ELSEVIER

Recent advances in assessing gene flow between diverging populations and species

Jody Hey

The evolutionary process of divergence, which ultimately leads to the generation of new species, is thought to occur usually without any gene exchange between the diverging populations. However, until the recent growth of multi-locus datasets, and the development of new population genetic methods, it has been very difficult to assess whether or not closely related species have, or have not, exchanged genes during their divergence. Several recent studies have found significant signals of gene flow during species formation, calling into question the conventional wisdom that gene flow is absent during speciation.

Addresses
Department of Genetics, Rutgers University, Nelson Biological Laboratories, Room B326A, 604 Allison Road, Piscataway, NJ 08854, USA

Corresponding author: Hey, Jody (hey@biology.rutgers.edu)

Current Opinion in Genetics & Development 2006, **16**:592–596

This review comes from a themed issue on
Genomes and evolution
Edited by Chris Tyler-Smith and Molly Przeworski

Available online 19th October 2006

0959-437X/$ – see front matter

DOI 10.1016/j.gde.2006.10.005

Introduction

If one discovers that two species are *not* reproductively isolated but have actually been hybridizing and exchanging genes, what is the first question to ask? For some practical concerns, the relevant question is the semantic one of 'should they be identified as separate species?'. But a more basic evolutionary question is 'how did the species separate from each other if they have been exchanging genes?' Species are nearly always identified because some kind of divergence (e.g. morphological or genetic) from other species has been discovered. We also know that it only takes a small amount of gene flow to keep two populations from diverging [1]. So if divergence has happened *and* gene exchange has also been happening, then there is a conundrum.

The general answer to the puzzle is that divergence can happen at some genes, even if there is gene exchange for other genes. Hybrids carry a full set of genes from each population, but backcross hybrids do not, and so it is possible for some genes to pass between populations if backcross hybrids vary in their fitness depending on which genes they carry. In this way, natural selection, acting differently in two diverging populations, can prevent gene flow at some genes (i.e. the genes at which divergence is occurring) and can enable other genes to pass between the populations. In other words, divergence in the presence of gene exchange implies that natural selection is playing an active role in the divergence process [2–4].

This review covers recent developments and applications in the detection and estimation of gene flow between species that have recently diverged. Given that a number of studies have found evidence for gene flow, it appears that divergence and speciation may often occur in the presence of gene flow.

Population genetics and divergence

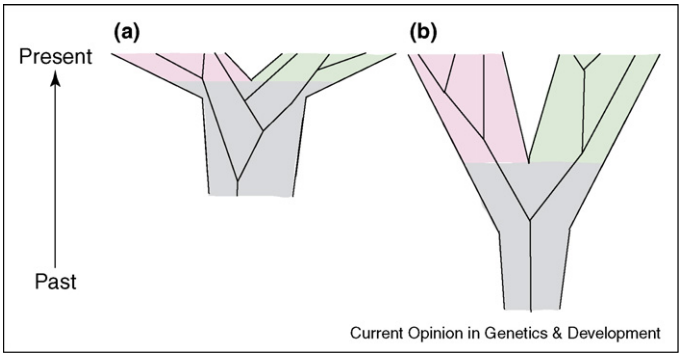
When one population separates into two, genetic variation will be shared for some period of time even in the absence of gene exchange [5–7]. Figure 1 shows an example of how genealogies are more likely to coalesce within species if separation times are longer ago. If the sizes of both populations are large, and gene trees are deep within populations, then genealogies and genetic variation might be shared at some genes for very long periods of time, possibly even after the populations have diverged and become reproductively isolated species. This means that it is possible to sequence a copy of a gene from one species and find that it is more similar to a gene from a closely related species than it is to another copy of the gene from the same species. This will happen simply by chance if the populations separated only recently, even if there is no gene exchange. So the challenge is to determine whether or not genetic variation that is shared by both populations is simply a remnant of variation in the common ancestor or if it is due to gene exchange after the population started to separate.

Gene flow revealed by differences among genes

If population genetic data are available from multiple loci, for each of two divergent populations or species, then the patterns of variation at the different genes can be contrasted with one another. A history of divergence with gene flow is generally indicated if some loci show little divergence and others show a large amount of divergence, such that the variation in divergence among the different genes is greater than expected under a model without gene flow [8].

In recent years, findings of this sort have mostly come from insect species, including *Drosophila* species [9–11],

Figure 1



Population-splitting events are depicted as an ancestral (gray) population splitting into green and red descendant populations. **(a)** Recent population splitting. **(b)** Older population splitting. In each figure a genealogy for a set of sampled gene copies is shown to demonstrate how genealogies can be intermingled. This occurs more for samples from two closely related populations (a), than for populations that have been separated for a longer time (b).

Hawaiian crickets [12], and the larch budmoth [13]. A particularly illuminating case arises among the 'M' and 'S' forms of the mosquito *Anopheles gambiae*, which are distinguished and identified on the basis of diagnostic single nucleotide polymorphisms (SNPs) and on being partially reproductively isolated from each other. The two forms belong to the same species, and they have broadly overlapping geographic ranges. A genome scan found just three small portions of the genome — on three different chromosomes — that are differentiated between the two types [14•]. The pattern fits very well with what might be expected in the early stages of speciation by sympatry. The two sympatric populations have diverged at multiple points in the genome, and this divergence causes partial reproductive isolation. Importantly, the pattern is also consistent with modeling work and data that suggest that the genes in which divergence does happen, when there is gene flow, are likely to be in areas of restricted recombination [15,16]. Two of the three genome-regions that are divergent in *A. gambiae* are located near centromeres, where crossing over is greatly reduced.

In some of these cases, the contrast is so great between the apparent histories of different genes that a conclusion of differential gene flow seems inescapable. However, the variance among loci in the degree to which they share alleles can be surprisingly large, even under a model of no gene flow [7,17]. This point raises the need for a statistical approach in which the null model has zero gene flow for the sampled loci.

Statistical approaches to inferring gene flow

Figure 2 shows a graphical model that represents an ancestral population that split, at a specific point in time, into two populations, after which there might have been gene exchange in one or both directions. The model can be represented in mathematical terms by six parameters, including the effective population sizes of the ancestral and descendant populations (N_A, and N_1 and N_2, respectively), the splitting time (t), and two gene-flow parameters (m_1 and m_2). Called the 'isolation with migration' [18], it is a form of 'divergence with gene-flow' model [4] that permits a fully mathematical, coalescent treatment of the problem [19,20]. For questions on gene flow, two general approaches have been developed to use this model statistically.

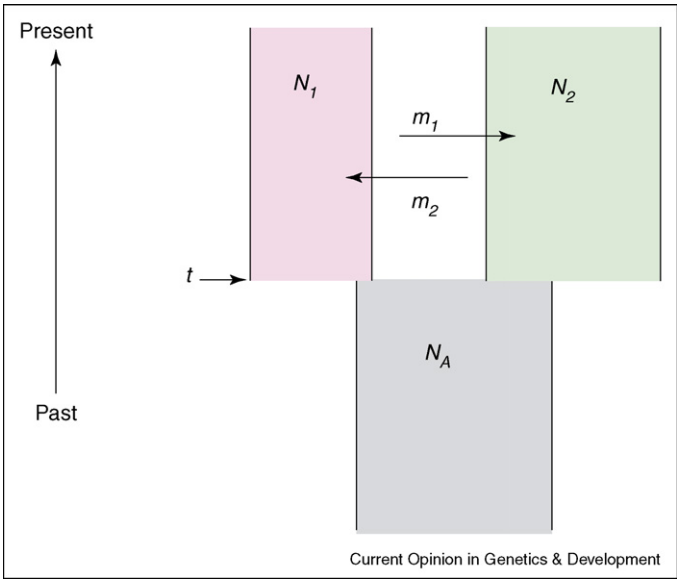
The simplest approach is to fit the data to the model, assuming no gene flow, and then to test whether the quality of the fit between data and model is so bad that the assumption of a zero gene flow model must be rejected [8,21].

Recently a much fuller likelihood-based approach was developed by extending Felsenstein's classic equation to the isolation with migration model. The Felsenstein equation, which was a conceptual breakthrough — not a mathematical one — shows how a dataset and the parameters of a population genetic model can be related to one another, by treating the unknown genealogy (G) as a dummy variable that is removed by integration [22]:

$$\Pr(Data|Parameters) = \sum_G \Pr(Data|G)\Pr(G|Parameters)$$

Actually approximating the solution to this requires an appropriate mutation model for the data, for calculation of $\Pr(Data|G)$, a probabilistic model of genealogies (i.e. a coalescent model) suitable to the problem for the calculation of $\Pr(G|Parameters)$, and an efficient way to sample from the possible set of genealogies. Nielsen and Wakeley [23] figured out how to adapt this framework to the isolation with migration model by casting the problem in a Bayesian framework. Their method is not fast, because it relies on a Markov chain Monte Carlo simulation, but it is flexible, and recently it has been extended to multiple

Figure 2



The 'isolation with migration' model. An ancestral population of effective size N_A separated at time t into two descendant populations of effective sizes N_1 and N_2. After the population separation, genes are exchanged at rates m_1 and m_2.

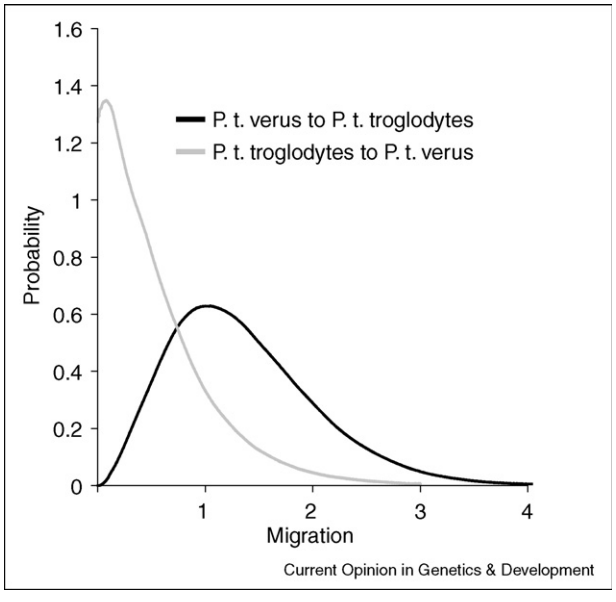
loci for the full six-parameter model in Figure 2 [24•], as well as to including a variety of mutation models [25,26]. The method is implemented in a computer program, called IM for 'isolation with migration' (http://lifesci.rutgers.edu/~heylab/HeylabSoftware.htm#IM).

The output of the IM program is a curve — one for each of the six model parameters — that is an estimate of the posterior probability of that parameter, given the data. In most cases, the curve can be treated as a likelihood surface from which the location of the peak identifies the maximum-likelihood estimate of that particular parameter, and the fall-off away from the peak can be used to estimate confidence intervals of that estimate. For example, if the migration parameters have the highest probability at zero, then the estimated migration rate is zero, and if the estimated probability of zero migration is itself near zero, then one can reject a migration of zero.

An example from chimpanzees

Figure 3 shows examples of migration parameter curves for the central and western subspecies of common chimpanzee, *Pan troglodytes troglodytes* and *Pan troglodytes verus*, respectively [26]. In the case of gene flow from *P. t. troglodytes* to *P. t. verus*, the curve has a peak near zero, and the probability of the peak is very near to what the probability is at zero. In this case, we cannot reject a migration rate of zero. However, in the reverse direction, the peak is far from zero, and the estimated probability of zero gene flow is zero, and so we can reject a model of no gene flow from *P. t. verus* to *P. t. troglodytes*. The full analysis of these subspecies suggests a separation time of about 400 000 years ago, whereas the divergence between the common chimpanzees and the bonobo (*Pan bonobo*) was estimated to be about 900 000 years ago [27].

Figure 3



Output from the isolation with migration (IM) program for the two migration parameters for an analysis of two chimpanzee subspecies, *Pan troglodytes verus* and *Pan troglodytes troglodytes*. For each value of each parameter, the graph shows the estimated probability that that is the true value of the parameter, given the data.

The IM approach is now in frequent use and has yielded some surprising findings that are not just restricted to questions on gene flow. For example, three pairs of *Heliconius* butterfly species [28] where found to have been exchanging genes, but also to have speciation times estimated in the hundreds of thousands of years. The time depth in years is surprising because of the short generation time of these species (i.e. less than two months). Therefore, even though the butterflies diverged from each other on a time-scale similar to that experienced by the chimpanzees, in population genetic terms the time depth for the butterflies is many times greater. Not all studies find evidence of gene flow, however. For example, pair-wise analyses of three populations of rainforest skinks estimated gene flow to be zero in all but one case, and in the non-zero case the rate was not significantly different from zero [29].

Separate migration rates for different genes

A variation on the basic six-parameter method is to fit a model in which every locus in the dataset is provided with a pair of gene-flow parameters [30]. This model can explicitly incorporate the differential gene flow that is expected if gene flow *and* divergence are both occurring. Bull *et al.* [31•] studied two of the three *Heliconius* species (*Heliconius cydno* and *Heliconius melpomene*) studied by Kronforst *et al.* [28], and also found evidence for gene flow. However, in this case, by running a model in which each locus has distinct gene-flow parameters, they could reject zero gene flow at one nuclear locus, *Mpi* (*mannose-6-phosphate isomerase*), but not at any of three other loci. Another recent study using the locus-specific migration parameters looked at two subspecies of the European rabbit (*Oryctolagus cuniculus*) [32] and, again, loci differed in their apparent histories of gene flow. The migration rate estimates for two loci near centromeres were not significantly different from zero, but zero migration could be rejected for two other loci.

Conclusions: understanding speciation

These recent findings of gene flow between divergent populations and species seem to suggest that gene flow is a common feature of the early stages of the divergence process. This is surprising for two reasons. First, these examples come not from populations that have recently separated but from different populations or species that have been clearly identified on the basis of divergence. Second, it has at times been claimed that gene flow is rare or non-existent between populations that continue to diverge and become species [33]. The recent findings that have come from applying population genetic methods to the divergence process undermine this view, although it is too soon to appreciate just how frequently divergence and speciation happen with, and without, gene flow. It is also important to recognize that these population genetic findings fit well within other considerations of the speciation process that emphasize processes other than strict biogeographic separation. In recent years, speciation discussions have focused increasingly on the roles of selective agents, such as ecological factors [34–37], competition within sexes (i.e. sexual selection) [38], and reinforcement [39].

References and recommended reading

Papers of particular interest, published within the annual period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Wright S: **Evolution in Mendelian populations**. *Genetics* 1931, **16**:97-159.
2. Maynard Smith J: **Sympatric speciation**. *Am Nat* 1966, **100**:637-650.
3. Endler JA: *Geographic Variation, Speciation, and Clines*. Princeton, NJ: Princeton University Press; 1977.
4. Rice WR, Hostert EF: **Laboratory experiments on speciation: what have we learned in 40 years**. *Evolution Int J Org Evolution* 1993, **47**:1637-1653.
5. Avise JC: **Systematic value of electrophoretic data**. *Syst Zool* 1975, **23**:465-481.
6. Pamilo P, Nei M: **Relationships between gene trees and species trees**. *Mol Biol Evol* 1988, **5**:568-583.
7. Hey J: **The structure of genealogies and the distribution of fixed differences between DNA sequence samples from natural populations**. *Genetics* 1991, **128**:831-840.
8. Wakeley J, Hey J: **Estimating ancestral population parameters**. *Genetics* 1997, **145**:847-855.
9. Machado CA, Hey J: **The causes of phylogenetic conflict in a classic *Drosophila* species group**. *Proc Biol Sci* 2003, **270**:1193-1202.
10. Bachtrog D, Thornton K, Clark A, Andolfatto P: **Extensive introgression of mitochondrial DNA relative to nuclear genes in the *Drosophila yakuba* species group**. *Evolution Int J Org Evolution* 2006, **60**:292-302.
11. Llopart A, Lachaise D, Coyne JA: **Multilocus analysis of introgression between two sympatric sister species of *Drosophila*: *Drosophila yakuba* and *D. santomea***. *Genetics* 2005, **171**:197-210.
12. Shaw KL: **Conflict between nuclear and mitochondrial DNA phylogenies of a recent species radiation: what mtDNA reveals and conceals about modes of speciation in Hawaiian crickets**. *Proc Natl Acad Sci USA* 2002, **99**:16122-16127.
13. Emelianov I, Marec F, Mallet J: **Genomic evidence for divergence with gene flow in host races of the larch budmoth**. *Proc Biol Sci* 2004, **271**:97-105.
14. • Turner TL, Hahn MW, Nuzhdin SV: **Genomic islands of speciation in Anopheles gambiae**. *PLoS Biol* 2005, **3**:e285.
A unique study of what may be the early stages of sympatric speciation, as revealed by a divergence at three separate regions of the genome.
15. Noor MA, Grams KL, Bertucci LA, Reiland J: **Chromosomal inversions and the reproductive isolation of species**. *Proc Natl Acad Sci USA* 2001, **98**:12084-12088.
16. Rieseberg LH: **Chromosomal rearrangements and speciation**. *Trends Ecol Evol* 2001, **16**:351-358.
17. Degnan JH, Salter LA: **Gene tree distributions under the coalescent process**. *Evolution Int J Org Evolution* 2005, **59**:24-37.
18. Wakeley J, Hey J: **Testing speciation models with DNA sequence data**. In *Molecular Approaches to Ecology and*

Evolution. Edited by DeSalle R, Schierwater B. Birkhäuser Verlag; 1998:157-175.

19. Kingman JFC: **The coalescent**. *Stochastic Process Appl* 1982, **13**:235-248.

20. Takahata N: **The coalescent in two partially isolated diffusion populations**. *Genet Res* 1988, **52**:213-222.

21. Wang RL, Wakeley J, Hey J: **Gene flow and natural selection in the origin of *Drosophila pseudoobscura* and close relatives**. *Genetics* 1997, **147**:1091-1106.

22. Felsenstein J: **Phylogenies from molecular sequences: inference and reliability**. *Annu Rev Genet* 1988, **22**:521-565.

23. Nielsen R, Wakeley J: **Distinguishing migration from isolation. A Markov chain Monte Carlo approach**. *Genetics* 2001, **158**:885-896.

24. • Hey J, Nielsen R: **Multilocus methods for estimating population sizes, migration rates and divergence time, with applications to the divergence of *Drosophila pseudoobscura* and *D. persimilis***. *Genetics* 2004, **167**:747-760.
The authors describe the development of a multi-locus Bayesian analysis of the 'isolation with migration' model.

25. Hey J, Won Y-J, Sivasundar A, Nielsen R, Markert JA: **Using nuclear haplotypes with microsatellites to study gene flow between recently separated Cichlid species**. *Mol Ecol* 2004, **13**:909-919.

26. Palsbøll PJ, Berube M, Aguilar A, Notarbartolo-Di-Sciara G, Nielsen R: **Discerning between recurrent gene flow and recent divergence under a finite-site mutation model applied to North Atlantic and Mediterranean Sea fin whale (*Balaenoptera physalus*) populations**. *Evolution Int J Org Evolution* 2004, **58**:670-675.

27. Won YJ, Hey J: **Divergence population genetics of chimpanzees**. *Mol Biol Evol* 2005, **22**:297-307.

28. Kronforst MR, Young LG, Blume LM, Gilbert LE: **Multilocus analyses of admixture and introgression among hybridizing *Heliconius* butterflies**. *Evolution Int J Org Evolution* 2006, **60**:1254-1268.

29. Dolman G, Moritz C: **A multilocus perspective on refugial isolation and divergence in rainforest skinks (*Carlia*)**. *Evolution Int J Org Evolution* 2006, **60**:573-582.

30. Won YJ, Sivasundar A, Wang Y, Hey J: **On the origin of Lake Malawi cichlid species: a population genetic analysis of divergence**. *Proc Natl Acad Sci USA* 2005, **102**:6581-6586.

31. • Bull V, Beltran M, Jiggins CD, McMillan WO, Bermingham E, Mallet J: **Polyphyly and gene flow between non-sibling Heliconius species**. *BMC Biol* 2006, **4**:11.
The first study to use a likelihood framework to reveal different levels of gene flow at different genes.

32. Geraldes A, Ferrand N, Nachman MW: **Contrasting patterns of introgression at X-linked loci across the hybrid zone between subspecies of the European rabbit (*Oryctolagus cuniculus*)**. *Genetics* 2006.

33. Mayr E: *Animal Species and Evolution*. Cambridge, MA: Belknap Press of Harvard University Press; 1963.

34. Smith TB, Wayne RK, Girman DJ, Bruford MW: **A role for ecotones in generating rainforest biodiversity**. *Science* 1997, **276**:1855-1857.

35. Schluter D: **Ecology and the origin of species**. *Trends Ecol Evol* 2001, **16**:372-380.

36. Graham CH, Ron SR, Santos JC, Schneider CJ, Moritz C: **Integrating phylogenetics and environmental niche models to explore speciation mechanisms in dendrobatid frogs**. *Evolution Int J Org Evolution* 2004, **58**:1781-1793.

37. Berlocher SH, Feder JL: **Sympatric speciation in phytophagous insects: moving beyond controversy?** *Annu Rev Entomol* 2002, **47**:773-815.

38. Edwards SV, Kingan SB, Calkins JD, Balakrishnan CN, Jennings WB, Swanson WJ, Sorenson MD: **Speciation in birds: genes, geography, and sexual selection.** *Proc Natl Acad Sci USA* 2005, in press.

39. Coyne JA, Orr HA: **Patterns of speciation in *Drosophila***. *Evolution Int J Org Evolution* 1989, **43**:362-381.