THE ROYAL SOCIETY

Published online 21 May 2003

Effects of inbreeding on the genetic diversity of populations

Deborah Charlesworth

Institute of Cell, Animal and Population Biology (ICAPB), University of Edinburgh, Ashworth Laboratory, King's Buildings, West Mains Road, Edinburgh EH9 3JT, UK (*deborah.charlesworth@ed.ac.uk*)

The study of variability within species is important to all biologists who use genetic markers. Since the discovery of molecular variability among normal individuals, data have been collected from a wide range of organisms, and it is important to understand the major factors affecting diversity levels and patterns. Comparisons of inbreeding and outcrossing populations can contribute to this understanding, and therefore studying plant populations is important, because related species often have different breeding systems. DNA sequence data are now starting to become available from suitable plant and animal populations, to measure and compare variability levels and test predictions.

Keywords: diversity; inbreeding; linkage disequilibrium; balancing selection; *Arabidopsis*; isozymes

1. INTRODUCTION

The study of molecular variability within species has long been recognized as important by population geneticists, and since the discovery of variability among normal individuals of many species, using allozyme electrophoresis, the development and use of genetic markers have become widespread in biology. Population genetics concepts, such as diversity and linkage disequilibrium, are therefore now used in studies of population biology (which includes the study of gene flow and local adaptation) and phylogeography (which includes the study of historical events that have led to the present distributions of species). Understanding these concepts, and understanding the patterns of variability that may be found within species, are therefore important to all biologists who use genetic markers, not just to population geneticists.

The ideal markers for many population biology studies are neutral variants, i.e. variants without any phenotypic effects, and more precisely without any effects on the fitness of their carriers. The idea of neutrality was put on a rigorous basis of population genetics theory, and explicitly related to the process by which neutral or weakly selected substitutions will spread through finite populations by genetic drift, causing prolonged neutral polymorphism within populations. This theory can explain variant frequencies often being higher than would be seen under a balance between mutation to deleterious alleles and their removal by natural (purifying) selection. Abundant isozyme variants are indeed often found at high frequencies in natural populations. However, these variants involve amino acid differences in the enzymes, and might instead be maintained in populations by balancing selection. The neutrality/selection of allozymes has long been vigorously debated (Lewontin 1974), and remains unresolved even today. In practice, the absence of strong selection will often allow accurate inferences of population structure and breeding systems, and allozymes have been extremely informative about such questions. Neutrality of DNA sequence variants, such as microsatellite variants and variants detected by approaches such as AFLP, which are often located in non-coding regions of genomes, is often a reasonable assumption.

Molecular evolutionary studies, including tests for selection, are, however, increasingly important in analyses of genome sequence data. It may become possible to test for molecular adaptations underlying physiological adaptations of plants and animals to different environments, and to assess whether changes in genes and other sequences are driven by selection, for example to ask whether gene duplications lead to adaptation to new or to specialized functions (Lynch *et al.* 2001). The interest in using evidence of selection as evidence for functions, and in distinguishing selection from neutrality, has generated tests using neutrality as a null hypothesis, and it should be possible to detect the action of natural selection maintaining polymorphisms (balancing selection) as well as selection driving substitutions between related species (reviewed by Fay *et al.* 2001, 2002; Kreitman 2001).

In testing for selection, population structure cannot be ignored. This is true even for human populations, in which (as in many outcrossing species; see below) variability within populations is the largest proportion of total variability. Even in such cases, patterns of variability may be strongly affected by population structure. For example, linkage disequilibrium may be important, and may form important parts of the evidence for historical events, such as human migration from ancestral African populations (Tishkoff *et al.* 1996). Thus, in addition to the long established concern of population biologists with subdivision and restricted gene flow, because they affect populations' ability to adapt to local conditions (reviewed by Barton 2000), there is now great interest in understanding how to test for selection in subdivided populations.

One contribution of 21 to a Discussion Meeting Issue 'Mechanisms regulating gene flow in flowering plants'.

2. BREEDING SYSTEM EVOLUTION IN PLANTS

The study of populations of non-model species with differing breeding systems has been useful in many aspects of the study of diversity and its selective importance. An important piece of evidence was that species with a largely haploid life cycle, such as mosses, had high isozyme diversity (e.g. Wyatt *et al.* 1989), making balancing selection due to heterozygote advantage unlikely to account for most isozyme variability.

Breeding system differences, such as differences in self-fertilization rates, are probably among the factors with major effects on genetic variability, clear enough to be discernible even in the presence of other factors. H. G. Baker was one of the first to recognize that inbreeders often have lower genetic variability than outcrossing species, and that variability tends to be chiefly found between populations in inbreeding species rather than within them (Baker 1953, 1959), and he understood, in intuitive terms, several of the most important causes. Modern work using molecular markers, including allozymes and DNA-sequence differences, confirms this difference. Both populations' selfing rates and their diversity can now be quantified using markers, and this has led to the development of theoretical models that can rigorously predict diversity levels and the causes of differences.

Plant mating systems are varied, and transitions from inbreeding to outcrossing have occurred many times in angiosperm evolution (Stebbins 1957), making it possible to test whether such changes have repeatable effects on diversity patterns. The evolution of breeding systems is starting to be traced using phylogenetic estimates of relationships between species, providing a framework for determining whether outcrossing or inbreeding is the ancestral state in a set of plant taxa, and even sometimes to assign transitions to lineages. However, breeding system data are not available for many plant species, and at present we still have little information about variability in inbreeding within species, beyond the knowledge that populations, and even individual plants within populations, may differ in their propensity to inbreed (e.g. Lloyd 1965).

It is widely thought that self-fertilization in plants is probably often of recent origin, suggesting that inbreeding species persist for shorter than average evolutionary times (Barrett *et al.* 1996; Schoen *et al.* 1997; Bena *et al.* 1998), though the phylogenetic support for this is controversial (Takebayashi & Morrell 2001). SI has certainly often broken down (Stebbins 1957). Although several different pistil and pollen proteins are involved in recognition in different SI systems, so that a single ancestral system is unlikely, and we must conclude that SI probably evolved several times (Uyenoyama 1995), it nevertheless seems to evolve rarely, and its history in the angiosperms is probably ancient. For instance, several recent phylogenetic studies show that 'S-RNases' involved in gametophytic SI in the three families Solanaceae, Rosaceae and Scrophulariaceae form a cluster distinct from other RNases; if S-RNases were recently evolved paralogues of other RNase genes, the sequences would be expected to cluster within plant families (Steinbachs & Holsinger 2002). In the genus *Lycopersicon*, detailed studies of the evolutionary breakdown of SI suggest its independent loss in different groups of species (Kondo *et al.* 2002); no large self-compatible groups of species were found, so selfing lineages often seem to have short durations.

I deal with three aspects of the general topic of the effects of inbreeding on the genetic diversity of populations: measuring diversity, understanding its patterns, particularly the effects of inbreeding, and testing for the effects of different proposed processes. After describing some important measures of diversity, I consider the patterns expected for neutral variants, for three kinds of population model. Although it is generally not known very precisely how long inbreeding has persisted in any lineage, and thus whether there has been enough time for diversity differences to evolve, I shall suggest that some trends are clear. To attempt to convey an understanding of the diversity trends, models of panmictic populations are dealt with first, to establish some of the basic concepts, and then models with population subdivision, and finally the effects of inbreeding are introduced into these models. After discussing predictions assuming neutrality, some relevant effects of selection will be described. In both inbreeding and outcrossing populations, but particularly when there is inbreeding, it is necessary to understand the effects of selection, not only on loci that are themselves under various forms of selection, but also on neutral variants in the same populations. For each kind of model discussed, I will illustrate the predictions with some empirical observations from plant or other populations, and use these to discuss how the proposed effects can be tested using genetic markers and DNA sequence data.

3. MEASURES OF DIVERSITY WITHIN AND BETWEEN POPULATIONS

Diversity data have been collected from many species, including vast amounts of information from isozymes and microsatellites, and RFLPs. For markers that are scored by the presence or absence of bands generated by many PCR-based methods (RAPD, AFLP, inter-simple sequence repeats; Newton *et al.* 1999), the presence of a band is dominant. Although these markers can detect diversity, their dominance makes it difficult to quantify diversity, since allele frequencies cannot be determined directly but must be estimated assuming Hardy–Weinberg equilibrium (Ouborg *et al.* 1999; Krauss 2000; Mougel *et al.* 2002). This is unsuitable for inbreeding populations. For many types of marker, understanding diversity is further complicated because only the states of alleles can be observed, while the relationships between alleles (and the changes involved in transitions between alleles) are unknown.

I therefore concentrate on DNA sequence data and patterns of DNA sequence variability. DNA sequences have several important advantages that often outweigh the cost of obtaining the data. One advantage is that DNA variants are often neutral or reasonably close to neutral. Testing for deviations from expected patterns requires a model that can yield expectations of diversity and other aspects of diversity data, such as frequencies of alleles or variants. Many such models used in thinking about population structure assume selective neutrality of variants (Kimura 1983). Neutrality is therefore an important null hypothesis, and many tests are now available for departures from

neutral equilibrium caused by various biological processes, including aspects of population history and demography (Kreitman 2001). The neutral theory also provides measures of diversity that allow parameters of the models to be estimated. In addition, DNA sequences contain information about the relationships between alleles and about the associations of individual variants that are present in the sequences (linkage disequilibrium). This can allow valuable inferences about the source of variants, for example making it possible to detect variants that are shared between populations, to test whether gene flow between populations is contributing to high diversity. Most importantly, recent developments in coalescence theory have provided insights and results for many biologically realistic situations that were previously difficult to model. There has therefore been a great increase in our understanding.

Diversity can be measured in several different ways, depending on the type of variant of interest. For markers including allozymes and markers such as RAPDs and AFLPs (i.e. markers classified according to the presence/absence of bands of different mobility on gels), the proportion of loci that are polymorphic is often used. This can be used to compare populations' diversity, though if the populations are not closely related the bands amplified, or the enzymes detected, may not be the same.

The number of alleles per locus can be used to measure diversity for co-dominant markers, such as allozymes and microsatellites, and mean values can usefully be compared between populations, but this measure neglects information from allele frequencies. For DNA sequences, an analogous measure is the number of haplotypes, i.e. the number of distinct sequences found in a sample from a given locus. The chance of detecting a nucleotide polymorphism depends on the length of sequence, so this measure is not good for quantifying diversity, but it provides valuable information about whether haplotypes tend to carry different variants (i.e. the variants are in linkage disequilibrium, with many fewer haplotypes than single nucleotide polymorphisms). The accumulation of different sequence substitutions in isolated populations is one possible cause of such linkage disequilibrium.

An important diversity measure is based on comparing pairs of alleles in a sample at a given locus, and calculating the probability that they are different. For allozymes, this estimate uses the allele frequencies, p_i, to calculate 'gene diversity' as

$$H_e = 1 - \sum_i p_i^2,$$

where p_i is the frequency of the ith allelic type, so that p^2 gives the probability that two alleles in a sample will be of this type. By taking allele frequencies into account, we can obtain information about natural populations. For instance, bottlenecks in population size tend to cause preferential loss of alleles at the lowest frequencies. Thus, gene diversity (which depends strongly on the commoner alleles) may be only mildly affected even when allele numbers are strongly reduced (Nei *et al.* 1975). After such an event, allele frequencies will be detectably higher than expected in an equilibrium population (Luikart *et al.* 1998*a*,*b*). For allozymes and microsatellites, models that assume neutrality and specify a mutation process (for instance the infinite-alleles model for allozymes, in which each mutation creates a new allele, or stepwise mutation models in which microsatellite alleles can arise by addition of one or more repeat units) have been analysed in detail to provide tests for bottlenecks that are based on differences in allele numbers and diversity taking allele frequencies into account.

For DNA sequences, the neutral theory (Kimura 1983) often assumes the infinite-sites model, in which each mutation causes a polymorphism at a different site in the sequence. This is reasonable if mutation rates are low (so that a given site can be assumed never to experience a second mutation; if alternative alleles are maintained for very long periods in a lineage, or if there are mutation hotspots, this model may not be biologically accurate). An important classical result from this model is that the probability of polymorphism per nucleotide site is $\theta \approx 4N_e\mu$ (see Kimura 1983), where μ is the neutral mutation rate per nucleotide site, and N_e is the effective population size. As will be described below, many biologically important effects on diversity levels and patterns, including the effects of inbreeding and of population subdivision, can be understood in terms of factors that influence effective population sizes. θ can be estimated in two ways. The first measure, nucleotide diversity, π, often denoted by k, is the analogue of gene diversity defined above for allelic variants; it is based on the probabilities of site differences found in pairwise comparisons among a sample of sequences (Nei 1987; Tajima 1993). Another diversity measure is based on counting the number of variable nucleotide sites in a sample of n sequences (S_n). Such counts of polymorphisms are evidently sample size dependent. To take this into account, a correction is used to generate the measure M, which is equal to

$$S_n \Big/ \sum_{i=1}^{n-1} \frac{1}{i}.$$

Diversity can be expressed per nucleotide site by dividing M by the sequence length (Tajima 1993).

As for allele frequencies at isozyme loci, bottlenecks and other processes that reduce population sizes predominantly cause loss of rare variants at polymorphic loci, while common polymorphisms will often persist into the descendant population, so that M is more strongly affected than π. This has led to the development of tests based on differences between the two diversity measures, such as Tajima's D test (Tajima 1989*b*), that can detect population bottlenecks and other departures from the expected neutral equilibrium that reduce effective population sizes.

These diversity measures can be applied to subsets of sequence data, for instance to synonymous or non-synonymous sites in coding sequences, or to non-coding sites. Frequently, sites that are not expected to be under strong selection, i.e. synonymous and non-coding sites, are used as estimates of neutral diversity. In subdivided populations, it is also often of interest to estimate diversity separately within subpopulations. This can provide information about their history, for instance absence of variants in a subpopulation might suggest either recent colonization involving a founder event, or strong isolation (so that diversity has been lost due to genetic drift with a small effective size). The hypothesis of recent colonization can

be tested by sampling other populations of the species (the subpopulation would be expected to have a subset of the variants found in other populations, particularly alleles that are common in the putative source population). Allozymes are excellent for such tests (Husband & Barrett 1993; Allen *et al.* 1996; Broyles 1998).

It is also informative to compare mean within-population diversity, usually weighted across the populations studied, and often denoted by H_S for allozymes (Nei 1987) and microsatellites and π_S for nucleotide diversity (Hudson *et al.* 1992), with the diversity estimated using the entire species-wide sample, H_T or π_T. A quantity of interest is F_{ST}, the ratio of between-population to total species-wide diversity (Nei 1987); for DNA sequences, F_{ST} can be estimated by $(\pi_T - \pi_S)/\pi_T$ (e.g. Hudson *et al.* 1992; Charlesworth 1998). High F_{ST} can be due to isolation and large between-population differences. In some simple models of subdivided populations (particularly the island model in which alleles in a subpopulation have the same chance of having arrived by migration from any other deme, regardless of distance or other factors affecting gene flow between demes) it can be used to estimate migration rates (Whitlock & McCauley 1999). However, F_{ST} will also be high whenever π_S is low (Charlesworth *et al.* 1997; Charlesworth 1998), and F_{ST} values decrease when allelic diversity is very high, as is often found for microsatellite loci (Nagylaki 1998; Hedrick 2002).

4. THEORETICAL PREDICTIONS

The simplest situation to consider first is neutral diversity in a DNA sequence in a panmictic population of size N, which has the expected value already mentioned, $\theta \approx 4N\mu$. This result can readily be derived using coalescent theory to derive the expected times to common ancestors in samples of pairs of alleles from a population (Hudson 1990). The coalescent approach has several important advantages. By separating the ancestry of gene lineages from the mutations that occur on these lineages, it clarifies the important role of times for allelic sequences to trace back to their common ancestors. These 'coalescence times' are major determinants of the population's neutral diversity. In addition, this approach has made possible the treatment of mutation models for different types of genetic diversity, including nucleotide polymorphisms and microsatellites. It has made it possible to understand the behaviour of non-panmictic populations, by deriving effective sizes (N_e) that relate the results in these more realistic situations to those expected in a panmictic population with size N_e (Nordborg & Donnelly 1997; Wakeley 2000; Nordborg & Innan 2002). By providing reasonably simple methods for generating simulated samples of alleles, the coalescent approach also provides a way to test null hypotheses, by comparing a property of interest in an empirical dataset with its value in a large number of samples simulated under the null hypothesis. This is particularly important when dealing with diversity within species, because diversity has great variance, due both to the stochasticity of mutations and especially to the very large number of possible outcomes of the genealogical process (the evolutionary variance). These sources of variance must be taken into account in statistical testing (reviewed by Hudson 1990). It is even possible to simulate samples of recombining sequences (Hudson 1990).

Even in outcrossing species, real populations are not panmictic. Population subdivision is very common, in plant populations as well as those of many other organisms. Furthermore, as will be discussed below, the effects of inbreeding on diversity cannot be understood without taking into account how inbreeding interacts with population subdivision. Before considering the effects of inbreeding, it is therefore necessary to understand how isolation affects diversity, and the different expectations for diversity within populations and species-wide.

Isolation allows differentiation between subpopulations (often called demes), so high diversity can develop in the species as a whole. In a system of n demes, provided the populations are not completely isolated, migration is a much faster process than mutation. As a result, the within-deme neutral diversity measured from pairwise comparisons of sequences (the nucleotide diversity, π_S) is predicted to be increased by subdivision, to $4Nn\mu$, where N is the effective size per deme, or $4N_T\mu$, where N_T is the effective size of the entire set of demes, i.e. the size of a panmictic population with the same total number of breeding individuals (Slatkin 1987; Strobeck 1987). In such a species, F_{ST} will be expected to be low, and this is found in many studies of outcrossing species (figure 1*a*). This result applies in this form only to certain models of population structure and migration, such as the island or stepping-stone models, in which migration pressure does not change deme sizes (Nagylaki 2000). However, it is an important result because it shows that the interconnectedness of demes in a species prevents loss of diversity within demes, so that π_S is high, relative to π_T, even though the demes are small compared with the species' total effective size (Nordborg 1997). The demes' own effective sizes do not determine their diversity unless they are cut off from the gene flow.

Several biologically realistic situations can, however, reduce effective sizes of species with subdivided populations. For instance, in 'source–sink' situations, including situations with fluctuations in subpopulation sizes, alleles in different demes will mostly be recently descended from the source population (they will have short coalescent times). Both π_T and π_S will then be determined mainly by the source population's effective size, and F_{ST} will be low (Whitlock & Barton 1997; Nagylaki 2000). Local extinction and recolonization similarly shortens coalescent times and can greatly reduce diversity. The population turnover in such situations means that sequences from different demes trace quickly back to the same deme, and rapid growth after colonization causes low coalescence times within demes (Slatkin 1977; Whitlock & McCauley 1993; Pannell & Charlesworth 1999). Both π_T and π_S are lowered by local extinction and recolonization and by source–sink situations, so that species-wide diversity could be low. The net effect on F_{ST}, however, is dependent on the balance between colonization and immigration into demes (Whitlock & McCauley 1993; Pannell & Charlesworth 1999, 2000; Wakeley & Aliacar 2001). Gene flow through colonization from a variety of sources leads to low F_{ST}. If, however, new demes are colonized by small numbers of founders (less than roughly twice the number of immigrants per generation into established populations),

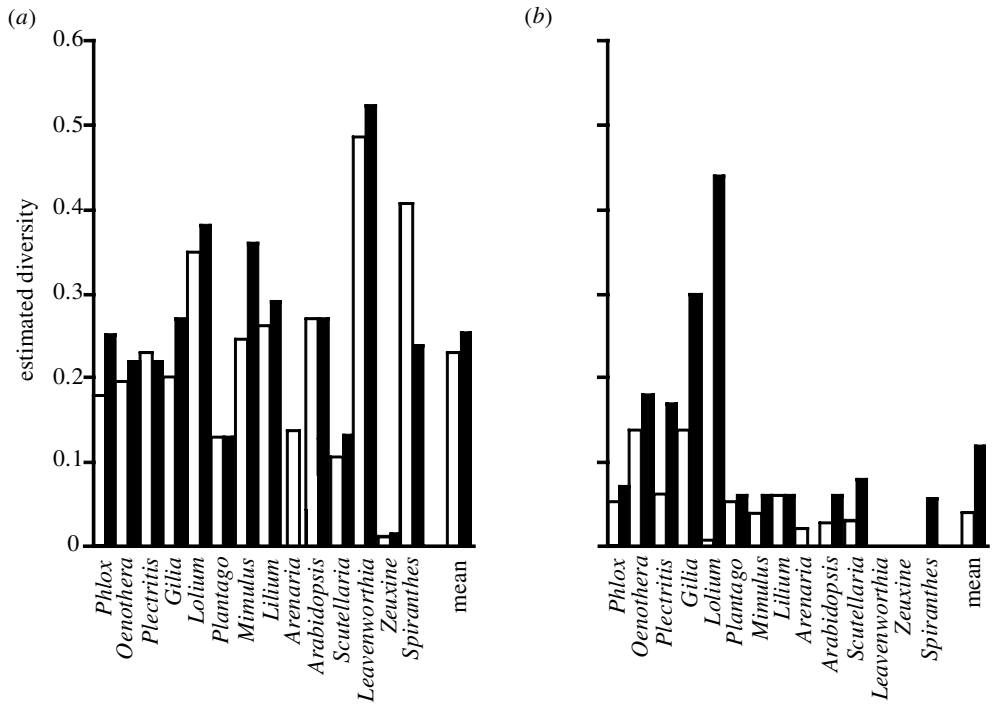

Figure 1. Allozyme diversity in samples from within subpopulations (open bars; H_S) and species-wide (closed bars; H_S) samples of (*a*) outcrossing and (*b*) inbreeding angiospem species, for 15 genera where paired comparisons are possible. Sources of data: *Phlox* (Levin 1978); *Oenothera* (Ellstrand & Levin 1980); *Plectritis* (Layton & Ganders 1984); *Gilia* (Schoen 1982); *Lolium* (Loos 1993); *Plantago* (Wolff 1991); *Mimulus* (Fenster & Ritland 1992); *Lilium* (Linhart & Premoli 1994); *Arenaria* (Wyatt *et al.* 1992); *Arabidopsis* (van Treuren *et al.* 1997; Abbott & Gomes 1988); *Scutellaria* (Olmstead 1989); *Leavenworthia* (Charlesworth & Yang 1997); *Zeuxine* (Sun & Wong 2001); *Spiranthes* (Sun 1995).

we expect chance differences in allele content of demes, and thus high differentiation and high F_{ST}. This corresponds to observations on *Silene dioica* populations: F_{ST} was high for newly founded populations, but was reduced for older established ones by gene flow (Ingvarsson & Giles 1999).

These processes will affect the diversity of both nuclear genes and those in the organelles. Chloroplast and mitochondrial diversity will generally be lower than for nuclear genes, because, as explained above, neutral diversity depends on the neutral mutation rate. Based on comparisons of silent site divergence between plant species, the mutation rate for chloroplast genes is severalfold lower than that of nuclear genes, and that of mitochondrial genes lower still (Wolfe *et al.* 1987; Gaut 1998). In addition, the effective size for maternally transmitted genes (such as mitochondria in many angiosperms) is lower than that of nuclear genes. In an outcrossing hermaphroditic population, it is half of the nuclear value, since genetic drift of this genome behaves like that in a haploid population (Birky *et al.* 1983), or one-quarter in dioecious species, since only half the population transmits the genome. Genetic drift will thus lead to more rapid loss of diversity for these genomes than for autosomal nuclear genes. Nevertheless, chloroplast and mitochondrial genome polymorphisms are found in many plants (reviewed in Forcioli *et al.* 1998; Terachi *et al.* 2001). Diversity is, as expected, generally lower, and differentiation between populations clearer, for chloroplast variants than nuclear ones (McCauley *et al.* 1996; Tarayre *et al.* 1997; Newton *et al.* 1999).

A low effective size is also expected for Y chromosomes, which are present in some dioecious plant species, and different effective sizes apply to X-linked versus autosomal genes in such species, since these have different probabilities of being in pollen and seeds (reviewed in Laporte & Charlesworth 2002). In *Silene latifolia* and *S. dioica*, some Y-linked genes have been discovered, and their sequences do indeed have low diversity (Filatov *et al.* 2000, 2001; Atanassov *et al.* 2001), and show more differentiation between populations, compared with their X-linked homologues. Further differences in N_e can be caused by variance in reproductive success. A large variance between different plants in male reproductive success, for example, reduces N_e generally, but particularly strongly for the Y chromosome (Laporte & Charlesworth 2002).

In plants, these effects are combined with important differences in the ability of different genes to migrate (Ennos 1994; McCauley 1994; McCauley *et al.* 1996; Hu & Ennos 1997; Laporte & Charlesworth 2002). Colonization must occur by seed migration, while immigration into established populations can involve seed and/or pollen flow, so that the effects of these processes may differ for genes in the nuclear genome, which (except for Y-linked genes) are transmitted by both seeds and pollen, and maternally inherited genes transmitted mainly by seeds (as is the case for mitochondrial genes in many angiosperms), and predominantly pollen-transmitted genes, such as the chloroplast genes of many gymnosperms. Even if seed migration is restricted, and maternally inherited variants are patchily distributed, pollen flow can reduce genetic differentiation between populations.

In a species with subdivided populations, non-equilibrium states with respect to diversity are likely. Recurrent population bottlenecks resulting from extinction and recolonization will reduce the frequencies of rare alleles within local populations so that alleles will tend to be at higher frequencies than expected at equilibrium (Luikart *et al.* 1998*a*). If a bottleneck is not too severe or prolonged, and some common variants from the progenitor population pass to the descendant one, many variants will be at intermediate frequencies, so Tajima's *D* values for DNA sequences will initially be positive (Tajima 1989*b*). If variability has been severely reduced (by a severe bottleneck, or if a small descendant population is isolated for long enough for loss of variability by genetic drift), new variants will subsequently arise by mutation, and the population will return towards the equilibrium frequencies of variants. Initially, mutations will cause an excess of low-frequency polymorphisms, resulting in negative Tajima's *D* values (Tajima 1989*a*,*b*). Different independent loci should be affected similarly by population bottlenecks, so that these can potentially be distinguished from other departures from the expected neutral equilibrium that reduce effective population sizes (Galtier *et al.* 2000). The transient effects on polymorphisms are not the same for all genes, however. Because of the lower N_e for maternally transmitted than nuclear genes, a given reduction in size behaves like a bottleneck of longer duration, and causes a more pronounced loss of rare variants for some time. This may account for the more negative Tajima's *D* values for human mitochondrial than nuclear loci (Fay & Wu 1999).

Population subdivision also affects haplotype numbers. If individuals in different populations do not hybridize, their gene sequences will not recombine, so that the number of haplotypes within demes will be low, relative to the number of polymorphisms in the species, and this can allow subdivision to be detected (Strobeck 1987). However, as has already been explained, species whose populations are subdivided into isolated demes may also often show the effects of population growth, which occurs as new demes are colonized. During episodes of population growth, the ancestry of sequences will tend to be a star-like phylogeny. New (rare) variants arising during the growth period will thus generally occur in different haplotypes, and there will be few associations among variants, even in non-recombining genomes (Slatkin 1994; Przeworski & Wall 2001; Pritchard & Przeworski 2001). The low linkage disequilibrium of variants due to mutations during the growth period might sometimes obscure linkage disequilibrium due to subdivision.

5. EFFECTS OF SELECTION ON DIVERSITY

Patterns of diversity in plant populations are, of course, likely to be affected by selection. Balancing selection due to overdominance (heterozygote advantage), or to frequency-dependent selection, may maintain variants in populations, and environmental differences may select for different genotypes in different populations. Purifying selection, however, removes deleterious variants that arise by mutation; such variants are expected to be present at frequencies lower than predicted for the neutral equilibrium. Another form of directional selection occurs when advantageous mutations rapidly reach high frequencies, whether they spread throughout a species to fixation or just within a population undergoing adaptation to its local environment (such as tolerance to copper or other heavy metals, e.g. Macnair *et al.* 1989; Vekemans & Lefebvre 2001), or establish a polymorphic equilibrium under balancing selection. When selection on a variant occurs, other sites (including neutral sites) in the gene are affected.

(a) *Balancing selection*

The SI loci of some plant species are a classical example of frequency-dependent selection, in which the effects of balancing selection on neutral diversity are evident. Many *S*-alleles are maintained within natural populations by a fertility advantage to rare alleles, which rarely encounter the same allele in a pollination event, and are thus rarely rejected (Wright 1939). Such alleles are expected to remain polymorphic for long evolutionary times (Vekemans & Slatkin 1994), and very similar alleles may be found in related species (Shiba *et al.* 2002). Variation at non-synonymous sites in such loci will include the sites that control the amino acid differences that are the targets of selection. Synonymous sites closely linked to the targets of balancing selection will have their diversity increased, because of the long times during which different functional alleles maintained, allowing sequence differentiation of alleles, which can only exchange variants through rare recombination or gene conversion events (Strobeck 1972; Takahata 1990). This effect is seen in *S*-allele sequences, which have extremely high diversity at synonymous sites (Richman *et al.* 1996*b*) and in the introns, which cannot themselves be under selection (Nishio *et al.* 1997; Charlesworth & Awadalla 1998; Schierup *et al.* 2001). Amino acid variants may also accumulate between alleles, adding differences at sites that do not cause specificity differences. In *S*-loci, non-synonymous diversity is pronounced (e.g. Richman *et al.* 1996*a*; Nasrallah & Nasrallah 1989; Schierup *et al.* 2001).

Different functional alleles will thus form haplotypes with distinctive variants across the sequence, and the high diversity and linkage disequilibrium will decline where recombination has allowed exchanges between the allelic alternatives in the region. In agreement with this prediction, genes located close to the *Brassica S*-locus, which may recombine at low frequency with the *S*-locus, appear to have low diversity (Hinata *et al.* 1995), though detailed studies with large samples of alleles of these loci have not yet been carried out. Local peaks of diversity are thus expected at loci under balancing selection (Nordborg *et al.* 1996) and excess common variants, so that Tajima's *D* values will be positive (Tajima 1989*b*). Less extreme diversity than at the *S*-loci is expected if alleles are maintained polymorphic for less evolutionary times, but high silent-site diversity has been observed at loci encoding some allozymes (Filatov & Charlesworth 1999). The observation of haplotype structure exceeding that expected under neutrality can suggest balancing selection (Wall 1999). However, for recombination and/or gene conversion rates that are plausible for *Drosophila*, the region affected by balancing selection is expected to be small, confined to the selected locus (Andolfatto & Nordborg 1998).

To detect balancing selection in a subdivided population, it is important to sample within demes. Subdivision

itself leads to differentiation, and positive Tajima's D values are expected for neutral loci (see above). However, in samples taken from a subdivided population as a whole, sequences linked to a locus under balancing selection may have D values lower than unlinked loci, though D is still positive; D increases for regions extremely tightly linked to the selected locus, but if migration is very restricted it may still be too low to detect the selection (Schierup *et al.* 2000*a*).

Balancing selection causes reduced differentiation between populations, compared with neutral loci, which can be measured by F_{ST} (Schierup *et al.* 2000*b*). This is chiefly because an incoming migrant allele that is not already present in a deme has a higher chance of establishment compared with a neutral allele, so that its effective migration rate is increased. The few relevant empirical observations are consistent with this prediction. In the fungus *Schizophyllum commune*, low differentiation for incompatibility loci (Raper *et al.* 1958) contrasts with high differentiation for a reference locus (James *et al.* 1999). Such differences in diversity patterns may allow balancing selection to be detected, particularly now that data from multiple loci can be compared (Lewontin & Krakauer 1973; Baer 1999; Bamshad *et al.* 2002).

A form of balancing selection that affects some plant populations is the maintenance of CMS factors in gynodioecious species. It is not known how long these polymorphisms are maintained. Gynodioecious populations with CMS can be invaded by nuclear factors that restore male fertility of individuals with the sterility cytoplasm, which can cause reversion to hermaphroditism and loss of CMS, but sometimes the population may remain gynodioecious, polymorphic for sterility and fertility cytoplasms, and for the restorer (Charlesworth 1981; Frank 1989; Gouyon *et al.* 1991). If restorers often become fixed, and new CMSs then arise, mitochondrial alleles will constantly 'turn over'; so cytoplasmic diversity should be low, restricted to variants accumulated since the last mitochondrial genome replacement. By contrast, if CMS polymorphisms persist for long evolutionary times, high DNA sequence diversity is expected. High diversity has now been documented within two gynodioecious species *Silene acaulis* and *S. vulgaris* (Ingvarsson & Taylor 2002; Städler & Delph 2002). It will be interesting to study patterns of chloroplast and nuclear genetic diversity within and between populations of gynodioecious species, because the maintenance of the CMS factors may reduce cytoplasmic gene differentiation relative to nuclear genes, altering the patterns from those reviewed above.

(b) *Directional selection*

Diversity is also affected by directional selection. The spread of advantageous mutations reduces the effective size for sites within the selected locus, and for linked genes. This important form of 'hitchhiking' (Maynard Smith & Haigh 1974; Kaplan *et al.* 1989) is now usually called a 'selective sweep'. If an advantageous mutation spreads quickly, diversity at sites close to the site under selection may be greatly reduced, and the selected allele (haplotype) will have low diversity. The size of the genome region whose diversity is reduced depends on how quickly the advantageous mutation spreads, and on the local recombination frequency. Cases have been detected in domesticated plants (Wang *et al.* 1999; Purugganan *et al.* 2000; Terachi *et al.* 2001) and in natural populations (Filatov *et al.* 2001). Recent increases in frequency, suggesting directional selection, are particularly clear when a genetically uniform high-frequency haplotype is found among other more diverse haplotypes, as seen in human populations where thalassaemia and β-globin alleles have spread in malarial regions (Fullerton *et al.* 1994; Currat *et al.* 2002) or at the *Sod* locus of *Drosophila melanogaster* (Hudson *et al.* 1997).

The reduced diversity resulting from a selective sweep can be detected for some time following the event, but eventually the equilibrium level of diversity is re-established. During the recovery period, the new mutations will mostly be at lower than the expected equilibrium frequencies, so that recent selective sweeps may be detected in DNA sequences, by the excess number of low-frequency variants. This non-equilibrium frequency spectrum of variants causes diversity estimated from the number of polymorphic sites (θ) to exceed the nucleotide diversity (π), as detected by tests such as Tajima's D or Fu and Li's F_s (Tajima 1989*b*; Fu & Li 1993; Braverman *et al.* 1995; Kreitman 2001).

Adaptation to specific local environmental conditions is a prominent feature of plant populations (Clausen *et al.* 1947; Baker 1953; Linhart & Grant 1996) and may cause local selective sweeps, which may be distinguishable from bottlenecks due to founder events at the time of colonization, since only individual loci will be affected. No examples in plants have so far been published. It is not yet clear whether different alleles of plant disease-resistance genes are maintained within populations, or whether selection maintains resistance to different pathogen genotypes in different populations (Parker 1994; Kaltz & Shykoff 1998; Tian *et al.* 2002). These genes are often highly polymorphic within species, with high amino acid variability suggesting diversifying selection (Bergelson *et al.* 2001). Genome-wide scans for unusually low microsatellite diversity are being developed to search for loci that may have been involved in adaptation to European conditions in *Drosophila* (Harr *et al.* 2002).

Genetic differences between populations can cause haplotype structure that may be difficult to distinguish from balancing selection (e.g. Charlesworth *et al.* 1997). Neutral variants at marker and other loci that migrate into a subpopulation by seed or pollen flow from another population will tend to carry alleles at the selected locus that differ from the local allele. They will thus be maladapted, and less likely to establish outside their source population. Selection thus creates a barrier to genetic exchange between locally adapted populations (Barton & Bengtsson 1986), which can generate stable frequency differences for neutral variants, and non-zero F_{ST} values, particularly at loci linked to the targets of selection (Charlesworth *et al.* 1997). Observing genetic differentiation at a locus does not, therefore, imply selective differences affecting that particular locus (see Hedrick & Holden 1979). Directional selection at a locus causes negative Tajima's D values, but only within demes that have recently experienced selection. If species-wide samples are tested, evidence for selection may be obscured by genetic differences between demes. Tajima's test will then not distinguish between balancing selection and dif-

ferentiation between populations, because both cause intermediate allele frequencies in samples from the entire population, and thus positive D values (Tajima 1989*b*). Local differences in alleles at a flowering-time gene, *FRIGIDA*, in *A. thaliana* suggest local adaptation, and the observation that most of the variants are non-synonymous suggests that differentiation may truly be due to selection at this locus (Le Corre *et al.* 2002).

Another form of hitchhiking occurs when selection removes deleterious mutations with moderately strong effects, so that they quickly disappear from populations, reducing the number of lineages that leave descendants, thus lowering the population's effective size and diversity (Charlesworth *et al.* 1993). If non-recombining regions exist in a genome with enough genes to have a substantial deleterious mutation rate, this 'background selection' can detectably reduce diversity. Background selection affects variant frequencies similarly to selective sweeps, but much less strongly (Charlesworth *et al.* 1995).

Differences in recombination frequencies are a major cause of differences in diversity levels within genomes. In centromeric regions of genomes of outcrossing species, recombination is absent, but genes are present both in animals such as *Drosophila* (Charlesworth & Guttman 1996) and in plants (Copenhaver *et al.* 1999). Low-recombination regions of the *D. melanogaster* genome regions tend to have low diversity (Aguadé *et al.* 1989; Begun & Aquadro 1992), most probably caused by hitchhiking effects of selection occurring at loci within these regions, though there is debate about the relative importance of selective sweeps and background selection (Andolfatto *et al.* 2001).

6. EFFECTS OF INBREEDING

In thinking about diversity, we can distinguish differences in the amount of diversity and differences in diversity patterns (the distribution of variants between individuals and populations), and this helps an understanding of the effects of inbreeding on diversity. In the short term, inbreeding should affect diversity patterns only. The frequency of homozygotes is increased, but initially neutral variants should not be lost. However, the rarity of heterozygotes means that variants are less likely to be maintained by overdominance (heterozygote advantage) in inbred populations, and these would be expected to be lost within a relatively small number of generations of inbreeding (Kimura & Ohta 1971; Charlesworth & Charlesworth 1995). In the longer term, inbreeding represents an important deviation from panmixia, and has several effects that can strongly affect amounts of diversity. First, high homozygosity reduces effective sizes relative to those of outbreeding populations with the same number of individuals. Second, inbreeding reduces the effective frequency of recombination throughout the genome. Finally, inbreeding increases isolation between individuals and populations.

High homozygosity affects neutral diversity because it reduces effective sizes relative to those of outbreeding populations. Using the coalescent reasoning, it is simple to see why this occurs for a completely inbreeding population (Nordborg & Donnelly 1997). For nuclear loci, the expected time to common ancestry (and thus N_e) in such a population is halved compared with a panmictic population of the same size. This is because a completely inbreeding population consists of homozygous genotypes, and therefore the time to common ancestry of the alleles in a sample is half that in a diploid population. The reduction in N_e is less than twofold in populations with less than complete inbreeding, which are thus predicted to have more than half the neutral diversity of comparable outbreeding populations. Assuming the 'mixed-mating' model with a population of N individuals assumed to reproduce with the same self-fertilization rate S (Brown 1979), the expected equilibrium inbreeding coefficient is $F = S/(2 - S)$, and the effective population size is $N_e = N/(1 + F)$ (Pollak 1987; Nordborg 2000).

A highly inbred hermaphroditic species should have equal N_e for the nuclear and organelle genomes, assuming maternal transmission of organelles, since both values will be half the outcrosser's nuclear N_e. Organelle genes of a selfing hermaphrodite should have similar N_e to that of an outcrosser with a comparable population number, but their nuclear N_e is halved; comparing a dioecious outcrosser with a selfing hermaphrodite, the organelle N_e is twice as high in the selfer. Thus, relative neutral diversity values can be predicted.

The second effect acting to reduce diversity in inbreeding populations is that the high homozygosity of individuals in inbreeding populations reduces the effective frequency of recombination throughout the genome. This affects the diversity of inbred populations in several ways. Without recombination, half of the time for a sample of sequences to coalescence to their common ancestor is the time for the last two sequences to coalesce. Even in an undivided population, the sequences would thus have a tendency to fall into two subsets, corresponding to the two branches of the gene tree, i.e. the diversity will be structured into two haplotypes. Recombination obliterates this dichotomy in the sequences (Hudson 1990). The quantitative effect of inbreeding on linkage disequilibrium of polymorphic sites in DNA sequences is expressed in terms of the population recombination parameter $4N_er$, where r is the recombination rate per base pair. For a population at equilibrium with a self-fertilization rate S, r in this expression becomes $r(1 - F)$, where F is the inbreeding coefficient, and the effective size is that of a population with the relevant selfing rate (Nordborg 2000). In inbreeders, closely linked sites, such as sites within a gene, will thus often show haplotype structure (i.e. there will be fewer haplotypes than expected, given the number of polymorphic sites; this can be detected as high linkage disequilibrium, or low numbers of recombination events; Hudson & Kaplan 1985; Wall 1999).

There is clear evidence of low effective recombination in *A. thaliana*. Linkage disequilibrium extends for more than 100 kb (Hagenblad & Nordborg 2002; Nordborg *et al.* 2002), whereas in *D. melanogaster* (Miyashita *et al.* 1993) and maize (Tenaillon *et al.* 2001) sequences it decays within a few hundred nucleotides. Linkage disequilibrium has not yet been compared in orthologous gene sequences of related inbreeders and outbreeders, and comparison may be difficult, because inbreeding populations often seem to evolve higher rates of crossing over (reviewed in Charlesworth *et al.* 1979) so that one cannot assume that recombination rates are the same per base pair. It is, however, clear that haplotype structure is often

observed in *A. thaliana* gene sequences (Aguadé 2001) and, for at least some loci in the highly inbreeding *Hordeum spontaneum* (Lin *et al.* 2002), is higher than in maize. Haplotype structure may suggest balancing selection or locally adapted alleles, but for inbreeding populations such interpretations cannot be accepted without independent supporting evidence.

A second implication of low effective recombination rates in inbreeders is that, if hitchhiking processes are important, they will further reduce inbreeding populations' effective sizes, and thus sequence diversity within such populations (Hedrick & Holden 1979; Hedrick 1980; Charlesworth *et al.* 1993). The lower effective recombination frequency implies that in highly inbreeding species regions throughout the entire genome are non-independent in their transmission. In the extreme, hitchhiking effects of both kinds, including selective sweeps occurring at any locus, may thus affect all other loci. The diversity of genes in the organelle genomes, as well as nuclear genes, would be predicted to be affected (Maruyama & Birky 1991; Charlesworth *et al.* 1993). An effect of inbreeding on numbers of chloroplast haplotypes, as well as isozymes, was detected in *Mimulus* species (Fenster & Ritland 1992). However, effects on organelle sequence diversity are difficult to study in plants, because of their low mutation rates (Wolfe *et al.* 1987), so that very long sequences will be needed to detect enough variants to estimate diversity and detect any differences.

When inbreeding evolves from an initially outcrossing population, a gene that causes the change must spread (or several genes in succession), so that one or more selective sweeps must occur. For example, a non-functional incompatibility allele may spread (e.g. Kondo *et al.* 2002), or a gene causing anther–stigma separation to be small, promoting self-pollination (e.g. Macnair *et al.* 1989). If an advantageous mutation causing selfing spreads rapidly, the entire resulting inbreeding population will effectively have been subjected to a bottleneck in population size and will, for a period of time, have low diversity, possibly so low that it is not possible to apply tests for selective sweeps, which require large samples of allele sequences within populations (Braverman *et al.* 1995). If, however, the evolution of selfing is slow, recombination may separate the loci involved in the phenotypic change from the neutral or marker loci studied. Polymorphisms from the ancestral outcrossing population may then persist in the selfing species.

(a) *Subdivided inbreeding populations*

Inbreeding interacts with population subdivision in its effects on diversity (figure 2). Many inbreeders are weedy colonizing species, and may frequently undergo bottlenecks and founder events, which will cause low within-deme diversity (Schoen & Brown 1991). Several differences from outcrossers increase isolation between subpopulations of inbreeders. Inbreeders' flowers often develop rapidly, with small anther–stigma separation, and their pollen and pistils mature simultaneously, so that self-pollination within the flowers may occur autogamously before pollinators visit, or during their visits, pre-empting ovules from being outcrossed (e.g. Macnair *et al.* 1989; Dole 1992). Inbred species also generally evolve reduced pollen output (Lloyd 1965; Cruden 1977), which lowers gene flow through pollen. Seed migration in inbreeders could be higher or lower than in related outcrossers. Nevertheless inbreeding populations will generally be more isolated than outcrossers. Local deme effective sizes may thus be small, and diversity low, so that F_{ST} values are high (Jain 1983; Hamrick & Godt 1990). Isolation also means that π_T may be high, possibly even higher than that for an otherwise comparable outcrosser, despite the reduced effective size caused by homozygosity. No general prediction can therefore be made about relative species-wide diversity in inbreeders and outbreeders, unless migration patterns are known in detail and quantified. The predicted reduction in diversity caused by lower effective size in inbreeders should therefore be tested using samples of multiple plants within demes, and it is not surprising that species-wide samples often show little or no reduction in the diversity of inbreeders, compared with related outbreeders. Strong isolation between demes of highly inbreeding species could lead to species-wide diversity being reduced by less than one-half. However, frequent colonizing events, combined with isolation may lead to low diversity species-wide (Charlesworth & Pannell 2001; Ingvarsson 2002).

Lower diversity in inbreeding than outcrossing populations is clear from the vast amount of allozyme data from plants (Jain 1983; Hamrick & Godt 1990, 1996; Berge *et al.* 1998; Crawford *et al.* 2001) and the lesser body of animal data (Jarne & Staedler 1995), and also from comparisons of populations of the same species with contrasting levels of inbreeding (Rick *et al.* 1977; Barrett & Husband 1990; Charlesworth & Yang 1997; Sun & Wong 2001), and in paired comparisons in plant genera (figure 1). Because some allozymes may be subject to balancing selection, it is important to compare silent diversity using DNA sequence data, and results are starting to appear from natural populations and crop species. However, few comparisons have been made for orthologous genes in related outcrossers, and data from within natural populations (the ideal type of data, as explained above) are scarce.

Two inbreeding species, the animal *Caenorhabditis elegans* and the plant *A. thaliana*, are among the small number of intensively studied 'model species', with many gene sequences available and, recently complete genome sequences, making it straightforward to carry out surveys of diversity. *C. elegans* has low species-wide diversity, and very high linkage disequilibrium (Koch *et al.* 2000). Diversity in this species therefore seems to behave as predicted for a highly inbreeding species. However, its breeding system has not been determined in nature, and it is not clear whether the observations can be quantitatively accounted for assuming an undivided inbreeding population. To test the extent of diversity reduction, comparisons are needed with related outcrossers. Recently, two highly inbreeding *Caenorhabditis* species were compared with their dioecious congener *C. remanei* (Graustein *et al.* 2002). Species-wide diversity estimates of two nuclear genes are reduced by about four–sixfold in the inbreeders. The diversity of a mitochondrial gene was halved in the inbreeder, which is roughly the expected effect relative to that estimated for the nuclear genes (see above). The extent of the reduction in diversity suggests that either population history has reduced N_e for this species as a whole, or else hitchhiking processes have been occurring. It will be helpful to have

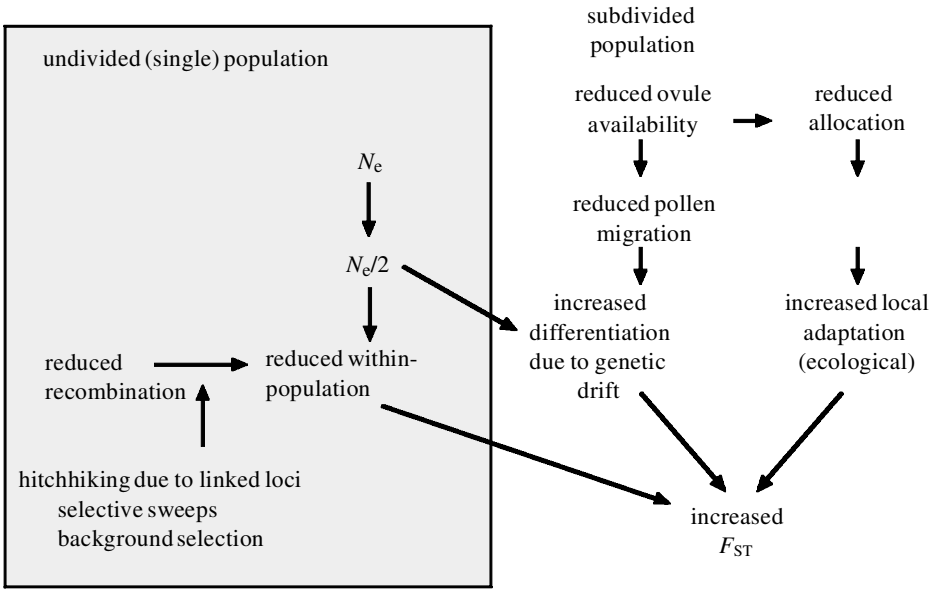

Figure 2. Summary of some of the various effects of inbreeding on diversity in undivided and subdivided populations.

larger samples from within populations of these species, to assess diversity patterns and test whether the low diversity could be accounted for by extinction and recolonization, and to obtain some evidence about migration rates between populations. At present, all that is known is that there is some diversity within local populations, though the sample sizes so far studied are very small and F_{ST} has not been estimated (Graustein *et al.* 2002).

The genus *Arabidopsis* should be excellent for comparing diversity in inbreeders and outbreeders. *Arabidopsis thaliana* is a weedy annual with a high self-fertilization rate (Abbott & Gomes 1988; Berge *et al.* 1998), but the closely related species *A. lyrata* and *A. halleri* are self-incompatible and highly outcrossing. The divergence times of these species are not accurately known, but are probably *ca.* 5 Myr (Koch *et al.* 2000, 2001), and cannot be very short, since introns in 11/18 genes surveyed are smaller in *A. thaliana* than *A. lyrata*, mainly because of accumulation of small insertions and deletions and simple sequence repeats; such differences could not arise quickly (Wright *et al.* 2002).

In *A. thaliana*, species-wide diversity levels (reviewed in Aguadé 2001) are lower than in maize (Tenaillon *et al.* 2001; Whitt *et al.* 2002), which has lost diversity, presumably owing to a bottleneck during domestication (Doebley 1989; Eyre-Walker *et al.* 1998). *Arabidopsis thaliana* is not, however, depauperate in diversity. As has been explained, species-wide data are difficult to interpret, and high species-wide diversity is quite likely in inbreeders, so an understanding of the causes of reduced diversity in inbreeders requires data on within-population samples. Very low levels of within-population allozyme (Abbott & Gomes 1988; Berge *et al.* 1998) and microsatellite polymorphism (Todokoro *et al.* 1995) are found within *A. thaliana* populations, consistent with estimates based on RFLP surveys (Bergelson *et al.* 1998). The outcrosser, *A. lyrata*, has much higher allozyme and microsatellite diversity; for allozymes, diversity values were estimated to be 0.27 and 0.40 for two populations of the subspecies *petraea*, and slightly lower (0.23 and 0.25) for the subspecies *lyrata* (van Treuren *et al.* 1997). A survey of DNA sequence polymorphism (Savolainen *et al.* 2000) provided preliminary evidence for higher within-population diversity at the *A. lyrata Adh* locus compared with diversity estimated by RFLP studies in *A. thaliana* (Bergelson *et al.* 1998), though species-wide polymorphism was higher in *A. thaliana*. A study of five further nuclear loci in these two species (Wright *et al.* 2003) confirms higher within-population silent-site polymorphism for *A. lyrata* than *A. thaliana* (figure 3). Interestingly, the European subspecies *A. lyrata* ssp. *petraea* has much higher variability overall, compared with *A. thaliana*, while the North American subspecies, *lyrata*, has less diversity than *A. thaliana*.

In the genus *Lycopersicon* (Solanaceae), inbreeding species have much lower diversity than outcrossers, based on RFLP studies (Miller & Tanksley 1990; Stephan & Langley 1998). In surveys of DNA sequence variability at five loci, estimated within-population silent-site diversity was reduced more than fourfold in both of the inbreeding species compared with any of the three outcrossing species; for *L. chmielewskii*, the reduction was 40-fold, compared with the self-incompatible species with the lowest diversity, *L. hirsutum* (Baudry *et al.* 2001). Some of these diversity differences may be attributable to differences in numbers of individuals, but it would be surprising if inbreeders systematically have lower numbers in each genus studied.

Greatly reduced diversity has also been found in the genus *Leavenworthia* (Brassicaceae), which includes self-incompatible species, but in which inbreeding has evolved several times (Lloyd 1965). An *Adh* locus has high synonymous and intron site variability in both outcrossing taxa studied, *L. stylosa*, and, to a lesser degree, in self-incompatible *L. crassa* populations, but no variation was found within populations of the inbreeding species *L. uniflora* and *L. torulosa*, and in self-incompatible *L. crassa* populations, little or no diversity was found (Liu *et al.* 1998). The diversity difference is highly significant (Innan & Tajima 2002). Surveys of five further loci in *L. stylosa*, *L. uniflora* and *L. torulosa* yielded the same extreme contrast in diversity (figure 4). In *L. crassa*, whose inbreeding probably evolved recently, the difference is less clear

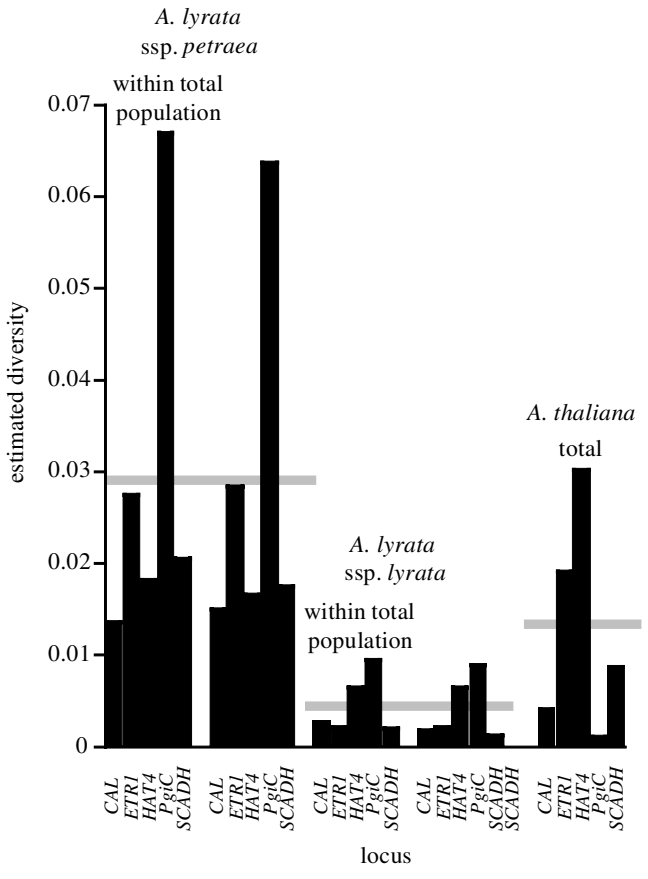

Figure 3. Silent-site polymorphism for five nuclear loci in *Arabidopsis lyrata* and *A. thaliana*. Within-population estimates are shown for the two *A. lyrata* subspecies and total diversity for each subspecies, together with the mean for each taxon (grey bars). For *A. thaliana*, only species-wide (total) diversity estimates are available.

(Liu 1998); diversity tends to be reduced within the inbred populations, but is not greatly reduced in the inbreeding populations treated as a distinct race (figure 4).

These results suggest strong general support for the predicted diversity reduction in inbreeding populations, compared with related outcrossers, with inbreeders' variability commonly reduced by more than twofold within populations, and often by approximately twofold, or somewhat more, overall. It is less easy to test which of the possible processes reviewed above are responsible. In weedy species, extinction and recolonization may be the major cause of low diversity. For example, this type of population structure may contribute to very low nucleotide diversity levels in *Plasmodium falciparum* (Volkman *et al.* 2001), in combination with inbreeding in populations where infection rates are low (Anderson *et al.* 2000). In addition, selective sweeps when drug resistance evolves will have genome-wide effects, and such advantageous mutations can potentially spread throughout the species, even if populations are subdivided. There are few data from plants, but extinction and recolonization could be the explanation for the very low apparent species-wide diversity levels of *L. uniflora* and *L. torulosa* (Liu 1998), despite *L. uniflora* having a more widespread distribution than other species in the genus, and a large total plant number in the species (Rollins 1963).

The high total diversity in many inbreeding species suggests, however, that extinction/recolonization is not generally responsible for their reduced within-population diversity. High species-wide diversity in inbreeders is observed for allozyme markers (Hamrick & Godt 1990). In *Lycopersicon*, some inbreeding species have high allozyme diversity (figure 1; see also Rick *et al.* 1977), suggesting that population history does not explain the inbreeders' low within-population diversity. The greatly reduced silent nucleotide site diversity within inbreeding *Lycopersicon* populations suggests that some process must be operating in addition to a reduced effective size due to homozygosity. DNA sequence data from multiple populations are not yet available to test for extinction and recolonization or other forms of source–sink population history. Tajima's D values in the inbreeding populations were low or negative, which might suggest expansion from a smaller size, but the tests are not significant. In *L. crassa*, there is substantial diversity in the inbreeding race. Although F_{ST} estimates are high, $\pi_T - \pi_S$ is also high (figure 4). In this case, however, we cannot exclude the possibility that the diversity may be due to the recent evolution of inbreeding in this race, so that ancestral variants have not yet been lost. Nevertheless, we clearly cannot account for the low diversity within these populations by population history, or by extreme bottlenecks or selective sweeps affecting the entire inbreeding race.

In addition to reduced within-population diversity, the effect of local adaptation in retarding genetic exchange can be very strong in inbreeders, because of their low effective recombination rates. Inbreeding populations are therefore even more isolated (relative to comparable outbreeding ones) than their reduced pollen movement would predict. Large allele frequency differences, and high F_{ST} values, may thus be seen at a high proportion of marker loci (Charlesworth *et al.* 1997). This can account for the strong differentiation between certain inbreeding populations at multiple loci (e.g. Brown *et al.* 1980; Nevo *et al.* 1988; Lin *et al.* 2002; Volis *et al.* 2002). An important implication is that in inbreeders it may be particularly difficult to identify loci that are the targets of selection and distinguish them from genes whose diversity is affected by selection at other loci. Another implication is that haplotype structure found in species-wide samples of sequences does not imply that balancing selection at the locus is acting to maintain variability within populations. When testing for selection, it is thus illuminating to use samples from within local populations, which may be able to distinguish within-population balanced polymorphism from local adaptation.

In *A. thaliana*, within-population diversities appear to be low and, together with high between-population diversity, this leads to high F_{ST}. Mean F_{ST} values based on multiple allozyme loci were estimated to be 0.61 for English *A. thaliana* populations (Abbott & Gomes 1988) and 0.71 in Norway (Berge *et al.* 1998) and a mean of 0.64 was estimated in a study of RFLP variants in North American populations (Bergelson *et al.* 1998). These values contrast with more moderate F_{ST} values for populations of the outcrossing *A. lyrata*, despite significant differentiation between populations. An estimate for allozymes was 0.41 (van Treuren *et al.* 1997), and from sequences of nuclear loci 0.16–0.54; these values are

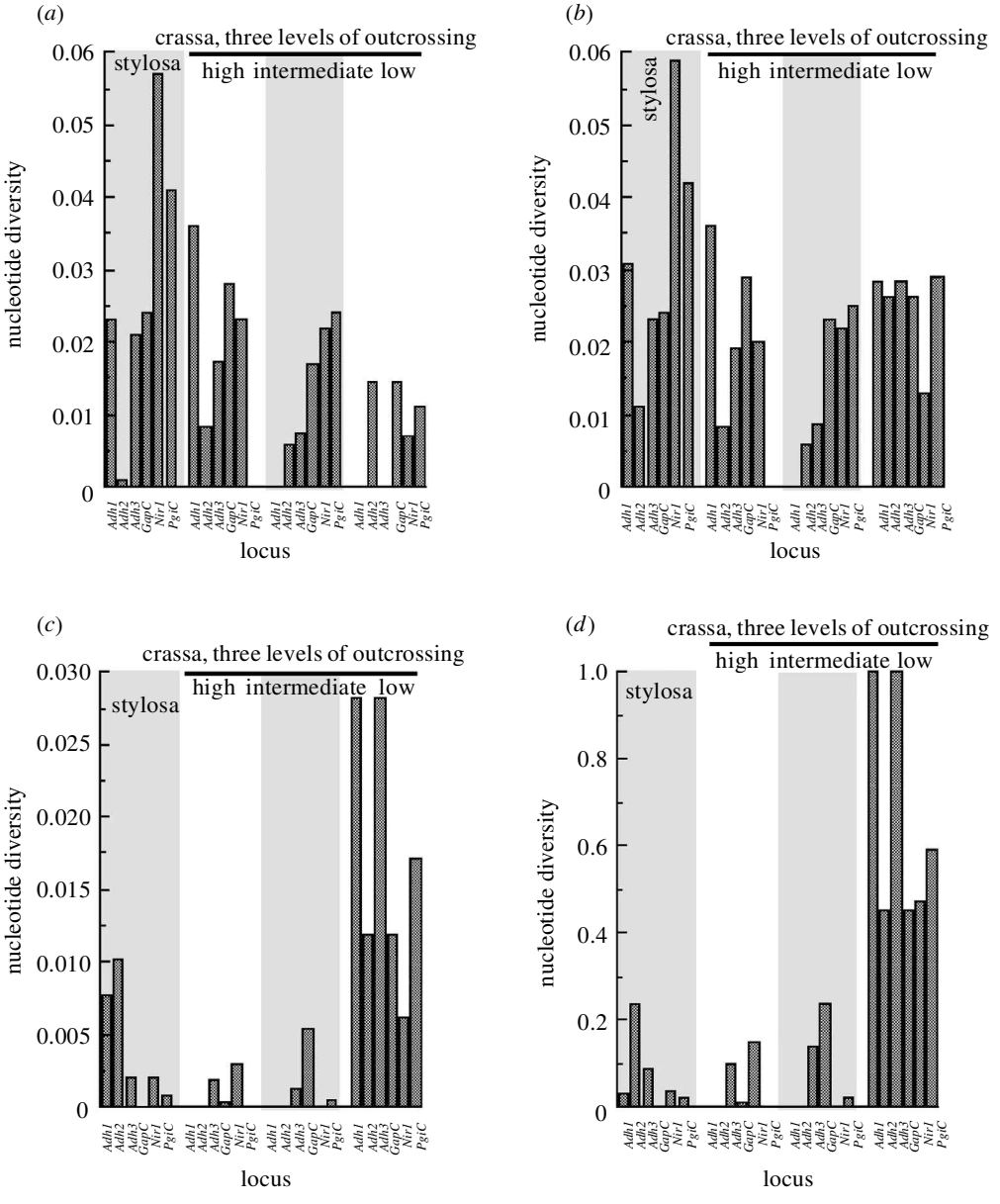

Figure 4. Mean DNA sequence diversity for six loci in *Leavenworthia* populations with contrasting breeding systems. *Leavenworthia stylosa* is self-incompatible. Its two highly inbred relatives, *L. uniflora* and *L. torulosa*, have no diversity and are not shown in the figure. For *L. crassa*, three sets of populations are shown separately (self-incompatible, intermediate outcrossing, and highly self-fertilizing). The figure shows, for each locus (*a*) mean over the populations of the within-population silent (intron and synonymous site) nucleotide diversity values (π_S), (*b*) total diversity for the entire sample, from all populations (π_T), (*c*) the difference between these ($D_{ST} = \pi_T - \pi_S$), and (*d*) F_{ST} ($\pi_T - \pi_S/\pi_T$).

inflated by the very low diversity within the two North American populations (Wright *et al.* 2003). The low diversity in *A. lyrata* subspecies *lyrata* might be due to recent colonization of North America, and the possibility that *A. lyrata* as a whole is not at equilibrium is suggested by excess linkage disequilibrium at several of the loci studied (Wright *et al.* 2003). This makes it difficult to compare diversity values in these two *Arabidopsis* species. Studies of more loci are needed from both *A. lyrata* and *A. thaliana*, particularly if any substantial fraction of loci are affected by selection, and studies of more populations are also needed.

The high F_{ST} values in *A. thaliana* suggest either strong isolation between populations, or isolation intensified by local adaptation, as explained above, and species-wide absence of diversity is excluded. It is difficult to explain the very low diversity within local populations, compared with *A. lyrata*, without any hitchhiking process. The only alternative would seem to be to propose that *A. thaliana* populations have been totally isolated for enough time for genetic drift to have allowed almost all within-deme variation to be lost, while between-deme differences have evolved. This possibility should be verifiable using sequence data from within populations to assess the degree to which variants are shared between demes, and to test for migration. Under the hypothesis of long-term isolation, one might expect to find isolation by distance, since when populations become extinct they should mostly be replaced by colonists from nearby populations, but there is little sign of such a pattern (Sharbel *et al.* 2000) so hitchhiking processes within local populations appear more plausible.

It is also possible that high diversity in outcrossers could be due to introgression from other species. Certainly, diversity is remarkably high in many plant species, including maize, where gene flow is not implausible (Tenaillon *et al.* 2001). This could be a further cause of differences in within-deme diversity between species with different breeding systems. Many plants can form hybrids, sometimes across considerable genetic distances (e.g. Crawford 1989), and several situations that are plausible in plants may produce high diversity from this cause. For instance, populations may often be isolated in refugia during one climatic period for long enough to allow differentiation, after which population expansion or migration occurs so that the differentiated populations introgress or merge. This might sometimes lead to higher diversity in outbreeding than inbreeding populations, since the greater isolation of inbreeding than outcrossing populations may mean that in some species pairs, pollen flow will go largely from inbreeding into outcrossing populations. This could be detectable from a geographical pattern in which high diversity is confined to populations where hybridization is possible. Self-incompatible populations generally reject pollen from related species, including inbreeders that have lost incompatibility, even when the reciprocal pollinations succeed (Lewis & Crowe 1958; Martin 1967); therefore introgression is unlikely for some species, but this rule is not always obeyed. For *L. stylosa*, gene flow cannot account for the high diversity levels, because the species is not cross-compatible with other extant species (Rollins 1963), but there is evidence that it could contribute to the high diversity of *L. chilense*, which shares polymorphisms with *L. peruvianum* (Baudry *et al.* 2001). For *A. lyrata* ssp. *petraea*, hybridization with *A. halleri* is possible, and a detailed dataset from many populations will be needed to test this. The isolation of inbreeding populations is also not absolute, but allows some introgression, and this could sometimes cause high diversity within such populations and species. Gene flow may sometimes be detectable from the presence of shared polymorphisms in DNA sequences, particularly when blocks of sequence are shared between two species but are sufficiently rare to be recognizable as 'foreign', or from the higher variance of pairwise nucleotide differences between sequences in two populations between which there is gene flow (Wakeley 1996).

7. SEQUENCE EVOLUTION IN INBREEDING POPULATIONS

Sequence evolution in inbreeders may also differ from that in outcrossing populations. Homozygosity increases the fixation probabilities of recessive advantageous alleles in inbreeding populations, so that newly arisen selfers might evolve rapidly, assuming that genetic variation exists from the outcrossing ancestral population (Charlesworth 1992). However, the fixation probabilities of deleterious mutations are much less affected. Inbreeding slightly increases the probabilities for dominant mutations (by less than twofold, even for the most deleterious mutations, for which the effect was the largest), and slightly decreases the probabilities for recessive mutations (Charlesworth 1992). If, however, hitchhiking effects lead to effective population sizes of inbreeders being much lower than those of outbreeders, natural selection will be less effective, since the product $N_e s$ determines the efficacy of selection, where s is the selection coefficient (Kimura 1983). Slightly deleterious mutations that are rapidly eliminated in a large randomly mating population may reach high frequencies, leading to excess segregating amino acid polymorphisms, and fixation in inbreeding populations. An excess of segregating amino acid polymorphisms has been observed for some loci in several inbreeding plants (Cummings & Clegg 1998), including *A. thaliana* (Savolainen *et al.* 2000; Bustamente *et al.* 2002), and it will be important to test whether this is a general property of genes in this species. At present, however, it is not clear whether these amino acid polymorphisms represent selected differences in different populations, or reflect a weakened efficacy of selection in these species.

The possibility of weakened efficacy of selection in inbreeding species is important, because it could increase the rate of sequence evolution (and/or lead to less optimal usage of synonymous codons). A faster rate of sequence evolution has been observed in annual than perennial plants (Bousquet *et al.* 1992; Gaut *et al.* 1996; Eyre-Walker & Gaut 1997), and is generally attributed to a generation-time effect. Since annuals are often inbreeding species (Hamrick & Godt 1990), inbreeding and life history are confounded, but both may contribute, and it will be necessary to compare species with contrasting breeding systems, but matching life histories, to test this. Such rate differences may need to be taken into account in phylogenetic inferences.

Comparisons between sequence evolution in inbreeders and outbreeders can now be made using the genus *Arabidopsis*. Ideally, *A. thaliana* and *A. lyrata* can be compared using an outgroup species to determine the numbers of substitutions on the two lineages. A recent study tested for increased non-synonymous/synonymous (K_a/K_s) ratios of substitutions since the *A. thaliana* and *A. lyrata* lineages diverged, using *Capsella rubella* sequences as outgroups where possible (or, failing that, sequences from the more distant relative, *Brassica*). With the sample of 25 loci studied (figure 5), there was no evidence for a difference in the efficacy of selection. Studies of more loci are, however, needed (Wright *et al.* 2002).

Levels of codon bias in *C. elegans* and *A. thaliana* are both substantially lower than in *D. melanogaster* (Marais *et al.* 2001), suggesting that in the two inbreeding species some slightly deleterious fixations may occur. However, in the comparisons between *A. thaliana* and *A. lyrata* genes, there was no strong sign of loss of bias in synonymous codon usage (Wright *et al.* 2002). Numbers of non-preferred relative to preferred substitutions do not differ significantly between the two species, though *A. lyrata* shows significantly higher levels of major codon usage for low-biased genes. A larger sample of loci has been compared between *A. thaliana* and *Brassica rapa* (Tiffin & Hahn 2002). There was some evidence for less optimal usage of synonymous codons in *A. thaliana*, but this conclusion is based on the assumption that the optimal codons do not differ in the two species, which are highly diverged (K_s values for 218 genes ranged from 7.6% to greater than 1, with an overall mean of 47%). The high divergence also means that the evolutionary time representing the highly inbreeding lineage is probably only a small fraction of the total time separating the species, so that clear signs of a

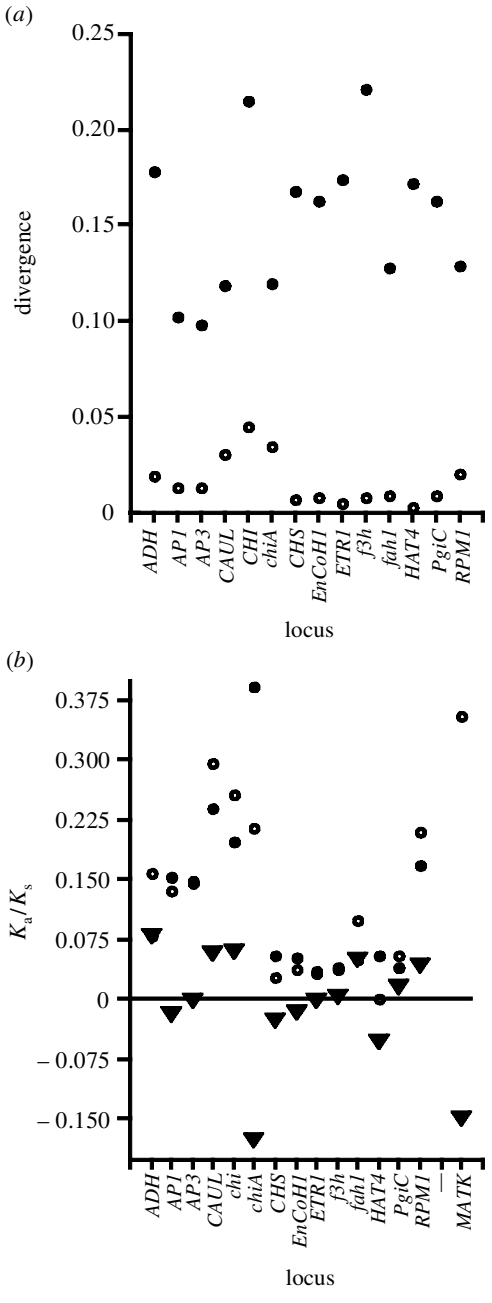

Figure 5. Divergence between *Arabidopsis thaliana* and *A. lyrata* for synonymous and non-synonymous sites in 15 nuclear loci. (*a*) K_a (open circles) and K_s (filled circles) values, and (*b*) the K_a/K_s ratios estimated for the lineages since the two species separated, as well as the K_a/K_s differences between the two lineages. No *A. lyrata* result is shown for MATK because it exceeds the maximum value on the *y*-axis. *Arabidopsis thaliana* (open circles); *A. lyrata* (filled circles); *A. thaliana*−*A. lyrata* (filled inverted triangles).

changed pattern of sequence evolution in *A. thaliana* might not be detectable. Indeed, *A. thaliana* may have been inbreeding for too little time for the slow process of genetic drift to change codon usage from that in *A. lyrata*.

8. CONCLUSIONS

The expected low diversity in inbreeding populations is clearly observed for allozymes, and observations of low DNA sequence diversity in inbreeders from the four well-studied genera *Caenorhabditis*, *Arabidopsis*, *Lycopersicon* and *Leavenworthia*, together with the much larger amount of allozyme data, suggest that a robust pattern exists. However, more species comparisons are needed. Many factors can affect effective population sizes, and thus diversity. At present, DNA sequence diversity data from plants are from loci whose recombination rates are unknown (Baudry *et al.* 2001), though recombination rates may affect genetic variation in plants similarly to *Drosophila*. An effect has been detected in populations of *Beta vulgaris* (Kraft *et al.* 1998). Mutation rates are another important factor, and it is not yet clear how much these may differ between species, and between different loci within a species, but some cases of low diversity may be due to low mutation rates (Dvornyk *et al.* 2002). It is now well established that diversity is often low in species and populations with small actual numbers of individuals, and in endemic species (Karron 1987; Hamrick & Godt 1990; van Treuren *et al.* 1991; Sun 1995; Gitzendanner & Soltis 2000). It will therefore not be clear whether there is a repeatable trend for low nucleotide diversity in inbreeders unless consistent differences are found in multiple species pairs of a variety of organisms.

In attempting to understand the causes of low diversity of inbreeding populations, it will be important to study polymorphisms in both inbreeding species and related outbreeding reference species. Loss of polymorphisms maintained by selection is one possible cause of the diversity differences. If differences between inbreeders and outbreeders are confirmed for silent variability at multiple loci, this will suggest that its cause is not loss of polymorphisms maintained within local populations by heterozygote advantage. It would be very surprising if a high proportion of the loci that have been studied had such polymorphisms. Investigations should nevertheless be conducted into the outbreeding reference populations, and diversity data should include sequences under little selection, such as pseudogenes, as well as coding sequences. To date, most diversity studies have not used samples large enough to have a high chance of detecting balancing selection, but there is no evidence for such selection at most loci studied in outcrossing populations. In *L. stylosa* populations, polymorphism at the *PgiC* locus may be maintained by selection (Filatov & Charlesworth 1999), and this locus is not polymorphic in the related inbreeders *L. uniflora* and *L. torulosa* but it is polymorphic in *A. thaliana* (Kawabe *et al.* 2000). If this possibility, as well as extinction and recolonization, could be ruled out, and if gene flow were known not to be cut off between inbreeding populations, the observed greater than twofold reduction in diversity in many such populations would suggest that one or both of the hitchhiking processes described above (selective sweeps and/or background selection) may be important. Because selective sweeps are a plausible cause of low diversity, studies should be large enough to detect these events, so we need larger samples of alleles within populations than in most studies of plant populations to date.

There is also a need for more theoretical studies. The effects of balancing selection on diversity of neutral variants at nearby sites has so far been studied only in terms of a single selected site (Hudson *et al.* 1987; Nordborg *et al.* 1996; Charlesworth *et al.* 1997; Takahata & Satta 1998), whereas some balanced polymorphisms, such as SI

loci (Charlesworth & Awadalla 1998) and disease resistance loci (Bergelson *et al.* 2001), may have several functionally different alleles. In such cases, one may expect that diversity is increased, not just in a narrow region but more widely throughout the locus (Navarro & Barton 2002; Nordborg & Innan 2003). The expected behaviour of weakly selected mutations in structured populations also needs to be better understood, as the effects of hitchhiking processes on neutral diversity are not clear in such populations (Slatkin & Wiehe 1998). Without such understanding, we will not know whether the apparent excess of segregating amino acid polymorphisms often seen in inbreeding plants (see above) could be caused by local adaptation to different environmental conditions.

REFERENCES

Abbott, R. J. & Gomes, M. F. 1988 Population genetic structure and outcrossing rate of *Arabidopsis thaliana* (L.) Heynh. *Heredity* **62**, 411–418.

Aguadé, M. 2001 Nucleotide sequence variation at two genes of the phenylpropanoid pathway, the FAH1 and F3H genes, in *Arabidopsis thaliana*. *Mol. Biol. Evol.* **18**, 1–9.

Aguadé, M., Miyashita, N. & Langley, C. H. 1989 Reduced variation in the yellow-achaete-scute region in natural populations of *Drosophila melanogaster*. *Genetics* **122**, 607–615.

Allen, G. A., Antos, J. A., Worley, A. C., Suttill, T. A. & Hebda, R. J. 1996 Morphological and genetic variation in disjunct populations of the avalanche lily *Erythronium montanum*. *Can. J. Bot.* **74**, 403–412.

Anderson, T. J. C., Paul, R. E. L., Donnelly, C. A. & Day, K. P. 2000 Do malaria parasites mate non-randomly in the mosquito midgut? *Genet. Res. Camb.* **75**, 285–296.

Andolfatto, P. & Nordborg, M. 1998 The effect of gene conversion on intralocus associations. *Genetics* **148**, 1397–1399.

Andolfatto, P., Wall, J. D. & Przeworski, M. 2001 Adaptive hitchhiking effects on genome variability. *Curr. Opin. Genet. Dev.* **11**, 635–641.

Atanassov, I., Delichère, C., Filatov, D. A., Charlesworth, D., Negrutiu, I. & Monéger, F. 2001 A putative monofunctional fructose-2,6-bisphosphatase gene has functional copies located on the X and Y sex chromosomes in white campion (*Silene latifolia*). *Mol. Biol. Evol.* **18**, 2162–2168.

Baer, C. F. 1999 Among-locus variation in Fst: fish, allozymes and the Lewontin–Krakauer test revisited. *Genetics* **152**, 653–659.

Baker, H. G. 1953 Race formation and reproductive method in flowering plants. *SEB Symp.* **7**, 114–145.

Baker, H. G. 1959 Reproductive methods as a factor in speciation in flowering plants. *Cold Spring Harb. Symp. Quant. Biol.* **24**, 177–191.

Bamshad, M. J. (and 10 others) 2002 A strong signature of balancing selection in the 5′ cis-regulatory region of CCR5. *Proc. Natl Acad. Sci. USA* **99**, 10 539–10 544.

Barrett, S. C. H. & Husband, B. C. 1990 Variation in outcrossing rates in *Eichhornia paniculata*: the role of demographic and reproductive factors. *Pl. Species Biol.* **5**, 41–55.

Barrett, S. C. H., Harder, L. D. & Worley, A. C. 1996 Comparative biology of plant reproductive traits. *Phil. Trans. R. Soc. Lond.* B **351**, 1272–1280.

Barton, N. H. 2000 The evolutionary consequences of gene flow and local adaptation: future approaches. In *Dispersal: individual, population and community* (ed. E. D. J. Clobert, A. A. Dhondt & J. D. Nichols), pp. 329–340. Oxford University Press.

Barton, N. H. & Bengtsson, B. O. 1986 The barrier to genetic exchange between hybridizing populations. *Heredity* **57**, 357–376.

Baudry, E., Kerdelhué, C., Innan, H. & Stephan, W. 2001 Species and recombination effects on DNA variability in the tomato genus. *Genetics* **158**, 1725–1735.

Begun, D. J. & Aquadro, C. F. 1992 Levels of naturally occurring DNA polymorphism correlate with recombination rates in *D. melanogaster*. *Nature* **356**, 519–520.

Bena, G., Prosperi, J. M., Lejeune, B. & Olivieri, I. 1998 Evolution of annual species of the genus *Medicago*: a molecular phylogenetic approach. *Mol. Phylogenet. Evol.* **9**, 552–559.

Berge, G., Nordal, I. & Hestmark, G. 1998 The effect of breeding systems and pollination vectors on the genetic variation of small plant populations within an agricultural landscape. *Oikos* **81**, 17–29.

Bergelson, J., Stahl, E. A., Dudek, S. & Kreitman, M. 1998 Genetic variation within and among populations of *Arabidopsis thaliana*. *Genetics* **148**, 1311–1323.

Bergelson, J., Kreitman, M., Stahl, E. A. & Tian, D. 2001 Evolutionary dynamics of plant R-genes. *Science* **292**, 2281–2285.

Birky, C. W., Maruyama, T. & Fuerst, P. 1983 An approach to population and evolutionary genetic theory for genes in mitochondria and chloroplasts, and some results. *Genetics* **103**, 513–527.

Bousquet, J., Strauss, S. H., Doerksen, A. H. & Price, R. A. 1992 Extensive variation in evolutionary rate of *rbcL* gene sequences among seed plants. *Proc. Natl Acad. Sci. USA* **89**, 7844–7848.

Braverman, J. M., Hudson, R. R., Kaplan, N. L., Langley, C. H. & Stephan, W. 1995 The hitchhiking effect on the site frequency spectrum of DNA polymorphisms. *Genetics* **140**, 783–796.

Brown, A. H. D. 1979 Enzyme polymorphism in plant populations. *Theor. Pop. Biol.* **15**, 1–42.

Brown, A. H. D., Feldman, M. W. & Nevo, E. 1980 Multilocus structure of natural populations of *Hordeum spontaneum*. *Genetics* **96**, 523–536.

Broyles, S. B. 1998 Post-glacial migration and the loss of allozyme variation in *Asclepias exaltata* (Asclepiadaceae). *Am. J. Bot.* **85**, 1091–1097.

Bustamente, C., Nielsen, R., Sawyer, S. A., Olsen, K. M., Purugganan, M. D. & Hartl, D. L. 2002 The cost of inbreeding in *Arabidopsis*. *Nature* **416**, 531–534.

Charlesworth, B. 1992 Evolutionary rates in partially self-fertilizing species. *Am. Nat.* **140**, 126–148.

Charlesworth, B. 1998 Measures of divergence between populations and the effect of forces that reduce variability. *Mol. Biol. Evol.* **15**, 538–543.

Charlesworth, B. & Guttman, D. S. 1996 Reductions in genetic variation in *Drosophila* and *Escherichia coli* caused by selection at linked sites. *J. Genet.* **75**, 49–61.

Charlesworth, B., Morgan, M. T. & Charlesworth, D. 1993 The effect of deleterious mutations on neutral molecular variation. *Genetics* **134**, 1289–1303.

Charlesworth, B., Nordborg, M. & Charlesworth, D. 1997 The effects of local selection, balanced polymorphism and background selection on equilibrium patterns of genetic diversity in subdivided inbreeding and outcrossing populations. *Genet. Res.* **70**, 155–174.

Charlesworth, D. 1981 A further study of the problem of the maintenance of females in gynodioecious species. *Heredity* **46**, 27–39.

Charlesworth, D. & Awadalla, P. 1998 The molecular population genetics of flowering plant self-incompatibility polymorphisms. *Heredity* **81**, 1–9.

Charlesworth, D. & Charlesworth, B. 1995 Quantitative genetics in plants: the effect of breeding system on genetic variability. *Evolution* **49**, 911–920.

Charlesworth, D. & Pannell, J. R. 2001 Mating systems and population genetic structure in the light of coalescent theory.

In *Integrating ecology and evolution in a spatial context. British Ecol. Soc. Special Symp. 2000* (ed. J. Silvertown & J. Antonovics), pp. 73–95. Oxford: Blackwell Science.

Charlesworth, D. & Yang, Z. 1997 Allozyme diversity in *Leavenworthia* populations with different inbreeding levels. *Heredity* **81**, 453–461.

Charlesworth, D., Charlesworth, B. & Strobeck, C. 1979 Selection for recombination in self-fertilising species. *Genetics* **93**, 237–244.

Charlesworth, D., Charlesworth, B. & Morgan, M. T. 1995 The pattern of neutral molecular variation under the background selection model. *Genetics* **141**, 1619–1632.

Clausen, J., Keck, D. D. & Hiesey, W. M. 1947 Heredity of geographically and ecologically isolated races. *Am. Nat.* **81**, 114–133.

Copenhaver, G. P. (and 13 others) 1999 Genetic definition and sequence analysis of *Arabidopsis* centromeres. *Science* **286**, 2468–2474.

Crawford, D. J. 1989 Enzyme electrophoresis and plant systematics. In *Isozymes in plant biology* (ed. D. E. Soltis & P. S. Soltis), pp. 146–164. Portland, OR: Dioscorides Press.

Crawford, D. J. (and 11 others) 2001 Allozyme diversity in endemic flowering plants endemic of the Juan Fernandez archipelago, Chile: ecological and historical factors with implications for conservation. *Am. J. Bot.* **88**, 2195–2203.

Cruden, R. W. 1977 Pollen–ovule ratios: a conservative index of breeding systems in flowering plants. *Evolution* **31**, 32–46.

Cummings, M. P. & Clegg, M. T. 1998 Nucleotide sequence diversity at the alcohol dehydrogenase locus in wild barley (*Hordeum vulgare* ssp. *spontaneum*): an evaluation of the background selection hypothesis. *Proc. Natl Acad. Sci. USA* **95**, 5637–5642.

Currat, M., Trabuchet, G., Rees, D., Perrin, P., Harding, R. M., Clegg, J. B., Langaney, A. & Excoffier, L. 2002 Molecular analysis of the beta-globin gene cluster in the Niokholo Mandenka population reveals a recent origin of the β^S Senegal mutation. *Am. J. Hum. Genet.* **70**, 207–223.

Doebley, J. 1989 Isozyme evidence and the evolution of crop plants. In *Isozymes in plant biology* (ed. D. E. Soltis & P. S. Soltis), pp. 165–191. Portland, OR: Dioscorides Press.

Dole, J. A. 1992 Reproductive assurance mechanisms in three taxa of the *Mimulus guttatus* complex (Scrophulariaceae). *Am. J. Bot.* **79**, 650–659.

Dvornyk, V., Sirviö, A., Mikkonen, M. & Savolainen, O. 2002 Low nucleotide diversity at the pal1 locus in the widely distributed *Pinus sylvestris*. *Mol. Biol. Evol.* **19**, 179–188.

Ellstrand, N. C. & Levin, D. A. 1980 Recombination system and population structure in *Oenothera*. *Evolution* **34**, 923–933.

Ennos, R. A. 1994 Estimating the relative rates of pollen and seed migration among plant-populations. *Heredity* **72**, 250–259.

Eyre-Walker, A. & Gaut, B. S. 1997 Correlated rates of synonymous site evolution across plant genomes. *Mol. Biol. Evol.* **14**, 455–460.

Eyre-Walker, A., Gaut, R. L., Hilton, H., Feldman, D. L. & Gaut, B. S. 1998 Investigation of the bottleneck leading to the domestication of maize. *Proc. Natl Acad. Sci. USA* **95**, 4441–4446.

Fay, J. C. & Wu, C. I. 1999 A human population bottleneck can account for the discordance between patterns of mitochondrial versus nuclear DNA variation. *Mol. Biol. Evol.* **16**, 1003–1005.

Fay, J. C., Wyckoff, G. J. & Wu, C. I. 2001 Positive and negative selection on the human genome. *Genetics* **158**, 1227–1234.

Fay, J. C., Wyckoff, G. J. & Wu, C. I. 2002 Testing the neutral theory of molecular evolution with genomic data from *Drosophila*. *Nature* **415**, 1024–1026.

Fenster, C. B. & Ritland, K. 1992 Chloroplast DNA and isozyme diversity in two *Mimulus* species (Scrophulariaceae) with contrasting mating systems. *Am. J. Bot.* **79**, 1440–1447.

Filatov, D. A. & Charlesworth, D. 1999 DNA polymorphism, haplotype structure and balancing selection in the *Leavenworthia* PgiC locus. *Genetics* **153**, 1423–1434.

Filatov, D. A., Monéger, F., Negrutiu, I. & Charlesworth, D. 2000 Evolution of a plant Y-chromosome: variability in a Y-linked gene of *Silene latifolia*. *Nature* **404**, 388–390.

Filatov, D. A., Laporte, V., Vitte, C. & Charlesworth, D. 2001 DNA diversity in sex linked and autosomal genes of the plant species *Silene latifolia* and *S. dioica*. *Mol. Biol. Evol.* **18**, 1442–1454.

Forcioli, D., Saumitou-Laprade, P., Valero, M., Vernet, P. & Cuguen, J. 1998 Distribution of chloroplast DNA diversity within and among populations in gynodioecious *Beta vulgaris* ssp. *maritima* (Chenopodiaceae). *Mol. Ecol.* **7**, 1193–1204.

Frank, S. A. 1989 The evolutionary dynamics of cytoplasmic male sterility. *Am. Nat.* **133**, 345–576.

Fu, Y.-X. & Li, W.-H. 1993 Statistical tests of neutrality of mutations. *Genetics* **133**, 693–709.

Fullerton, S. M., Harding, R. M., Boyce, A. J. & Clegg, J. B. 1994 Molecular and population genetic analysis of allelic sequence diversity at the human beta-globin locus. *Proc. Natl Acad. Sci. USA* **91**, 1805–1809.

Galtier, N., Depaulis, F. & Barton, N. H. 2000 Detecting bottlenecks and selective sweeps from DNA sequence polymorphism. *Genetics* **155**, 981–987.

Gaut, B. S. 1998 Molecular clocks and nucleotoide substitution rates in higher plants. *Evol. Biol.* **30**, 93–120.

Gaut, B. S., Morton, B. R., McCaig, B. C. & Clegg, M. T. 1996 Substitution rate comparisons between grasses and palms: synonymous rate differences at the nuclear gene *Adh* parellel rate differences at the plastid gene *rbcL*. *Proc. Natl Acad. Sci. USA* **93**, 10 274–10 279.

Gitzendanner, M. A. & Soltis, P. S. 2000 Patterns of variation in rare and widespread plant congeners. *Am. J. Bot.* **87**, 783–792.

Gouyon, P. H., Vichot, F. & Damme, J. v. 1991 Nuclear-cytoplasmic male sterility: single point equilibria versus limit cycles. *Am. Nat.* **137**, 498–514.

Graustein, A., Gaspar, J. M., Walters, J. R. & Palopoli, M. F. 2002 Levels of DNA polymorphism vary with mating system in the nematode genus *Caenorhabditis*. *Genetics* **161**, 99–107.

Hagenblad, J. & Nordborg, M. 2002 Sequence variation and haplotype structure surrounding the flowering time locus *FRI* in *Arabidopsis thaliana*. *Genetics* **161**, 289–298.

Hamrick, J. L. & Godt, M. J. 1990 Allozyme diversity in plant species. In *Plant population genetics, breeding, and genetic resources* (ed. A. H. D. Brown, M. T. Clegg, A. L. Kahler & B. S. Weir), pp. 43–63. Sunderland, MA: Sinauer.

Hamrick, J. L. & Godt, M. J. W. 1996 Effects of life history traits on genetic diversity in plant species. *Phil. Trans. R. Soc. Lond.* B **351**, 1291–1298.

Harr, B., Kauer, M. & Schlotterer, C. 2002 Hitchhiking mapping: a population-based fine-mapping strategy for adaptive mutations in *Drosophila melanogaster*. *Proc. Natl Acad. Sci. USA* **99**, 12 949–12 954.

Hedrick, P. W. 1980 Hitch-hiking: a comparison of linkage and partial selfing. *Genetics* **94**, 791–808.

Hedrick, P. W. 2002 Highly variable loci and their interpretation in evolution and conservation. *Evolution* **53**, 313–318.

Hedrick, P. W. & Holden, L. 1979 Hitch-hiking: an alternative to coadaptation for the barley and slender wild oat examples. *Heredity* **43**, 79–86.

Hinata, K., Watanabe, M., Yamakawa, S., Satta, Y. & Isogai, A. 1995 Evolutionary aspects of the *S*-related genes of the *Brassica* self-incompatibility system: synonymous and non-synonymous base substitutions. *Genetics* **140**, 1099–1104.

Hu, X. S. & Ennos, R. A. 1997 On estimation of the ratio of pollen to seed flow among plant populations. *Heredity* **79**, 541–552.

Hudson, R. R. 1990 Gene genealogies and the coalescent process. *Oxf. Surv. Evol. Biol.* 7, 1–45.

Hudson, R. R. & Kaplan, N. L. 1985 Statistical properties of the number of recombination events in the history of a sample of DNA sequences. *Genetics* **111**, 147–164.

Hudson, R. R., Kreitman, M. & Aguadé, M. 1987 A test of neutral molecular evolution based on nucleotide data. *Genetics* **116**, 153–159.

Hudson, R. R., Boos, D. D. & Kaplan, N. L. 1992 A statistical test for detecting geographic subdivision. *Mol. Biol. Evol.* **9**, 138–151.

Hudson, R. R., Saez, A. G. & Ayala, F. J. 1997 DNA variation at the *Sod* locus of *Drosophila melanogaster*: an unfolding story of natural selection. *Proc. Natl Acad. Sci. USA* **94**, 7725–7729.

Husband, B. C. & Barrett, S. C. H. 1993 Multiple origins of self-fertilization in tristylous *Eichhornia paniculata* (Pontederiaceae): inferences from style morph and isozyme variation. *J. Evol. Biol.* **6**, 591–608.

Ingvarsson, P. K. 2002 A metapopulation perspective on genetic diversity and differentiation in partially self-fertilizing plants. *Evolution* **56**, 2368–2373.

Ingvarsson, P. K. & Giles, B. E. 1999 Kin-structured colonization and small-scale genetic differentiation in *Silene dioica*. *Evolution* **53**, 605–611.

Ingvarsson, P. K. & Taylor, D. R. 2002 Genealogical evidence for epidemics of selfish genes. *Proc. Natl Acad. Sci. USA* **99**, 11 265–11 269.

Innan, H. & Tajima, F. 2002 A statistical test for the difference in amounts of DNA variation between two populations. *Genet. Res. Camb.* **80**, 15–25.

Jain, S. K. 1983 Breeding systems and the dynamics of plant populations. In *Genetics: new frontiers* (ed. V. L. Chopra, B. C. Joshi, R. P. Sharma & H. C. Bansal), pp. 291–316. New Delhi: Oxford and IBH Publishing Co.

James, T. Y., Porter, D., Hamrick, J. L. & Vilgalys, R. 1999 Evidence for limited intercontinental gene flow in the cosmopolitan mushroom *Schizophyllum commune*. *Evolution* **53**, 1665–1677.

Jarne, P. & Staedler, T. 1995 Population genetic structure and mating system in freshwater pulmonates. *Experientia* **51**, 482–497.

Kaltz, O. & Shykoff, J. A. 1998 Local adaptation in host–parasite systems. *Heredity* **81**, 361–370.

Kaplan, N. L., Hudson, R. R. & Langley, C. H. 1989 The 'hitch-hiking effect' revisited. *Genetics* **123**, 887–899.

Karron, J. D. 1987 A comparison of levels of genetic polymorphism and self-compatibility in geographically restricted and widespread plant congeners. *Evol. Ecol.* **1**, 47–58.

Kawabe, A., Yamane, K. & Miyashita, N. T. 2000 DNA polymorphism at the cytosolic phosphoglucose isomerase (*PgiC*) locus of the wild plant *Arabidopsis thaliana*. *Genetics* **156**, 1339–1347.

Kimura, M. 1983 *The neutral theory of molecular evolution*. Cambridge University Press.

Kimura, M. & Ohta, T. 1971 *Theoretical topics in population genetics*. Princeton University Press.

Koch, M., Weisshaar, B., Kroymann, J., Haubold, B. & Mitchell-Olds, T. 2001 Comparative genomics and regulatory evolution: conservation and function of the *Chs* and *Apetala3* promoters. *Mol. Biol. Evol.* **18**, 1882–1891.

Koch, R., van Luenen, H. G. A. M., Horst, M., Thijssen, K. L. & Plasterk, R. H. 2000 Single nucleotide polymorphisms in wild isolates of *Caenorhabditis elegans*. *Genome Res.* **10**, 1690–1696.

Kondo, K., Yamamoto, M., Itahashi, R., Sato, T., Egashira, H., Hattori, T. & Kowyama, Y. 2002 Insights into the evolution of self-compatibility in *Lycopersicon* from a study of stylar factors. *Plant J.* **30**, 143–153.

Kraft, T., Säll, T., Magnusson-Rading, I., Nilsson, N. O. & Halldén, C. 1998 Positive correlation between recombination rates and levels of genetic variation in natural populations of sea beet (*Beta vulgaris* subsp. *maritima*). *Genetics* **150**, 1239–1244.

Krauss, S. L. 2000 Accurate gene diversity estimates from amplified fragment length polymorphism (AFLP) markers. *Mol. Ecol.* **9**, 1241–1245.

Kreitman, M. 2001 Methods to detect natural selection in populations with applications to the human. *A. Rev. Genomics Hum. Genet.* **1**, 539–559.

Laporte, V. & Charlesworth, B. 2002 Effective population size and population subdivision in demographically structured populations. *Genetics* **162**, 501–519.

Layton, C. R. & Ganders, F. R. 1984 The genetic consequences of contrasting breeding systems in *Plectritis* (Valerianaceae). *Evolution* **38**, 1308–1325.

Le Corre, V., Roux, F. & Reboud, X. 2002 DNA polymorphism at the FRIGIDA gene in *Arabidopsis thaliana*: extensive nonsynonymous variation is consistent with local selection for flowering time. *Mol. Biol. Evol.* **19**, 1261–1271.

Levin, D. A. 1978 Genetic variation in annual Phlox: self-compatible vs. self-incompatible species. *Evolution* **32**, 245–263.

Lewis, D. & Crowe, L. H. 1958 Unilateral incompatibility in flowering plants. *Heredity* **12**, 233–256.

Lewontin, R. C. 1974 *The genetic basis of evolutionary change*. New York: Columbia University Press.

Lewontin, R. C. & Krakauer, J. 1973 Distribution of gene frequencies as a test of the theory of the selective neutrality of polymorphisms. *Genetics* **74**, 175–195.

Lin, J.-Z., Morrell, P. L. & Clegg, M. T. 2002 The influence of linkage and inbreeding on patterns of nucleotide sequence diversity at duplicate alcohol dehydrogenase loci in wild barley (*Hordeum vulgare* ssp. *spontaneum*). *Genetics* **162**, 2007–2015.

Linhart, Y. B. & Grant, M. C. 1996 Evolutionary significance of local genetic differentiation in plants. *A. Rev. Ecol. Syst.* **27**, 237–277.

Linhart, J. & Premoli, A. C. 1994 Genetic variation in central and disjunct populations of *Lilium parryi*. *Can. J. Bot.* **72**, 79–85.

Liu, F. 1998 Genetic diversity in *Leavenworthia* populations with different inbreeding levels. The effect of breeding system on the level and pattern of molecular variation in plant populations. PhD thesis, University of Chicago, IL, USA.

Liu, F., Zhang, L. & Charlesworth, D. 1998 Genetic diversity in *Leavenworthia* populations with different inbreeding levels. *Proc. R. Soc. Lond.* B **265**, 293–301. (DOI 10.1098/rspb.1998.0295.)

Lloyd, D. G. 1965 Evolution of self-compatibility and racial differentiation in *Leavenworthia* (Cruciferae). *Contrib. Gray Herbarium Harv. Univ.* **195**, 3–134.

Loos, B. P. 1993 Allozyme variation within and between populations in *Lolium* (Poaceae). *Plant Syst. Evol.* **188**, 101–113.

Luikart, G., Allendorf, F. W. & Cornuet, J. M. 1998*a* Empirical evaluation of a test for identifying recently bottlenecked populations from allele frequency data. *Conserv. Biol.* **12**, 228–237.

Luikart, G., Allendorf, F. W., Cornuet, J. M. & Sherwin, W. B. 1998*b* Distortion of allele frequency distributions provides a test for recent population bottlenecks. *J. Heredity* **89**, 238–247.

Lynch, M., O'Hely, M., Walsh, B. & Force, A. 2001 The probability of preservation of a newly arisen gene duplicate. *Genetics* **159**, 1789–1804.

McCauley, D. E. 1994 Contrasting the distribution of chloroplast DNA and allozyme polymorphism among local populations of *Silene alba*: implications for studies of gene flow in plants. *Proc. Natl Acad. Sci. USA* **91**, 8127–8131.

McCauley, D. E., Stevens, J. W., Peroni, P. A. & Raveill, J. A. 1996 The spatial distribution of chloroplast DNA and allozyme polymorphisms within a population of *Silene alba* (Caryophyllaceae). *Am. J. Bot.* **83**, 727–731.

Macnair, M. R., Macnair, V. E. & Martin, B. E. 1989 Adaptive speciation in *Mimulus*: an ecological comparison of *M. cupriphilus* with its presumed progenitor *M. guttatus*. *New Phytol.* **112**, 269–279.

Marais, G., Mouchiroud, D. & Duret, L. 2001 Does recombination improve selection on codon usage? Lessons from nematode and fly complete genomes. *Proc. Natl Acad. Sci. USA* **156**, 1661–1669.

Martin, F. W. 1967 The genetic control of unilateral incompatibility between two tomato species. *Genetics* **56**, 391–398.

Maruyama, T. & Birky, C. W. 1991 Effects of periodic selection on gene diversity in organelle genomes and other systems without recombination. *Genetics* **127**, 449–451.

Maynard Smith, J. & Haigh, J. 1974 The hitchhiking effect of a favorable gene. *Genet. Res. Camb.* **219**, 1114–1116.

Miller, J. C. & Tanksley, S. D. 1990 RFLP analysis of phylogenetic relationships and genetic variation in the genus *Lycopersicon*. *Theor. Appl. Genet.* **80**, 437–448.

Miyashita, N. T., Aguadé, M. & Langley, C. H. 1993 Linkage disequilibrium in the white locus region of *Drosophila melanogaster*. *Genet. Res.* **62**, 101–109.

Mougel, C., Thioulouse, J., Perriere, G. & Nesme, X. 2002 A mathematical method for determining genome divergence and species delineation using AFLP. *Int. J. Syst. Evol. Microbiol.* **52**, 573–586.

Nagylaki, T. 1998 Fixation indices in subdivided populations. *Genetics* **148**, 1325–1332.

Nagylaki, T. 2000 Geographical invariance and the strong-migration limit in subdivided populations. *J. Math. Biol.* **41**, 123–142.

Nasrallah, M. E. & Nasrallah, J. B. 1989 The molecular genetics of self-incompatibility in *Brassica*. *A. Rev. Genet.* **23**, 121–139.

Navarro, A. & Barton, N. H. 2002 The effects of multilocus balancing selection on neutral variability. *Genetics* **161**, 849–863.

Nei, M. 1987 *Molecular evolutionary genetics*. New York: Columbia University Press.

Nei, M., Maruyama, T. & Chakraborty, R. 1975 The bottleneck effect and genetic variability in populations. *Evolution* **29**, 1–10.

Nevo, E., Beiles, A. & Krugman, T. 1988 Natural selection of allozyme polymorphisms: a microgeographic climatic differentiation in wild emmer wheat (*Triticum dicoccoides*). *Theor. Appl. Genet.* **75**, 529–538.

Newton, A. C., Allnutt, T. R., Gillies, A. C. M., Lowe, A. J. & Ennos, R. A. 1999 Molecular phylogeography, intraspecific variation and the conservation of tree species. *Trends Ecol. Evol.* **14**, 140–145.

Nishio, T., Kusaba, M., Sakamoto, K. & Ockendon, D. 1997 Polymorphism of the kinase domain of the *S*-locus receptor kinase gene (*SRK*) in *Brassica oleracea* L. *Theor. Appl. Genet.* **95**, 335–342.

Nordborg, M. 1997 Structured coalescent processes on different time scales. *Genetics* **146**, 1501–1514.

Nordborg, M. 2000 Linkage disequilibrium, gene trees and selfing: an ancestral recombination graph with partial self-fertilization. *Genetics* **154**, 923–929.

Nordborg, M. & Donnelly, P. 1997 The coalescent process with selfing. *Genetics* **146**, 1185–1195.

Nordborg, M. & Innan, H. 2002 Molecular population genetics. *Curr. Opin. Plant Biol.* **15**, 69–73.

Nordborg, M. & Innan, H. 2003 The genealogy of sequences containing multiple sites subject to strong selection in a subdivided population. *Genetics* **163**, 1201–1213.

Nordborg, M., Charlesworth, B. & Charlesworth, D. 1996 Increased levels of polymorphism surrounding selectively maintained sites in highly selfing species. *Proc. R. Soc. Lond.* B **163**, 1033–1039.

Nordborg, M. (and 11 others) 2002 The extent of linkage disequilibrium in *Arabidopsis thaliana*. *Nature Genet.* **30**, 190–193.

Olmstead, R. G. 1989 Phylogeny, phenotypic evolution, and biogeography of the *Scutellaria angustifolia* complex (Lamiaceae): inference from morphology and molecular data. *Syst. Bot.* **14**, 320–338.

Ouborg, N. J., Piquot, Y. & Groenendael, J. M. V. 1999 Population genetics, molecular markers and the study of dispersal in plants. *J. Ecol.* **87**, 551–568.

Pannell, J. R. & Charlesworth, B. 1999 Neutral genetic diversity in a metapopulation with recurrent local extinction and recolonization. *Evolution* **53**, 664–676.

Pannell, J. R. & Charlesworth, B. 2000 Effects of metapopulation processes on measures of genetic diversity. *Phil. Trans. R. Soc. Lond.* B **355**, 1851–1864. (DOI 10.1098/rstb.2000.0740.)

Parker, M. A. 1994 Pathogens and sex in plants. *Evol. Ecol.* **8**, 560–584.

Pollak, E. 1987 On the theory of partially inbreeding finite populations. I. Partial selfing. *Genetics* **117**, 353–360.

Pritchard, J. K. & Przeworski, M. 2001 Linkage disequilibrium in humans: models and data. *Am. J. Hum. Genet.* **69**, 1–14.

Przeworski, M. & Wall, J. D. 2001 Why is there so little intragenic linkage disequilibrium in humans? *Genet. Res. Camb.* **77**, 143–151.

Purugganan, M. D., Boyles, A. L. & Suddith, J. I. 2000 Variation and selection at the CAULIFLOWER floral homeotic gene accompanying the evolution of domesticated *Brassica oleracea*. *Genetics* **155**, 855–862.

Raper, J. R., Krongelb, G. S. & Baxter, M. G. 1958 The number and distribution of incompatibility factors in *Schizophyllum*. *Am. Nat.* **92**, 221–232.

Richman, A. D., Uyenoyama, M. K. & Kohn, J. R. 1996*a* Allelic diversity and gene genealogy at the self-incompatibility locus in the Solanaceae. *Science* **273**, 1212–1216.

Richman, A. D., Uyenoyama, M. K. & Kohn, J. R. 1996*b* S-allele diversity in a natural population of ground cherry *Physalis crassifolia* assessed by RT-PCR. *Heredity* **76**, 497–505.

Rick, C. M., Fobes, J. F. & Holle, M. 1977 Genetic variation in *Lycopersicon pimpinellifolium*: evidence of evolutionary change in mating systems. *Plant Syst. Evol.* **127**, 139–170.

Rollins, R. C. 1963 The evolution and systematics of *Leavenworthia* (Cruciferae). *Contr. Gray Herb. Harv.* **192**, 3–98.

Savolainen, O., Langley, C. H., Lazzaro, B. P. & Freville, H. 2000 Contrasting patterns of nucleotide polymorphism at the alcohol dehydrogenase locus in the outcrossing *Arabidopsis lyrata* and the selfing *Arabidopsis thaliana*. *Mol. Biol. Evol.* **17**, 645–655.

Schierup, M. H., Vekemans, X. & Charlesworth, D. 2000*a* The effect of hitchhiking on genes linked to a balanced polymorphism in a subdivided population. *Genet. Res. Camb.* **76**, 63–73.

Schierup, M. H., Vekemans, X. & Charlesworth, D. 2000*b* The effect of subdivision on variation at multi-allelic loci under balancing selection. *Genet. Res. Camb.* **76**, 51–62.

Schierup, M. H., Mable, B. K., Awadalla, P. & Charlesworth, D. 2001 Identification and characterization of a polymorphic

receptor kinase gene linked to the self-incompatibility locus of *Arabidopsis lyrata*. *Genetics* **158**, 387–399.
Schoen, D. J. 1982 Genetic variation and the breeding system of *Gilia achilleifolia*. *Evolution* **36**, 361–370.
Schoen, D. J. & Brown, A. H. D. 1991 Intraspecific variation in population gene diversity and effective population size correlates with the mating system in plants. *Proc. Natl Acad. Sci. USA* **88**, 4494–4497.
Schoen, D. J., L'Heureux, A.-M., Marsolais, J. & Johnston, M. O. 1997 Evolutionary history of the mating system in *Amsinckia* (Boraginaceae). *Evolution* **51**, 1090–1099.
Sharbel, T. F., Haubold, B. & Mitchell-Olds, T. 2000 Genetic isolation by distance in *Arabidopsis thaliana*: biogeography and postglacial colonisation of Europe. *Mol. Ecol.* **9**, 2109–2118.
Shiba, H. (and 10 others) 2002 The dominance of alleles controlling self-incompatibility in *Brassica* pollen is regulated at the RNA level. *Plant Cell* **14**, 491–504.
Slatkin, M. 1977 Gene flow and genetic drift in a species subject to local extinctions. *Theor. Pop. Biol.* **12**, 253–262.
Slatkin, M. 1987 The average number of sites separating DNA sequences drawn from a subdivided population. *Theor. Pop. Biol.* **32**, 42–49.
Slatkin, M. 1994 Linkage disequilibrium in growing and stable populations. *Genetics* **137**, 331–336.
Slatkin, M. & Wiehe, T. 1998 Genetic hitchhiking in a subdivided population. *Genet. Res. Camb.* **71**, 155–160.
Städler, T. & Delph, L. F. 2002 Ancient mitochondrial haplotypes and evidence for intragenic recombination in a gynodioecious plant. *Proc. Natl Acad. Sci. USA* **99**, 11 730–11 735.
Stebbins, G. L. 1957 Self fertilization and population variation in the higher plants. *Am. Nat.* **91**, 337–354.
Steinbachs, J. E. & Holsinger, K. E. 2002 S-Rnase-mediated gametophytic self-incompatibility is ancestral in eudicots. *Mol. Biol. Evol.* **19**, 825–829.
Stephan, W. & Langley, C. H. 1998 DNA polymorphism in *Lycopersicon* and crossing-over per physical length. *Genetics* **150**, 1585–1593.
Strobeck, C. 1972 Heterozygosity in pin–thrum plants or with partial sex linkage. *Genetics* **72**, 667–678.
Strobeck, C. 1987 Average number of nucleotide differences in a sample from a single population: a test for population subdivision. *Genetics* **117**, 149–153.
Sun, M. 1995 Effects of population size, mating system, and evolutionary origin on genetic diversity in *Spiranthes sinensis* and *S. hongkingensis*. *Conserv. Biol.* **10**, 785–795.
Sun, M. & Wong, K. C. 2001 Genetic structure of three orchid species with contrasting breeding systems using RAPD and allozyme markers. *Am. J. Bot.* **88**, 2180–2188.
Tajima, F. 1989*a* The effect of change in population size on DNA polymorphism. *Genetics* **123**, 597–601.
Tajima, F. 1989*b* Statistical method for testing the neutral mutation hypothesis. *Genetics* **123**, 585–595.
Tajima, F. 1993 Measurement of DNA polymorphism. In *Mechanisms of molecular evolution* (ed. N. Takahata & A. G. Clark), pp. 37–59. Sunderland, MA: Sinauer.
Takahata, N. 1990 A simple genealogical structure of strongly balanced allelic lines and trans-species polymorphism. *Proc. Natl Acad. Sci. USA* **87**, 2419–2423.
Takahata, N. & Satta, Y. 1998 Footprints of intragenic recombination at *HLA* loci. *Immunogenetics* **47**, 430–441.
Takebayashi, N. & Morrell, P. P. 2001 Is self-fertilization an evolutionary dead end? Revisiting an old hypothesis with genetic theories and a macroevolutionary approach. *Am. J. Bot.* **88**, 1143–1150.
Tarayre, M., Saumitou-Laprade, P., Cuguen, J., Couvet, D. & Thompson, J. D. 1997 The spatial structure of cytoplasmic (cpDNA) and nuclear (allozyme) markers within and among populations of the gynodioecious *Thymus vulgaris* (Labiatae) in southern France. *Am. J. Bot.* **84**, 1675–1684.
Tenaillon, M. I., Sawkins, M. C., Long, A. D., Gaut, R. L., Doebley, J. F. & Gaut, B. S. 2001 Patterns of DNA sequence polymorphism along chromosome 1 of maize (*Zea mays* ssp. *mays* L.). *Proc. Natl Acad. Sci. USA* **98**, 9161–9166.
Terachi, T., Yamaguchi, K. & Yamagishi, H. 2001 Sequence analysis on the mitochondrial orfB locus in normal and Ogura male-sterile cytoplasms from wild and cultivated radishes. *Curr. Genet.* **40**, 276–281.
Tian, D., Araki, H., Stahl, E. A., Bergelson, J. & Kreitman, M. 2002 Signature of balancing selection in *Arabidopsis*. *Proc. Natl Acad. Sci. USA* **99**, 11 525–11 530.
Tiffin, P. & Hahn, M. W. 2002 Coding sequence divergence between two closely related plant species: *Arabidopsis thaliana* and *Brassica rapa* ssp. *pekinensis*. *J. Mol. Evol.* **54**, 746–753.
Tishkoff, S. A. (and 14 others) 1996 Global patterns of linkage disequilibrium at the CD4 locus and modern human origins. *Science* **271**, 1380–1387.
Todokoro, S., Terauchi, R. & Kawano, S. 1995 Microsatellite polymorphisms in natural populations of *Arabidopsis thaliana* in Japan. *Jpn. J. Genet.* **70**, 543–554.
Uyenoyama, M. K. 1995 A generalized least-squares estimate for the origin of self-incompatibility. *Genetics* **139**, 975–992.
van Treuren, R., Bijlsma, R., Van Delden, W. & Ouborg, N. J. 1991 The significance of genetic erosion in the process of extinction. 1. Genetic differentiation in *Salvia pratensis* and *Scabiosa columbaria* in relation to population-size. *Heredity* **66**, 181–189.
van Treuren, R., Kuittinen, H., Karkkainen, K., Baena-Gonzalez, E. & Savolainen, O. 1997 Evolution of microsatellites in *Arabis petraea* and *Arabis lyrata*, outcrossing relatives of *Arabidopsis thaliana*. *Mol. Biol. Evol.* **14**, 220–229.
Vekemans, X. & Lefebvre, C. 2001 On the evolution of heavy-metal tolerant populations in *Armeria maritima*: evidence from allozyme variation and reproductive barriers. *J. Evol. Biol.* **10**, 175–191.
Vekemans, X. & Slatkin, M. 1994 Gene and allelic genealogies at a gametophytic self-incompatibility locus. *Genetics* **137**, 1157–1165.
Volis, S., Mendlinger, S., Turuspekov, Y. & Esnazarov, U. 2002 Phenotypic and allozyme variation in Mediterranean and desert populations of wild barley, *Hordeum spontaneum* Koch. *Evolution* **56**, 1403–1415.
Volkman, S. K., Barry, A. E., Lyons, E. J., Nielsen, K. M., Thomas, S. M., Choi, M., Thakore, S. S., Day, K. P., Wirth, D. F. & Hartl, D. L. 2001 Recent origin of *Plasmodium falciparum* from a single progenitor. *Science* **293**, 482–484.
Wakeley, J. 1996 The variance of pairwise nucleotide differences in two populations with migration. *Theor. Pop. Biol.* **49**, 39–57.
Wakeley, J. 2000 The effects of subdivision on the genetic divergence of populations and species. *Evolution* **54**, 1092–1101.
Wakeley, J. & Aliacar, N. 2001 Gene genealogies in a metapopulation. *Genetics* **159**, 893–905.
Wall, J. D. 1999 Recombination and the power of statistical tests of neutrality. *Genet. Res. Camb.* **74**, 65–79.
Wang, R.-L., Stec, A., Hey, J., Doebley, J. F. & Lukens, L. 1999 The limits of selection during maize domestication. *Nature* **398**, 236–239.
Whitlock, M. C. & Barton, N. H. 1997 The effective size of a subdivided population. *Genetics* **146**, 427–441.
Whitlock, M. C. & McCauley, D. E. 1993 Some population genetic consequences of colony formation and extinction—

genetic correlations within founding groups. *Evolution* **44**, 1717–1724.
Whitlock, M. C. & McCauley, D. E. 1999 Indirect measures of gene flow and migration: F_{ST} not equal to $1/(4Nm + 1)$. *Heredity* **82**, 117–125.
Whitt, S. R., Wilson, L. M., Tenaillon, M. I., Gaut, B. S. & Buckler, E. S. 2002 Genetic diversity and selection in the maize starch pathway. *Proc. Natl Acad. Sci. USA* **99**, 12 959–12 962.
Wolfe, K. H., Li, W.-H. & Sharp, P. M. 1987 Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. *Proc. Natl Acad. Sci. USA* **84**, 9054–9058.
Wolff, K. 1991 Analysis of allozyme variability in three *Plantago* species and a comparison to morphological variability. *Theor. Appl. Genet.* **81**, 119–126.
Wright, S. 1939 The distribution of self-sterility alleles in populations. *Genetics* **24**, 538–552.
Wright, S. I., Lauga, B. & Charlesworth, D. 2002 Rates and patterns of molecular evolution in inbred and outbred *Arabidopsis*. *Mol. Biol. Evol.* **19**, 1407–1420.
Wright, S. I., Lauga, B. & Charlesworth, D. 2003 Subdivision and haplotype structure in natural populations of *Arabidopsis lyrata*. *Mol. Ecol.* (In the press.)
Wyatt, R., Odrzykoski, I. J. & Stoneburner, A. 1989 High levels of genetic variability in the haploid moss *Plagiomnium ciliare*. *Evolution* **43**, 1085–1096.
Wyatt, R., Evans, E. A. & Sorenson, J. C. 1992 The evolution of self-pollination in granite outcrop species of *Arenaria* (Caryophyllaceae). VI. Electrophoretically detectable genetic variation. *Syst. Bot.* **17**, 201–209.

Discussion

T. R. Meagher (*School of Biology, University of St Andrews, St Andrews, UK*). Inbreeding has an evolutionary origin within a lineage. Does it seem to you that a phylogenetic perspective might be useful when deciding when inbreeding originated? For example, in your *Leavenworthia* examples of recent origin of inbreeding, there may not have been sufficient time for the decay of genetic variance.

D. Charlesworth. I do not think genetic variance would decay quickly enough in inbred lines; rather, selective sweeps over a shorter time-frame are more likely to account for loss of genetic variation. In the case of a genus like *Leavenworthia* one might apply Tajima's test to test for selective sweeps. Our sample sizes to date are not sufficient to do this.

GLOSSARY

AFLP: amplified fragment length polymorphism
CMS: cytoplasmic male sterility
RAPD: random amplification of polymorphic DNA
RFLP: restriction fragment length polymorphism
SI: self-incompatibility