Letter to the Editor

Evolutionary Rates of Duplicate Genes in Fish and Mammals

Marc Robinson-Rechavi and Vincent Laudet

Centre National de la Recherche Scientifique UMR 5665, Laboratoire de Biologie Moléculaire et Cellulaire, Ecole Normale Supérieure de Lyon, Lyon, France

Recently, much attention has been attracted by the abundance of duplicate genes in teleost fish (Amores et al. 1998; Wittbrodt, Meyer, and Schartl 1998). It has been suggested that this abundance reflected an ancestral genome duplication and that it may be related to the great diversity of this group (Vogel 1998). Emphasizing the importance of gene duplication in evolution, Ohno (1970) pointed out that at least one of the two copies may become less constrained by selection and thus be able to evolve toward a new function. Hughes and Hughes (1993) tested this hypothesis in the recent tetraploid *Xenopus laevis* and showed that both duplicate copies evolve at the same rate, with evidence of negative selection on both.

Our aim was to characterize whether duplicated genes have allowed a further exploration of the adaptative landscape in teleost fish and mammalian genomes. For this, we compared all genes for which at least one duplication was characterized either in a teleost fish or in a eutherian mammal. Although pseudogenes are relevant to the overall duplication rate, they do not participate in an eventual adaptative role of gene or genome duplication. Since we were interested here in functional evolution, we did not take into account duplications which clearly led to the formation of pseudogenes. This was done by the exclusion in the HOVERGEN database (Duret, Mouchiroud, and Gouy 1994) of all sequences explicitly declared to be pseudogenes.

For each of these genes, we compared rates (1) between the two duplicated copies, as in Hughes and Hughes (1993), and (2) between the lineage with the duplication and the lineage without the duplication. Indeed, if gene duplication is followed by a relaxation of selective constraint on genes, the resulting copies should evolve faster than their orthologs in other species which were not duplicated. The second test should be able to detect a decrease in selection on either or both copies, unlike the first. In practice, we build a 2×2 contingency table to compare the observed and expected numbers of genes under the hypothesis of independence between the lineage with the duplication and the lineage with the highest substitution rate (mammals or teleost fish).

Genes were recovered from the HOVERGEN database (Duret, Mouchiroud, and Gouy 1994), allowing for selection of orthologs and of lineage-specific duplications. For rate tests, only genes with a clear outgroup were retained. We thus obtained a data set of 19 genes, of which 5 showed a specific duplication in the mammalian lineage, and 14 showed a specific duplication in the teleost lineage (table 1). Amino acid sequences were aligned by CLUSTAL W (Thompson, Higgins, and Gibson 1994) and corrected by eye in with Seaview (Galtier, Gouy, and Gautier 1996). DNA sequences were not directly used, as synonymous substitutions were saturated over the evolutionary distances we considered. Phylogenetic analyses were done with Phylo_win (Galtier, Gouy, and Gautier 1996) using neighbor joining (Saitou and Nei 1987) with Poisson-corrected distances. Rates were compared between lineages using the relative-rate test on all available sequences (Wilson, Carlson, and White 1977; Robinson et al. 1998), with a Poisson correction for multiple substitutions, implemented in RRTree (Robinson-Rechavi and Huchon 2000). Although the relative-rate test normally relies on an assumption of constant rates over time, the test remains valid under a fractal model of rate variation over time (Bickel 2000). In any case, we were interested in testing whether or not observed rates variation are linked to gene duplications, and random error eventually linked to nonconstant evolution rates may introduce some noise but probably not a systematic bias.

Four genes out of 19 exhibited significant rate differences between the two (or more) duplicate genes at the 5% level (table 1), but none exhibited such differences at the 0.26% level (5%, 19 repetitions), which should be used to take into account test repetition. If indeed duplication releases protein evolution, we expect faster evolution in the lineage in which the duplication happened; it is immediately apparent from table 2 that there is no relationship between gene duplication and evolution rate. Indeed, the expected numbers of genes in each category are almost identical to those expected by chance: evolution is not faster for duplicated genes.

As expected under the ''more genes in fish'' hypothesis (Wittbrodt, Meyer, and Schartl 1998), there were more duplicated genes in the fish lineage than in the mammalian lineage. A less expected result was that most genes by far (16/19) had higher rates in fish than in mammals, irrespective of duplication, including in the five cases where the test was significant at the 5% level. Choosing mice to represent mammals, and zebrafish to represent teleost fish, we concatenated all genes, choosing one copy randomly every time there was a duplication. With this alignment of 7,483 amino acids, the zebrafish lineage evolved very significantly faster than the mouse lineage (rate difference $\Delta K = 0.058 \pm 0.009$, significant at a 10^{-7} threshold), even though genes are known to evolve faster in murids than in other mammals

Key words: selection, generation time, relative-rate test, molecular clock, genome duplication.

Address for correspondence and reprints: Marc Robinson-Rechavi, Centre National de la Recherche Scientifique UMR 5665, Laboratoire de Biologie Moléculaire et Cellulaire, Ecole Normale Supérieure de Lyon, 46 Allée d'Italie, 69364 Lyon cedex 07, France. E-mail: marc.robinson@ens-lyon.fr.

Mol. Biol. Evol. 18(4):681–683. 2001

Table 1
Genes Used in this Study

Gene Family	Duplication[a]	Outgroup[b]	Hovergen Family No.	Duplicate Rate Difference[c]
α globin	Rat	Cartilaginous fishes	FAM000215	0.027 ± 0.046
Aromatase	Euteleost fish	Stingray	FAM000502	0.018 ± 0.029
BMP-2	Zebrafish	BMP-4	FAM000306	0.048 ± 0.020*
BMP-4	Zebrafish	BMP-2	FAM000306	0.023 ± 0.015
Claudin-1	Mammals	Claudin-6 and -9	FAM002254	0.076 ± 0.048
CLIM-2	Zebrafish	CLIM-1	FAM005331	0.017 ± 0.014
Complement factor B	Zebrafish	Lamprey	FAM001595	0.032 ± 0.043
Connexin-45, -43.4, and -44.2	Zebrafish	Connexin-46.6	FAM000242	0.049 ± 0.041
Cytokeratin	Zebrafish	Cartilaginous fishes	FAM001329	0.0023 ± 0.024
Delta-1	Zebrafish	Delta-2	FAM003932	0.035 ± 0.025
DLX-2	Zebrafish	DLX-4 and -5	FAM002888	0.11 ± 0.048*
Eya-1	Mouse	Eya-2	FAM005524	0.024 ± 0.011*
HSP90	Zebrafish	HSP90 paralogs	FAM001357	0.0030 ± 0.0041
LMP2	Euteleost fish	Proteasome Y	FAM000800	0.0082 ± 0.034
Nodal	Zebrafish	Xnr-1, -2, and -3, fugacin	FAM000308	0.063 ± 0.083
Pax-6	Zebrafish	Pax-4	FAM002894	0.014 ± 0.0081
PBX-3	Mouse	Drosophila	FAM000822	0.029 ± 0.012*
Ser/Thr phosphatase 2C	Mouse	Mg phosphatase 2C	FAM002257	0.00069 ± 0.0047
TCF-3	Zebrafish	LEF-1	FAM001499	0.014 ± 0.053

[a] Largest possible group in which the duplication has been characterized.
[b] Used to root the vertebrate gene tree and compare rates between teleost fish and mammals.
[c] Rate difference between the two duplicate genes ± standard deviation. The nearest lineage without a duplication was used as an outgroup for these tests.
* Significant difference at the 5% level between the two duplicate copies.

(Wu and Li 1985). This clearly confirms that genes evolve faster in fish than in mammals.

Various life history parameters have been suggested to correlate with substitution rate, but metabolic rates (Martin and Palumbi 1993) are clearly lower in fish than in mammals, while it is hardly possible to define generation time (Wu and Li 1985) over groups as vast as mammals and teleost fish. The main hypothesis to explain the origin of so many duplicate genes in teleost fish is an ancestral genome duplication (Amores et al. 1998). However, systematic phylogenetic analysis leads us to propose rather a model of many independent gene or chromosome duplications (data not shown). This is consistent with a proposed model of frequent single *hox* cluster duplications and losses (Stellwag 1999). Spliceosomal introns also appear to have been gained many times independently in different fish lineages (Venkatesh, Ning, and Brenner 1999). Thus, high substitution rates are probably another manifestation of the dynamism of fish genomes, and it seems that, genetically at least, fish are anything but ''primitive'' vertebrates.

There does not appear to be any specific relaxation of selection on duplicate genes leading to higher rates of evolution and thus further exploration of the adaptive space. Hughes (1994) suggested that after duplication, both copies retain the same activity but gain different specificities, notably in their regulation. An alternative model is that duplication allows divergence of a few otherwise highly constrained amino acids, leading to functional divergence. Although no systematic study seems to have been conducted, most experimental results for fish indicate that divergence of duplicate genes in fish affects expression patterns more than protein activity, in support of Hughes' model.

Acknowledgments

We thank Hector Escriva and two anonymous reviewers for critical reading, and Association de Recherche sur le Cancer, Centre National pour la Recherche Scientifique, the European Molecular Biology Organisation, Fondation pour la Recherche Médicale, Ligue

Table 2
Rate Comparisons for Duplicated Genes

	Mammal Rate > Fish Rate		Mammal Rate < Fish Rate		
	Observed	Expected	Observed	Expected	Total
Duplication in mammals	1 (0)	0.79	4 (2)	4.21	5
Duplication in fish	2 (0)	2.21	12 (3)	11.79	14
Total	3		16		19

Note.—Values in the table are the numbers of genes meeting the requirements of the category. For example, there is one gene with a specific duplication in mammals and a higher substitution rate in mammals than in fish. In parentheses, the number of significant rate differences at the 5% level between the two lineages (mammales and fish) is shown. The number of expected genes in each category is calculated treating this as a contingency table under the hypothesis of independence between the lineage with the duplication and the lineage with the highest substitution rate.

Nationale contre le Cancer, and Ministère de l'Education Nationale for financial support.

LITERATURE CITED

AMORES, A., A. FORCE, Y. L. YAN et al. (13 co-authors). 1998. Zebrafish *hox* clusters and vertebrate genome evolution. Science **282**:1711–1714.

BICKEL, D. R. 2000. Implications of fluctuations in substitution rates: impact on the uncertainty of branch lengths and on relative-rate tests. J. Mol. Evol. **50**:381–390.

DURET, L., D. MOUCHIROUD, and M. GOUY. 1994. HOVERGEN: a database of homologous vertebrate genes. Nucleic Acids Res. **22**:2360–2365.

GALTIER, N., M. GOUY, and C. GAUTIER. 1996. SEAVIEW and PHYLO_WIN: two graphic tools for sequence alignment and molecular phylogeny. Comput. Appl. Biosci. **12**:543–548.

HUGHES, A. L. 1994. The evolution of functionally novel proteins after gene duplication. Proc. R. Soc. Lond. B Biol. Sci. **256**:119–124.

HUGHES, M. K., and A. L. HUGHES. 1993. Evolution of duplicate genes in a tetraploid animal, *Xenopus laevis.* Mol. Biol. Evol. **10**:1360–1369.

MARTIN, A. P., and S. R. PALUMBI. 1993. Body size, metabolic rate, generation time, and the molecular clock. Proc. Natl. Acad. Sci. USA **90**:4087–4091.

OHNO, S. 1970. Evolution by gene duplication. Springer-Verlag, New York.

ROBINSON, M., M. GOUY, C. GAUTIER, and D. MOUCHIROUD. 1998. Sensitivity of the relative-rate test to taxonomic sampling. Mol. Biol. Evol. **15**:1091–1098.

ROBINSON-RECHAVI, M., and D. HUCHON. 2000. RRTree: relative-rate tests between groups of sequences on a phylogenetic tree. Bioinformatics **16**:296–297.

SAITOU, N., and M. NEI. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. **4**:406–425.

STELLWAG, E. J. 1999. Hox gene duplication in fish. Semin. Cell Dev. Biol. **10**:531–540.

THOMPSON, J. D., D. G. HIGGINS, and T. J. GIBSON. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. **22**:4673–4680.

VENKATESH, B., Y. NING, and S. BRENNER. 1999. Late changes in spliceosomal introns define clades in vertebrate evolution. Proc. Natl. Acad. Sci. USA **96**:10267–10271.

VOGEL, G. 1998. Doubled genes may explain fish diversity. Science **281**:1119–1121.

WILSON, A. C., S. S. CARLSON, and T. J. WHITE. 1977. Biochemical evolution. Annu. Rev. Biochem. **46**:573–639.

WITTBRODT, J., A. MEYER, and M. SCHARTL. 1998. More genes in fish? Bioessays **20**:511–515.

WU, C. I., and W. H. LI. 1985. Evidence for higher rates of nucleotide substitution in rodents than in man. Proc. Natl. Acad. Sci. USA **82**:1741–1745.

NARUYA SAITOU, reviewing editor

Accepted November 14, 2000