NEWS AND COMMENTARY

Genome Evolution

Gene history repeats itself

V Pereira

Heredity (2004) 93, 3–4, advance online publication, 14 April 2004; doi:10.1038/sj.hdy.6800457

Over 3 billion years of evolution have produced the remarkable diversity of form and complexity of life on Earth. Two central questions in evolution are how much of life's diversity is the result of adaptation, and how does the process of adaptation occur. At the molecular level, adaptive evolution has been particularly difficult to study; yet, examples of genes that evolved new functions via positive Darwinian selection can provide insight into molecular adaptation. What if we could run the course of evolution again, reset a gene to its ancestral state and submit it to the same selective pressure a second time? In a recent paper, Francisco Rodríguez-Trelles, Rosa Tarrío and Francisco Ayala report on a 'naturally replicated experiment': two duplication events of the same ancestral gene, a billion years apart, that resulted in the duplicates evolving the same novel function via positive selection (Rodríguez-Trelles *et al.*, 2003).

Duplication of segments of (or whole) DNA molecules has long been thought to be the main source of genetic novelty (Ohno, 1970). New gene copies (paralogues) are created which, if multiplicity leads to a degree of redundancy, experience relaxed selective constraint and may evolve new functions (neofunctionalization). Another possible fate for duplicated genes is subfunctionalization, when each duplicate specializes in part of the original function(s). But most duplicate genes rapidly become pseudogenes, since null mutations are far more common than advantageous ones (non-functionalization). Nevertheless, gene duplication can still be a major source of functional novelty, as it occurs at a very high rate. In eukaryotes, this rate has been estimated to average at one duplication (and fixation) event per gene per 100 million years (Lynch and Conery, 2000), which is comparable to the rate of point substitutions per nucleotide site.

Xanthine dehydrogenase (*Xdh*) and aldehyde oxidase (*Ao*) are genes that encode structurally complex enzymes (XDH and AOX, respectively) of the same family. *Xdh* is present in all cellular organisms and XDH is essential for the catabolism of purines. *Ao* is ubiquitous

among multicellular eukaryotes. (AOX is less well studied than XDH, but is known to catalyze the oxidation of aldehydes into acids.) XDH and AOX can be easily aligned along their whole sequences (>1330 codons), and it is likely that *Ao* evolved from *Xdh* before the origin of multicellularity. However, it is known that the primary sequences of mammalian AOXs are more similar to those of mammalian XDHs than to invertebrate AOXs. Rodríguez-Trelles, Tarrío and Ayala performed a combined phylogenetic analysis of both proteins, showing that in fact all vertebrate AOXs (which they labeled AOX') are more closely related to vertebrate XDHs than to invertebrate AOXs. (This points to the importance, for evolutionary studies, of protein functional assays, which established that AOX' and AOX perform the same function, distinct from that of XDH.) These evolutionary relationships are reinforced by the fact that the vertebrate *Ao* gene possesses 17 out of 28 intron positions in the *Xdh* gene of a sea squirt (a protochordate), but only two out of 27 intron positions in the sea

squirt *Ao*. Intron positions are highly conserved in vertebrate genes. So, the authors make a strong case for the evolution of *Ao* via recurrent neofunctionalization of two different paralogues of *Xdh*: the first time before the origin of multicellularity, and a second time between the origin of protochordates and the vertebrate radiation (see Figure 1).

Functional divergence following gene duplication does not necessarily occur as a result of positive selection. Alternatively, random fixation of neutral mutations could lead to a new function when the genetic background or environment changes. These two evolutionary regimes would be expected to lead to different values of the ratio (ω) of nonsynonymous (i.e. amino-acid altering) to synonymous nucleotide substitutions (per site) accumulating between the duplicates. Rodríguez-Trelles *et al.* used recent, maximum-likelihood methods capable of detecting lineage-specific adaptation at particular codons (Nielsen and Yang, 1998; Yang and Nielsen, 2002) to infer that both AOX and AOX' experienced periods of positive selection. If ω had been calculated from an average over the entire sequences and over the whole time since duplication, the signature of positive selection ($\omega > 1$) could have remained hidden. The power of these novel methods revealed that adaptive evolution of the *Xdh* paralogues occurred mainly before the diversification of multicellular eukaryotes (for AOX) and of the vertebrates (AOX'). Furthermore, the

Figure 1 Convergent adaptive evolution after ancient, recurrent gene duplications (Rodríguez-Trelles *et al.*, 2003). *Xdh* is a gene present in both prokaryotes and eukaryotes, while *Ao* is present in multicellular eukaryotes. Before the origin of multicellularity, *Xdh* was duplicated (marked by star, above) and one of the copies evolved a new function (*Ao*) by positive selection. A billion years later, *Xdh* was again duplicated (marked by star, below) in the lineage leading to vertebrates, and once again one copy (marked *Ao'*) evolved the *Ao* function by positive selection. The ancestral *Ao* gene was lost (marked by black dot) from the lineage leading to vertebrates before their diversification.

individual amino-acid substitutions responsible for the convergence of function were not all shared between the two proteins. Gene history repeated itself, but not exactly: AOX and AOX' are variant solutions to the same problem.

Other cases of molecular parallelism are known. For example, fish antifreeze proteins have evolved independently, many times, in response to freezing conditions in the Antarctic and Arctic Oceans, from C-type lectin genes and also from completely unrelated DNA (Logsdon and Doolittle, 1997). It is clear that neofunctionalization via positive Darwinian selection does occur, though it is less clear why there is selection for the (presumably) serial changes that take a protein from its ancestral to its current sequence and function. Many sites were identified as having experi-

enced positive selection in both AOX and AOX', which evolved their new function via different routes in sequence space. What were the roles of the intermediate proteins on the way? It is unknown when AOX was lost along the lineage leading to vertebrates, and therefore how much redundancy the new *Xdh* paralogue experienced. More generally, how often does positive selection determine the fate of a gene duplicate? Analyses of several eukaryotic (Lynch and Conery, 2000) and many prokaryotic (Kondrashov *et al.*, 2002) genomes suggest that, excluding the vast majority of duplicates that rapidly become pseudogenes, gene duplicates experience a period of relaxed constraint (with $\omega < 1$) as a rule, followed by strong purifying selection ($\omega \ll 1$). Thus, it seems that neofunctionalization via positive selection is the exception.

Nevertheless, this interpretation is based on averaged ω statistics of relatively low power for detecting adaptive molecular evolution. More general patterns may only emerge when more sophisticated analyses can be deployed in genomic surveys.

Vini Pereira is at Imperial College London, Silwood Park Campus, Buckhurst Road, Ascot, Berks SL5 7PY, UK.

e-mail: vini.pereira@imperial.ac.uk

- Kondrashov FA *et al.* (2002). *Genome Biol* **3**: research 0008.1–0008.9.
 Logsdon Jr JM, Doolittle WF (1997). *Proc Natl Acad Sci USA* **94**: 3485–3487.
 Lynch M, Conery JS (2000). *Science* **290**: 1151–1155.
 Nielsen R, Yang Z (1998). *Genetics* **148**: 929–936.
 Ohno S (1970). *Evolution by Gene Duplication*. Springer: Heidelberg.
 Rodríguez-Trelles F, Tarrío R, Ayala FJ (2003). *Proc Natl Acad Sci USA* **100**: 13413–13417.
 Yang Z, Nielsen R (2002). *Mol Biol Evol* **19**: 908–917.