Genetic estimates of population structure and gene flow: limitations, lessons and new directions

J.L. Bossart

D. Pashley Prowell

Indirect methods using genetic markers are the primary measure of gene flow levels among interbreeding populations. Results from studies employing these methods are often ambiguous and open to multiple interpretation. This is primarily due to low resolution of molecular markers and low precision of model-based estimates. Studies of paternity, kinship and phylogeography generate the most reliable results. Future studies should employ more powerful analytical methods, analyse loci independently, and attempt to distinguish confounding contributions of vicariance to isolation-by-distance studies. Moreover, direct assessment of movement remains the most valid approach to many key issues in population biology.

J.L. Bossart and D.P. Prowell are at 402 Life Sciences Building, Dept of Entomology, Louisiana State University, Baton Rouge, LA 70803, USA.

Studies of gene flow are integral to interpretation of microevolutionary patterns and geographic structure. Through such studies, we strive to gain insights into evolutionary independence and potential for population diversification, differentiation, and ultimately speciation. Indirect methods, which infer gene flow from genetic data on measurements of population structure, are relied upon almost exclusively to calculate this important parameter (see Box 1 for definitions of relevant terms). This widespread popularity reflects the ease with which genetic data can generally be collected and analysed[1], and the belief that such estimates may be more valid than those obtained through direct studies of movement[2] because they give a temporal perspective. However, many indirect estimates of gene flow may be of limited interpretive value and reveal little about the degree of genetic cohesion and evolutionary potential of contemporary populations, for two reasons: (1) molecular markers appear to lack the resolving power required to distinguish contemporary patterns of gene flow with current analytical methods; (2) limitations imposed by conventional population genetics models lead to low precision of gene flow estimates and a general lack of biological realism.

Others have emphasized the difficulties associated with interpreting patterns of molecular variation and model-based estimates (for example, Ref. 3). Yet these cautions have not been widely embraced by the scientific community. Conclusions often are drawn about genetic structure and levels of gene flow even though there are multiple, equally viable interpretations

of results. Moreover, these methods are advocated for use in ecological studies[4], such as population demography, metapopulation processes, and interspecific interactions, but there is much uncertainty about the success of molecular methods in ecological time scales. Indeed, present-day patterns of genetic structure may often reflect effects of Pleistocene glaciations and post-glacial range expansions[5,6].

Problems with interpretation of indirect estimates

Most population genetic studies using indirect methods seek to establish levels of gene flow. Statistics that estimate this parameter generally include variance among populations in allele frequencies (F_{ST}, Wright[7]; G_{ST}, Nei[8]; and Θ, Weir and Cockerham[9]; and their analogues, e.g. R_{ST}, Slatkin[10]) and Nm, the effective number of migrants per generation[7]. Values are interpreted as indications of relative levels of gene flow among populations. For instance, high Nm and low F_{ST} indicate high migration. Additional analyses are often conducted based on spatial relationships predicted by isolation-by-distance models. Correlations between geographic distance and genetic distance are calculated, and geographic patterns visualized by clustering algorithms. Three possible results obtained from these studies are: (1) no genetic structure or differentiation among populations; (2) significant genetic structure but no geographic pattern to the structure; or (3) significant genetic differentiation and geographically structured populations. Difficulties arise in the interpretation of all of these results.

The first case can produce ambiguous results because there are multiple explanations for homogeneity that cannot be readily distinguished. These include current gene flow among populations, balancing selection on markers, and lack of resolution of markers (i.e. retention of shared ancestral polymorphisms). Scientists generally favor one of these explanations over the others. Occasionally, a follow-up study with a second marker detects significant heterogeneity among populations and implicates lack of resolution for the first marker (for example, Zink[11] detected no pattern in fox sparrows with allozymes but later detected a pattern with mtDNA[12]). Unless lack of resolution is documented, distinguishing among explanations is conjectural.

Results are also ambiguous in the second case where population structure is detected by a significant F_{ST}, yet no correlation with geography or geographic structuring is observed. Despite this result, estimates of Nm often are taken at face value as the approximate number of migrants moving among populations. However, these estimates are not necessarily indicative of gene flow for at least three reasons. First, if these measures provide meaningful estimates of gene flow, there should be hierarchical structure that is associated with geography. Second, significant discrepancies among the various statistics used to partition genetic variation suggest problems in general with the methods and their underlying assumptions. For example, F_{ST}, G_{ST}, and private allele analyses do not always generate comparable Nm values[13,14]. Third, demographic instability or heterogeneity among populations can cause loose correlations between F_{ST} and Nm (Ref. 15). F_{ST} may fluctuate over time and space, resulting in a poor estimation of genetic structure and gene flow.

The third outcome, significant genetic and geographic structure, can be ambiguous because patterns can result from two distinct processes. Structure may result when vicariant events restrict gene flow among a subset of populations (i.e. geographic isolation). Sharp discontinuities in gene frequencies are usually evident on either side of the vicariant event with genetic homogeneity among populations on the same side. Alternatively, gene flow levels among populations may vary and follow an isolation-by-distance model with neighboring populations exchanging more migrants than distant ones. Isolation-by-distance theoretically results in a direct relationship between genetic distance and geographic distance. In reality, these two alternatives cannot be easily distinguished because the scatter of points in the distance regression is usually large. Furthermore, it is likely that many studies

of isolation-by-distance are, in fact, confounded by vicariance because vicariance is more likely to be detected with increasing geographic distance. Irrespective of the term used to define such isolated taxa, one is not studying levels of gene flow in contemporary populations (i.e. *Nm* has been zero over some period of evolutionary time), but instead, geographic variation among allopatric taxa (i.e. phylogeography). There are a few studies that appear to support isolation-by-distance, usually in very 'sedentary' organisms (e.g. Refs 14,16,17). However, in these studies, it is not possible to discount the fact that samples contributing to a significant genetic–geographic correlation may be completely isolated.

Many of these interpretational problems have been recognized previously[3,18], and have fostered an expanding interest in the development of high resolution molecular markers. Population genetic studies have moved beyond allozymes and now routinely include data from DNA regions. In general, allozymes evolve at a slower rate than mtDNA and nuclear DNA such as microsatellites[19,20] and rarely provide enough resolution to track contemporary processes and assess gene flow levels. Allozymes are good at detecting vicariance[1]. Although DNA markers evolve at faster rates and thus potentially reveal more recent evolutionary events, their ability to resolve levels of gene flow remains uncertain. In a cursory review of the recent literature, we found little evidence that DNA markers give better estimates of gene flow (Table 1). In summary, allozymes and DNA markers produced similar results except in cases where a vicariant event was suggested by the DNA but not the allozymes (for example, Ref. 22). The challenge thus remains to separate genetic history from present-day patterns. Moreover, although indirect methods clearly reveal the history of gene frequency distributions up to some 'recent' point, the precise location of that point is unknown. The presumed advantage of molecular approaches is their temporal perspective on genetic structure, but this is only an advantage to the extent the evolutionary time scale of this temporal perspective can be determined.

Problems with population genetic models

A second issue regarding indirect methods concerns limitations imposed by population genetics models (for example, Ref. 29, p. 863). These limitations are sufficiently stringent that it is difficult to have confidence in estimates obtained. As Lewontin[3] cautioned, model-based estimates of evolutionary parameters, such as gene flow, will almost never be valid because parameters that define these models are spatially

---

Box 1. Definitions of terms

Contemporary gene flow: present-day pattern of movement of genes (via individuals) among interbreeding populations of a species.

Direct measures of movement: use of genetic (constructed) or physical markers that allow tracking of dispersal of individuals among populations.

Indirect estimates of gene flow: movement of genes (via individuals) inferred from genetic data on measurement of population structure.

Isolation-by-distance: decreasing gene flow and thus genetic similarity between populations with increasing geographic distance.

Marker resolution: relative power to reveal meaningful patterns of genetic structure and track contemporary gene flow; largely a function of mutation rate.

Vicariance: historical process of range fragmentation that leads to geographic isolation of populations within an ancestral species.

Table 1. Comparison of congruence and resolution of allozyme and DNA markers used to assess gene flow levels in some recent studies

Taxon	Marker	Significant heterogeneity	Genetic–geographic correlation	Geographic pattern
Anopheline mosquito[21]	Allozymes	N?	NA	NA
(*Anopheles gambiae*)	Microsatellites	N?	NA	NA
American oyster[22]	Allozymes	N	NA	N
(*Crassostrea virginica*)	mtDNA	Y	NA	Y*
	Nuclear RFLPs	Y	NA	Y*
Sea beet[23]	Allozymes	Y	N	NA
(*Beta vulgaris maritima*)	Nuclear RFLPs	Y	Y	NA
Atlantic salmon[24]	Allozymes	Y	Y	Y*
(*Salmo salar*)	Microsatellites	Y	Y	Y*
Atlantic cod[25]	Allozymes	Y	Y	Y*
(*Gadus morhua*)	Nuclear RFLPs	Y	Y	Y*
Cyprinid minnow[26]	Allozymes	Y	NA	Y*/N
(*Tiaroga cobitis*)	mtDNA	Y	NA	Y*/N
Cyprinid minnows[27]	Allozymes	Y	NA	Y*
(*Meda fulgida*)	mtDNA	Y	NA	Y*
Fox sparrow	Allozymes[11]	N	NA	N
(*Passerella iliaca*)	mtDNA[12]	Y	NA	Y*
Threespine stickleback	Allozymes[27]	Y	NA	Y*
(*Gasterosteus aculeatus*)	mtDNA[28]	Y	NA	Y*

N=No, Y=Yes; NA=value not calculated; *=vicariance probably responsible for observed pattern; Y/N=some inconsistencies in geographic pattern; N?=some loci significant, significance of overall heterogeneity not tested.

Box 2. Abiotic and biotic factors that promote genetic and demographic instability and non-uniformity across a species' range

- Weather-related effects, both catastrophic and deterministic
- Habitat patch location, e.g. edge or central
- Variable patch size
- Variable distance between habitat patches

- Variable resource availability and quality
- Rates of gene flow that evolve in response to habitat patch attributes
- Variable brood numbers and size
- Interactions with other organisms

---

and temporally variable and too sensitive to forces of evolution. More generally, population genetics models ignore the complexity that defines biological systems[30]. Conventional models assume that population structure and demographic parameters such as population size and dispersal rates are uniform and constant over space and time. Assumptions of demographic and genetic equilibrium and uniformity are unrealistic and violated as a matter of course. Demographic parameters are a function of abiotic and biotic factors that regularly change (Box 2), often in unpredictable ways (e.g. Ref. 31). These deterministic and stochastic elements and their interactions routinely perturb and introduce variability into natural systems, as commonly evidenced in the literature. Population genetic studies, however, only rarely address basic model assumptions of population structure and equilibrium with

---

<div style="border:1px solid #000; padding:10px;">

Box 3. Potential sources of error and ambiguity surrounding estimates of genetic structure and gene flow

- Choice of analytical method used to partition genetic variation
- Choice of model to estimate gene flow
- Choice of marker
- Inherent error associated with second order F-statistics
- Violations of assumptions of analytical methods, genetic models, or both
- Variation among locus specific estimates and averaging over loci
- Variation among allele specific estimates and averaging over alleles
- Inadequate or improper sampling of species' range
- Confounding of contemporary patterns with historical associations
- Cumulative averages that confound gene flow and other evolutionary determinants of gene frequencies among populations (e.g. demographic fluctuations and genetic bottlenecks)

</div>

respect to populations of interest. Moreover, discussions rarely include explanations of how violations of these assumptions affect results.

From a purely statistical perspective, effects of genetic and demographic non-uniformity and stochasticity on estimates of genetic structure and model-based parameters include scale dependence of F_{ST} (e.g. Refs 14,32), fluctuation of F_{ST} over space and time (e.g. Ref. 15), wide variance of F_{ST} values among loci (e.g. Ref. 33), and a loose correlation between F_{ST} and Nm (Ref. 15). Thus, estimates of Nm will have large standard errors. 'Snapshot' estimates assessed at one (or a few) points in space and time will be unlikely to reveal actual levels of interpopulation genetic exchange. Note that population genetic estimates based on F-statistics inherently have large variances[29], irrespective of variance attributable to genetic and demographic factors.

More importantly, demographic instability determines, in part, the evolution of genetic variation among populations[30,34,35,36]. Fluctuating or novel conditions of genetic variance that occur as a consequence may be integral to population diversification and speciation, and temporarily provide conditions conducive to kin selection or shifting balance evolution[15,30,33,35]. Conventional indirect approaches implicitly assume that gene frequency distributions among populations are caused by gene flow (e.g. Ref. 29). Contributions of other evolutionary forces to genetic structure are not distinguished and estimates obtained are interpreted as weighted averages of effective 'gene flow' over time. Such cumulative weighted averages however may provide few insights into the issues we would most like to address as evolutionary biologists, that is, the microevolutionary processes that promote or hinder genetic differentiation and diversification of populations.

As Strong[37] and others have previously emphasized '...mathematical theory of equilibrium and stability has proven disappointingly sterile for ecology'. Dispersal and rates of gene flow will likely be unique

characteristics of different populations and most appropriately viewed as a dynamic, interconnected network of differential values. Moreover, to the extent that individuals differ genetically in their propensity to disperse and wander[38], attributes of different habitat patches will selectively act on this underlying variation such that rates of gene flow will 'evolve' like any other quantitative trait. Just as the evolution and spread of adaptive traits can generally not be assessed from studies of population means, but rather require detailed analysis of (co)variance structure, the importance of gene flow as a determinant of genetic pattern across environmental landscapes will be difficult to evaluate from cumulative weighted averages.

Future studies

After more than 25 years of using molecular methods to estimate genetic structure and gene flow, enough data have accumulated to assess where the markers have succeeded and where they have not. Their most successful use is in phylogeography and species level analyses[1,39]. Allozymes are generally sufficient for detecting isolation caused by vicariant events, but in some groups higher resolution DNA markers are required. Some of the more variable markers such as microsatellites may be excellent tools for establishing kinship relationships among related individuals (e.g. Ref. 16) and even allozymes have proven useful in this respect (e.g. Ref. 40). There appear to be many fewer successes in areas between these two extremes, that is, evaluating levels of gene flow among interbreeding populations and in population demographic studies. Here, efforts are undermined by the many sources of error and ambiguity associated with indirect approaches (Box 3).

Despite shortcomings summarized here, genetic approaches are a useful starting point for evolutionary studies when results are properly analysed statistically and interpreted critically, and when the question of interest is an appropriate application of their use. In this regard, we offer four specific recommendations for consideration.

Analyses of population structure rely almost exclusively on antiquated methodology. From the beginning of 1996 to mid 1997, 21 articles published in the journal *Evolution* used indirect methods to measure population structure. Of these, almost all based interpretation on F-statistics analysis (18 of 21 studies) and allozyme markers (15 of 21 relied solely on allozymes). Although most population geneticists acknowledge that Nm imparts minimal information, this value is still widely reported (14 of 21 studies) and often literally interpreted. Theoretical population genetics methodology has sufficiently progressed that we should be hesitant to base measures of genetic structure on F-statistics analysis, especially of allozyme variation, and simple genetic models. More realistic models and analytical methods are available and should be evaluated for their precision and statistical power (e.g. Refs 41,42). Many of these new methods rely upon DNA data and genealogical analysis of haplotypes or alleles and have increased capacity for distinguishing genetic/spatial associations including vicariance, isolation-by-distance, and range expansion.

Second, variance among single locus estimates of genetic structure is common. This suggests that either uninformative loci have been included in the analysis (i.e. those exhibiting low levels of polymorphism and thus unlikely to reveal genetic structure) or that basic assumptions of genetic equilibrium or marker neutrality or both have been violated[7,32]. Our current approach of combining data across discrepant loci introduces error in the estimation procedure. More importantly, when locus-to-locus variance cannot be explained by disparate levels of allelic diversity, estimates obtained will be biologically misleading. Such estimates confound gene flow with other evolutionary determinants of genetic variance among populations and thus do not take full advantage of the information we have available. If our goal is to distinguish relative effects of different evolutionary forces in population diversification, discrepant loci should be emphasized for their power to reveal underlying mechanisms. Compatibility assessments across loci would be more informative[42] and locus-specific estimates should routinely be reported and interpreted in the context of allelic diversity, model assumptions, locus-specific evolutionary forces, and observed patterns of genetic structure.

Our third suggestion concerns analysis of isolation-by-distance. Such studies generally employ regression analysis of data derived from estimates of genetic distance (or gene flow) among pairwise population contrasts. Effects of single populations or clusters of populations are rarely

distinguished. Statistical significance tests of each pairwise value should routinely be conducted to assess whether certain populations are largely responsible for isolation-by-distance correlations. If statistically significant values are associated with a single population or cluster of populations, vicariance should be discussed as a viable explanation. If significance is scattered throughout population pairs, a stronger case can be made for isolation-by-distance.

Our final recommendation concerns direct studies of movement. Even our most enthusiastic supporters of genetic approaches have long advocated the key role such studies play[18]. Yet data from direct observational studies have accumulated at a much slower rate than data from indirect studies. Admittedly, direct studies of movement are logistically difficult and suffer from their own set of limitations[18]. The most severe of which is that such studies may tell us little about the importance of gene flow over evolutionary time. However, direct observational studies of movement can provide data on the degree of spatial and temporal heterogeneity of movement patterns across a species' range, the impact of environmental attributes such as habitat patchiness and resource quality on movement patterns, and the extent of quantitative genetic variation for propensity to stay or wander. Moreover, direct assessment of movement among contemporary populations remains the only valid approach to the study and interpretation of gene flow in an ecological context. Technological advances will undoubtedly facilitate such studies in the future. The use of uniquely marked or tagged individuals tracked by satellite or radar will allow unbiased assessments of direct movement over short and long distances for organisms amenable to such tags (e.g. Ref. 2). Indeed, the use of genetic tags to track the movement of marine organisms has proven to be a successful arena for applying genetic approaches to the study of movement among present-day populations[43,44].

In summary, our understanding of gene flow will progress more significantly through candid discussion of limitations, more refined statistical analysis, and a better balance between indirect and direct approaches. Our reliance on easy to apply, conventional indirect methods to study population structure and gene flow is no longer advancing the field. Technological advances in DNA analysis over the past decade have brought about a new era of data gathering that has potential for enhanced resolution. The challenge now is to develop and implement analytical methods that keep pace with technology and allow more precise estimates of gene flow in contemporary time.

References

1 Avise, J.C. (1994) *Molecular Markers, Natural History and Evolution*, Chapman & Hall
2 Koenig, W.D., van Vuren, D. and Hooge, P.N. (1997) **Detectability, philopatry and the distribution of dispersal distances in vertebrates**, *Trends Ecol. Evol.* 11, 514–517
3 Lewontin, R.C. (1985) **Population genetics**, in *Evolution: Essays in Honour of John Maynard Smith* (Greenwood, P.J., Harvey, P.H. and Slatkin, M., eds), pp. 3–18, Cambridge University Press
4 Roderick, G.K. (1996) **Geographic structure of insect populations: gene flow, phylogeography and their uses**, *Annu. Rev. Entomol.* 41, 325–352
5 Larson, A., Wake, D.B. and Yanev, K.P. (1984) **Measuring gene flow among populations having high levels of genetic fragmentation**, *Genetics* 106, 293–308
6 Merila, J., Bjorklund, M. and Baker, A.J. (1997) **Historical demography and present day population structure of the greenfinch, *Carduelis chloris* – an analysis of mtDNA control-region sequences**, *Evolution* 51, 946–956
7 Wright, S. (1951) **The genetical structure of populations**, *Ann. Eugen.* 15, 323–354
8 Nei, M. (1973) **Analysis of gene diversity in subdivided populations**, *Proc. Natl. Acad. Sci. U. S. A.* 70, 3321–3323
9 Weir, B.S. and Cockerham, C.C. (1984) **Estimating *F*-statistics for the analysis of population structure**, *Evolution* 38, 1358–1370
10 Slatkin, M. (1995) **A measure of population subdivision based on microsatellite allele frequencies**, *Genetics* 139, 457–462
11 Zink, R.M. (1986) **Patterns and evolutionary significance of geographic variation in the Schistacea group of the fox sparrow (*Passerella iliaca*)**, *Ornithol. Monogr.* 40
12 Zink, R.M. (1994) **The geography of mitochondrial DNA variation, population structure, hybridization and species limits in the Fox Sparrow (*Passerella iliaca*)**, *Evolution* 48, 96–111
13 Slatkin, M. and Barton, N. (1989) **A comparison of three methods for estimating average levels of gene flow**, *Evolution* 43, 1349–1368
14 Husband, B.C. and Barrett, S.C.H. (1995) **Estimates of gene flow in *Eichhornia paniculata* (Pontederiaceae): effects of range substructure**, *Heredity* 75, 549–560
15 Whitlock, M.C. (1992) **Temporal fluctuations in demographic parameters and the genetic variance among populations**, *Evolution* 46, 608–615

16 Peterson, M.A. (1996) **Long-distance gene flow in the sedentary butterfly, *Euphilotes enoptes* (Lepidoptera: Lycaenidae)**, *Evolution* 50, 1990–1999
17 Hellberg, M.E. (1996) **Dependence of gene flow on geographic distance in two solitary corals with different larval dispersal capabilities**, *Evolution* 50, 1167–1175
18 Slatkin, M. (1985) **Gene flow in natural populations**, *Annu. Rev. Ecol. Syst.* 16, 393–430
19 Mitton, J. (1994) **Molecular approaches to population biology**, *Annu. Rev. Ecol. Syst.* 25, 45–69
20 Queller, D.C., Strassman, J. and Hughes, C.R. (1993) **Microsatellites and kinship**, *Trends Ecol. Evol.* 8, 285–288
21 Lehmann, T. *et al.* (1996) **Genetic differentiation of *Anopheles gambiae* populations from East and West Africa; comparison of microsatellite and allozyme loci**, *Heredity* 77, 192–208
22 Karl, S.A. and Avise, J.C. (1992) **Balancing selection at allozyme loci in oysters: implication from nuclear RFLP's**, *Science* 256, 100–102
23 Raybould, A.F., Mogg, R.J. and Clarke, R.T. (1996) **The genetic structure of *Beta vulgaris* ssp. *Armitima* (sea beet) populations: RFLP's and isozymes show different patterns of gene flow**, *Heredity* 77, 245–250
24 Sánchez, J.A. *et al.* (1996) **Protein and microsatellite single locus variability in *Salmo salar* L. (Atlantic salmon)**, *Heredity* 77, 423–432
25 Pogson, G.H., Mesa, K.A. and Boutilier, R.G. (1995) **Genetic population structure and gene flow in the Atlantic Cod *Gadus morhus*: a comparison of allozyme and nuclear RFLP loci**, *Genetics* 139, 375–387
26 Tibbets, C.A. and Dowling, T.E. (1996) **Effects of intrinsic and extrinsic factors on population fragmentation in three species of North American minnows (Teleostei: Cyprinidae)**, *Evolution* 50, 1280–1292
27 Haglund, T.R., Buth, D.G. and Lawson, R. (1992) **Allozyme variation and phylogenetic relationships of Asian, North American, and European populations of the threespine stickleback, *Gasterosteus aculeatus*, *Copeia* 1992, 432–443
28 Orti, G. *et al.* (1994) **Global survey of mitochondrial DNA sequences in Threespine Stickleback: evidence for recent migrations**, *Evolution* 48, 608–622
29 Cockerham, C.C. and Weir, B.S. (1993) **Estimation of gene flow from *F*-statistics**, *Evolution* 47, 855–863
30 Slatkin, M. (1989) **Population structure and evolutionary progress**, *Genome* 31, 196–202
31 Andrewartha, H.G. and Birch, L.C. (1954) *The Distribution and Abundance of Animals*, University of Chicago Press
32 Johannsen, J., Veith, M. and Seitz, A. (1996) **Structure of the butterfly *Melitaea didyma* (Nymphalidae) along a northern distribution range border**, *Mol. Ecol.* 5, 259–267
33 Levin, D.A. (1988) **Consequences of stochastic elements in plant migration**, *Am. Nat.* 132, 643–651
34 Wright, S. (1948) **On the roles of directed and random changes in gene frequency in the genetics of populations**, *Evolution* 2, 279–294

35 Endler, J. (1977) *Geographical Variation, Speciation and Clines*, Princeton University Press

36 Loveless, M.D. and Hamrick, J.L. (1984) **Ecological determinants of genetic structure in plant populations,** *Annu. Rev. Ecol. Syst.* 15, 65–95

37 Strong, D.R. (1986) **Density-vague population change,** *Trends Ecol. Evol.* 1, 39–42

38 Swingland, I.R. (1983) **Intraspecific differences in movement,** in *The Ecology of*

Animal Movement (Swingland, I.R. and Greenwood, P.J., eds), pp. 102–115, Clarendon Press

39 Berlocher, S.H. (1984) **Insect molecular systematics,** *Annu. Rev. Entomol.* 29, 403–434

40 Friesen, V.L. *et al.* (1996) **Molecular evidence for kin groups in the absence of large-scale genetic differentiation in a migratory bird,** *Evolution* 50, 924–930

41 Neigel, J.E. (1997) **A comparison of**

alternative strategies for estimating gene flow from genetic markers, *Annu. Rev. Ecol. Syst.* 28, 105–128

42 Templeton, A.R. (1997) **Nested clade analyses of phylogeographic data: testing hypotheses about gene flow and population history,** *Mol. Ecol.* (in press)

43 Bowen, B.W. (1995) **Tracking marine turtles with genetic markers,** *Bioscience* 45, 528–534

44 Palsbøll, P.J. *et al.* (1997) **Genetic tagging of humpback whales,** *Nature* 388, 767–769

'Lamarckian' mechanisms in darwinian evolution

Eva Jablonka

Marion J. Lamb

Eytan Avital

Since the Modern Synthesis, evolutionary biologists have assumed that the genetic system is the sole provider of heritable variation, and that the generation of heritable variation is largely independent of environmental changes. However, adaptive mutation, epigenetic inheritance, behavioural inheritance through social learning, and language-based information transmission have properties that allow the inheritance of induced or learnt characters. The role of induced heritable variation in evolution therefore needs to be reconsidered, and the evolution of the systems that produce induced variation needs to be studied.

Eva Jablonka and Marion Lamb are at the Cohn Institute for the History and Philosophy of Science and Ideas, Tel-Aviv University, Tel-Aviv 69978, Israel; Eytan Avital is at the Dept of Natural Sciences, David Yelin Teachers College, POB 3578, Jerusalem, Israel; Eva Jablonka is also at the Wissenschaftskolleg zu Berlin, Wallotstrasse 19, D-14193 Berlin, Germany (jablonka@wiko-berlin.de).

'**L**amarckism' and 'darwinism' are traditionally seen as alternative theories trying to account for evolutionary change (Box 1). The verdict of history is that Lamarck got it wrong–evolutionary change does not occur through the inheritance of acquired characters. Acquired characters are the outcome of *instructive processes*, such as those seen in embryonic induction, transcriptional regulation, and learning, all of which involve highly specific and usually adaptive responses to factors external to the responding system. The inheritance of the outcomes of instructive processes is deemed to be impossible. Adaptive evolutionary change is assumed to be based on darwinian (or more accurately neo-darwinian) evolution in which guidance comes exclusively from selective processes. The production and nature of heritable variation is assumed to be uninformed by the environment or by previous history. The future is open-ended, determined solely by the contingencies of life. It is neither foretold nor intimated.

General selection theory makes no assumptions about the origin of heritable variation. It maintains that evolution by natural selection will occur in any system with entities manifesting the properties of multiplication, heredity and heritable variation affecting reproductive success[1]. In the current version of biological darwinism, it is assumed that information is digital and encoded in DNA base sequences, that multiplication of information occurs through DNA replication, and that variation, which is generated by mutation and recombination, is random with respect to the selecting environment and the developmental history of the organism and the lineage. However, this version of evolution – 'genic neodarwinism' – is incomplete: it gives natural selection an exclusive deterministic role in the evolution of all conceivable complex adaptations, but until recently it has had rather little to say about the evolution of new systems for acquiring, storing and transferring information, and even less about the evolutionary effects

of such systems once they are in place. Natural selection leads not only to the evolution of eyes, wings, and sonars, but also to the evolution of new evolutionary rules. Many of these rules undermine the assumption that variation is random. Mechanisms allowing the inheritance of acquired characters have evolved several times during the history of life, and understanding their evolution is crucial to understanding the transitions to new levels of individuality[2-4].

Evolved 'lamarckian' heredity systems

The heredity systems that we consider here are all complex mechanisms for the acquisition, storage and transfer of information. All evolved through natural selection, but they differ from each other in the type of information they transmit, in their evolutionary history, and in their evolutionary effects. They include adaptive mutational systems involving non-random changes in DNA, cellular heredity systems in which information is acquired and transmitted through intracellular structures and biochemical mechanisms, the transfer of patterns of behaviour through social learning coupled with certain types of social organization, and the transmission of information using symbolic languages. All of these systems allow certain outcomes of the interaction between the organism and its environment to be incorporated into and maintained within the information-carrying system, and the information to be transmitted to future generations. All therefore allow the inheritance of acquired or learnt characters.

Adaptive mutational systems: the intelligent genome

In the genetic inheritance system, information for making RNA and proteins is stored in DNA base sequences; an elaborate enzyme system enables this information to be replicated and transmitted to the next generation. Physico-chemical damage to the DNA and errors occurring during its replication can be removed by a battery of repair processes. Errors that remain, and sequence changes that are

206 *TREE vol. 13, no. 5 May 1998*