

Secret life of genes

Günter Theißen

'Homology' is one of the most important terms in biology. Features are homologous if they share a common evolutionary origin — for example, bat wings and bird wings are homologous as tetrapod forelimbs, but they are not homologous as organs of flight. The definition of homology has changed over time and, although still unresolved, functional similarity has never been part of it (bat wings and bird wings certainly have similar functions).

When homology is applied to genes, at least two fundamentally different subclasses must be distinguished: paralogy, the relationship between genes that have originated by gene duplication; and orthology, which refers to genes that originated by speciation. Phylogenetic reconstructions of organisms created using information from the nucleotide sequences of genes require orthologous, rather than paralogous, genes, so the distinction between these two gene classes is important for practical reasons. More fascinating is the observation that orthologues and paralogues usually have

very different evolutionary fates. Orthologues often take over the function of the precursor gene in the species of origin and thus tend to be conserved. In contrast, young paralogues have redundant functions, which is an evolutionarily unstable situation. Thus, in the long run — with a few exceptions — paralogues either diverge functionally, or all but one of the versions is lost.

Although the theoretical definition of orthology seems straightforward, the term is misused or misunderstood in many ways, particularly in developmental biology and genomics. An example is the romantic, but incorrect, idea that every gene in a species' genome has exactly one orthologue in the genome of another species. Genes can be lost during evolution, meaning that a given gene may or may not have surviving orthologues in other species. Moreover, gene duplications can follow a speciation event, generating orthologous 'clades' of paralogues.

Orthology between individual genes does not therefore exist; rather, one-to-many or many-to-many orthologous relationships are formed. The situation is further complicated by the fact that gene duplication, gene loss and speciation can be frequent events in the history of a group of organisms. Thus, complex gene relationships are established which cannot be described in simple terms.

A related error is the naive assumption that a gene with the greatest sequence similarity to one from another genome is 'the' orthologue of that gene. But in view of gene loss and other events, the most similar genes in two species' genomes may be paralogues, or not homologous at all. Thus, even a comparative analysis of entire genome sequences does not guarantee the identification of orthologues.

In my view, the most serious misconception is to confuse orthology with functional similarity, or to use functional similarity as the decisive criterion by which to identify 'the' orthologue from a group of homologous genes. As a subclass of homology, orthology cannot be identified by functional similarity. Functional similarity may reflect orthology, but it could also be the result of convergent evolution. And it is clearly possible for orthologues to diverge functionally. Hence, orthology and functional equivalence must be tested independently, by phylogenetic reconstruction and comparative functional studies, respectively.

It follows that a reconstruction of gene phylogeny, and its superimposition on the

Orthology

The relationship between genes that originated through speciation is one of the most widely misunderstood — and misused — concepts of evolutionary biology.

clarified evolution of the taxa involved, is the only valid way in which true orthologues can be identified. As this is a laborious task, and the available data are often too limited to provide conclusive results, proven (rather than putative) orthologues are as rare in the literature as diamonds in bare rock. On the other hand, experimentally untested claims about orthology seem to be as numerous as grains of sand on a beach.

Of course, the definition of a term may change if this reflects an improved understanding — a good example is the definition of homology, which, unsurprisingly, has taken common ancestry to be the decisive criterion only since Darwin's time. What is bad about using the term 'orthology' to denote the relationship between homologous genes that are functionally equivalent in different species is that this habit undermines a better understanding of some of the most fascinating aspects of nature. Because the complexity and diversity of living beings are largely encoded in their genes, the mechanism and timing of how genes originate and evolve is of the utmost importance.

For a precise description of gene evolution, we need a controlled vocabulary that distinguishes strictly between gene genealogy and gene function. Only then can we properly describe — and eventually understand — the conservation and change of gene function during evolution, and the consequences of these processes for the phenotype. Outside evolutionary biology, 'orthology' is becoming a fashionable, yet fuzzy, buzzword. But terms are tools — care is required to ensure they are not blunted by over-frequent use. ■

Günter Theißen is at the Genetics Lehrstuhl, Friedrich Schiller University, Philosophenweg 12, D-07743 Jena, Germany.

FURTHER READING

- Fitch, W. M. *Syst. Zool.* **19**, 99–113 (1970).
- Fitch, W. M. *Trends Genet.* **16**, 227–231 (2000).
- Graur, D. & Li, W.-H. *Fundamentals of Molecular Evolution* (Sinauer, Sunderland, Massachusetts, 2000).
- Tatusov, R. L., Koonin, E. V. & Lipman, D. J. *Science* **278**, 631–637 (1997).

CAGC
ACCA
CAGC
ACC
CA
AC



AGCG
ACCAU
GCG
CAU
GCG
CAU

