonable costs; an investigator needs only to input the desired oligonucleotide probe sequences into an online database (these are often designed based on an initial NGS sequencing run of the transcriptome), and arrays are printed using an automated system. Once identified, the expression of individual candidate genes in various individuals and/or tissues can be verified using quantitative PCR (qPCR).

### 3. Genetic Differentiation

One of the most important determinants of microevolutionary change is gene flow between populations, because migrants typically increase  $N_e$  by introducing novel alleles, whereas isolated populations are more susceptible to the effects of genetic drift and therefore loss of alleles. Gene flow can therefore be considered an evolutionary facilitator because it increases the gene pool upon which selection can act. Conversely, gene flow can be viewed as an evolutionary deterrent because the continued introduction of alleles may counter local adaptation; the latter has been proposed as one explanation for the limits of species' ranges ([80], and references therein). Thus, there exists the potential for tension between adaptation and gene flow, particularly at range margins, and the outcome will partly depend on the strength of selection pressure *versus* the extent of gene flow. This may result in different patterns of differentiation in adaptive *versus* non-adaptive genes, although before exploring that possibility, it is necessary to consider how we may determine which genes appear adaptive across a landscape.

#### 3.1. Identifying Adaptively Divergent Genes

Migration and drift are expected to have approximately equal effects on all neutral loci, whereas the effects of selection will vary between neutral and non-neutral loci. All neutral loci may therefore show



similar levels of genetic divergence among populations (once variable mutation rates have been accounted for), whereas non-neutral loci (or loci linked to non-neutral loci) are expected to show anomalous levels of divergence. These anomalous levels may be either unusually high or unusually low, depending on the type of selection that the relevant genes have been subjected to; for example, directional selection will increase population differentiation if different alleles are selected for in different populations, whereas balancing selection may decrease population differentiation by maintaining the same suite of alleles in multiple populations. A comparison of multiple measures of population differentiation, each based on a different locus, may reveal a marker with unusual levels of differentiation; this is often referred to as an outlier, and if the marker is found within a coding region, the latter may be considered a candidate gene [81,82]. An outlier may be used to identify a genetic region that is either directly under selection, or is linked to a gene that is under selection [83,84]. Approaches for using genome scans to identify markers of potential adaptive significance are comprehensively reviewed in [85]. However, an element of caution must be introduced to this approach because differentiating between adaptive and neutral genes can be problematic in expanding populations: expansions can impact neutral allele frequencies in ways that are similar to the effects of directional selection [86]. In addition, false positives are common even with the most rigorous analytical methods [87].

The ease with which data can now be simultaneously collected for many markers means that studies of discordant genetic differentiation (*i.e.*, the identification of outlier loci) have increased in recent years. A growing number of these studies are based on a genome scanning approach, which means that hundreds or even thousands of markers are used to sample broadly from across the genome (as opposed to a handful of microsatellite loci), and this increases the likelihood of identifying markers linked to genes that are under the influence of natural selection. A number of such studies have been based on dominant markers, specifically amplified fragment length polymorphisms (AFLPs) (e.g., [88–90]). More recently, studies have also been taking advantage of the advent of high-throughput SNP genotyping technologies which can generate data from thousands of markers [91,92] (but see [93] for a discussion of some of the challenges associated with using SNPs).

Adaptive genes can also be inferred from clinal gradients in allele frequencies, which arise when allele frequencies vary along an environmental cline in a seemingly adaptive manner. Studies that have identified such clines sometimes target specific genes that may be expected to show signatures of natural selection. One example of this was the discovery of a latitudinal gradient in Chinook salmon (*Oncorhynchus tshawytscha*) clock gene allele frequencies which corresponded to latitudinal variation in reproductive timing; because clock genes are known to be involved with the regulation of circadian rhythm, the authors of this study had an a priori reason to expect that such a cline may exist [94]. Another approach is to use genome scans to search for genes that may be correlated with environmental clines; as with outlier detections, these scans are increasingly based on high-throughput genotyping of hundreds or thousands of SNP markers. This formed the basis of a study of loblolly pine (*Pinus taeda*) sampled across its range: the frequencies of several SNPs, identified from a total of 1730 loci, corresponded with aspects of geography, temperature, growing degree-days, precipitation and aridity [95]. The authors were then able to assign putative function to a number of SNPs by using annotated orthologs from *Arabidopsis*. Several SNPs that were correlated with climatic variables

(such as temperature and precipitation) were located within abiotic stress response genes ranging from transmembrane proteins to proteins involved in sugar metabolism.

A note of caution about using clinal patterns to infer patterns of selection is that random events or processes such as founder effects, isolation by distance, or secondary contact of populations that have previously differentiated by genetic drift can create an illusion of an adaptive cline [96]. As with outliers, conclusions may be strengthened by common garden experiments or geographically independent replicates. The latter approach revealed that an insulin signalling gene, the Insulin-like Receptor (InR), had replicate latitudinal clines in allele frequencies among *Drosophila melanogaster* populations in both Australia and North America [97]. Replicate findings also strengthened conclusions regarding parallel temperature-associated clines in SNPs which were found in Atlantic cod (*Gadus morhua*) populations in the eastern and western north Atlantic: in both regions, allele frequencies at temperature-associated loci were significantly correlated with the ocean temperature, whereas neutral markers showed no such correlation [98]. See also Box 2 for other approaches that can be used to infer natural selection from genetic data; these are summarized in Table 2.

#### **Box 2. Genetic signatures of selection.**

Even in the absence of broad geographical sampling, evidence for natural selection may be found in patterns of mutation. The rate of evolution in protein-coding genes is commonly assessed using two quantities:  $dN$  (rate of nonsynonymous substitutions per nonsynonymous site, also called  $Ka$ ) and  $dS$  (rate of synonymous substitutions per synonymous site, also called  $Ks$ ; [99]). Synonymous substitutions usually occur in the third position of each codon within a gene, and do not alter the encoded amino acid. In contrast, nonsynonymous substitutions (which usually result from a mutation in the first or second position within a codon) alter the encoded amino acid, and are therefore more likely to be deleterious; thus, nonsynonymous substitutions are more likely to be purged from the gene pool via purifying selection. As a result, genes under the influence of purifying selection are expected to have a relatively low number of nonsynonymous substitutions relative to synonymous substitutions (this is referred to as the  $dN:dS$  ratio), while genes not under the influence of selection are expected to have a  $dN:dS$  ratio of approximately 1:1. Conversely, if  $dN:dS$  is greater than 1, positive selection may be acting on the coding region in question (first proposed by [100]; see also [101–103]).

The strategy of  $dN:dS$  comparisons was originally developed to compare sequence evolution between orthologs from different lineages, and polymorphisms within lineages were ignored [104]. For example, Nam, *et al.* [105] estimated rates of non-synonymous and synonymous substitutions in the chicken and zebra finch genomes using one lizard and three mammalian species as outgroups. The authors identified 11,225 orthologs between the two avian genomes. Overall the ratio  $dN:dS$  was 0.152 according to the pairwise comparison between chicken and finch, indicating widespread purifying selection. The authors then sought to identify genes (and their associated functional categories) that were under positive selection in only the chicken or finch genome. Nine hundred and thirty-six genes showed signatures of positive signatures of selection ( $dN > dS$ ) in the finch lineage, and 883 in the chicken lineage.

Extended models of nucleotide evolution have been used to investigate possible instances of more recent selection acting within and among populations. Instead of considering only differences between species, these models incorporate data on synonymous and nonsynonymous polymorphism within populations, compared to the number of fixed differences between populations. Under such models, balancing selection and diversifying selection are both expected to maintain an excess of mid-frequency alleles within populations. If balancing selection is acting across populations of interest and their outgroups, a deficit of fixed differences between lineages is expected. Conversely, in the case of diversifying selection, no shared polymorphisms are expected between lineages. For example, Ersoz, *et al.* [106] sequenced 41 candidate genes (thought to be involved in plant-pathogen interactions) from 32 haploid seed megagametophytes of loblolly pine (*Pinus taeda*), using two Scots pine (*Pinus sylvestris*) seed samples as outgroups. The authors proposed various expectations regarding patterns of nucleotide diversity within these candidate gene regions, based on various models of co-evolution between plants and their pathogens. They predicted, for example, that if loblolly pine and their pathogens were engaged in an evolutionary arms race, plant genes involved in resistance would continually develop novel nonsynonymous alleles that would subsequently be fixed, resulting in successive selective sweeps over evolutionary time. Under this scenario, the authors expected to see an excess of nonsynonymous substitutions involved in resistance (directional selection), and a low level of nucleotide diversity (indicative of a selective sweep; the implications of selection on genetic diversity are discussed more extensively in [107]). The authors found that four of the 41 candidate genes examined met the expectations of the arms race hypothesis.

Caution needs to be used in the application of neutrality tests based on  $dN:dS$  estimates for population level inferences because the inferences that can be drawn from the data are not always clear [108,109]. For example, negative selection against slightly deleterious nonsynonymous mutations can lead to a relative excess of rare variants in a population, and this can be confused with balancing or diversifying selection. Also, spurious signals of selection can be detected because demographic processes (for example small population sizes or population bottlenecks followed by expansion) can sometimes lead to the fixation of slightly deleterious alleles as a result of genetic drift.

Signatures of selection may also be inferred by measuring linkage disequilibrium (LD) across the genome. Selective sweeps are expected to be associated with a high degree of linkage disequilibrium around the locus under selection, and (based on the principle of genetic hitchhiking; [110]) long haplotypes are expected to reflect recent selective sweeps; in other words, adaptive alleles are swept rapidly to fixation, and there is insufficient time for recombination to break up surrounding nucleotide combinations (e.g., [111–113]). However, long haplotypes are expected to break up relatively quickly over evolutionary time, so older selective sweeps may not be easy to detect using this approach.

Studies of the human genome provide some of the most widely cited examples of the use of linkage disequilibrium for the identification of regions under selection. For example, Sabeti *et al.* [113] first used this approach to study LD around two loci implicated in human resistance to malaria. The authors compared actual patterns of LD around these loci to expectations that were generated based on simulations that accounted for demographic patterns under neutral models of evolution. They found that haplotypes in the regions of these two loci were much longer than expected according to neutral models.

**Table 2.** A summary of approaches used to identify genomic regions under the influence of selection.

Approach	Target Region	Strategy
Identify regions under selection based on genomic information only	Protein-coding regions	Compare rate of nonsynonymous ( $dN$ ) <i>versus</i> synonymous ( $dS$ ) substitutions (neutrality test) between species or populations
	Whole Genome	Linkage-disequilibrium based approaches Population Differentiation Levels Comparisons of Nucleotide Diversity between different genome regions (not discussed in detail here)
Search for correlations between phenotypes and allele frequencies	Whole Genome	QTL mapping Genome-Wide Marker Association (GWA) Studies
Search for correlations between environmental variables and allele frequencies.	Whole Genome	Genome Scanning combined with the identification of outlier loci.

### 3.2. Model-Based Advances

As noted above, one large stumbling block in the identification of non-neutral markers has been the difficulty in accounting for complex population demography, including historical patterns of population expansion and contraction, unequal migration rates between populations, and inbreeding. A failure to account for such demographic processes can lead to either spurious signatures of selection at loci that are in fact neutral [86], or to a lack of power to detect loci under selection. Advances in the sophistication of statistical tools and models available for the analysis of molecular data are facilitating a much more intricate and comprehensive understanding of the processes that shape neutral and adaptive genetic variation. For example, currently available Bayesian approaches [114,115] and coalescence models (e.g., [116,117]) incorporate relatively realistic scenarios in which the migration rate can differ between pairs of subpopulations, and multiple historical population bottlenecks and expansions can be accounted for. Additionally, a number of software packages incorporate spatial and environmental data with genetic data to identify loci that are associated with specific environmental variables (e.g., [118]). Many of these model-based advances are reviewed in detail by [119].

### 3.3. Isolation by Adaptation

Finally, when examining patterns of population differentiation at neutral *versus* non-neutral loci, it is important to keep in mind that gene flow will not necessarily ensure that non-adaptive genes are continually exchanged between even proximate populations. Although the differentiation of neutral markers is driven primarily by stochastic processes, whereas that of non-neutral markers is driven by both selective and stochastic processes (e.g., [120]), natural selection can also influence the distribution of markers that are neither being directly selected, nor are linked to regions under selection [121]. This arises if divergent selection is sufficiently strong to promote reproductive isolation between populations. In these cases, a reproductive barrier will then create a barrier to gene

flow that results in the potentially genome-wide differentiation of populations following genetic drift. This will lead to an inverse correlation between gene flow and the adaptive divergence of populations, and thus a positive association between the phenotypic divergence and neutral molecular genetic differentiation of populations following a pattern that is known as isolation by adaptation (IBA) [122,123]. In other words, populations will not only diverge at adaptive loci as a direct result of selection, but will also diverge at neutral loci as a direct result of drift, which is indirectly a result of selection via a reproductive barrier. This pattern was identified in a semi-natural experiment in which adjacent populations of sweet vernal grass (*Anthoxanthum odoratum*) had diverged from one another as a result of adaptation to different nutrient additions. Genetic differentiation was evident at outlier loci, and also across a wider survey of putatively neutral loci. This was interpreted as evidence that the selection pressures from varying combinations of nutrient additions in different plots was strong enough to cause reproductive isolation between populations, which in turn has led to neutral genetic differentiation as a result of genetic drift [89]. Studies such as this, which have identified candidate adaptive molecular markers, should in the future find it increasingly feasible to take the next step and characterize the phenotypic outcomes of alternative genotypes (Box 3).

**Box 3. Linking genotype to phenotype.**

The identification of outlier loci, or candidate genes under selection, must always retain an element of speculation until the genetic region in question has been directly linked to a phenotype that is subject to selection. Both quantitative trait locus (QTL) mapping and genome-wide association studies are used to identify correlations between specific marker alleles and phenotypic traits of interest. QTL mapping is perhaps the oldest form of genome scanning, and has been widely used in studies of genetic model organisms and commercially important species for at least two decades (e.g., [124]). The aim of QTL analysis is to use a large number of individuals from a known pedigree that show considerable variation in phenotypic trait(s) of interest, and to genotype them across a large number of loci using a set of markers that cover the whole genome. Usually QTL mapping is carried out using an F<sub>2</sub> or backcrossed family (BC) from a known cross, or sometimes using recombinant inbred lines (RILs). A linkage map is constructed based on observed rates of recombination between markers in the mapping population, and measurements of phenotypic traits are made from the mapping population under standardized conditions. Various statistical methods are used to calculate estimated recombination rates between marker loci and the QTL that control for the phenotypic trait(s) of interest. Depending on the genetic architecture of the trait and the experimental design, one or more QTL can be identified and the relative proportion of phenotypic variation explained by each QTL can be calculated. Also, interactions between different QTL can sometimes be identified (e.g., epistasis or pleiotropy). For example, Latta *et al.* [125] developed 179 RILs from a cross between moist- and dry-associated ecotypes of *Avena barbata* (wild oats). Two loci accounted for more than half of the variation in plant fitness across both moist and dry environments, and no genotype-by-environment interactions were detected with regard to the direction of selection at these loci.

Studies that link phenotypes to genotypes have recently been extended on a more widespread scale to populations of unrelated individuals, in the form of genome-wide association (GWA) studies. GWA studies have been widely used in the study of human disease (recently reviewed by [126]), but have only more recently been applied to other organisms (e.g., *Arabidopsis*, [127,128]; dogs, [129]). GWA studies generally provide higher resolution than traditional QTL studies, because recombination between loci is generally greater in large populations of unrelated individuals than that in F<sub>2</sub> or BC families. A recent GWA study of 15 different morphological traits in barley incorporated 500 different cultivars that were genotyped with 1536 SNPs [130]. The authors identified 18 genomic regions associated with the 15 traits (most of the traits were associated with a single genetic locus). Based on these results, the authors selected one phenotypic trait—anthocyanin pigmentation, which is involved in determining seed colour—for more detailed fine-scale mapping using a QTL approach after crossing two of the cultivars included in the original GWA study to create a mapping population. This allowed them to identify the specific mutation involved in generating the variation explained by the candidate locus identified in their original GWA study.

### 3.4. Future Work

There are a number of exciting avenues for future research that will allow researchers to increasingly incorporate data on adaptive genes into studies of molecular ecology. The sequencing of a greater number of genomes from non-model organisms will be among the most obvious and rapid advancement in genomics over the coming years, and this will provide opportunities for the identification of non-neutral markers in numerous and diverse species. Furthermore, cross-validation using a combination of different approaches will lead to a greater understanding of the interaction between demographic processes and selection, the interaction between selection at linked loci in the genome, and fine-scale patterns of molecular evolution. Studies of signatures of selection in the human genome have led the way in this regard (reviewed by [131]), and can serve as models for similar studies in other organisms.

There are two widespread challenges that arise in many studies that target genes that are under the influence of selection. First, studies in non-model organisms now frequently hone in on relatively broad genomic regions that are under selection, but it remains difficult to actually identify the genes (or the mutations) that are subject to selection. Increasing the density of markers in genome scans is paramount to overcoming this problem, and validating signals of selection from particular genes using multiple methods should also help. Second, once a candidate gene has been identified, it may have no known annotated function. This occurs because annotated functional genes from model organisms may not overlap with genes that are under the influence of selection in non-model organisms that are being studied in the context of ecological and evolutionary genomics. Advances in identifying the functional significance of genes subject to selection will require ongoing integration between genomics methods and functional experiments that provide mechanistic insights into molecular pathways controlled by candidate genes (reviewed by [132]). Studies of gene expression, which can be carried out using microarrays, quantitative PCR (qPCR), and comparative sequencing of the transcriptome, can also provide evidence of differential expression of candidate genes. Genome wide scans based on complete genomic data—already close to fruition in humans—will permit a much more detailed understanding

of fine scale processes involved in genome evolution (reviewed by [131]). Finally, future work on adaptive genes will likely also focus on epigenetic modifications in DNA methylation and DNA-associated proteins such as histones, which can vary among individuals and populations of the same species. The heritability of some of these modifications is now widely accepted, and means that heritable variation in ecologically important phenotypic traits may be apparent even in the absence of DNA polymorphisms (see [63]). We are therefore entering a truly exciting time in molecular ecology, in which we seem poised to make numerous important discoveries about the interactions between genotypes and phenotypes in varied—and often rapidly changing—environmental conditions.

#### 4. Conclusions

To date, the vast majority of studies in the field of molecular ecology have been based on neutral molecular markers, in other words genetic regions that do not directly influence fitness. These markers have given us invaluable insights into parameters such as genetic diversity within populations, genetic differentiation among populations, inbreeding, and demographic events; however, they provide limited insight into adaptive evolution and evolutionary potential. In recent years, developments such as next-generation sequencing mean that we have become increasingly able to develop non-neutral markers by targeting genetic regions that are directly influenced by natural selection, which means that a growing number of studies have been able to use molecular genetic data to directly study natural selection and local adaptation of natural populations from a wide range of taxonomic groups. In addition, researchers are increasingly able to link genotypes to phenotypes under a range of environmental conditions. More specifically, these data have provided numerous examples of how local adaptation shapes the genetic diversity and differentiation of populations, and have also provided insight into some of the mechanistic processes behind inbreeding depression, and some of the demographic processes that are associated with adaptive evolutionary change. Although researchers will continue to use neutral molecular markers because of their ease of use and their relatively straightforward histories (which can allow more accurate inferences of past demographic events), future studies will be increasingly likely to supplement data from neutral markers with data from markers that are influenced by natural selection.

#### References

1. Freeland, J.R.; Kirk, H.; Petersen, S. *Molecular Ecology*, 2nd ed.; Wiley & Sons: Chichester, UK, 2011.
2. Freeland, J.R.; Gillespie, J.; Ciotir, C.; Dorken, M.E. Conservation genetics of Hill's thistle (*Cirsium hillii*). *Botany* **2010**, *88*, 1073–1080.
3. Silvertown, J.; Biss, P.M.; Freeland, J. Community genetics: Resource addition has opposing effects on genetic and species diversity in a 150-year experiment. *Ecol. Lett.* **2009**, *12*, 165–170.
4. Storfer, A.; Murphy, M.A.; Spear, S.F.; Holderegger, R.; Waits, L.P. Landscape genetics: Where are we now? *Mol. Ecol.* **2010**, *19*, 3496–3514.
5. Waser, P.M.; Hadfield, J.D. How much can parentage analyses tell us about precapture dispersal? *Mol. Ecol.* **2011**, *20*, 1277–1288.

6. Ferriol, M.; Pichot, C.; Lefevre, F. Variation of selfing rate and inbreeding depression among individuals and across generations within an admixed *Cedrus* population. *Heredity* **2011**, *106*, 146–157.
7. Freeland, J.R.; Lodge, R.J.; Okamura, B. Sex and outcrossing in a sessile freshwater invertebrate. *Freshwater Biol.* **2003**, *48*, 301–305.
8. Raye, G.; Miquel, C.; Coissac, E.; Redjadj, C.; Loison, A.; Taberlet, P. New insights on diet variability revealed by DNA barcoding and high-throughput pyrosequencing: Chamois diet in autumn as a case study. *Ecol. Res.* **2011**, *26*, 265–276.
9. Zeale, M.R.K.; Butlin, R.K.; Barker, G.L.A.; Lees, D.C.; Jones, G. Taxon-specific PCR for DNA barcoding arthropod prey in bat faeces. *Mol. Ecol. Res.* **2011**, *11*, 236–244.
10. Bottger-Schnack, R.; Machida, R.J. Comparison of morphological and molecular traits for species identification and taxonomic grouping of oncaeid copepods. *Hydrobiologia* **2011**, *666*, 111–125.
11. Hawlitschek, O.; Porch, N.; Hendrich, L.; Balke, M. Ecological niche modelling and nDNA sequencing support a new, morphologically cryptic beetle species unveiled by DNA barcoding. *PLoS One* **2011**, *6*, e16662. doi:10.1371/journal.pone.0016662.
12. Freeland, J.R.; Rimmer, V.K.; Okamura, B. Evidence for a residual post-glacial founder effect in a highly dispersive freshwater invertebrate. *Limnol. Oceanog.* **2004**, *49*, 879–883.
13. Pepper, M.; Fujita, M.K.; Moritz, C.; Keogh, J.S. Palaeoclimate change drove diversification among isolated mountain refugia in the Australian arid zone. *Mol. Ecol.* **2011**, *20*, 1529–1545.
14. Ballentine, B.; Greenberg, R. Common garden experiment reveals genetic control of phenotypic divergence between swamp sparrow subspecies that lack divergence in neutral genotypes. *PLoS One* **2010**, *5*, e10229. doi:10.1371/journal.pone.0010229.
15. Kawakami, T.; Morgan, T.J.; Nippert, J.B.; Ocheltree, T.W.; Keith, R.; Dhakal, P.; Ungerer, M.C. Natural selection drives clinal life history patterns in the perennial sunflower species, *Helianthus maximiliani*. *Mol. Ecol.* **2011**, *20*, 2318–2328.
16. Richter-Boix, A.; Quintela, M.; Segelbacher, G.; Laurila, A. Genetic analysis of differentiation among breeding ponds reveals a candidate gene for local adaptation in *Rana arvalis*. *Mol. Ecol.* **2011**, *20*, 1582–1600.
17. Cassel-Lundhagen, A.; Tammaru, T.; Windig, J.J.; Ryrholm, N.; Nylin, S. Are peripheral populations special? Congruent patterns in two butterfly species. *Ecography* **2009**, *32*, 591–600.
18. Pampoulie, C.; Danielsdottir, A.K.; Storr-Paulsen, M.; Hovgard, H.; Hjørleifsson, E.; Steinarsson, B.A. Neutral and Nonneutral Genetic Markers Revealed the Presence of Inshore and Offshore Stock Components of Atlantic Cod in Greenland Waters. *Trans. Am. Fish. Soc.* **2011**, *140*, 307–319.
19. Jump, A.S.; Marchant, R.; Penuelas, J. Environmental change and the option value of genetic diversity. *Trends Plant Sci.* **2009**, *14*, 51–58.
20. Jump, A.S.; Penuelas, J. Running to stand still: Adaptation and the response of plants to rapid climate change. *Ecol. Lett.* **2005**, *8*, 1010–1020.
21. Eckert, C.G.; Samis, K.E.; Loughheed, S.C. Genetic variation across species' geographical ranges: The central-marginal hypothesis and beyond. *Mol. Ecol.* **2008**, *17*, 1170–1188.



22. Gebremedhin, B.; Ficetola, G.F.; Naderi, S.; Rezaei, H.R.; Maudet, C.; Rioux, D.; Luikart, G.; Flagstad, O.; Thuiller, W.; Taberlet, P. Frontiers in identifying conservation units: From neutral markers to adaptive genetic variation. *Anim. Conserv.* **2009**, *12*, 107–109.
23. Piertney, S.B.; Webster, L.M.I. Characterising functionally important and ecologically meaningful genetic diversity using a candidate gene approach. *Genetica* **2010**, *138*, 419–432.
24. Abzhanov, A.; Kuo, W.P.; Hartmann, C.; Grant, B.R.; Grant, P.R.; Tabin, C.J. The calmodulin pathway and evolution of elongated beak morphology in Darwin's finches. *Nature* **2006**, *442*, 563–567.
25. Ben-Shahar, Y. The foraging gene, behavioral plasticity, and honeybee division of labor. *J. Compar. Phys.* **2005**, *191*, 987–994.
26. Case, R.A.J.; Hutchinson, W.F.; Hauser, L.; Buehler, V.; Clemmesen, C.; Dahle, G.; Kjesbu, O.S.; Moksness, E.; Ottera, H.; Paulsen, H.; Svasand, T.; Thorsen, A.; Carvalho, G.R. Association between growth and Pan I genotype within Atlantic cod full-sibling families. *Trans. Am. Fish. Soc.* **2006**, *135*, 241–250.
27. Haag, C.R.; Saastamoinen, M.; Marden, J.H.; Hanski, I. A candidate locus for variation in dispersal rate in a butterfly metapopulation. *Proc. Biol. Sci.* **2005**, *272*, 2449–2456.
28. Baudry, E.; Desmadril, M.; Werren, H. Rapid adaptive evolution of the tumor suppressor gene Pten in an insect lineage. *J. Mol. Evol.* **2006**, *62*, 738–744.
29. Voss, S.R.; Prudic, K.L.; Oliver, J.C.; Shaffer, H.B. Candidate gene analysis of metamorphic timing in ambystomatid salamanders. *Mol. Ecol.* **2003**, *12*, 1217–1223.
30. Kronforst, M.R.; Young, L.G.; Kapan, D.D.; McNeely, C.; O'Neill, R.J.; Gilbert, L.E. Linkage of butterfly mate preference and wing color preference cue at the genomic location of wingless. *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 6575–6580.
31. Gratten, J.; Beraldi, D.; Lowder, B.V.; McRae, A.F.; Visscher, P.M.; Pemberton, J.M.; Slate, J. 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* **2007**, *274*, 619–626.
32. Gotzek, D.; Ross, K.G. Genetic regulation of colony social organization in fire ants: An integrative overview. *Q. Rev. Biol.* **2007**, *82*, 201–226.
33. Ficetola, G.F.; Garner, T.W.J.; Wang, J.L.; de Bernardi, F. Rapid selection against inbreeding in a wild population of a rare frog. *Evol. Appl.* **2010**, *4*, 30–38.
34. Kupper, C.; Kosztolanyi, A.; Augustin, J.; Dawson, D.A.; Burke, T.; Szekely, T. Heterozygosity-fitness correlations of conserved microsatellite markers in Kentish plovers *Charadrius alexandrinus*. *Mol. Ecol.* **2010**, *19*, 5172–5185.
35. Szulkin, M.; Bierne, N.; David, P. Heterozygosity-fitness correlations: A time for reappraisal. *Evolution* **2010**, *64*, 1202–1217.
36. Chapman, J.R.; Nakagawa, S.; Coltman, D.W.; Slate, J.; Sheldon, B.C. A quantitative review of heterozygosity-fitness correlations in animal populations. *Mol. Ecol.* **2009**, *18*, 2746–2765.
37. Grueber, C.E.; Waters, J.M.; Jamieson, I.G. The imprecision of heterozygosity-fitness correlations hinders the detection of inbreeding and inbreeding depression in a threatened species. *Mol. Ecol.* **2010**, *20*, 67–79.

38. Thoss, M.; Imonen, P.; Musolf, K.; Penn, D.J. Major histocompatibility complex heterozygosity enhances reproductive success. *Mol. Ecol.* **2011**, *20*, 1546–1557.
39. Paige, K.N. The Functional Genomics of Inbreeding Depression: A New Approach to an Old Problem. *Bioscience* **2010**, *60*, 267–277.
40. Kristensen, T.N.; Sorensen, P.; Kruhoffer, M.; Pedersen, K.S.; Loeschcke, V. Genome-wide analysis on inbreeding effects on gene expression in *Drosophila melanogaster*. *Genetics* **2005**, *171*, 157–167.
41. Kristensen, T.N.; Sorensen, P.; Pedersen, K.S.; Kruhoffer, M.; Loeschcke, V. Inbreeding by environmental interactions affect gene expression in *Drosophila melanogaster*. *Genetics* **2006**, *173*, 1329–1336.
42. Crnokrak, P.; Roff, D.A. Inbreeding depression in the wild. *Heredity* **1999**, *83*, 260–270.
43. Demontis, D.; Pertoldi, C.; Loeschcke, V.; Mikkelsen, K.; Axelsson, T.; Kristensen, T.N. Efficiency of selection, as measured by single nucleotide polymorphism variation, is dependent on inbreeding rate in *Drosophila melanogaster*. *Mol. Ecol.* **2009**, *18*, 4551–4563.
44. Kristensen, T.N.; Pedersen, K.S.; Vermeulen, C.J.; Loeschcke, V. Research on inbreeding in the “omic” era. *Trends Ecol. Evol.* **2009**, *25*, 44–52.
45. Pujol, B.; Pannell, J.R. Reduced responses to selection after species range expansion. *Science* **2008**, *321*, 96.
46. Markert, J.A.; Champlin, D.M.; Gutjahr-Gobell, R.; Gear, J.S.; Kuhn, A.; McGreevy, T.J.; Roth, A.; Bagley, M.J.; Nacci, D.E. Population genetic diversity and fitness in multiple environments. *BMC Evol. Biol.* **2010**, *10*, 205. doi:10.1186/1471-2148-10-205.
47. Wright, S. Evolution in mendelian populations. *Genetics* **1931**, *16*, 97–159.
48. Wright, S. Size of population and breeding structure in relation to evolution. *Science* **1938**, *87*, 430–431.
49. Karasov, T.; Messer, P.W.; Petrov, D.A. Evidence that adaptation in *Drosophila* is not limited by mutation at single sites. *PLoS Genet.* **2010**, *6*, e1000924. doi:10.1371/journal.pgen.1000924.
50. Spurgin, L.G.; Richardson, D.S. How pathogens drive genetic diversity: MHC, mechanisms and misunderstandings. *Proc. R. Soc. B* **2010**, *277*, 979–988.
51. Charlesworth, D.; Bartolome, C.; Schierup, M.H.; Mable, B.K. Haplotype structure of the stigmatic self-incompatibility gene in natural populations of *Arabidopsis lyrata*. *Mol. Biol. Evol.* **2003**, *20*, 1741–1753.
52. Cho, S.C.; Huang, Z.Y.; Green, D.R.; Smith, D.R.; Zhang, J.Z. Evolution of the complementary sex-determination gene of honey bees: Balancing selection and trans-species polymorphisms. *Genome Res.* **2006**, *16*, 1366–1375.
53. Radwan, J.; Biedrzycka, A.; Babik, W. Does reduced MHC diversity decrease viability of vertebrate populations? *Biol. Conserv.* **2010**, *143*, 537–544.
54. Hohenlohe, P.A.; Bassham, S.; Etter, P.D.; Stiffler, N.; Johnson, E.A.; Cresko, W.A. Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. *PLoS Genet.* **2010**, *6*, e1000862. doi:10.1371/journal.pgen.1000862.
55. Martinez-Fernandez, M.; Bernatchez, L.; Rolan-Alvarez, E.; Quesada, H. Insights into the role of differential gene expression on the ecological adaptation of the snail *Littorina saxatilis*. *BMC Evol. Biol.* **2010**, *10*, 356.

56. Merila, J.; Crnokrak, P. Comparison of genetic differentiation at marker loci and quantitative traits. *J. Evol. Biol.* **2001**, *14*, 892–903.
57. Takahata, N.; Nei, M. Fst and Gst Statistics in the Finite Island Model. *Genetics* **1984**, *107*, 501–504.
58. Kohn, M.H.; Shapiro, J.; Wu, C.I. Decoupled differentiation of gene expression and coding sequence among Drosophila populations. *Genes Genet. Syst.* **2008**, *83*, 265–273.
59. Luikart, G.; England, P.R.; Tallmon, D.; Jordan, S.; Taberlet, P. The power and promise of population genomics: From genotyping to genome typing. *Nat. Rev. Genet.* **2003**, *4*, 981–994.
60. Storz, J.F.; Nachman, M.W. Natural selection on protein polymorphism in the rodent genus *Peromyscus*: Evidence from interlocus contrasts. *Evolution* **2003**, *57*, 2628–2635.
61. Slate, J.; David, P.; Dodds, K.G.; Veenvliet, B.A.; Glass, B.C.; Broad, T.E.; McEwan, J.C. Understanding the relationship between the inbreeding coefficient and multilocus heterozygosity: theoretical expectations and empirical data. *Heredity* **2004**, *93*, 255–265.
62. Vali, U.; Einarsson, A.; Waits, L.; Ellegren, H. To what extent do microsatellite markers reflect genome-wide genetic diversity in natural populations? *Mol. Ecol.* **2008**, *17*, 3808–3817.
63. Ouborg, N.J.; Pertoldi, C.; Loeschcke, V.; Bijlsma, R.; Hedrick, P.W. Conservation genetics in transition to conservation genomics. *Trends Genet.* **2010**, *26*, 177–187.
64. Hudson, M.E. Sequencing breakthroughs for genomic ecology and evolutionary biology. *Mol. Ecol. Res.* **2008**, *8*, 3–17.
65. Shendure, J.; Ji, H.L. Next-generation DNA sequencing. *Nat. Biotechnol.* **2008**, *26*, 1135–1145.
66. Stapley, J.; Reger, J.; Feulner, P.G.D.; Smadja, C.; Galindo, J.; Ekblom, R.; Bennison, C.; Ball, A.D.; Beckerman, A.P.; Slate, J. Adaptation genomics: the next generation. *Trends Ecol. Evol.* **2010**, *25*, 705–712.
67. McCarthy, A. Third generation DNA sequencing: Pacific Biosciences' single molecule real time technology. *Chem. Biol.* **2010**, *17*, 675–676.
68. Ragoussis, J. Genotyping technologies for genetic research. *Ann. Rev. Genomics Hum. Genet.* **2009**, *10*, 117–133.
69. Fan, J.B.; Chee, M.S.; Gunderson, K.L. Highly parallel genomic assays. *Nat. Rev. Genet.* **2006**, *7*, 632–644.
70. Pertoldi, C.; Bijlsma, R.; Loeschcke, V. Conservation genetics in a globally changing environment: Present problems, paradoxes and future challenges. *Biodivers. Conserv.* **2007**, *16*, 4147–4163.
71. Squirrell, J.; Hollingsworth, P.M.; Woodhead, M.; Russell, J.; Lowe, A.J.; Gibby, M.; Powell, W. How much effort is required to isolate nuclear microsatellites from plants? *Mol. Ecol.* **2003**, *12*, 1339–1348.
72. Zane, L.; Bargelloni, L.; Patarnello, T. Strategies for microsatellite isolation: A review. *Mol. Ecol.* **2002**, *11*, 1–16.
73. Molodstova, D.; Crowe, E.; Olson, A.; Yee, J.; Freeland, J.R. Conserved flanking microsatellite sequences (ReFS) differentiate between Lepidoptera species, and provide insight into microsatellite evolution. *Syst. Entomol.* **2011**, *36*, 371–376.
74. Boguski, M.S.; Lowe, T.M.J.; Tolstoshev, C.M. DbEST—Database for Expressed Sequence Tags. *Nat. Genet.* **1993**, *4*, 332–333.

75. Ellis, J.R.; Burke, J.M. EST-SSRs as a resource for population genetic analyses. *Heredity* **2007**, *99*, 125–132.
76. Chagne, D.; Chaumeil, P.; Ramboer, A.; Collada, C.; Guevara, A.; Cervera, M.T.; Vendramin, G.G.; Garcia, V.; Frigerio, J.M.M.; Echt, C.; Richardson, T.; Plomion, C. Cross-species transferability and mapping of genomic and cDNA SSRs in pines. *Theor. Appl. Genet.* **2004**, *109*, 1204–1214.
77. Gutierrez, M.V.; Patto, M.C.V.; Huguet, T.; Cubero, J.I.; Moreno, M.T.; Torres, A.M. Cross-species amplification of *Medicago truncatula* microsatellites across three major pulse crops. *Theor. Appl. Genet.* **2005**, *110*, 1210–1217.
78. Pashley, C.H.; Ellis, J.R.; McCauley, D.E.; Burke, J.M. EST databases as a source for molecular markers: Lessons from *Helianthus*. *J. Heredity* **2006**, *97*, 381–388.
79. Kalia, R.K.; Rai, M.K.; Kalia, S.; Singh, R.; Dhawan, A.K. Microsatellite markers: An overview of the recent progress in plants. *Euphytica* **2011**, *177*, 309–334.
80. North, A.; Pennanen, J.; Ovaskainen, O.; Laine, A.L. Local adaptation in a changing world: The roles of gene-flow, mutation, and sexual reproduction. *Evolution* **2010**, *65*, 79–89.
81. Hansen, M.M.; Meier, K.; Mensberg, K.L.D. Identifying footprints of selection in stocked brown trout populations: A spatio-temporal approach. *Mol. Ecol.* **2010**, *19*, 1787–1800.
82. Williams, L.M.; Oleksiak, M.F. Signatures of selection in natural populations adapted to chronic pollution. *BMC Evol. Biol.* **2008**, *8*, 282. doi:10.1186/1471-2148-8-282.
83. Narum, S.R.; Hess, J.E. Comparison of FST outlier tests for SNP loci under selection. *Mol. Ecol. Resour.* **2011**, *11*, 184–194.
84. Prunier, J.; Laroche, J.; Beaulieu, J.; Bousquet, J. Scanning the genome for gene SNPs related to climate adaptation and estimating selection at the molecular level in boreal black spruce. *Mol. Ecol.* **2011**, *20*, 1702–1716.
85. Holderegger, R.; Herrmann, D.; Poncet, B.; Gugerli, F.; Thuiller, W.; Taberlet, P.; Gielly, L.; Rioux, D.; Brodbeck, S.; Aubert, S.; Manel, S. Land ahead: Using genome scans to identify molecular markers of adaptive relevance. *Plant Ecol. Divers.* **2008**, *1*, 273–283.
86. Excoffier, L.; Hofer, T.; Foll, M. Detecting loci under selection in a hierarchically structured population. *Heredity* **2009**, *103*, 285–298.
87. Perez-Figueroa, A.; Garcia-Pereira, M.J.; Saura, M.; Rolan-Alvarez, E.; Caballero, A. Comparing three different methods to detect selective loci using dominant markers. *J. Evol. Biol.* **2010**, *23*, 2267–2276.
88. Apple, J.L.; Grace, T.; Joern, A.; Amand, P.S.; Wisely, S.M. Comparative genome scan detects host-related divergent selection in the grasshopper *Hesperotettix viridis*. *Mol. Ecol.* **2010**, *19*, 4012–4028.
89. Freeland, J.R.; Biss, P.; Conrad, K.F.; Silvertown, J. Selection pressures have caused genome-wide population differentiation of *Anthoxanthum odoratum* despite the potential for high gene flow. *J. Evol. Biol.* **2010**, *23*, 776–782.
90. Nunes, V.L.; Beaumont, M.A.; Butlin, R.K.; Paulo, O.S. Multiple approaches to detect outliers in a genome scan for selection in ocellated lizards (*Lacerta lepida*) along an environmental gradient. *Mol. Ecol.* **2011**, *20*, 193–205.

91. Gomez-Uchida, D.; Seeb, J.E.; Smith, M.J.; Habicht, C.; Quinn, T.P.; Seeb, L.W. Single nucleotide polymorphisms unravel hierarchical divergence and signatures of selection among Alaskan sockeye salmon (*Oncorhynchus nerka*) populations. *BMC Evol. Biol.* **2011**, *11*, 48.
92. Renaut, S.; Nolte, A.W.; Rogers, S.M.; Derome, N.; Bernatchez, L. SNP signatures of selection on standing genetic variation and their association with adaptive phenotypes along gradients of ecological speciation in lake whitefish species pairs (*Coregonus* spp.). *Mol. Ecol.* **2011**, *20*, 545–559.
93. Helyar, S.J.; Hemmer-Hansen, J.; Bekkevold, D.; Taylor, M.I.; Ogden, R.; Limborg, M.T.; Cariani, A.; Maes, G.E.; Diopere, E.; Carvalho, G.R.; Nielsen, E.E. Application of SNPs for population genetics of nonmodel organisms: New opportunities and challenges. *Mol. Ecol. Res.* **2011**, *11*, 123–136.
94. O'Malley, K.G.; Ford, M.J.; Hard, J.J. Clock polymorphism in Pacific salmon: Evidence for variable selection along a latitudinal gradient. *Proc. R. Soc. B* **2010**, *277*, 3703–3714.
95. Eckert, A.J.; Bower, A.D.; Gonzalez-Martinez, S.C.; Wegrzyn, J.L.; Coop, G.; Neale, D.B. Back to nature: Ecological genomics of loblolly pine (*Pinus taeda*, Pinaceae). *Mol. Ecol.* **2010**, *19*, 3789–3805.
96. Gilchrist, A.S.; Meats, A.W. The genetic structure of populations of an invading pest fruit fly, *Bactrocera tryoni*, at the species climatic range limit. *Heredity* **2010**, *105*, 165–172.
97. Paaby, A.B.; Blacket, M.J.; Hoffmann, A.A.; Schmidt, P.S. Identification of a candidate adaptive polymorphism for *Drosophila* life history by parallel independent clines on two continents. *Mol. Ecol.* **2010**, *19*, 760–774.
98. Bradbury, I.R.; Hubert, S.; Higgins, B.; Borza, T.; Bowman, S.; Paterson, I.G.; Snelgrove, P.V.R.; Morris, C.J.; Gregory, R.S.; Hardie, D.C.; Hutchings, J.A.; Ruzzante, D.E.; Taggart, C.T.; Bentzen, P. Parallel adaptive evolution of Atlantic cod on both sides of the Atlantic Ocean in response to temperature. *Proc. R. Soc. B* **2010**, *277*, 3725–3734.
99. Yang, Z.H.; Bielawski, J.P. Statistical methods for detecting molecular adaptation. *Trends Ecol. Evol.* **2000**, *15*, 496–503.
100. McDonald, J.H.; Kreitman, M. Adaptive Protein Evolution at the Adh Locus in *Drosophila*. *Nature* **1991**, *351*, 652–654.
101. Bustamante, C.D.; Fledel-Alon, A.; Williamson, S.; Nielsen, R.; Hubisz, M.T.; Gnanowski, S.; Tanenbaum, D.M.; White, T.J.; Sninsky, J.J.; Hernandez, R.D.; Civello, D.; Adams, M.D.; Cargill, M.; Clark, A.G. Natural selection on protein-coding genes in the human genome. *Nature* **2005**, *437*, 1153–1157.
102. Petersen, L.; Bollback, J.P.; Dimmic, M.; Hubisz, M.; Nielsen, R. Genes under positive selection in *Escherichia coli*. *Genome Res.* **2007**, *17*, 1336–1343.
103. Yang, Z.H.; Wong, W.S.W.; Nielsen, R. Bayes empirical Bayes inference of amino acid sites under positive selection. *Mol. Biol. Evol.* **2005**, *22*, 1107–1118.
104. Goldman, N.; Yang, Z.H. Codon-Based Model of Nucleotide Substitution for Protein-Coding DNA-Sequences. *Mol. Biol. Evol.* **1994**, *11*, 725–736.
105. Nam, K.; Mugal, C.; Nabholz, B.; Schielzeth, H.; Wolf, J.B.W.; Backstrom, N.; Kunstner, A.; Balakrishnan, C.N.; Heger, A.; Ponting, C.P.; Clayton, D.F.; Ellegren, H. Molecular evolution of genes in avian genomes. *Genome Biol.* **2010**, *11*, R68. doi:10.1186/gb-2010-11-6-r68.

106. Ersoz, E.S.; Wright, M.H.; Gonzalez-Martinez, S.C.; Langley, C.H.; Neale, D.B. Evolution of Disease Response Genes in Loblolly Pine: Insights from Candidate Genes. *PLoS One* **2010**, *5*, e14234. doi:10.1371/journal.pone.0014234.
107. Oleksyk, T.K.; Zhao, K.; de la Vega, F.M.; Gilbert, D.A.; O'Brien, S.J.; Smith, M.W. Identifying selected regions from heterozygosity and divergence using a light-coverage genomic dataset from two human populations. *PLoS One* **2008**, *3*, e1712. doi:10.1371/journal.pone.0001712.
108. Eyre-Walker, A. Changing effective population size and the McDonald-Kreitman test. *Genetics* **2002**, *162*, 2017–2024.
109. Eyre-Walker, A.; Keightley, P.D. Estimating the Rate of Adaptive Molecular Evolution in the Presence of Slightly Deleterious Mutations and Population Size Change. *Mol. Biol. Evol.* **2009**, *26*, 2097–2108.
110. Maynard Smith, J.; Haigh, J. The hitch-hiking effect of a favourable gene. *Genet. Res.* **1974**, *23*, 23–35.
111. Kim, Y.; Stephan, W. Detecting a local signature of genetic hitchhiking along a recombining chromosome. *Genetics* **2002**, *160*, 765–777.
112. McVean, G. The structure of linkage disequilibrium around a selective sweep. *Genetics* **2007**, *175*, 1395–1406.
113. Sabeti, P.C.; Reich, D.E.; Higgins, J.M.; Levine, H.Z.P.; Richter, D.J.; Schaffner, S.F.; Gabriel, S.B.; Platko, J.V.; Patterson, N.J.; McDonald, G.J.; Ackerman, H.C.; Campbell, S.J.; Altshuler, D.; Cooper, R.; Kwiatkowski, D.; Ward, R.; Lander, E.S. Detecting recent positive selection in the human genome from haplotype structure. *Nature* **2002**, *419*, 832–837.
114. Beaumont, M.A.; Balding, D.J. Identifying adaptive genetic divergence among populations from genome scans. *Mol. Ecol.* **2004**, *13*, 969–980.
115. Foll, M.; Gaggiotti, O. A genome-Scan method to identify selected loci appropriate for both dominant and codominant markers: A bayesian perspective. *Genetics* **2008**, *180*, 977–993.
116. Excoffier, L.; Foll, M.; Petit, R.J. Genetic Consequences of Range Expansions. *Ann. Rev. Ecol. Evol. Syst.* **2009**, *40*, 481–501.
117. Kuhner, M.K. LAMARC 2.0: maximum likelihood and Bayesian estimation of population parameters. *Bioinformatics* **2006**, *22*, 768–770.
118. Joost, S.; Kalbermatten, M.; Bonin, A. Spatial analysis method(SAM): A software tool combining molecular and environmental data to identify candidate loci for selection. *Mol. Ecol. Res.* **2008**, *8*, 957–960.
119. Siol, M.; Wright, S.I.; Barrett, S.C.H. The population genomics of plant adaptation. *New Phytol.* **2010**, *188*, 313–332.
120. Galindo, J.; Moran, P.; Rolan-Alvarez, E. Comparing geographical genetic differentiation between candidate and noncandidate loci for adaptation strengthens support for parallel ecological divergence in the marine snail *Littorina saxatilis*. *Mol. Ecol.* **2009**, *18*, 919–930.
121. Charlesworth, B.; Nordborg, M.; Charlesworth, D. The effects of local selection, balanced polymorphism and background selection on equilibrium patterns of genetic diversity in subdivided populations. *Genet. Res.* **1997**, *70*, 155–174.
122. Nosil, P.; Funk, D.J.; Ortiz-Barrientos, D. Divergent selection and heterogeneous genomic divergence. *Mol. Ecol.* **2009**, *18*, 375–402.

123. Thibert-Plante, X.; Hendry, A.P. Five questions on ecological speciation addressed with individual-based simulations. *J. Evol. Biol.* **2009**, *22*, 109–123.
124. Mackay, T.F.C.; Langley, C.H. Molecular and Phenotypic Variation in the Achaete-Scute Region of *Drosophila melanogaster*. *Nature* **1990**, *348*, 64–66.
125. Latta, R.G.; Gardner, K.M.; Staples, D.A. Quantitative trait locus mapping of genes under selection across multiple years and sites in *Avena barbata*: Epistasis, pleiotropy, and genotype-by-environment interactions. *Genetics* **2010**, *185*, 375–385.
126. Stranger, B.E.; Stahl, E.A.; Raj, T. Progress and promise of genome-wide association studies for human complex trait genetics. *Genetics* **2011**, *187*, 367–383.
127. Aranzana, M.J.; Kim, S.; Zhao, K.Y.; Bakker, E.; Horton, M.; Jakob, K.; Lister, C.; Molitor, J.; Shindo, C.; Tang, C.L.; Toomajian, C.; Traw, B.; Zheng, H.G.; Bergelson, J.; Dean, C.; Marjoram, P.; Nordborg, M. Genome-wide association mapping in *Arabidopsis* identifies previously known flowering time and pathogen resistance genes. *PLoS Genet.* **2005**, *1*, 531–539.
128. Li, Y.; Huang, Y.; Bergelson, J.; Nordborg, M.; Borevitz, J.O. Association mapping of local climate-sensitive quantitative trait loci in *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 21199–21204.
129. Akey, J.M.; Ruhe, A.L.; Akey, D.T.; Wong, A.K.; Connelly, C.F.; Madeoy, J.; Nicholas, T.J.; Neff, M.W. Tracking footprints of artificial selection in the dog genome. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 1160–1165.
130. Cockram, J.; White, J.; Zuluaga, D.L.; Smith, D.; Comadran, J.; Macaulay, M.; Luo, Z.W.; Kearsey, M.J.; Werner, P.; Harrap, D.; Tapsell, C.; Liu, H.; Hedley, P.E.; Stein, N.; Schulte, D.; Steuernagel, B.; Marshall, D.F.; Thomas, W.T.B.; Ramsay, L.; Mackay, I.; Balding, D.J.; Waugh, R.; O’Sullivan, D.M. Genome-wide association mapping to candidate polymorphism resolution in the unsequenced barley genome. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 21611–21616.
131. Oleksyk, T.K.; Smith, M.W.; O’Brien, S.J. Genome-wide scans for footprints of natural selection. *Phil. Trans. R. Soc. B* **2010**, *365*, 185–205.
132. Storz, J.F.; Wheat, C.W. Integrating evolutionary and functional approaches to infer adaptation at specific loci. *Evolution* **2010**, *64*, 2489–2509.