Admixture as the basis for genetic mapping

C. Alex Buerkle[1] and Christian Lexer[2,3]

[1] Department of Botany, University of Wyoming, Laramie, WY 82071, USA
[2] Jodrell Laboratory, Royal Botanic Gardens, Richmond, Surrey TW9 3DS, UK
[3] Current address: Department of Biology, University of Fribourg, CH-1700 Fribourg, Switzerland

Genetic mapping in natural populations is increasing rapidly in feasibility and accessibility. As with many areas in genetics, advances in molecular techniques and statistics are drastically altering how we can investigate inheritance in wild organisms. For ecology and evolution, this is particularly significant and promising, because many of the organisms of interest are not amenable to conventional genetic approaches. Admixture mapping falls within a family of statistical approaches that use natural recombination and linkage disequilibrium between genetic markers and phenotypes as the basis for mapping. Our aim in this review is to provide a snapshot of previous and ongoing research, existing methods and challenges, the nature of questions that can be investigated and prospects for the future of admixture mapping.

Genetics in species colliders

A greater understanding of the genetic basis of adaptations, and of reproductive isolation between species, is a long-standing goal for biologists, in part because the genetics can inform us about the history and processes of adaptation and speciation [1,2]. A variety of methods are utilized for associating phenotypes with particular genotypes, including traditional approaches to genetic mapping, which involve statistical analyses of experimental crosses or populations with known pedigrees. However, many of the taxa of interest in ecology and evolution are not amenable to experimental crosses or preclude estimation of pedigrees. Consequently, there has been increasing interest in using natural admixture and recombination between divergent taxa in the wild as the basis for genetic mapping.

Hybridization or admixture creates recombinants between divergent parental taxa where they come into geographic contact, at habitat ecotones or in hybrid zones. For any type of genetic mapping, recombination is required to break up combinations of parental genotypes. The physical processes of recombination and segregation mean that typical parental allele combinations across loci are disrupted and some independence between loci is introduced (i.e. linkage disequilibrium is reduced). As a result, the effects of individual genetic loci on a phenotype can be detected and neighboring regions can be excluded (i.e. phenotypic effects can be mapped to specific genomic regions). In the case of admixture mapping, researchers utilize naturally occurring, recombinant organisms to isolate the effects of individual loci on phenotypes of interest (including components of reproductive isolation and clinally varying, adaptive phenotypes). This is done by detecting informative, introgressed genotypes at focal loci and by quantifying excess admixture, the extent to which genotypes are in a foreign genomic background (Box 1).

For mapping loci that affect isolation between taxa, the over- and underrepresentation of certain genotypes in different genomic backgrounds (quantified by excess admixture) are of interest, without direct reference to traits that might be subject to selection [3–5]. In this application, loci and genotypes are identified that might be subject to positive or negative selection at a barrier to gene flow between divergent species or ecotypes (e.g. an environmental cline or ecotone, or a barrier due to intrinsic incompatibility). Similarly, excess admixture at individual loci can be used as a predictor of phenotypic variation [6–8]

Glossary

Admixture: the mixing of genomes of divergent parental taxa; meant as a more general term than 'hybridization,' so as to include taxa that we would not typically refer to as hybridizing (i.e. mating between human populations and between subspecific populations). The mixing of parental genomes can occur at the level of individuals (individual admixture; the focus in this paper) and at the level of populations through mixing of individuals from different source populations (population admixture; possibly without matings among ancestry lineages and the resulting individual admixture).

Association mapping: category of genetic analysis of phenotypic variation in natural populations. Involves a search for significant associations between phenotypes and nucleotide variation in candidate genes or whole-genome scans. Also known as linkage disequilibrium mapping.

Genetic and physical maps maps: of molecular markers are a representation of the linear order of markers on chromosomes. The distance between markers on a genetic map is given by the rate of meiotic recombination between them (recombination distance), whereas the distance between markers on a physical map is the number of nucleotides separating them in the genome sequence. Variation in the recombination rate along chromosomes will cause the relative distance (genetic divided by physical distance) between markers to vary.

Genetic architecture: the number and genomic location of loci that contribute to variation in a trait, as well as the allelic effect sizes and direction, the genotypic effects (additivity and dominance) and the extent of epistatic interactions among loci.

Introgression: movement of alleles of one taxon into the genetic background of another. This requires successful reproduction of the F1 and either mating with other F1 s or backcrosses to one of the parental taxa. Recombination in the F1 begins to break up parental allele combinations.

Linkage disequilibrium: statistical association between genotypes at different loci, or between a phenotype and a focal locus, such that one can predict probabilistically the genotype of the second locus (or phenotype) on the basis of the genotype at the first.

Population structure: subdivision of a larger, potentially interbreeding population into smaller units, the members of which are more similar to one another than to members of other subpopulation units.

Corresponding author: Buerkle, C.A. (buerkle@uwyo.edu)

and as a basis for mapping traits involved in adaptive divergence between taxa. In both applications of excess admixture, an existing genetic map of molecular markers (either for the focal taxa or closely related species) will

Box 1. Concept of excess admixture

Excess admixture quantifies the extent to which alleles and genotypes are introgressed into a foreign genetic background or are found in their native background. The contribution of each parental lineage to the genome of admixed individuals is estimated by a hybrid index (h [3,74]) or an admixture proportion [69,75,76]. In the case of parental taxa that are fully differentiated and do not share alleles, the index is simply a proportion of alleles across the genome that originate from one of the parental taxa. Individual loci are expected to have genotypes that are consistent with the ancestry of the remainder of the genome. Departures of individual loci from genome-wide ancestry are referred to as excess admixture or ancestry.

Excess admixture ($x - h$) is illustrated in Figure I, where a focal locus (at the far left) is shown in combination with various genetic backgrounds at 30 other loci. A score of 1 for excess admixture corresponds to a pure genomic background of one parental species ($h = 0$; light purple in Figure I), with the exception of the introgression of a homozygous genotype at a focal locus (genotypic score, $x = 1$; bottom row in Figure I). Similarly, a value of excess admixture of -1 corresponds to an individual with a genomic composition of the other parental species ($h = 1$; dark purple in Figure I), with the exception of an introgressed homozygous genotype at the focal locus ($x = 0$; top row in Figure I). Individuals with parental genotypes at all loci, without introgression and excess admixture ($x - h = 0$), are in the middle two rows of Figure I. In Figure I, the intermediate shade of purple indicates heterozygotes, but for simplicity no heterozygotes are depicted at the focal locus (they would have $x = 0.5$).

Excess ancestry is a measure of introgression and of the residual from the prediction of the focal locus genotype based on the ancestry of the remainder of the genome, as quantified by a hybrid index (h; Figure II). In Figure II, the genotype 'aa' is typical of the parental taxon with $h = 0$, and genotype 'AA' is typical of the parental taxon with $h = 1$. Individuals with genotypes that fall off of the diagonal line ($x = h$) possess an introgressed, recombinant genotype and exhibit excess admixture at the focal locus.

Measures of excess admixture form the basis for mapping phenotypic variation and loci that contribute to isolation between taxa [4,28], although different implementations of admixture mapping parameterize and utilize excess admixture in different ways.
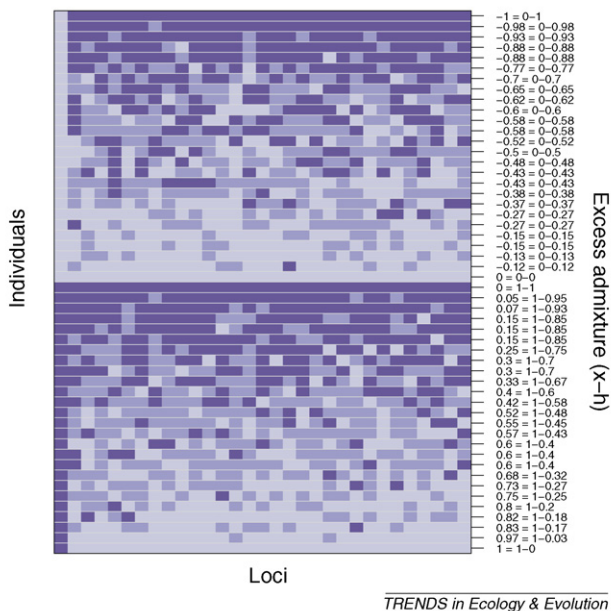

Figure II. Quantification of excess admixture.

Figure I. Introgressed genotypes in a foreign genetic background.

allow much stronger inferences about genetic architectures than would be possible if the linkage relationships of markers were unknown. Admixture can be thought of as an opportunity to observe the consequences of the 'collision' of the genomes of species, to note the regions that resist or experience introgression and to utilize genetic recombinants for mapping, in a manner roughly analogous to using atom collisions to reveal their particular nature.

In contrast to commonly used and simple crossing schemes for genetic mapping (e.g. BC1 and F2) that involve only one generation of recombination, admixed populations can harbor a range of recombinant individuals that result from one or more generations of hybridization and recombination, depending on the area of range overlap, the extent of reproductive isolation and the length of time the populations have been in contact. Consequently, many admixed populations will contain early-generation hybrids that exhibit the necessary balance between recombination and linkage disequilibrium for initial, coarse-scale mapping. Ideally, an admixed population will also contain a large number of individuals that have a history of several generations of hybridization, with sufficient recombination for fine-scale mapping (Figure 1).

The potential value of admixed populations for mapping has been noted before in the context of ecology and evolution [9,10], and several studies have now utilized this approach to mapping (see below). In the following paragraphs, we place admixture mapping into the context of different methods to study the genetic basis, functional role and evolution of fitness-related traits. The uniqueness of admixture mapping lies in its ability to embrace two related but different methodologies commonly known as 'association mapping' and 'population genomics.'

Association mapping uses recombination events accumulated in natural populations over many generations to detect statistical associations between genetic loci of known genomic location and phenotypic traits of interest [11]. Association mapping generally requires the absence of population structure in the sample or statistical methods to account for structure [12–15]. Similarly, admixture
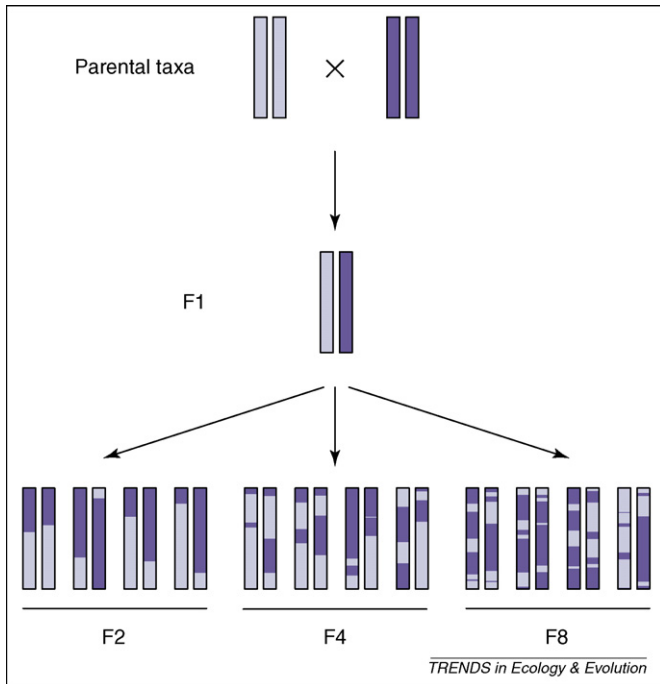
Figure 1. A schematic illustrating chromosomes of diploid recombinants that are suitable for coarse- and fine-scale mapping. Parental taxa hybridize and form F1 progeny. Crosses between F1 individuals, or backcrosses to parental taxa, result in recombinant chromosomes through meiosis in F1 individuals. F2 progeny are the result of crossing F1 individuals, and each chromosome is expected to have one recombination breakpoint on average (one crossover per meiosis); recombination within chromosomes and segregation of chromosomes makes mapping possible, even at this coarse level in the F2. Later-generation intercrosses (F4 and F8) are also illustrated and these possess more recombination breakpoints, which facilitate finer-scale mapping. Regions of each chromosome with different colors represent genomic regions with ancestry in each of the parental taxa. The size of these ancestry blocks and the linkage disequilibrium between neighboring molecular markers decays with additional generations (t) of recombination (θ) at a rate proportional to $(1 - \theta)^t$ [64–66].

mapping incorporates population structure through its measure of excess ancestry (Box 1). The ability to account for population structure, in spite of the absence of controlled crosses or pedigrees, is a common feature of association and admixture mapping (Box 2). The absence of confounding population structure allows true, developmental genetic associations to be recognized, rather than the large number of spurious associations between allelic and phenotypic variation that are expected in the presence of population structure, where the effects of individual loci are not separated from their genetic background [12–14,16,17].

Population genomics typically refers to the use of many molecular markers with known genomic locations to identify regions under selection, by distinguishing the locus-specific effects of selection from the shared demographic history of the entire genome [18–20]. Population genomics can help identify fitness-related genetic variation even without *a priori* information on the actual phenotypes under selection [21], but it cannot replace experiments to identify and understand the traits targeted by selection [20].

The promise of admixture mapping for ecologists and evolutionary biologists lies in its combination of association mapping and population genomics. Admixture mapping in natural populations can facilitate studies of genetic

Box 2. Relationship between admixture and association mapping

Until recently, the distinction between association and admixture mapping was clearer. Traditionally, association mapping has been applied to polymorphisms within candidate genes (e.g. [77–80]), rather than on genome scans of variation, whereas admixture mapping was typically applied to markers spanning a genome. With the rapid increase of molecular markers for model organisms (e.g. mouse, *Arabidopsis*, human), association mapping has become feasible at the scale of the entire genome, with a high likelihood of markers in linkage disequilibrium with genes underlying traits of interest (e.g. [14,81,82]). Consequently, both approaches now can involve mapping across the genome.

Despite the similarities between admixture and association mapping, both in terms of purpose and methods, it is worth recognizing the differences that remain between them, and the tradeoffs associated with each. For one, admixture mapping will often utilize the known axis of variation among hybrids of two parental forms and focus on populations that are likely to contain recombinant individuals that are the substrate for all approaches for mapping. By contrast, association mapping often involves analysis of more complex ancestries among a larger number of more closely related source populations, and proceeds without *a priori* knowledge of which are likely to contain informative, recombinant individuals. Admixture mapping will often entail genetic recombination between comparatively simple parental genomic contributions, and substantial and consequential phenotypic variation that exists between more divergent parental taxa [6–8], and therefore might be more likely to be successful.

Second, high-resolution studies of the extent of linkage disequilibrium in populations of *Arabidopsis*, mouse and human suggest that a high density of molecular markers is required for association mapping (intraspecific, population-level variation) at the scale of the whole genome [72,73,83]. Admixture mapping with individuals that are the result of a few generations of hybrid matings will require fewer markers, because linkage disequilibrium will extend over larger blocks of the genome, which should facilitate initial, coarse-scale mapping. In some admixed populations there will also be more advanced generation hybrids with linkage disequilibrium that extends over smaller regions, which should allow fine-scale mapping. Finally, many of the models and methods developed for association mapping are relevant to admixture mapping. However, it is not clear that the population histories that have been modeled and investigated for association mapping encompass the considerable population structure and linkage disequilibrium that result from crosses between substantially diverged species and populations [12–16,66]. For example, more work remains to be done to compare the performance of different methods of analysis when population structure is great (e.g. when a sample might contain many F1s between divergent parental populations).

architecture of phenotypes and the detection of regions or loci under selection. This potentially allows evolutionary biologists to study the genetic architecture of adaptation and speciation *in situ* [1], including long-lived or otherwise intractable taxa for which crosses or pedigrees would be difficult to obtain [4,9,10]. Similarly, admixture mapping can also aid studies of the phenotypic effects of recombination in microorganisms of medical or other direct relevance to humans [22,23], and the interpretation of data from metagenomics or community genomics research on microbial populations in their natural environments [22,24].

Applications in ecology and evolution

Although the potential of admixture mapping in evolutionary biology has been recognized for 15–20 years [25,26], the first successful applications were focused on human traits

of medical interest. Examples include recent studies of hypertension, susceptibility to multiple sclerosis, prostate cancer risk and white blood cell count [27–31]. However, multiple studies have identified genetic variants associated with lighter and darker skin pigmentation in humans [32–34], a trait that might be differentially selected in regions with low and high ultraviolet radiation (UVR), owing to known associations between skin pigmentation and vitamin D synthesis and protection from UVR [34,35]. These studies have identified several genes associated with skin pigmentation that displayed genetic signatures consistent with divergent positive selection [33,34] and with recent convergent evolution of a lighter pigmentation type in Europeans and East Asians [34]. Remarkably, variation at two of these loci is also responsible for pigment variation in the scales of fish [32,33].

There is an extensive history of using hybrid individuals to investigate the genetics of isolation between taxa, particularly analyses of population genetic variation along spatial clines (e.g. [36–39]). However, to our knowledge, the first application of admixture mapping to the ecology and evolution of non-human organisms was in the plant genus *Helianthus* (annual sunflowers [3]). Instead of spatial position along a cline, this research used the overall genomic composition of plants to predict the variation at individual mapped loci and to detect excess admixture. The study characterized the genetic architecture of the barrier to gene flow between two ecologically divergent species of sunflower, *Helianthus annuus* and *H. petiolaris*, using three replicate hybrid zones in the central USA. For 26 chromosomal regions, introgression across replicate hybrid zones was significantly reduced relative to neutral expectations from the remainder of the genome, and 16 of these were associated with pollen sterility, an important intrinsic isolation factor in plants. Also, ∼50% of the barrier to gene flow was attributable to chromosomal rearrangements segregating between the two species, thus indicating an important role for both genic and chromosomal factors in species isolation. A follow-up study revealed remarkably similar patterns of introgression for hybrid zones located as far apart as Nebraska and California, USA, thus confirming an important role for intrinsic incompatibilities in determining barriers to gene flow in these sunflowers (rather than extrinsic, environment-dependent factors [40]).

A similar study of admixed populations was carried out in members of the lake whitefish (*Coregonus* spp.) species complex in North America [5]. This species complex offers a tractable system for studying differentiation in traits associated with adaptation to contrasting ecological niches. Lake whitefish occur in multiple terrestrial lakes formed after Pleistocene glaciations and are the subject of studies of parallel selection during ecotype differentiation and ecological speciation [41,42]. In admixed populations, several loci were highly over- or underrepresented relative to the genomic background. Some regions showed consistent deviations in natural admixed populations and experimental crosses [5]. Research in this system is a good illustration of how studies of admixed natural populations contribute to our understanding of the genetics of population divergence and speciation, especially when com-

bined with data from experimental crosses and replicated contrasts between divergent populations or forms (in this case, multiple lakes).

Admixture mapping work also is under way in two ecologically divergent European members of the genus *Populus* (poplars, aspens, cottonwoods), *P. alba* and *P. tremula* (floodplain versus upland species) [4,43,44]. Natural admixed populations of these species contain a high proportion of recombinant, backcrossed genotypes and occur in multiple locations throughout Europe [4,43,45]. A first survey allowed the refinement of statistical approaches to detect locus-specific effects in genomic studies of hybrid zones [4]. This research also revealed a high potential to utilize genetic markers from *P. trichocarpa*, the first completely sequenced forest tree [46], for mapping studies across different sections of the genus. Admixture mapping in European species of *Populus* is now being used to study intraspecific variation for genomic isolation across the ranges of these species. This is feasible because 'replicate' hybrid zones between these species are found in several European river valleys, with largely independent phylogeographic histories [4,43].

Research with mice has also utilized admixture as a basis for studying the genetics of reproductive isolation and trait variation, and excellent opportunities exist for additional research [47]. Admixture between subspecies of the house mouse (*Mus musculus*) has been the subject of classical hybrid zone analyses [38,39] and offers great potential for future analysis, particularly given the genome resources for the species [48]. Similarly, existing laboratory mouse lines might provide a suitable recombination history for admixture mapping of isolating factors and other trait variation [49,50], as the genomes of laboratory mouse lines are largely derived from admixture between *domesticus* and *musculus* subspecies of *M. musculus* [51]. However, whether sufficient recombination for fine-scale mapping has occurred remains unclear [52–54].

Examples of other hybridizing taxa that show particular promise for admixture mapping include *Peromyscus* mice and the inheritance of coat color variation (Figure 2) [55,56]. Similarly, variation in male nuptial coloration in sticklebacks (*Gasterosteus aculeatus*) might be amenable to admixture mapping in a lake with hybrids (Figure 2) [57,58]. Hybridization between invasive and native sculpins (*Cottus*) in multiple tributaries of the Rhine offers the opportunity to map components of reproductive isolation in a replicated manner [59]. It is likely that variation in flower color, as well as other traits, can be mapped in hybrid populations of *Antirrhinum* [60] and of *Silene* [61]. Hybridization between species of *Heliconius* butterflies with dramatically different wing patterns and coloration offers a final example of a system with pervasive hybridization that might provide the substrate for mapping in natural populations [62].

Statistical methods and data acquisition

Progress in admixture mapping is being driven by its application to a growing number of biological systems, outlined above, as well as by the rapid development of new statistical approaches and modeling and of molecular markers with increasing coverage of the genome. Below we

provide a brief overview of advances in statistics and markers.

Individuals are expected to have genotypes at focal loci that are consistent with the ancestry of the remainder of their genome. Individual loci that depart from genome-wide ancestry possess excess admixture (Box 1) and provide the basis for genetic mapping. In genetic and practical terms, excess ancestry represents introgression of a genomic region of one parent into the genomic background of another and allows estimation of the isolated effects of this locus, independent of genetic background. Thus, in a general linear model (Box 3), phenotypic variation among individuals can be partitioned into variation owing to genotypes at a focal locus, estimates of parental admixture and any other appropriate predictors (e.g. coancestry, sex, environmental covariates [12,15,28]). The use of a general linear model makes these tests possible in standard statistical software (e.g. SAS, R, S-PLUS) and benefits from the wealth of knowledge about regression models, but specialized software applications for admixture mapping are also available [63].

In the context of mapping the genetic architecture of reproductive barriers between taxa, excess admixture can be observed separately for the genotypes at a locus (Box 1). The probability of observing hybrids with selectively neutral genotypes should be predicted by their hybrid indexes. If individual loci lack excess admixture, they can be considered neutral or nearly neutral, because positively or negatively selected genotypes will be over- or underrepresented relative to admixture in the remainder of the gen-
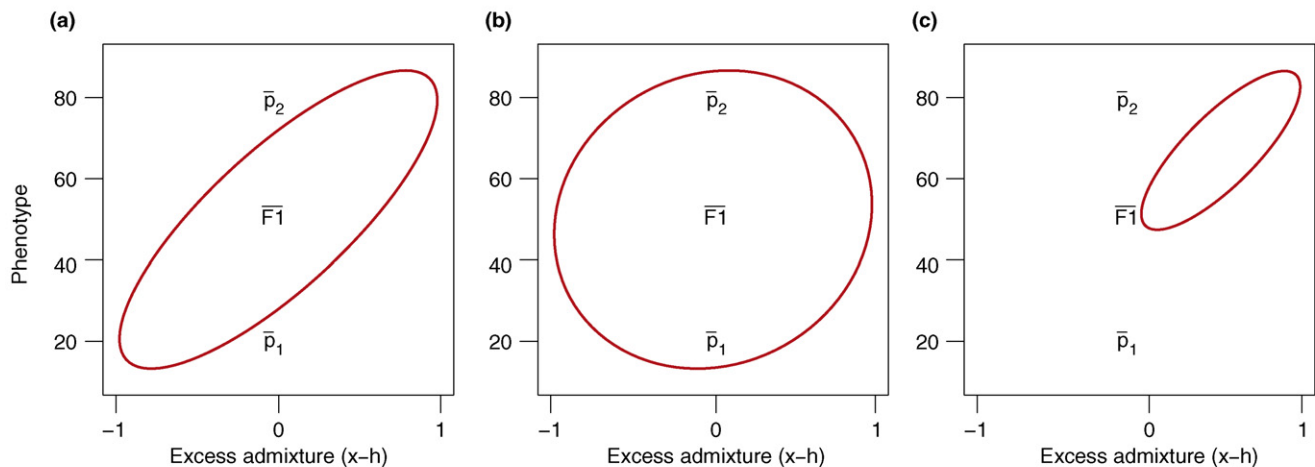
ome. This logic allows for a genome scan of differential introgression and the identification of genes that experience selection in hybrids [3,4].

Given the straightforward use of linear models for analysis in admixture mapping, much of the opportunity for statistical innovation relates to controlling for population structure and ancestry. At the extreme, an allele responsible for adaptation in one parental taxon will be in complete linkage disequilibrium with all markers that have fixed allele differences between parental lineages, even those not physically linked to the causal locus. Recombination in admixed individuals will reduce the extent of linkage disequilibrium [64–66]. A variety of different approaches have been developed to control statistically for genome-wide admixture and ancestry in modeling phenotypes [12,13,15,67]. At this relatively early stage in the development of methods, we need to know more about the utility of different approaches with different population and ancestry structures [12–14,16,66,67]. Empirical studies would be helpful, but there is a need for comprehensive modeling. One persistent question that might be answered by future modeling studies is the sample sizes that are required to achieve desirable levels of power in mapping studies. This is a problem that depends on several factors, including the information content of markers (i.e. the allele frequency differentials between parental lineages, dominance versus codominant markers), the phenotype distribution (Box 3), the hybrid index distribution (Box 4) and the recombination history and linkage disequilibrium in the sample (Figure 1). Additional areas of

Box 3. Relationship between phenotype and excess admixture

Phenotypic variation among admixed individuals can be modeled as a function of excess admixture in a linear regression [28,76]. In each of the panels in Figure I, ellipses encompass data points among hybrid individuals (parental individuals are not included) in the relationship between phenotypes (arbitrary magnitude on the y axis) and excess admixture at a single locus (see Box 1). In each panel, parental ($\bar{p}_1$ and $\bar{p}_2$) and F1 ($\overline{F1}$) phenotypic means are marked. Phenotypic variation among individuals can be a significant function of excess admixture at a focal locus (a), or can be largely independent thereof (b). Contrasts between significant (a) and

nonsignificant (b) loci are what make whole-genome mapping possible. (c) Illustration of the case of asymmetric introgression and a focal locus that is responsible for some of the variation in phenotype. In each of these cases, admixture mapping would involve constructing a linear model of phenotype as a function of excess admixture (and potentially other covariates). General linear models provide a flexible and powerful statistical framework to model phenotypes. Beyond this, clearly the frequency in a study sample of individuals with different levels of admixture and excess admixture will affect the prospects for mapping (see Box 4).

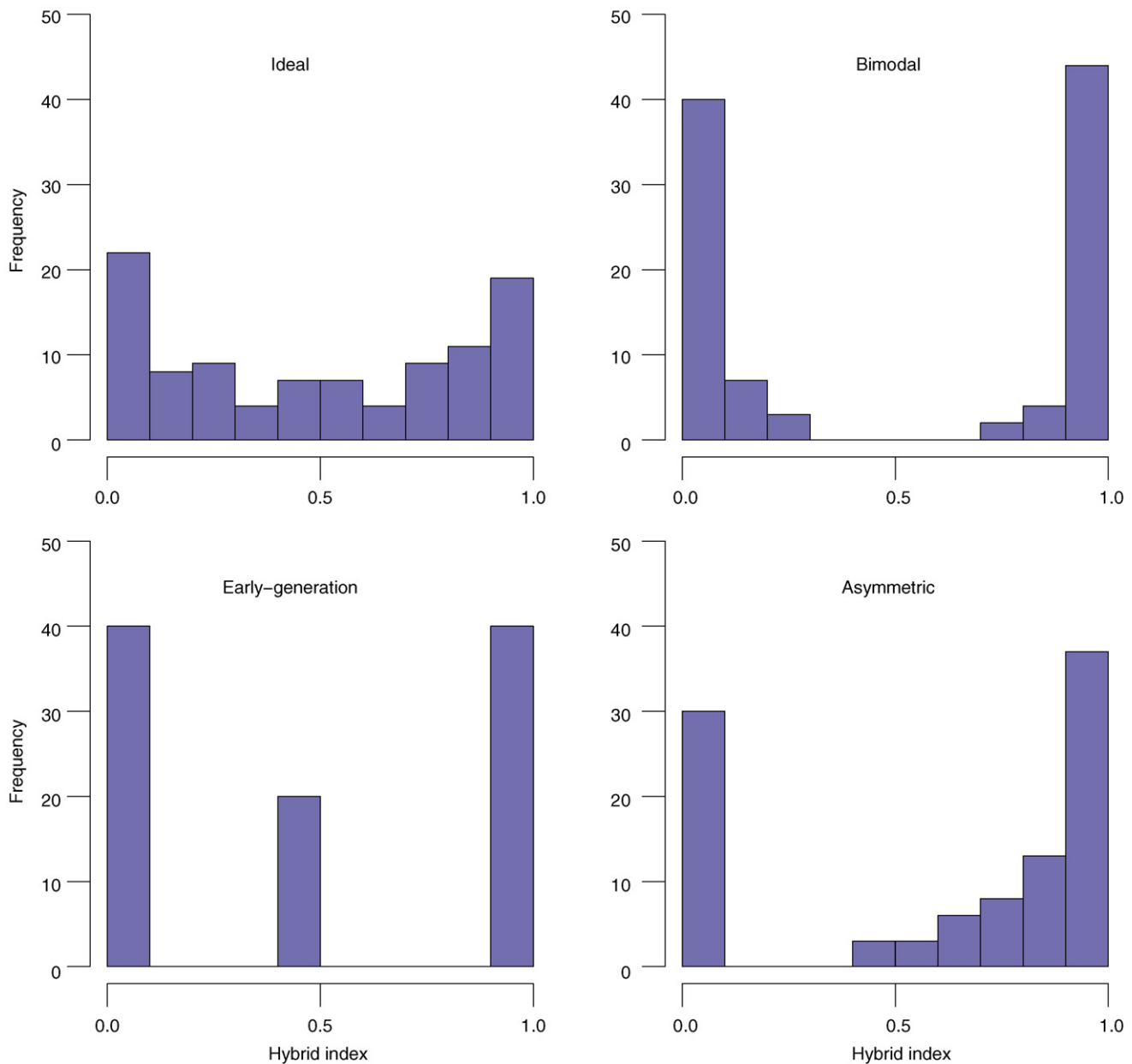
TRENDS in Ecology & Evolution

Figure I. Relationship of phenotype to loci with excess admixture.

Box 4. Sampling distributions of admixed individuals

A critical determinant of the prospects for mapping in any study in natural populations is the frequency of individuals with different levels of admixture. Figure I illustrates four distributions that might be encountered with 100 individuals sampled from a hybrid population. Ideally, a roughly even distribution of individuals will be found across the hybrid index, but this level of admixture between groups might be rare. Many hybrid populations might lack intermediate individuals and be bimodal, with a preponderance of parental and highly backcrossed individuals [84]. Some hybrid populations consist only of parental and F1 individuals and will be less suitable for mapping, given that the products of meiosis in the F1 are not observed. Finally, in some cases introgression will be highly asymmetric, with introgressed individuals only toward one of the parental taxa (e.g. [57]). Asymmetric introgression will permit analysis of the effects of introgressed alleles into one of the parental genotypic backgrounds, but will not yield estimates of the effects of the alternative alleles in the other genomic background (Box 3, Figure Ic).

These distributions were chosen to represent extremes and it is likely that admixed populations will fall between these cases. More studies of natural populations, with sufficient marker resolution to provide informative estimates of hybrid index, would be helpful in judging how frequently admixture mapping can be applied profitably. The distributions also illustrate an advantage of admixture mapping relative to other forms of association mapping. The existence of a single axis of variation between parental taxa allows potentially informative, intermediate individuals to be identified easily. The distribution of informative individuals will be much more difficult to establish for the typically greater number of potential ancestral populations in association mapping.

TRENDS in Ecology & Evolution

Figure I. Frequency of individuals with different types of admixture.
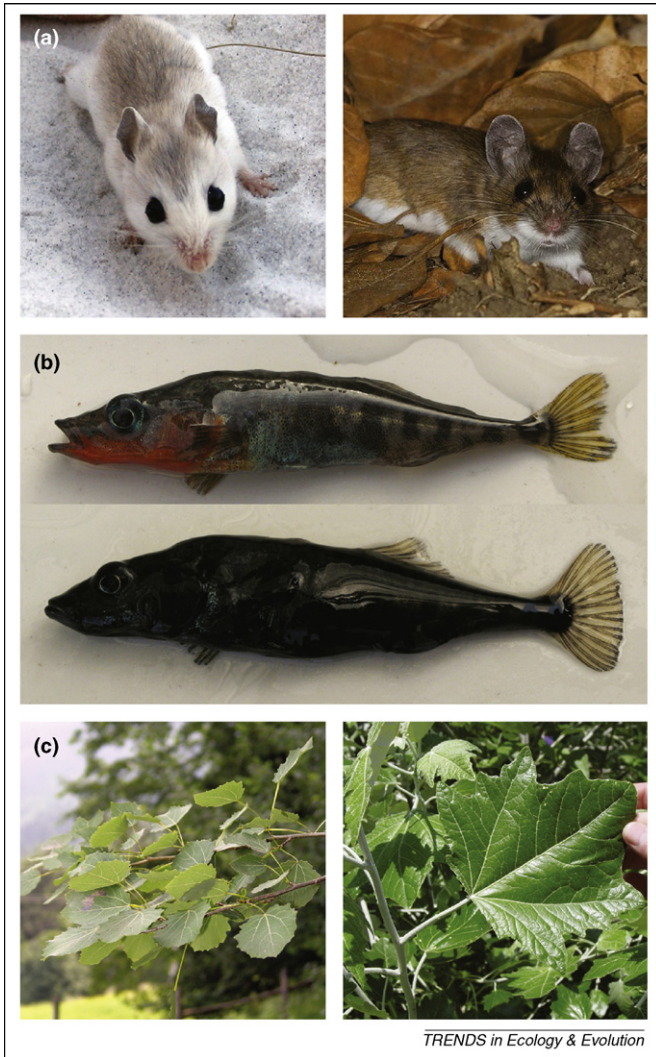
Figure 2. *Peromyscus*, sticklebacks and *Populus* are subjects of current studies that utilize admixture mapping to resolve the genetics of isolating barriers and trait differences. **(a)** Color variation between beach mice (*Peromyscus polionotus leucocephalus*, left) and mainland mice (*P. p. subgriseus*, right) has been the subject of candidate gene analysis and classical crossing studies [55,56]. Contact between color variants at multiple geographic locations should facilitate replicated analysis of the genetics of color variation and differential selection in natural populations. **(b)** Extreme variation in male nuptial coloration is being mapped by utilizing hybrids between limnetic and benthic sticklebacks (*Gasterosteus aculeatus*) in Enos Lake in British Columbia [57,58]. **(c)** The genetic architecture of trait differences and reproductive isolation is the subject of ongoing research in European *Populus* (*P. tremula*, left; *P. alba*, right) [4]. Leaf morphology is associated with differences in biomass accumulation in *Populus*. Photo credits: (a) Cynthia Steiner and Shawn Cary, (b) Tiffany Malek and (c) Christian Lexer.

ongoing statistical refinement address the pervasive issue of multiple tests in genome scans [68], and future methodological developments might allow for models that use information from neighboring markers (as in Ref. [69]) and for analysis of epistatic interactions among genes.

High-throughput methods for sequencing and nucleotide polymorphism detection are making it possible to detect variation at nearly 10^7 molecular markers in some species of interest [48,51]. Clearly, analysis at this scale generates new demands on computational efficiency and interpretability of results. Haplotype analysis might be feasible in some populations (of some taxa) so as to reduce the number of loci that need to be considered [70]. But it is likely that as the scale at which marker data are gathered

continues to increase, new methods for analysis will need to be developed. Another challenge arises as larger numbers of simple sequence repeat markers (SSRs) are becoming available for many species. These SSRs often are highly polymorphic, sometimes with 40–50 different alleles observed at a locus, which presents a challenge for mapping allelic and genotypic effects. One possibility in admixture mapping is to collapse alleles into allelic classes that best represent variation between parental lineages [4,69].

Whereas mapping in natural populations is clearly beneficial for studies of many organisms of interest, its applicability across taxa is limited because a genetic or physical map of markers is critical for mapping. A description of the genetic architecture of traits depends on knowledge of the likely independence of individual markers and contrasts among markers along a linkage group. Without linkage information, conclusions will be limited to the identification of markers associated with isolation or phenotype and estimates of gene action (e.g. dominance relationships of alleles).

Variation in chromosome number among species is a major determinant of the amount of recombination per meiosis (i.e. genetic map length), and of the number of markers that are required to cover a genetic map with a certain density of markers. Typically each chromosome pair has one crossover per meiosis (i.e. is 1 Morgan in length), and therefore the genetic map length of species increases linearly with chromosome number. Achieving a density of mapped markers of one marker every 10 centiMorgans will vary from ∼30 markers for a species like *Anopheles gambiae* with 3 chromosome pairs, to 170–200 markers for sunflowers, *Populus* and mouse with 17–20 chromosome pairs, and beyond for species with larger numbers of chromosomes.

Prospects

The list of organisms studied by admixture mapping is growing. The ongoing studies outlined in this paper are promising and will be good tests of the value of admixture mapping in diverse taxa. In addition to the methodological challenges discussed in the preceding section, the prospects for admixture mapping also depend on the genomic architecture of traits of interest. For example, the genomic architecture of reproductive isolation might be complex in a highly diverged species pair with many chromosomes and incomplete isolation; in this case, the effects of individual loci might be relatively weak and difficult to detect, or to disentangle from genes with which they interact epistatically. Similarly, given that reproductive isolation between taxa is a property that potentially results from the contributions of many phenotypes, particularly in a natural hybrid zone, one would expect many underlying genes and possibilities for epistasis. In this case, a powerful genome scan for contributions to isolation should detect many significant regions, each of relatively small effect. Whereas a small number of leading, large-effect candidates would facilitate additional study and characterization of the effect of individual genes, this genetic architecture might not be the reality. Furthermore, statistical detection of many genes of small effect will have lower power than of few genes of large effect [71], so the

likelihood of fully characterizing the genetics will be limited with admixture mapping, as with other approaches to mapping.

Additional empirical studies, with increasing marker resolution, will yield better descriptions of what patterns of recombination are found in admixed populations. Recent studies of wild mice and *Arabidopsis* provide estimates of linkage disequilibrium in natural populations without admixture [72,73], but studies with comparable marker resolution have not been conducted for admixed populations. Sampling of admixed populations in various species will identify populations with good power for mapping (balance of historical recombination and sufficient linkage disequilibrium for mapping; see Figure 1 and Box 4) and eventually will give better knowledge of the suitability of different taxa and admixture histories for mapping. A similar concern involves the expected rarity of negatively selected genotypes in sampled hybrids [9]. Whereas these will be detected as underrepresented genotypes and contribute to a map of isolating factors, depending on the pattern of introgression, their rarity might make it difficult to test for associations of these loci with phenotypes associated with isolating barriers.

Beyond fine-scale marker surveys of the genome, an important feature of future admixture mapping studies will be the sampling of other important covariates of phenotype. Given that these studies are in natural populations and are subject to uncontrolled variation in environmental influences on phenotypic variation, the measurement of relevant environmental variables and their incorporation into linear models [12] might be critical to successful mapping of genetic variation.

Whole-genome association mapping of intraspecific variation holds considerable promise [14], and we expect that this method will contribute to our understanding of local adaptation and diversification within taxa. Admixture mapping complements this area of research in that it is analytically simpler (two ancestral lineages rather than many) and it involves recombination between phenotypically divergent parental lineages (Box 2). Consequently, admixture mapping is particularly appropriate for the study of species boundaries and the role of adaptation and divergent phenotypes in isolation between taxa. It remains to be seen whether natural populations that experience admixture are more likely to provide the requisite recombinant individuals for mapping than are the more complex population surveys of variation in association mapping. Both approaches for mapping in natural populations yield candidates for future study of the genetics, molecular biology and development of phenotypes of consequence in ecology and evolution.

Acknowledgements

References

1 Coyne, J.A. and Orr, H.A. (2004) *Speciation*. Sinauer Associates

2 Orr, H.A. (2005) The genetic theory of adaptation: a brief history. *Nat. Rev. Genet.* 6, 119–127

3 Rieseberg, L.H. *et al.* (1999) Hybrid zones and the genetic architecture of a barrier to gene flow between two sunflower species. *Genetics* 152, 713–727

4 Lexer, C. *et al.* (2007) Admixture in European *Populus* hybrid zones makes feasible the mapping of loci that contribute to reproductive isolation and trait differences. *Heredity* 98, 74–84

5 Rogers, S.M. *et al.* (2001) Combining the analyses of introgressive hybridisation and linkage mapping to investigate the genetic architecture of population divergence in the lake whitefish (*Coregonus clupeaformis*, Mitchill). *Genetica* 111, 25–41

6 McKeigue, P.M. (2005) Prospects for admixture mapping of complex traits. *Am. J. Hum. Genet.* 75, 1–7

7 Darvasi, A. and Shifman, S. (2005) The beauty of admixture. *Nat. Genet.* 37, 118–119

8 Smith, M.W. and O'Brien, S.J. (2005) Mapping by admixture linkage disequilibrium: advances, limitations and guidelines. *Nat. Rev. Genet.* 6, 623–632

9 Rieseberg, L.H. and Buerkle, C.A. (2002) Genetic mapping in hybrid zones. *Am. Nat.* 159, S36–S50

10 Lexer, C. *et al.* (2004) Hybrid zones as a tool for identifying adaptive genetic variation in outbreeding forest trees: lessons from wild annual sunflowers (*Helianthus* spp.). *For. Ecol. Manage.* 197, 49–64

11 Thomas, D.C. *et al.* (2005) Recent developments in genomewide association scans: a workshop summary and review. *Am. J. Hum. Genet.* 77, 337–345

12 Redden, D.T. *et al.* (2006) Regional admixture mapping and structured association testing: conceptual unification and an extensible general linear model. *PLoS Genet.* 2, e137

13 Setakis, E. *et al.* (2006) Logistic regression protects against population structure in genetic association studies. *Genome Res.* 16, 290–296

14 Zhao, K. *et al.* (2007) An *Arabidopsis* example of association mapping in structured samples. *PLoS Genet.* 3, e4

15 Yu, J. *et al.* (2006) A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nat. Genet.* 38, 203–208

16 Rosenberg, N.A. and Nordborg, M. (2006) A general population-genetic model for the production by population structure of spurious genotype-phenotype associations in discrete, admixed or spatially distributed populations. *Genetics* 173, 1665–1678

17 Mackay, I. and Powell, W. (2007) Methods for linkage disequilibrium mapping in crops. *Trends Plant Sci.* 12, 57–63

18 Luikart, G. *et al.* (2003) The power and promise of population genomics: from genotyping to genome typing. *Nat. Rev. Genet.* 4, 981–994

19 Storz, J.F. (2005) Using genome scans of DNA polymorphism to infer adaptive population divergence. *Mol. Ecol.* 14, 671–688

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

21 Schlötterer, C. (2003) Hitchhiking mapping–functional genomics from the population genetics perspective. *Trends Genet.* 19, 32–38

22 Falush, D. *et al.* (2003) Traces of human migrations in *Helicobacter pylori* populations. *Science* 299, 1582–1585

23 Liti, G. *et al.* (2006) Sequence diversity, reproductive isolation and species concepts in *Saccharomyces*. *Genetics* 174, 839–850

24 Whitaker, R.J. and Banfield, J.F. (2006) Population genomics in natural microbial communities. *Trends Ecol. Evol.* 21, 508–516

25 Chakraborty, R. and Weiss, K. (1988) Admixture as a tool for finding linked genes and detecting that difference from allelic association between loci. *Proc. Natl. Acad. Sci. U. S. A.* 85, 9119–9123

26 Briscoe, D. *et al.* (1994) Linkage disequilibrium in admixed populations: applications in gene mapping. *J. Hered.* 85, 59–63

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

28 Zhu, X. *et al.* (2005) Admixture mapping for hypertension loci with genome-scan markers. *Nat. Genet.* 37, 177–181

29 Freedman, M.L. *et al.* (2006) Admixture mapping identifies 8q24 as a prostate cancer risk locus in African-American men. *Proc. Natl. Acad. Sci. U. S. A.* 103, 14068–14073

30 Deo, R.C. *et al.* (2007) A high-density admixture scan in 1,670 African Americans with hypertension. *PLoS Genet.* 3, e196

31 Nalls, M.A. *et al.* (2008) Admixture mapping of white cell count: genetic locus responsible for lower white blood cell count in the Health ABC and Jackson Heart studies. *Am. J. Hum. Genet.* 82, 81–87

32 Lamason, R.L. *et al.* (2005) SLC24A5, a putative cation exchanger, affects pigmentation in zebrafish and humans. *Science* 310, 1782–1786

33 Miller, C.T. *et al.* (2007) *cis*-Regulatory changes in Kit ligand expression and parallel evolution of pigmentation in sticklebacks and humans. *Cell* 131, 1179–1189

34 Norton, H.L. *et al.* (2007) Genetic evidence for the convergent evolution of light skin in Europeans and East Asians. *Mol. Biol. Evol.* 24, 710–722

35 Jablonski, N.G. and Chaplin, G. (2000) The evolution of human skin coloration. *J. Hum. Evol.* 39, 57–106

36 Harrison, R.G. (1990) Hybrid zones: windows on evolutionary process. *Oxf. Surv. Evol. Biol.* 7, 69–128

37 Barton, N.H. and Gale, K.S. (1993) Genetic analysis of hybrid zones. In *Hybrid Zones and the Evolutionary Process* (Harrison, R.G., ed.), pp. 13–45, Oxford University Press

38 Payseur, B.A. *et al.* (2004) Differential patterns of introgression across the X chromosome in a hybrid zone between two species of house mice. *Evolution* 58, 2064–2078

39 Teeter, K.C. *et al.* (2008) Genome-wide patterns of gene flow across a house mouse hybrid zone. *Genome Res.* 18, 67–76

40 Buerkle, C.A. and Rieseberg, L.H. (2001) Low intraspecific variation for genomic isolation between hybridizing sunflower species. *Evolution* 55, 684–691

41 Campbell, D. and Bernatchez, L. (2004) Genomic scan using AFLP markers as a means to assess the role of directional selection in the divergence of sympatric whitefish ecotypes. *Mol. Biol. Evol.* 21, 945–956

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

43 Lexer, C. *et al.* (2005) Barrier to gene flow between two ecological divergent *Populus* species, *P. alba* (white poplar) and *P. tremula* (European aspen): the role of ecology and life history in gene introgression. *Mol. Ecol.* 14, 1045–1057

44 van Loo, M. *et al.* (2008) Clonality and spatial genetic structure in *Populus × canescens* and its sympatric backcross parent *P. alba* in a Central European hybrid zone. *New Phytol.* 177, 506–516

45 Fossati, T. *et al.* (2004) Development of molecular markers to assess the level of introgression of *P. tremula* into *P. alba* natural populations. *Plant Breed.* 123, 382–385

46 Tuskan, G.A. *et al.* (2006) The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). *Science* 313, 1596–1604

47 Storz, J. and Hoekstra, H. (2007) The study of adaptation and speciation in the genomic era. *J. Mammal.* 88, 1–4

48 Frazer, K.A. *et al.* (2007) A sequence-based variation map of 8.27 million SNPs in inbred mouse strains. *Nature* 448, 1050–1053

49 Harr, B. (2006) Genomic islands of differentiation between house mouse subspecies. *Genome Res.* 16, 730–737

50 Payseur, B.A. and Place, M. (2007) Searching the genomes of inbred mouse strains for incompatibilities that reproductively isolate their wild relatives. *J. Hered.* 98, 115–122

51 Yang, H. *et al.* (2007) On the subspecific origin of the laboratory mouse. *Nat. Genet.* 39, 1100–1107

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

53 Cervino, A.C.L. *et al.* (2007) An integrated *in silico* gene mapping strategy in inbred mice. *Genetics* 175, 321–333

54 Payseur, B.A. and Place, M. (2007) Prospects for association mapping in classical inbred mouse strains. *Genetics* 175, 1999–2008

55 Hoekstra, H.E. *et al.* (2006) A single amino acid mutation contributes to adaptive beach mouse color pattern. *Science* 313, 101–104

56 Steiner, C.C. *et al.* (2007) Adaptive variation in beach mice produced by two interacting pigmentation genes. *PLoS Biol.* 5, e219

57 Gow, J.L. *et al.* (2006) Contrasting hybridization rates between sympatric three-spined sticklebacks highlight the fragility of reproductive barriers between evolutionarily young species. *Mol. Ecol.* 15, 739–752

58 Taylor, E.B. *et al.* (2006) Speciation in reverse: morphological and genetic evidence of the collapse of a three-spined stickleback (*Gasterosteus aculeatus*) species pair. *Mol. Ecol.* 15, 343–355

59 Nolte, A.W. *et al.* (2006) When invaders meet locally adapted types: rapid moulding of hybrid zones between sculpins (*Cottus*, Pisces) in the Rhine system. *Mol. Ecol.* 15, 1983–1993

60 Whibley, A.C. *et al.* (2006) Evolutionary paths underlying flower color variation in *Antirrhinum*. *Science* 313, 963–966

61 Minder, A.M. *et al.* (2007) Genetic structure of hybrid zones between *Silene latifolia* and *Silene dioica* (Caryophyllaceae): evidence for introgressive hybridization. *Mol. Ecol.* 16, 2504–2516

62 Mallet, J. *et al.* (2007) Natural hybridization in heliconiine butterflies: the species boundary as a continuum. *BMC Evol. Biol.* 7, 28

63 Montana, G. and Hoggart, C. (2007) Statistical software for gene mapping by admixture linkage disequilibrium. *Brief. Bioinform.* 8, 393–395

64 Chapman, N.H. and Thompson, E.A. (2003) A model for the length of tracts of identity by descent in finite random mating populations. *Theor. Popul. Biol.* 64, 141–150

65 MacLeod, A.K. *et al.* (2005) Marker densities and the mapping of ancestral junctions. *Genet. Res.* 85, 69–79

66 Pfaff, C.L. *et al.* (2001) Population structure in admixed populations: effect of admixture dynamics on the pattern of linkage disequilibrium. *Am. J. Hum. Genet.* 68, 198–207

67 Price, A.L. *et al.* (2006) Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.* 38, 904–909

68 Sha, Q. *et al.* (2006) Analytical correction for multiple testing in admixture mapping. *Hum. Hered.* 62, 55–63

69 Falush, D. *et al.* (2003) Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics* 164, 1567–1587

70 Mott, R. *et al.* (2000) A method for fine mapping quantitative trait loci in outbred animal stocks. *Proc. Natl. Acad. Sci. U. S. A.* 97, 12649–12654

71 Lynch, M. and Walsh, B. (1998) *Genetics and Analysis of Quantitative Traits.* Sinauer Associates

72 Kim, S. *et al.* (2007) Recombination and linkage disequilibrium in *Arabidopsis thaliana*. *Nat. Genet.* 39, 1151–1155

73 Laurie, C.C. *et al.* (2007) Linkage disequilibrium in wild mice. *PLoS Genet.* 3, e144

74 Buerkle, C.A. (2005) Maximum-likelihood estimation of a hybrid index based on molecular markers. *Mol. Ecol. Notes* 5, 684–687

75 Pritchard, J.K. *et al.* (2000) Inference of population structure using multilocus genotype data. *Genetics* 155, 945–959

76 Hoggart, C.J. *et al.* (2004) Design and analysis of admixture mapping studies. *Am. J. Hum. Genet.* 74, 965–978

77 Thornsberry, J.M. *et al.* (2001) *Dwarf8* polymorphisms associate with variation in flowering time. *Nat. Genet.* 28, 286–289

78 Caicedo, A.L. *et al.* (2004) Epistatic interaction between *Arabidopsis FRI* and *FLC* flowering time genes generates a latitudinal cline in a life history trait. *Proc. Natl. Acad. Sci. U. S. A.* 101, 15670–15675

79 Ingvarsson, P.K. *et al.* (2006) Clinal variation in *phyB2*, a candidate gene for day-length-induced growth cessation and bud set, across a latitudinal gradient in European aspen (*Populus tremula*). *Genetics* 172, 1845–1853

80 González-Martínez, S.C. *et al.* (2007) Association genetics in *Pinus taeda* L. I. Wood property traits. *Genetics* 175, 399–409

81 Aranzana, M.J. *et al.* (2005) Genome-wide association mapping in *Arabidopsis* identifies previously known flowering time and pathogen resistance genes. *PLoS Genet.* 1, e60

82 Fellay, J. *et al.* (2007) A whole-genome association study of major determinants for host control of HIV-1. *Science* 317, 944–947

83 The International HapMap Consortium (2007) A second generation human haplotype map of over 3.1 million SNPs. *Nature* 449, 851–861

84 Jiggins, C.D. and Mallet, J. (2000) Bimodal hybrid zones and speciation. *Trends Ecol. Evol.* 15, 250–255