

- 45 Harrison, R.G. (1990) Hybrid zones: windows on evolutionary process. *Oxf. Surv. Evol. Biol.* 7, 69–128
- 46 Rieseberg, L.H. *et al.* (1999) Hybrid zones and the genetic architecture of a barrier to gene flow between two wild sunflower species. *Genetics* 152, 713–727
- 47 Ting, C.T. *et al.* (2000) The phylogeny of closely related species as revealed by the genealogy of a speciation gene, *Odysseus*. *Proc. Natl. Acad. Sci. U. S. A.* 97, 5313–5316
- 48 Rieseberg, L.H. *et al.* (1995) Chromosomal and genic barriers to introgression in *Helianthus*. *Genetics* 141, 1163–1171
- 49 Greenbaum, I.F. and Reed, M.J. (1984) Evidence for heterosynaptic pairing of the inverted segment in pericentric inversion heterozygotes of the deer mouse (*Peromyscus maniculatus*). *Cytogenet. Cell Genet.* 38, 106–111
- 50 Hale, D.W. (1986) Heterosynapsis and suppression of chiasmata within heterozygous pericentric inversions of the Sitka deer mouse. *Chromosoma* 94, 425–432
- 51 Davisson, M.T. and Akeson, E.C. (1993) Recombination suppression by heterozygous Robertsonian chromosomes in the mouse. *Genetics* 133, 649–667
- 52 Felsenstein, J. (1981) Skepticism towards Santa Rosalia, or why are there so few kinds of animals. *Evolution* 35, 124–138
- 53 Trickett, A.J. and Butlin, R.K. (1994) Recombination suppressors and the evolution of new species. *Heredity* 73, 339–345
- 54 Haufler, H.C. and Searle, J.B. (1993) Extreme karyotypic variation in a *Mus musculus domesticus* hybrid zone – the tobacco mouse story revisited. *Evolution* 47, 1374–1395
- 55 White, M.J.D. (1973) *Animal Cytology and Evolution*, Cambridge University Press
- 56 Charlesworth, D. and Charlesworth, B. (1980) Sex differences in fitness and selection for centric fusions between sex chromosomes and autosomes. *Genet. Res.* 35, 205–214
- 57 Levin, D.A. (2000) *The Origin, Expansion, and Demise of Plant Species*, Oxford University Press
- 58 Rhymer, J.M. and Simberloff, D. (1996) Extinction by hybridization and introgression. *Annu. Rev. Ecol. Syst.* 27, 83–109
- 59 Rieseberg, L.H. (2000) Crossing relationships among ancient and experimental sunflower hybrid lineages. *Evolution* 54, 859–865
- 60 Andolfatto, P.A. (2001) Inversion polymorphisms and nucleotide variability in *Drosophila*. *Genet. Res.* 77, 1–8

# Gene trees and species trees are not the same

Richard Nichols

The relationship between species is usually represented as a bifurcating tree with the branching points representing speciation events. The ancestry of genes taken from these species can also be represented as a tree, with the branching points representing ancestral genes. The time back to the branching points, and even the branching order, can be different between the two trees. This possibility is widely recognized, but the discrepancies are often thought to be small. A different picture is emerging from new empirical evidence, particularly that based on multiple loci or on surveys with a wide geographical scope. The discrepancies must be taken into account when estimating the timing of speciation events, especially the more recent branches. On the positive side, the different timings at different loci provide information about the ancestral populations.

Molecular PHYLOGENETICS (see Glossary) is based on the principle that the number of substitutions that have accumulated between the DNA sequences of two species indicates the time since their common ancestor. There is a fundamental problem with this approach, widely acknowledged in the standard texts (e.g. Ref. 1): the time back to the common ancestor of the two DNA sequences is typically longer than the time back to the common ancestor of the two species. The difference between the two times is shown as  $T_1$  in Box 1. The molecules are most unlikely to have a common ancestor living at the very moment that the ancestral species split in two. Rather, the period  $T_1$  is the time back to the common ancestor of the two molecules within the single ancestral species.

The timing of the SPECIATION events estimated from molecular phylogenies must be corrected for this bias, corresponding to the average value of  $T_1$ , and for the

uncertainty owing to the variation around this average. These issues are important if the timing of speciation events is used to draw conclusions about the nature of the speciation process (e.g. Barraclough and Nee<sup>2</sup>, this issue). Although it is inconvenient for some applications, the variability in timings can also be informative. Differences between loci can be used to draw inferences about the past population size and population subdivision. This approach will become more important as comparisons between species are more routinely made at multiple loci. The results could provide clues about the demography of populations that have undergone speciation.

Under the assumption that the ancestral species had a population size similar to the current species, it is possible to make a crude correction for the bias in timings that result from  $T_1$ . This makes use of the similarity between times to a common ancestor for the genes at a locus within a species and  $T_1$  (Box 2). This and more sophisticated methods that deal with information from multiple loci are reviewed and developed by Edwards and Beerli<sup>3</sup>.

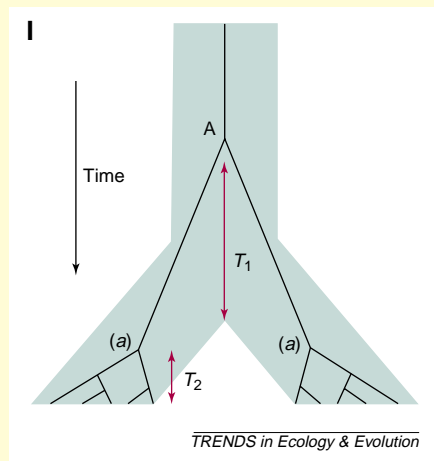
Figure I in Box 1 provides a simplified view of speciation. There is a single point at which the inverted 'Y' splits, implicitly indicating that the ancestral species divided instantaneously into two descendant species between which there was no gene flow. Many of the modes of speciation sketched by Turelli *et al.*<sup>4</sup> (this issue) would involve a more protracted interruption of gene flow. Populations diverging in ALLOPATRY could sporadically come into contact, the accumulation of REPRODUCTIVE ISOLATION in SYMPATRY or PARAPATRY might

Richard Nichols  
School of Biological  
Sciences, Queen Mary,  
University of London,  
London, UK E1 4NS.  
e-mail:  
r.a.nichols@qmw.ac.uk

### Box 1. Gene trees within species trees

The inverted Y (Fig. I) represents the splitting of one species into two. The single upper branch is the ancestral species; the two lower arms are the descendant species, which are alive today (at the base).

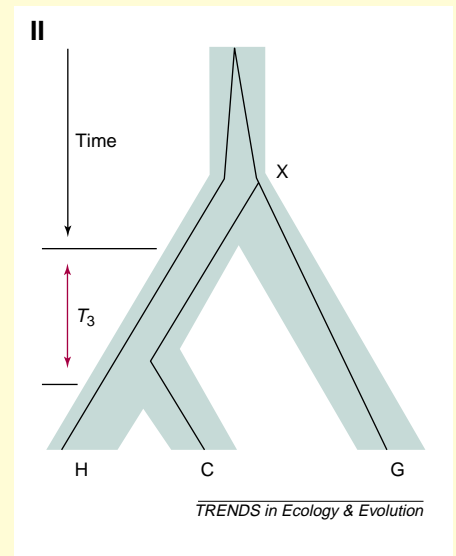
Within each of the extant species, the five lower tips of dark lines represent five genes. The ancestry of these genes can be



traced back in time to common ancestors (branch points). In this case, the common ancestor (a) for the genes in each species occurred more recently than did the speciation event (the branching of the inverted Y). The time back to these common ancestors will vary between loci and species. In this case, it is  $T_2$  generations in the left-hand species. The two a genes can be traced back to two genes in the ancestral species. Even further back, they share a common ancestor (A)  $T_1$  generations before the split between the two species.

Figure II shows an example in which the gene tree and species tree do not match. There are three species (H, C and G); although H and C share the most recent ancestry, the history of the genes is different and the genes of C and G share the more recent ancestor (X). This occurred because the lineages from H and C remained separate for the period  $T_3$ . Consequently, all three species have distinct ancestral genes in the top ancestral species. It happens that the

C and G lineages have the most recent ancestor, so the branching order of the genes differs from that of the species (they could, of course, have had the correct branching pattern by chance).



be incomplete for a period, and even taxa classified as distinct species can hybridize. This type of episodic gene flow might affect some loci and not others – if so, the differences between loci could provide information about these episodes.

Even in the case of instantaneous species splitting (Box 1), there will be differences between the gene trees at different loci because of the variation in the length of  $T_1$ . Are differences large enough to be confused with the signal from more protracted speciation? Is it unreliable to estimate the timing of the species split from one or two gene trees, such as that for mitochondrial DNA (mtDNA)? The answer to these questions is 'that depends'.

Here, I outline how the magnitude of the discrepancy depends on the effective population size ( $N_e$ ) of the species being studied. This sharpens the question. We need first to ask what is meant by the effective sizes, and second, whether they are generally large enough to affect phylogenetic inference. The second question has been addressed empirically, in a few well-studied species, by comparing results from many loci, and it gives some surprising insights into the demographic history of humans. I also consider the other aspects of phylogenetic history that might distort the estimates of timing from molecular data.

#### Timings and $N_e$

$T_1$  can be calculated for an idealized diploid random-mating population of size  $N_e$ . There are  $2N_e$  genes in

the population: the maternal and paternal copy in each of  $N_e$  individuals. I use the term 'gene' for these copies where some authors use 'allele' to make it clear that they are referring to copies at the same locus (whereas, in some branches of genetics, the 'number of alleles' refers to the number of variants found at a locus).

To calculate  $T_1$ , imagine starting at the time of the speciation event and tracing back the ancestry of the LINEAGE of a gene from each of the two descendant species. Because each lineage is equally likely to be descended from the  $2N_e$  genes in the previous generation, there is a probability of  $1/2N_e$  of arriving at the common ancestor (A in Box 1) in each generation. This is known as the coalescence rate from which the expected timings can be calculated<sup>5</sup> (Box 2). The average time (generations per coalescent event) is the inverse of the rate (coalescent events per generation),  $2N_e$  generations. If, at the time of speciation, there were more than two ancestors for the genes in the present two species, then the average time would increase towards  $4N_e$ , but the upper end of the distributions is quite similar. The discrepancy between the times for genes and species is unlikely to be much more than  $7N_e$  generations (Box 2). These calculations assume that the species was a single random-mating population. In reality, the situation will be more complicated (e.g. the population might be subdivided, Box 3), but the average times estimated from genetic data are often characterized by the corresponding  $N_e$  for an idealized single population.

### Box 2. The variation in the time to a common ancestor for different loci

The time back to a common ancestor will vary from locus to locus in the same species. It depends on the effective size of the population and the selection acting on the locus or closely linked loci. There is also a large contribution from chance leading to differences among loci (Fig. 1). In an idealized population, there are

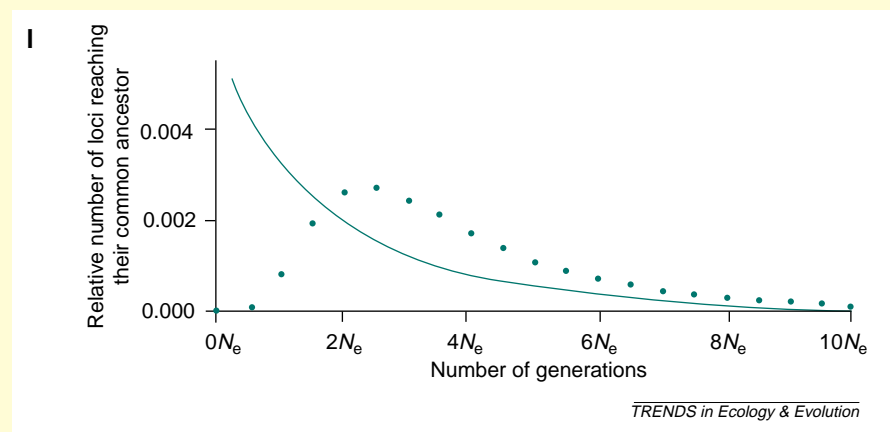
relatively simple distributions for the times back to a common ancestor of different loci where selection is weak<sup>a</sup>.

$T_1$  is the time back to an ancestor of a pair of genes within the same species (Fig. 1, solid curve). The average for an autosomal gene is  $2N_e$  generations, but times two- or threefold greater will be seen at many loci.

Time  $T_2$  is the time back to the common ancestor for a sample of genes from within a species (Fig. 1, dotted curve). Compared with the solid curve, only a few loci have a very recent common ancestor. This is because  $T_2$  is the time to the common ancestors for all of the genes in the sample. Although some pairs of genes in the sample will have a recent common ancestor, others will not. As the size of the sample increases, the distribution rapidly converges to the distribution shown in Fig. 1. The average time is twice that for  $T_1$  ( $4N_e$  generations), but the upper ranges are similar. These distributions are calculated by assuming a probability of  $1/2N_e$  that a pair of genes share a common ancestor in each preceding generation.

#### Reference

- <sup>a</sup> Hudson, R.R. (1990) Gene genealogies and the coalescent process. *Oxford Surv. Evol. Biol.* 7, 1–44



Is  $T_1$  large compared with the time between speciation events? Some of the evidence suggests not. Moore<sup>6,7</sup> surveyed the published within-species differences at the mitochondrial cytochrome *b* gene from birds. MtDNA is expected to provide particularly accurate timings because it has essentially one quarter of the effective population size of an autosomal locus. Effective population size is halved because mitochondria are only transmitted through females and halved again because each female transmits only one haplotype. Moore looked for evidence of small effective size in intraspecific DNA sequence comparisons. The largest difference observed within a species will give some indication of the maximum error of the between-species values (compare the two curves in Box 2). Although one intraspecific comparison showed 2.5% sequence divergence, most species had substantially smaller (maximum) divergence. Interspecies divergence tends to be larger.

Klicka and Zink<sup>8,9</sup> found divergence of around 5% in comparisons between North American songbirds, which they interpret to suggest that separation dated back to the Early Pleistocene–Late Pliocene. The substantially larger genetic differences between species than within seem to imply that there was little bias in the estimates of the timing of separation. However, Avise and Walker<sup>10</sup> explain these large differences as divergence that accumulated between relatively isolated regions in the ancestral species. They demonstrate that comparable divergence can be seen in present day species. In their survey, 76% of bird species contained lineages that are so distinct

that they appear to have been separate since the Pleistocene. This degree of divergence is detected in a wide range of species when the survey has sufficient geographical scope<sup>11</sup>. If reproductive isolation accumulated gradually between these allopatric populations, there will be no single speciation event that can be given a precise date. The phylogeny might, however, be revealing about the period through which the isolation persisted, in particular, whether isolation continued through the major climatic fluctuations.

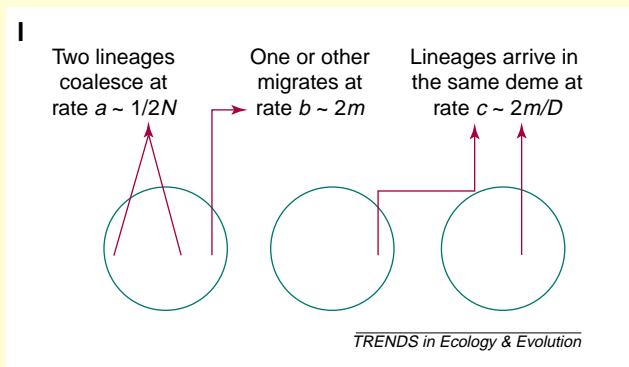
#### Comparisons at multiple loci

Given this rather equivocal evidence from studies of several species at a single locus, what more can be learnt from multiple loci? One approach is to assess the proportion of loci that show the same branching patterns. Branching patterns can differ because of the different timings of species trees and gene trees. Box 1 examines the chance of the genes and species sharing the same branching pattern (using three species). The chance of congruent branching patterns depends on the length of the internode shown as  $T_3$  (Refs 12, 13) (Box 1). If the internode is long enough relative to  $N_e$ , the genes for the two closest species will have the most recent common ancestor. For example, if  $T_3$  is longer than  $\sim 5N_e$ , then there is a 95% chance that the gene and species trees are congruent. The exact proportions also depend on the probability that appropriate mutations will have occurred so that the actual branching pattern is detected.

Chen and Li<sup>14</sup> used this reasoning in reverse<sup>15</sup> to work out the effective population size of the common ancestor of humans and chimps. They looked at

### Box 3. Effective population size in structured populations

One of the simplest models used to characterize effects of population subdivision is the Finite Island Model. Figure 1 illustrates the three rates that characterize its GENEALOGICAL (see Glossary) structure. The circles represent the panmictic demes that make up the population. The arrows show the different outcomes that occur when lineages are traced back through time. The population is composed of  $D$  demes, each of  $N$  diploid individuals. Each individual migrates with probability  $m$  per generation. The destination is equally likely to be any one of the other demes. For simplicity the expressions for rates have excluded multiple events in one generation. (Modified, with permission, from Ref. a.)



Genes in the same deme tend to have a recent common ancestor, except for those descended from recent migrants. To find the common ancestor of two genes from different demes, their lineages must be traced back to a time when their ancestors lived in the same deme. Therefore, they tend to have a more ancient common ancestor. It follows that the time to a common ancestor for the whole species depends largely on the rate at which genes in different demes coalesce. This can be approximated by the rate at which they arrive in the same deme ( $c$ ), multiplied by the probability that they then coalesce before migrating out again [which is given by the relative size of the two rates:  $a/(a + b)$ ].

In an idealized single population, the rate of coalescence is  $1/2N_e$ . By equating this to the long-term rate above, an expression for  $N_e$  in a subdivided population is obtained (Eqn 1):

$$\frac{1}{2N_e} = c \frac{a}{(a + b)},$$

$$\Rightarrow N_e = ND \left( 1 + \frac{1}{4Nm} \right) \quad [1]$$

#### Reference

a Nichols, R.A. *et al.* (2001) Sustaining genetic variation in a small population: evidence from the Mauritius kestrel. *Mol. Ecol.* 10, 593–602

53 autosomal human sequences, each of a non-coding region 2–20-kb long and 5 kb away from any suspected gene. This strategy was designed to minimize the effect of selection on linked loci on the patterns of coalescence. They did not include repeated elements that might have had unusual intragenomic dynamics.

Each sequence was obtained from a human, a chimpanzee, a gorilla and an orang-utan. When combined together, the total sequence emphatically confirmed the current consensus that, of these species, the humans and chimps have the most recent common ancestry (Box 1). When taken on their own, however, approximately 40% of the sequences gave different branching orders. Using the relationship between non-congruence and effective population size, Chen and Li calculated that the ancestral population size was 52 000–90 000 individuals. It is perhaps easier to follow the logic of the calculation in the opposite direction.  $T_3$  is thought to be around two million years, this represents roughly 100 000 generations (assuming a 20-year generation span for the common ancestor). This number of generations is one to two times the estimate of the ancestral size. A substantial proportion of loci would go through the internode without coalescing (Fig. 1, in Box 2) and would consequently produce the high frequency of incongruent gene trees.

A further indication of the poor reliability of estimates from single loci comes from the estimates of the relative age of the most recent human–chimp

common ancestor and the common ancestor for all three species. The combined estimate for the multiple loci is that  $T_3$  represents one third of the time to the common ancestor for humans and chimps. An individual estimate previously obtained for a globin pseudogene<sup>16</sup> was 10%, demonstrating the substantial error on estimates from single loci. An estimate from mtDNA (Ref. 17) was 60%. In this case, the difference from the combined average of 33% will also include a bias owing to the smaller effective population size of mtDNA.

Other species might be expected to have larger  $N_e$  than do primates and the timings estimated from their molecular phylogenies would, therefore, be subject to greater bias and error. A similar multilocus approach indicates that the ancient and modern effective sizes of some *Drosophila* spp. were of the order of three million (Ref. 18) and that large population size results in different species sharing the same polymorphism<sup>19</sup>. To extend the lessons to other species, it would be helpful to find some rules for assessing  $N_e$  from our knowledge of the biology of a species. It is particularly instructive to compare the estimate for the human–chimp ancestor with that for the human lineage. The relative size can be estimated from comparisons between and within species over multiple loci<sup>20</sup>. The effective size along the human lineage is an order of magnitude smaller than that for the human–chimp ancestor, around 10 000. This result probably does not indicate a decreased population size in humans, but, instead, can be

explained by the effect of increased migration in a subdivided population.

#### $N_e$ in a subdivided population

In Box 2, the times to a common ancestor were modelled under the assumption that the species was an idealized population of size  $N_e$ , which meant that lineages coalesced at the rate  $1/2N_e$ . Box 3 shows how, by quantifying the equivalent rate in a simple model of subdivided population (the 'Island Model'), it is possible to obtain an expression for the effective size of a subdivided population as  $N_e = ND(1 + [4Nm]^{-1})$ . The product  $ND$  is the total population size of the species. It follows that increasing subdivision (i.e. decreasing the migration rate  $m$ ) increases the effective size above that for an undivided species. This principle is long established<sup>20–23</sup> but can seem counterintuitive. For example, Wakely<sup>22</sup> used this equation to argue that the decrease in effective population size along the human lineage might be a result of the increased migration rate of humans.

One important improvement on the simple model in Box 3 is to allow for the possibility that individual demes can go extinct and be re-established. Slatkin<sup>24</sup> showed that the effect on genetic differentiation among demes depends on whether the colonists tend to be drawn from the same deme, or from a collection of demes. Nonetheless, the general effect of extinction and recolonization is, consistently, to reduce long-term effective population size<sup>25</sup>. Takahata<sup>23</sup> uses this principle to argue that the reduction in human effective population size was the result of increased extinction and recolonization rates.

Whitlock and Barton's<sup>26</sup> model was even more realistic and therefore, more complex than that in Box 3. They included the dynamics of population size and migration and found that the effect of increasing migration rate (i.e. decreasing subdivision) could be to increase effective population size. This is the diametrically opposite conclusion of the simpler analysis. How is this possible? One effect of increasing migration can be to cause the stochastic fluctuations in the size of different demes to synchronize, which evens out their contribution to migrants in the subsequent generation. This means that migrants are drawn from a wide range of demes, reducing the rate of coalescence. A second effect is to increase the proportion of localities that are occupied as a result of greater movement. This also contributes to the increased effective size. These effects occur for only some combinations of parameter values and it might be hard to predict which species will exhibit them.

In most species, gene flow will tend to be between adjacent populations. This means that gene exchange between distant populations will usually require a series of movements, over several generations, to span the intervening populations. The times to a common ancestor are therefore greater than in the Island Model of Box 3. The average time will depend

on the distance between populations<sup>27</sup>. Barton and Wilson<sup>28</sup> derived expressions for the distribution of times. The average times and variance are greater than for the simple Island Model, and in many species the common ancestors (Genes *A* or *a* in Box 1) will have lived before the last ice age when the distribution of the populations was radically different from that of the present day.

In addition to these demographic effects, effective population size can be affected by selection at linked loci and, consequently, can vary among different regions in the genome. This effect has been used to explain regions of unusually high or low genetic variation within the genome of *Drosophila*<sup>29,30</sup>. It appears, then, that a daunting depth of knowledge is required to predict the effective population size of a species from its biology. The more convincing evidence probably comes from interpreting the genetic diversity within species.

#### Diversity within species and postglacial expansions

In actual species, the nature of subdivided populations is even more complex than that of the Whitlock and Barton model<sup>26</sup>. The geographical distribution of genetic diversity in many species implies that their current distribution was strongly affected by postglacial recolonization of their current range<sup>31</sup>. The greatest genetic distances within species are often between regions that are thought to have served as glacial refugia. Away from these regions, there are often large ranges that are genetically more uniform and with less differentiation within each population. This pattern can be explained by the dramatic loss of variation that occurred during postglacial range expansion<sup>32</sup> out of the refugia. Expansion into virgin territory can lead to a series of population BOTTLENECKS, each establishing populations that spread to cover a large area. Surveys of within-species diversity in these regions, which might make up much of the species range, would lead to the erroneous conclusion of a very recent common ancestry.

The magnitude of the differentiation among putative refugia can, however, be surprisingly large. Taberlet *et al.*<sup>33</sup> collated information from several species that are thought to have been restricted to refugia in the same two regions: the Iberian peninsular and the Balkans. The authors compared mtDNA sequence divergence from populations of the same species between the two regions and found that the divergence was often as great as that found between distinct species. Furthermore, the values were quite disparate, for example: 2% for bears *Ursus arctos*, 6.5% for a shrew *Crocidura suaveolens*, and 7.6% for water voles *Arvicola* spp. Taberlet *et al.* argue that these values do not seem to correspond to particular climatic events splitting the species range, but might simply reflect divergence since they became established in northern Europe, in some cases, several million years ago.

This sort of history could help explain the patterns in bird divergence pointed out by Klicka and Zink<sup>8</sup>, in which species previously thought to be more recently diverged showed levels of differentiation similar to those thought to be older. Part of the explanation might be that the lineages of the ancestral genes have diverged for many thousands of years within the ancestral species in geographically isolated localities<sup>10,11</sup>. Even when there is such long-standing divergence at some loci, other loci in the same species might share a much more recent ancestor. This would require occasional gene flow between the two regions, so that the pattern of ancestry was similar to that of the simple subdivided populations described in Box 3. Studies of the hybrid zones between the descendants of different refugia suggest that such gene flow is possible<sup>33,34</sup>. Although gene flow has been detected at the zones themselves, evidence of genetic introgression is typically found within only a few kilometres or tens of kilometres of the zone. The distances between refugial areas are thousands of kilometres. It remains possible that there could be some gene flow over long periods, or that some genes are linked to selected alleles and have therefore spread rapidly throughout the species range. Such effects might be picked up by multilocus surveys.

#### Errors over longer evolutionary periods

Over longer evolutionary periods, a different set of problems limits the interpretation of trees inferred from single loci. In this issue, Barraclough and Nee<sup>2</sup> point out that evolution also involves hybridization between species, especially in plants. This process means that phylogenies will form reticulate (netlike) rather than simple bifurcating patterns. The various processes by which crosslinks on the net (representing hybridization) can form and stabilize are reviewed in Ref. 35. The resolution of such phylogenies will require both new analytical methods and data from many loci. This will be complicated by the interaction between the genomes of the hybridizing species. There is intriguing evidence that the creation of allopolyploid hybrids could induce a burst of rapid evolutionary changes in the first few generations. Using restriction fragment length polymorphisms (RFLPs), Song *et al.*<sup>36</sup> detected substantial genome changes after only the F2 generation in synthetic rapeseed *Brassica napus*

(allotetraploid). Similarly, Liu *et al.*<sup>37</sup> detected loss of RFLP variants within the first six generations of synthetic cultivars of allohexaploid wheat *Triticum aestivum*.

A second major problem also arises over longer periods. The assumption of constant rates of substitution can no longer be made confidently. Many studies show evidence of relatively constant rates of substitution, but the tests often lack power<sup>38</sup> or else compare closely related species. Careful comparisons do show variation in rate<sup>39</sup>. Strategies might be designed to overcome them; for example, species with very different generation times might have more similar rates of amino acid substitution compared with rates of synonymous substitutions<sup>40–42</sup>.

#### Conclusion

As the subject matures and sequence data increases, the issue of rate variation will probably be dealt with by more careful calibration of molecular clocks, and crossvalidation between species and loci. The uncertainty about the timing and the branching order for recent events can be reduced for the recent branches. Edwards and Beerli<sup>3</sup> argue that the way forward is to extend the analysis of ancestral population size to include multispecies trees. Information from long, linked sequences will become less attractive than that from multiple unlinked loci with essentially independent phylogenies (which will need to be short to reduce the effects of recombination). The most valuable outcome might not be the more precise dates, but an insight into the demography of the ancestral populations.

The estimates of the ancestral effective population size might show dramatic changes, as in the case of humans, or relative stability, as in the case of *Drosophila*. These ancestral size estimates can reflect the degree of population subdivision rather than reflecting the actual size of the population. The results might, therefore, help assess the relevance of models of speciation that require a particular population size or structure. There is a major challenge in interpreting the data from species showing reticulate evolution. The analysis of multiple gene trees might help us understand both the transmission of genes through a reticulate NETWORK and the interaction of the different genomes when they come together.

#### References

- 1 Nei, M. (1987) *Molecular Evolutionary Genetics*, Columbia University Press
- 2 Barraclough, T. and Nee, S. (2001) Phylogenetics and speciation. *Trends Ecol. Evol.* 16, 391–399
- 3 Edwards, S.V. and Beerli, P. (2000) Perspective: gene divergence, population divergence, and the variance in coalescence times in phylogeographic studies. *Evolution* 54, 1839–1854
- 4 Turelli, M. *et al.* (2001) Theory and speciation. *Trends Ecol. Evol.* 16, 330–343
- 5 Hudson, R.R. (1990) Gene genealogies and the coalescent process. *Oxford Surv. Evol. Biol.* 7, 1–44
- 6 Moore, W.S. (1995) Inferring phylogenies from mtDNA variation – mitochondrial-gene trees versus nuclear-gene trees. *Evolution* 49, 718–726
- 7 Moore, W.S. (1997) Mitochondrial-gene trees versus nuclear-gene trees, a reply to Hoelzer. *Evolution* 51, 627–629
- 8 Klicka, J. and Zink, R.M. (1997) The importance of recent ice ages in speciation: a failed paradigm. *Science* 277, 1666–1669
- 9 Klicka, J. and Zink, R.M. (1999) Pleistocene effects on North American songbird evolution. *Proc. R. Soc. London B Biol. Sci.* 266, 695–700
- 10 Avise, J.C. and Walker, D. (1998) Pleistocene phylogeographic effects on avian populations and the speciation process. *Proc. R. Soc. London B Biol. Sci.* 265, 457–463
- 11 Avise, J.C. (2000) *Phylogeography: the History and Formation of Species*, Harvard University Press
- 12 Hudson, R.R. (1992) Gene trees, species trees and the segregation of ancestral alleles. *Genetics* 131, 509–512
- 13 Wu, C.I. (1991) Inferences of species phylogeny in relation to segregation of ancient polymorphisms. *Genetics* 127, 429–435

- 14 Chen, F.C. and Li, W.H. (2001) Genomic divergences between humans and other hominoids and the effective population size of the common ancestor of humans and chimpanzees. *Am. J. Hum. Genet.* 68, 444–456
- 15 Ruvolo, M. (1997) Molecular phylogeny of the hominoids: inferences from multiple independent DNA sequence data sets. *Mol. Biol. Evol.* 14, 248–265
- 16 Bailey, W.J. *et al.* (1991) Molecular evolution of the psi-eta-globin gene locus: gibbon phylogeny and the hominoid slowdown. *Mol. Biol. Evol.* 8, 155–184
- 17 Horai, S. *et al.* (1992) Man's place in Hominoidea revealed by mitochondrial DNA genealogy. *J. Mol. Evol.* 35, 32–43
- 18 Li, Y.J. *et al.* (1999) Paleo-demography of the *Drosophila melanogaster* subgroup: application of the maximum likelihood method. *Genes Genet. Syst.* 74, 117–127
- 19 Hey, J. and Kliman, R.M. (1993) Population genetics and phylogenetics of DNA-sequence variation at multiple loci within the *Drosophila melanogaster* species complex. *Mol. Biol. Evol.* 10, 804–822
- 20 Takahata, N. and Satta, Y. (1997) Evolution of the primate lineage leading to modern humans: phylogenetic and demographic inferences from DNA sequences. *Proc. Natl. Acad. Sci. U. S. A.* 94, 4811–4815
- 21 Wright, S. (1969) *Evolution and the Genetics of Populations: the Theory of Gene Frequencies* (Vol. 2), University of Chicago Press
- 22 Wakeley, J. (1999) Nonequilibrium migration in human history. *Genetics* 153, 1863–1871
- 23 Takahata, N. (1995) A genetic perspective on the origin and history of humans. *Annu. Rev. Ecol. Syst.* 26, 343–372
- 24 Slatkin, M. (1977) Gene flow and genetic drift in a species subject to frequent local extinction. *Theor. Popul. Biol.* 12, 253–262
- 25 Pannell, J.R. and Charlesworth, B. (1999) Neutral genetic diversity in a metapopulation with recurrent local extinction and recolonization. *Evolution* 53, 664–676
- 26 Whitlock, M.C. and Barton, N.H. (1997) The effective size of a subdivided population. *Genetics* 146, 427–441
- 27 Slatkin, M. (1991) Inbreeding coefficients and coalescence time. *Genet. Res.* 58, 167–175
- 28 Barton, N.H. and Wilson, I. (1995) Genealogies and geography. *Philos. Trans. R. Soc. London Ser. B* 349, 49–59
- 29 Berry, A.J. *et al.* (1991) Lack of polymorphism on the *Drosophila* 4th chromosome resulting from selection. *Genetics* 29, 1111–1117
- 30 Kreitman, M.E. and Aguade, M. (1986) Excess polymorphism at the ADH locus in *Drosophila melanogaster*. *Genetics* 114, 93–110
- 31 Hewitt, G.M. (1999) Post-glacial re-colonization of European biota. *Biol. J. Linn. Soc.* 68, 87–112
- 32 Ibrahim, K.M. *et al.* (1996) Spatial patterns of genetic variation generated by different forms of dispersal during range expansion. *Heredity* 77, 282–291
- 33 Taberlet, P. *et al.* (1998) Comparative phylogeography and postglacial colonization routes in Europe. *Mol. Ecol.* 7, 453–464
- 34 Hewitt, G.M. (2000) The genetic legacy of the quaternary ice ages. *Nature* 405, 907–913
- 35 Rieseberg, L.H. and Noyes, R.D. (1998) Genetic map-based studies of reticulate evolution in plants. *Trends Plant Sci.* 3, 254–259
- 36 Song, K. *et al.* (1995) Rapid genome change in synthetic polyploids of *Brassica* and its implications for polyploid evolution. *Proc. Natl. Acad. Sci. U. S. A.* 92, 7719–7723
- 37 Liu, B. *et al.* (1998) Rapid genomic changes in newly synthesized amphiploids of *Triticum* and *Aegilops*. I. Changes in low-copy noncoding DNA. *Genome* 41, 272–277
- 38 Bromham, L. *et al.* (2000) The power of relative rates tests depends on the data. *J. Mol. Evol.* 50, 296–301
- 39 Rand, D.M. (1994) Thermal habit, metabolic rate and the evolution of mtDNA. *Trends Ecol. Evol.* 9, 125–131
- 40 Chao, L. and Carr, D.E. (1993) The molecular clock and the relationship between population-size and generation time. *Evolution* 47, 688–690
- 41 Gillespie, J.H. (1989) Lineage effects and the index of dispersion of molecular evolution. *Mol. Biol. Evol.* 6, 636–647
- 42 Li, W.H. *et al.* (1987) An evaluation of the molecular clock hypothesis using mammalian DNA-sequences. *J. Mol. Evol.* 25, 330–342

# Sexual selection and speciation

Tami M. Panhuis, Roger Butlin, Marlene Zuk and Tom Tregenza

The power of sexual selection to drive changes in mate recognition traits gives it the potential to be a potent force in speciation. Much of the evidence to support this possibility comes from comparative studies that examine differences in the number of species between clades that apparently differ in the intensity of sexual selection. We argue that more detailed studies are needed, examining extinction rates and other sources of variation in species richness. Typically, investigations of extant natural populations have been too indirect to convincingly conclude speciation by sexual selection. Recent empirical work, however, is beginning to take a more direct approach and rule out confounding variables.

Tami M. Panhuis\*  
Marlene Zuk  
Dept of Biology,  
University of California,  
Riverside, CA 92521, USA.  
\*e-mail:  
panhuist@citrus.ucr.edu

Roger Butlin  
Tom Tregenza  
Centre for Biodiversity  
and Conservation, School  
of Biology, University of  
Leeds, Leeds, UK LS2 9JT.

SEXUAL SELECTION (see Glossary) results from differential mating success among individuals within a population. Competition for fertilization occurs through direct competition between members of the same sex (e.g. male–male competition and sperm competition) or through the attraction of one sex to the other (e.g. female choice). Although long recognized as important in intrapopulation evolution, sexual selection has more recently been invoked as a driving force behind SPECIATION. Speciation, the splitting of one SPECIES into two or more, occurs by sexual selection when a parallel change in mate

preference and SECONDARY SEXUAL TRAITS within a population leads to PREZYGOTIC ISOLATION between populations, and when this is the primary cause of REPRODUCTIVE ISOLATION.

Classic models of speciation<sup>1,2</sup> recognized that reproductive isolation, and subsequent speciation, could be generated by differences in sexual traits (including behaviours). Divergence in sexual traits between allopatric populations was considered to result either from drift, PLEIOTROPY or adaptation to environmental conditions, or following SECONDARY CONTACT, because individuals benefited by avoiding heterospecific matings (i.e. by REINFORCEMENT). It became clear, however, that changes between populations in sexual traits could also result from sexual selection and that this might represent a distinct process of speciation<sup>3,4</sup>. Sexual selection has the potential to lead to rapid divergence between populations, it can be independent of environmental differences, and it is predisposed to generate reproductive isolation because of its direct effect on traits involved in mate recognition. It is important to point out that the rapid change between populations