Heredity (2004) 93, 405–406
© 2004 Nature Publishing Group All rights reserved 0018-067X/04 $30.00
www.nature.com/hdy

NEWS AND COMMENTARY

Evolutionary genetics

Transcriptome evolution – much ado about nothing?

F Rodríguez-Trelles

Heredity (2004) **93,** 405–406. doi:10.1038/sj.hdy.6800571
Published online 8 September 2004

Long-standing questions in evolutionary biology are being tackled with help from microarray technologies analyzing the transcriptome, namely the whole set of transcripts and their relative levels of expression in a cell or tissue type under defined conditions. In a recent paper, Khaitovich *et al* (2004) use these methods and conclude that the majority of the evolutionary changes in gene expression are of little or no adaptive significance, being ultimately determined by chance.

Khaitovich *et al* (2004) set off from a prediction of quantitative genetics according to which, if evolutionary changes in the levels of gene expression are caused by selectively neutral alleles, the amount of gene expression divergence among populations or species should be proportional to the time since their common ancestry (Lande, 1976; Lynch and Hill, 1986). To test this hypothesis, Khaitovich *et al* (2004) assayed differences in the levels of expression of around 12 000 genes in samples of prefrontal cortex from humans and three other primates, chimp, orang-utan, and macaque, using oligonucleotide microarrays – also known by the trademark Affymetrix GeneChip® – designed for the human genome. Plotting average amounts of divergence between transcriptomes against published molecular estimates of species age, they discovered an approximate linear rate of accumulation of expression differences over evolutionary time.

Reliability of microarray analyses hinges critically on the suitability of the array probe set. Suboptimal fit between the array probes and their target genes can cause crosshybridization artifacts, meaning that the same transcript is recognized by probes from different genes, which can yield severely distorted expression readouts. Crosshybridization artifacts are expected particularly in oligonucleotide arrays when comparisons involve distantly related species and/or genome sequences, which are insufficiently known. These factors impose a limit on the sample size that can be handled. Khaitovich *et al* (2004) sampled only four species (but see below), of which the two more distantly related to humans, that is, orang-utan and macaque, have as yet unascertained genome sequences. To alleviate these drawbacks, Khaitovich *et al* (2004) first applied the same analyses to published gene expression data for the liver; secondly, they repeated the experiment using cDNA microarrays, considered to be less prone to crosshybridization artifacts than oligonucleotide arrays, because they are based on greater probe lengths; and thirdly, extended their analyses to three mice species, in all cases reproducing the original result.

Yet, linear accumulation of expression differences alone does not conclusively rule out selection. Therefore, Khaitovich *et al* (2004) conducted an additional battery of tests. They thus found that the variation in gene expression between humans and chimps was positively correlated with the variation in gene expression within humans, with the association being of a similar magnitude to the one reported in other studies for random, mostly noncoding, genomic DNA sequences from the same species. Further, the rates of expression divergence between humans and chimps do not differ significantly between intact genes and expressed pseudogenes, which are expected to not be under the direct influence of selection since they do not produce any functional gene products. Besides these findings, Khaitovich *et al* (2004) encountered that gene expression differences between various tissues of the same individual – be it human, chimp, or mouse – accumulated proportionally to the tissue-divergence times, irrespective of their differences in function, which led them to propose the use of gene expression differences as a molecular clock to date the evolutionary history of tissues.

Taken conjointly, the tests conducted by Khaitovich *et al* (2004) indicate that the majority of gene expression differences within and between species are not functional adaptations, but selectively neutral or nearly neutral. This conclusion agrees with results of previous evolutionary transcriptomics studies conducted in fish (Oleksiak *et al*, 2002) and fruitfly (Rifkin *et al*, 2003) species. Also, they are consistent with those recently obtained by Yanai *et al* (2004) in a comparison of 1350 orthologous gene pairs from human and mouse. In particular, Yanai *et al* (2004) observed that expression of human genes changes from one tissue to another in a manner that cannot be anticipated from the corresponding expression changes of their orthologous genes in mouse, despite having likely retained the same function. They explain their findings assuming widespread occurrence of mutations causing ectopic expression, which does not affect fitness. It would be of great interest to test whether the levels of gene expression divergence between human and mouse reported by Yanai *et al* (2004) fit those that might be projected from the linear trends within primates and rodents observed by Khaitovich *et al* (2004).

Inspired by the apparent uncoupling between the rates of phenotypic and protein evolution in humans and chimps, King and Wilson (1975) proposed that the key to understanding the often disparate differences between species in their anatomy and way of life should be searched for not in gene sequences, but rather in the DNA regions that regulate the levels, locations, and timing of gene expression (reviewed in Rodríguez-Trelles *et al*, 2003). On this conceptual substrate grew the conviction that genome-wide characterizations of gene expression variation would make the understanding of species differences straightforward. Deeply buried in this conception is the notion that gene activity is synonymous with gene function. Thus, the results of Khaitovich *et al* (2004) and Yanai *et al* (2004) are revealing and important. Yet, they should not be surprising in light of the modularity, redundancy, and high rate of turnover lately uncovered to be intrinsic properties of the sequences influencing gene expression (reviewed in Carroll *et al*, 2001; Davidson, 2001; Wray *et al*, 2003). Complex information encoding makes these sequences refractory to '*in silico*' identification. Analogously, the discovery of widespread neutral expression implies that distinguishing functionally significant expression from neutral expression will ultimately require empirical validation. The extent to which noise in transcriptome evolution echoes at the

proteome level remains to be ascertained.

F Rodriguez-Trelles is at the Fundación Pública Galega de Medicina Xenómica, Hospital Cliníco Universitario, Universidad de Santiago de Compostela, Spain.

e-mail: ftrelles@usc.es

Carroll SB *et al* (2001). *From DNA to Diversity: Molecular Genetics and the Evolution of Animal Design*. Blackwell Science: Malden.
Davidson EH (2001). *Genomic Regulatory Systems: Development and Evolution*. Academic Press: San Diego.
Khaitovich P *et al* (2004). *PloS Biol* **2**: 682–689.
King MC, Wilson AC (1975). *Science* **188**: 107–116.
Lande R (1976). *Evolution* **30**: 314–334.
Lynch M, Hill WG (1986). *Evolution* **40**: 915–935.
Oleksiak MF *et al* (2002). *Nat Genet* **32**: 261–266.
Rifkin SA *et al* (2003). *Nat Genet* **33**: 138–144.
Rodríguez-Trelles F *et al* (2003). *Int J Dev Biol* **47**: 665–673.
Wray GA *et al* (2003). *Mol Biol Evol* **20**: 1377–1419.
Yanai I, Graur D, Ophir R (2004). *OMICS* **8**: 15–24.