EMBO reports

analysis

A meeting at the gene

Biodiversity and natural history

Modern biology has many triumphs to celebrate, but a generally applicable species definition is not among them. This is no small debt, because classification of the differentness of organisms is a basic and undeniable human need inherited from our non-human ancestors. If for no other purpose than simple communication, we need words to describe the levels of differentness among the organisms that we observe, from the highest Linnéan ranks down to the things that taxonomists call species. But biologists have found it impossible to agree upon where, exactly, to position the upper and lower boundaries of species in the spectrum of variation observable for any given group of organisms. It is perhaps biology's most grotesque concession that 140 years after the publication of *The Origin of Species* (Darwin, 1859), we still do not know exactly what those things are whose origin the theory of evolution explains. We know that species reside somewhere within a continuum of genetic diversity extending from the individual to the kingdom, but we are wholly unable to pinpoint them. This circumstance has a far-reaching impact upon efforts to quantify biological diversity, because if species as units of diversity are an outdated concept (Bachmann, 1998), we are faced with a serious problem: how to measure biodiversity?

To illustrate the matter, consider the following estimates provided in a recent review: 14 million total contemporary species distributed across some 1–6 billion global populations with extinction rates surmised to encompass at least 27 000 species per year (Purvis and Hector, 2000). These values, like all current measures of biodiversity, involve in one way or another the concept of species numbers or richness. But how can we rely upon them when the unit of count is an undefined quantity? Conservationists are fully aware of the seemingly insurmountable difficulties posed by defining species and their boundaries in any universally applicable manner, but they have no alternative to the species concept when it comes to measuring biological diversity in ecosystems. Even worse, the problem of species definition becomes increasingly severe as one moves outside the realm of plants and animals into the world of single-celled eukaryotes—let alone the prokaryotes, who exchange chunks of their genomes with sufficient ease as to preclude any biologically meaningful species definition (Doolittle, 1999). We also know that to pinpoint species boundaries in any scale of morphological or genetic variation is purposeless, because it requires unattainable knowledge about the future fate of those boundaries as time and natural history progress. For these reasons, biologists direly need measures of biological diversity that are independent of the semantic strictures inherent to the species definition.

> Nucleotide substitution is quantifiable. It is the atom of genetics and a quantum of evolution

The concepts of 'biodiversity' and 'species' are thus related in that the words rarely evoke the same meaning when read by different people. Furthermore, the meanings of both are contingent upon our understanding of evolutionary process. Indeed, the fundamental issues that are germane to understanding biodiversity are the subjects of evolutionary biology: the differentness of organisms, their interactions in the environment, their success or failure in the face of ever-changing natural and man-made selection, the plenitude of their forms as well as the underlying history of that plenitude, and the measures that should be implemented to ensure the *preservation of* endangered (or *favoured*) *races in the struggle for life*—to paraphrase Darwin's subtitle. And how can we fully understand the diversity of life other than within an evolutionary context?

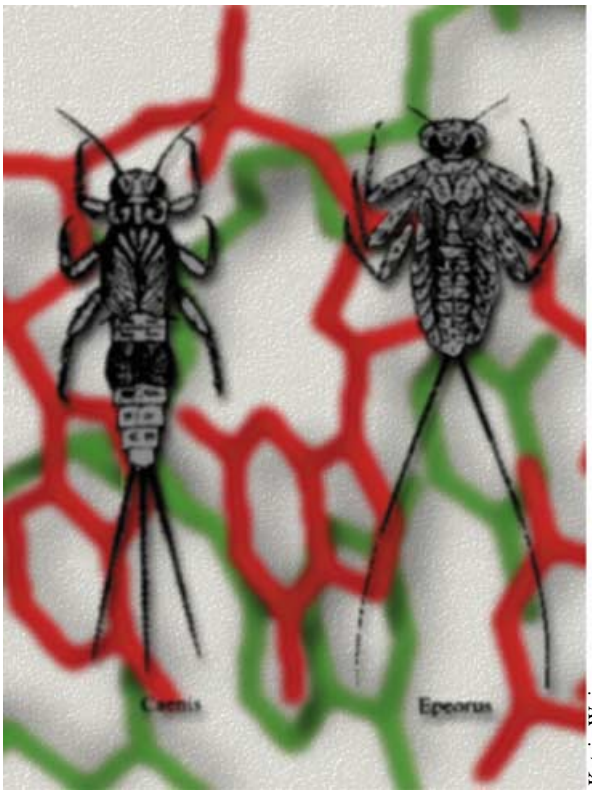
Our knowledge of evolution is imperfect, but it has been immeasurably enriched by gene sequencing technology (Nei, 1987; Avise, 1994; Graur and Li, 2000). Similarly, new technologies will enrich the study of biodiversity, in particular the study of its dimensions. Intuitively, biological diversity can be measured in terms of differences of DNA sequence. Nucleotide substitution is quantifiable. It is the atom of genetics and a quantum of evolution. But in the long term, gene sequence comparisons of one or a few loci from nuclear and organelle genomes, as is current practice, will not suffice. This is particularly true at the transitional territory between populations and species, where nuclear alleles tend to be older than the speciation process itself. Each gene thus gives a different tree of alleles and only the sum of all these allele-trees comes close to explaining the underlying historical separation of the emerging biological entities, be they species, populations or individuals. Clearly, we are in need of genome-wide measures of diversity that reveal the dynamic processes of natural populations and that provide us with estimates of nucleotide substitutions as a measure of genetic distinctness.

Of course, one could sequence the genome of every individual in a given ecosystem and examine the DNA and chromosomal polymorphisms so inferred. Such information would harbour roughly everything that we can observe about the genetic basis of intra- and infra-specific

biological diversity. Relating these differences to allele frequencies, population structures, biogeography, climate, habitat history, human influence, mutation rates, substitution rates, etc., we would obtain—under the premise that we marshal sufficient computational tools—the grandest possible view of the biological past preserved within contemporary genomes, right down to the sibling level. As a side product, by relating DNA diversity to the morphological differences that scientists familiar with their flora and fauna tend to call species, we would understand more fully what species among various groups of organisms are in terms of their genetic distinctness. This alone would be a substantial advance for biology.

But as our *Gedankenexperiment* cannot be carried out, the question emerges of whether the results that it would provide are altogether unattainable. Obviously, such a vista of evolution at work for every individual cannot be obtained. But an estimate of its contours for a given ecosystem can, in principle, be inferred by sampling sufficient individuals of sufficient taxa and studying them at representative numbers of chromosomal loci. This would be an ambitious but nevertheless tractable long-term undertaking with available technologies, employing *discontinuous* markers, which provide a measure of genetic difference dispersed across the genome.

Discontinuous markers currently encompass DNA fragment length polymorphisms (RFLPs, RAPDs, AFLPs) and nucleotide changes (SNPs, single nucleotide polymorphisms) (Schafer and Hawkins, 1998). Traditionally the tools of geneticists and breeders, these markers are diffusing into studies of natural diversity, and rightly so. Breeders have used them to assess the origins of cultivated crops, for example einkorn wheat (Heun *et al.*, 1997) and barley (Badr *et al.*, 2000). Genome-wide assessment of polymorphisms among cultivated plants compared with their wild relatives has revealed the history of these ancient cultivated crops, hence depicting the earliest migration of agriculture and a glimpse of mankind's first tinkering with biodiversity. Furthermore, such studies led to the identification of those wild populations from which the cultivars were initially selected (Heun *et al.*, 1997), and uncovered valuable reserves of genetic resources. But perhaps more importantly, the same discontinuous markers revealed the natural populations that were *not* the source of initial cultivation. These are even more valuable reserves of genetic diversity, because they contain hitherto untapped alleles that were not present in the small sample that humans initially selected for breeding. Finally, such markers can help to sort out the natural groupings even within seemingly intractable genera such as *Hordeum* (F. Salamini, in preparation), a result that previously seemed impossible.



Katrin Weigmann

The most promising discontinuous markers for biodiversity studies seem to be SNPs, which reveal the state of a nucleotide position somewhere in the genome and thus yield results similar to gene sequencing data. Using SNPs, sequence differences can be correlated to linkage relationships among polymorphic loci, while also summarizing overall DNA sequence divergence between genomes. Moreover, a group of tightly linked SNPs provides a basis to construct robust haplotypes that can be associated with morphological variation not yet attributable to single genes (Schafer and Hawkins, 1998). In addition, SNPs provide measures of nucleotide diversity (π), an important estimator of overall DNA polymorphism (Nei, 1987; Graur and Li, 2000).

Several platform technologies are available to detect wild-type and variant SNP alleles (McCarthy and Hilfiker, 2000): hybridization on filters, DNA chips, conformational polymorphisms revealed in gels, and primer extension. Large-scale SNP analysis with the help of automated, quick and inexpensive procedures can be envisaged to detect random nucleotide variation or to test the genome in orthologous positions of orthologous genes across many individuals and taxa. In the long term, this should lead to empirical measures of natural diversity with a firm foundation in genome data. Developing these procedures will not be simple. But if sufficiently distinct groups of SNP-based haplotypes can be detected by hybridization on microarrays or by other multiplex methods, this will be a major step towards understanding the continuum of natural variation.

Importantly, estimates of the genetic richness of ecosystems, in addition to species richness, would be the immediate result of such an undertaking. Furthermore, the former could be used to help quantify the latter in more meaningful terms than was previously possible. Many ecosystems that are apparently species-poor will be revealed to harbour particular genetic richness. Both for domesticated organisms, which man has learned to use for his immediate survival, and for those organisms in the wild that constitute the majority of biological diversity, knowledge of the natural genetic resources that support their survival will be the best possible compass guiding efforts to conserve them. After all, if the last line of defence against extinction is a targeted breeding programme, we had better know what we are doing when it comes to making the crosses.

Measures of genetic distinctness within and between species hold the key to understanding how nature has generated and preserved biological diversity. Striking examples of its utility can be found in studies of island colonization (Böhle *et al.*, 1996) or in studies of

After all, if the last line of defence against extinction is a targeted breeding programme, we had better know what we are doing when it comes to making the crosses

post-glacial recolonization in European and other habitats, a process that emanated from various southern refugia and that often involved crosses among conspecific populations in hybrid zones (Hewitt, 2000). But if there are no field biologists who know their flora and fauna, geneticists will neither have material to work on, nor will they know the biology of the organisms they are studying. In this sense, there is a natural predisposition towards a symbiosis between genetics and biodiversity—a union that current progress in DNA technology is forging.

Avise, J.C. (1994) *Molecular Markers, Natural History and Evolution*. Chapman and Hall, New York, NY.

Bachmann, K. (1998) Species as units of diversity: an outdated concept. *Theory Biosci.*, **117**, 213–230.

Badr, A., Muller, K., Schafer-Pregl, R., El Rabey, H., Effgen, S., Ibrahim, H.H., Pozzi, C., Rohde, W. and Salamini, F. (2000) On the origin and domestication history of barley (*Hordeum vulgare*). *Mol. Biol. Evol.*, **17**, 499–510.

Böhle, U.-R., Hilger, H.H. and Martin, W. (1996) Island colonisation and evolution of the insular woody habit in *Echium* L. (Boraginaceae). *Proc. Natl Acad. Sci. USA*, **93**, 11740–11745.

Darwin, C. (1859) *The Origin of Species by Means of Natural Selection or the Preservation of Favoured Races in the Struggle for Life*. Murray, London, UK.

Doolittle, W.F. (1999) Phylogenetic classification and the universal tree. *Science*, **284**, 2124–2128.

Graur, D. and Li, W.-H. (2000) *Fundamentals of Molecular Evolution*. Sinauer, Sunderland, UK.

Heun, M., Schafer-Pregl, R., Klawan, D., Castagna, R., Accerbi, M., Borghi, B. and Salamini, F. (1997) Site of einkorn wheat domestication identified by DNA fingerprinting. *Science*, **278**, 1312–1314.

Hewitt, G. (2000) The genetic legacy of the Quaternary ice ages. *Nature*, **405**, 907–913.

McCarthy, J. and Hilfiker, R. (2000) The use of single-nucleotide polymorphism maps in pharmacogenomics. *Nature Biotechnol.*, **18**, 505–508.

Nei, M. (1987) *Molecular Evolutionary Genetics*. Columbia University Press, New York, NY.

Purvis, A. and Hector, A. (2000) Getting the measure of biodiversity. *Nature*, **405**, 212–219.

Schafer, A.J. and Hawkins, J.R. (1998) DNA variation and the future of human genetics. *Nature Biotechnol.*, **16**, 33–39.

William Martin[1] and Francesco Salamini[2]

The authors are at [1]Institut für Botanik III at the Heinrich-Heine Universität Düsseldorf and [2]Max-Planck Institut für Züchtungsforschung in Cologne.
E-mail: w.martin@uni-duesseldorf.de

DOI: 10.1093/embo-reports/kvd055

Would you buy a tomato from this man?

How to overcome public mistrust in scientific advances

What is the difference between a genetically modified (GM) tomato and a mobile phone? One is a triumph of science and technology that benefits millions; the other is a GM tomato—an object of fear and suspicion. The popular press may tell you that mobile phones fry your brain, but tell that to an upwardly mobile city type who simply cannot live without this device. Then tell the same person that GM tomatoes taste better and last longer, and the reply will probably be 'so what?' GM foods have struck fear and distrust into the public without providing sufficiently attractive benefits. It seems that no amount of 'educating' the public will overcome the resistance to GM food; rather, the public wishes to be involved in a dialogue on new technologies that affect them. The GM debate is but the tip of an iceberg, whose treacherous depths threaten many areas of the life sciences if not explored. It has unambiguously sent the message that scientists must become more engaged with the public in order to demystify their research.

The GM debate is but the tip of an iceberg, whose treacherous depths threaten many areas of the life sciences if not explored

After a long and arduous debate, the Swiss finally agreed, in 1998, to support research involving the genetic modification of organisms, hence averting a decision that would doubtless have had pan-European consequences. In the meantime, public trust in GenSuisse, the Internet-based public information provider established in response to the threatened ban, has fallen. The reason is simple: GenSuisse was founded by the Basel pharmaceutical industries. Message one from the Swiss experience: the public is highly sceptical of anything that smells of industry—the Monsanto factor. Message two: the public wants a continuous dialogue, not merely information in times of need. Given that the EC aspires to co-ordinate and fund science on a European level, it would be fitting that it also concerns itself with a European strategy for public outreach. At present, the most efficient means of pan-European communication is the press, the irony being that it is also the least likely to communicate a balanced story, and most likely to serve its own interests.