

# Evolutionary significance of gene expression divergence

I. King Jordan\*, Leonardo Mariño-Ramírez, Eugene V. Koonin

*National Center for Biotechnology Information, National Institutes of Health 8600 Rockville Pike,  
Bldg 38A/Room 5N511-M, Bethesda, MD 20894, United States*

Received 15 September 2004; received in revised form 8 November 2004; accepted 15 November 2004

Available online 29 December 2004

Received by U. Bastolla

## Abstract

Recent large-scale studies of evolutionary changes in gene expression among mammalian species have led to the proposal that gene expression divergence may be neutral with respect to organismic fitness. Here, we employ a comparative analysis of mammalian gene sequence divergence and gene expression divergence to test the hypothesis that the evolution of gene expression is predominantly neutral. Two models of neutral gene expression evolution are considered: 1—purely neutral evolution (i.e., no selective constraint) of gene expression levels and patterns and 2—neutral evolution accompanied by selective constraint. With respect to purely neutral evolution, levels of change in gene expression between human–mouse orthologs are correlated with levels of gene sequence divergence that are determined largely by purifying selection. In contrast, evolutionary changes of tissue-specific gene expression profiles do not show such a correlation with sequence divergence. However, divergence of both gene expression levels and profiles are significantly lower for orthologous human–mouse gene pairs than for pairs of randomly chosen human and mouse genes. These data clearly point to the action of selective constraint on gene expression divergence and are inconsistent with the purely neutral model; however, there is likely to be a neutral component in evolution of gene expression, particularly, in tissues where the expression of a given gene is low and functionally irrelevant. The model of neutral evolution with selective constraint predicts a regular, clock-like accumulation of gene expression divergence. However, relative rate tests of the divergence among human–mouse–rat orthologous gene sets reveal clock-like evolution for gene sequence divergence, and to a lesser extent for gene expression level divergence, but not for the divergence of tissue-specific gene expression profiles. Taken together, these results indicate that gene expression divergence is subject to the effects of purifying selective constraint and suggest that it might also be substantially influenced by positive Darwinian selection. © 2004 Elsevier B.V. All rights reserved.

**Keywords:** Molecular evolution; Neutral theory; Human; Mouse; Genomics

## 1. Introduction

Changes in gene expression have long been recognized as fundamental to the process of evolution (Britten and Davidson, 1969; Britten and Davidson, 1971; King and Wilson, 1975). However, most molecular evolution studies have focused on gene and protein sequence divergence (Li, 1997). This can be attributed largely to the paucity of gene expression data that existed until relatively recently. Now, thanks to the development of techniques for high throughput

gene expression studies (Adams et al., 1991; Schena et al., 1995; Velculescu et al., 1995) and the creation of databases needed to store and disseminate the resulting deluge of expression data (Edgar et al., 2002; Gollub et al., 2003; Karolchik et al., 2003), enough gene expression data have accumulated to facilitate the evolutionary analysis of gene expression divergence on a large scale.

Initial attempts to study evolutionary changes in gene expression in a systematic way have led to the proposal that gene expression divergence might be predominantly neutral (Khaitovich et al., 2004; Yanai et al., 2004). The neutral model of evolution (Kimura, 1968, 1983; King and Jukes, 1969), with respect to gene expression, would hold that the vast majority of changes in expression do not affect fitness and so they accumulate rapidly and in a regular, clock-like

*Abbreviations:* CDS, Coding sequence; NCBI, National Center for Biotechnology Information; UCSC, University of California, Santa Cruz.

\* Corresponding author. Tel.: +1 301 594 5714; fax: +1 301 435 7794.

E-mail address: [jordan@ncbi.nlm.nih.gov](mailto:jordan@ncbi.nlm.nih.gov) (I.K. Jordan).

manner (Kimura and Ohta, 1974). Here, we examine patterns of mammalian gene expression to test whether the levels of gene expression divergence are indeed consistent with a neutral model of evolution.

### 1.1. Models of neutral evolution

Before proceeding, it is necessary to delineate two distinct models of neutral evolution and clarify the implications of each model for the evolution of gene expression. In the strictest sense, neutral evolution of genes can be taken as accumulation of divergence in the absence of any selective pressure. A classic example of such purely neutral evolution is the divergence of pseudogene sequences (Li et al., 1981). The observation that gene expression levels (Khaitovich et al., 2004) diverge at comparable rates for both pseudogenes and intact protein-coding genes has been taken as evidence for purely neutral evolution of gene expression. Such rapid divergence of gene expression patterns is inconsistent with any constraints on expression imposed by natural selection.

However, it should be noted that evolution according to the neutral theory does not imply the total absence of natural selection. From the neutralist perspective, the primary role of natural selection in evolution is the removal of variants that reduce fitness. The elimination of such deleterious variants is referred to as purifying or negative selection. Of course, the neutral theory does not deny the qualitative importance of positive, Darwinian selection (i.e., the fixation of variants that increase fitness) either; quantitatively, however, positive selection is construed to be rare compared to purifying selection. According to the neutral theory, evolutionary constraints on genes are due to purifying selection and the differences that are observed between genes reflect (nearly) neutral changes that have no, or only slight (Ohta, 1973), deleterious effects on fitness. This can be taken to mean that changes in gene expression with negative functional consequences are constrained by purifying selection, while neutral expression changes are free to accumulate. This model predicts a regular, clock-like accumulation of gene expression divergence over the course of evolution and such a pattern for changes in gene expression level has been observed for both primates and rodents (Khaitovich et al., 2004). In this work, we interrogate patterns of mammalian gene expression divergence with respect to the predictions of models of (i) purely neutral evolution, and (ii) neutral evolution accompanied by selective constraint.

## 2. Results and discussion

### 2.1. Purely neutral evolution

Protein-coding gene sequences (CDSs) do not evolve in a purely neutral manner as pseudogenes and some inter-

genic regions do (Li, 1997). On the contrary, purifying selection plays a critical role in constraining the levels of CDS divergence. Selective constraints on the evolution of CDSs show a broad range of variation and have been related to a number of functional characteristics of genes and the proteins that they encode (Krylov et al., 2003). These characteristics include protein dispensability (Hirsh and Fraser, 2001; Jordan et al., 2002) and interactivity (Fraser et al., 2002, 2003; Jordan et al., 2003) as well as gene expression level and breadth (Duret and Mouchiroud, 2000; Pal et al., 2001; Jordan et al., 2004; Zhang and Li, 2004). Of all of these factors, gene expression level is most strongly correlated with gene sequence divergence (Pal et al., 2003; Bloom and Adami, 2004; Rocha and Danchin, 2004), i.e., highly expressed genes tend to evolve slowly.

The correlation between gene expression levels and gene sequence divergence levels strongly suggests a connection between expression and natural selection. To more directly address whether gene expression levels change in a purely neutral manner or if they are constrained by purifying selection as CDS sequences are, the divergence of gene expression levels was compared to the level of CDS divergence between human and mouse. Gene expression data for human and mouse were taken from the mammalian gene expression atlas (Su et al., 2004) and gene sequence data were taken from the National Center for Biotechnology Information's (NCBI) Refseq database (Pruitt and Maglott, 2001) as described in the Methods section. Consistent with previous results, the level of gene expression is negatively correlated with the level of protein sequence divergence (Fig. 1a). Thus, more highly expressed genes are, on average, more constrained by purifying selection. More importantly, the amount of interspecific change in gene expression level between orthologs is positively correlated with the level of sequence divergence (Fig. 1b). In other words, pairs of genes with greater relative differences in their expression levels tend to encode more divergent proteins on average. This suggests that changes in both sequence and expression level are similarly constrained by purifying selection. In addition, changes in gene expression level between human–mouse orthologs are significantly lower than changes in gene expression levels between randomly chosen pairs of human–mouse genes (Fig. 1c). Changes in gene expression levels between randomly chosen pairs of human and mouse genes are used to approximate the rate of neutral expression level divergence (i.e., evolution with no selective constraint on gene expression level). This is because, under a purely neutral model of evolution, no significant detectable similarity due to shared common ancestry of expression profiles is expected remain during the ~100 million years of divergence between human and mouse (Gu et al., 2002), just as negligible similarity survives between neutrally evolving sequences over this time frame (Shabalina et al., 2001; Ogurtsov et al., 2004). Thus, taken together, the data shown

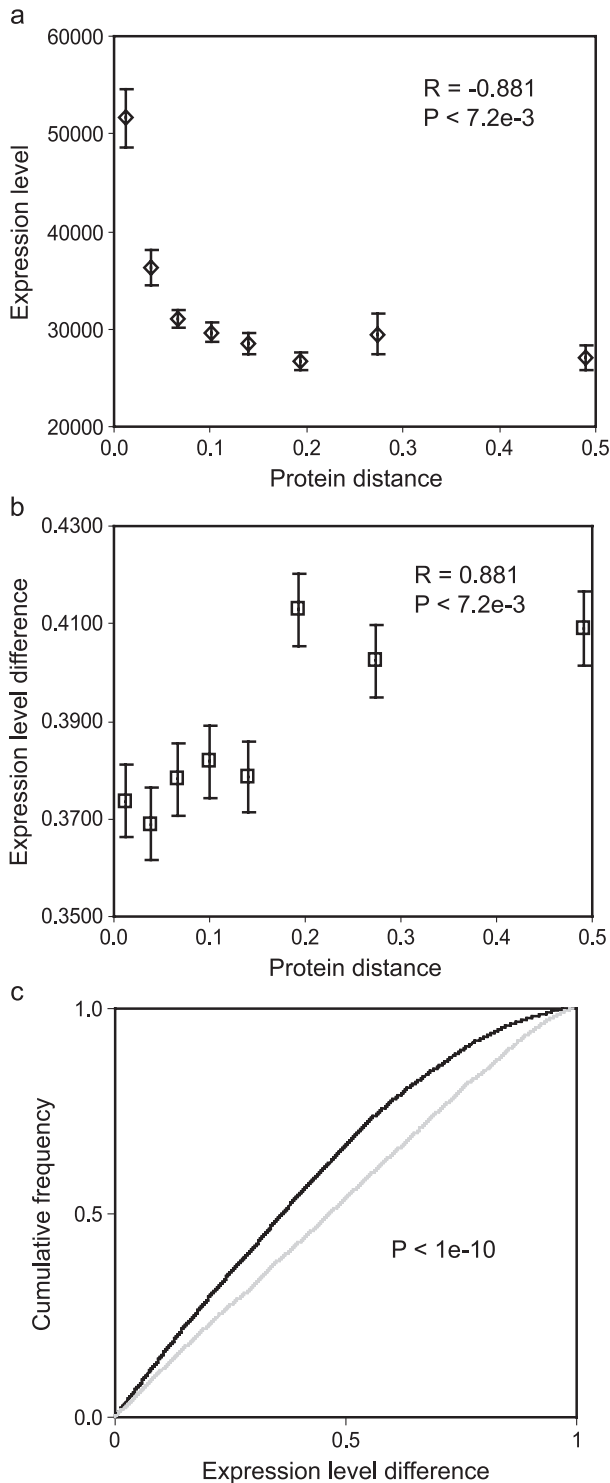


Fig. 1. Comparison between human and mouse orthologous protein sequence divergence and gene expression levels. 9059 pairs of human–mouse orthologs were compared. (a) Human–mouse ortholog pair expression levels plotted against human–mouse ortholog protein distances. (b) Relative differences between human–mouse expression levels plotted against human–mouse ortholog protein distances. For (a) and (b), average expression values are shown for eight bins of ascending protein distances. Spearman rank correlations ( $R$ ) and  $P$ -values for the correlations are shown. (c) Cumulative frequency distributions of the relative expression level differences for human–mouse orthologs (black line) and for 10,000 randomly chosen human–mouse gene pairs.

in Fig. 1 indicate that changes in gene expression level between species do not accumulate randomly as predicted by the purely neutral model.

The relationship between the number of tissues a gene is expressed in (i.e., the breadth of expression) and its substitution rate was assessed in the same way as described for gene expression levels. Consistent with previously published results (Duret and Mouchiroud, 2000; Jordan et al., 2004), gene expression breadth showed a significant negative correlation with gene sequence divergence. More broadly expressed genes tend to be more conserved, on average, than genes with a narrow range of expression (data not shown;  $r = -0.90$ ,  $P = 0.002$ ). As was also the case with gene expression level, the relative amount of interspecific change in expression breadth was significantly positively correlated with sequence divergence: orthologous gene pairs with more similar expression breadths tend to encode more conserved proteins (data not shown;  $r = 0.80$ ,  $P = 0.017$ ). These results indicate that changes in gene expression breadth, just like changes in the expression level, are also constrained by purifying selection.

It has been shown previously that changes in tissue-specific gene expression patterns between human and mouse are unrelated to levels of sequence divergence (Jordan et al., 2004; Yanai et al., 2004). This was taken as evidence for the purely neutral evolution of gene expression patterns (Yanai et al., 2004). Here, we reproduced this result with the new, expanded gene expression atlas (Su et al., 2004) by showing that the Euclidean distance between expression profiles of orthologous genes is not correlated with CDS divergence (Fig. 2a). However, there appears to be some non-linear trend in the relationship between gene expression profile divergence and sequence divergence (Fig. 2a). This is likely to be due to the use of Euclidean distances to compare expression profiles. Such distances may be inflated for gene pairs with high expression levels, and this is consistent with the high average Euclidean distance seen for the most conserved bin of gene sequences, which is known to contain numerous highly expressed genes (Fig. 1a). To control for this effect, Pearson correlation coefficients were also used to compare tissue-specific gene expression profiles. When this is done, there is no apparent trend (linear or otherwise) in the relationship between gene expression profile divergence and gene sequence divergence (data not shown;  $r = -0.31$ ,  $P = 0.45$ ).

Despite the lack of correlation between gene expression profile and gene sequence divergence, and also in agreement with previous results (Jordan et al., 2004; Yanai et al., 2004), human–mouse orthologs have tissue-specific gene expression patterns that are significantly more similar ( $t = 10.8$ ,  $P < 1e-10$ , Student's  $t$ -test) than those for randomly chosen human–mouse gene pairs (Fig. 2b). Therefore, as with gene expression levels, interspecific divergence of gene expression patterns is constrained by

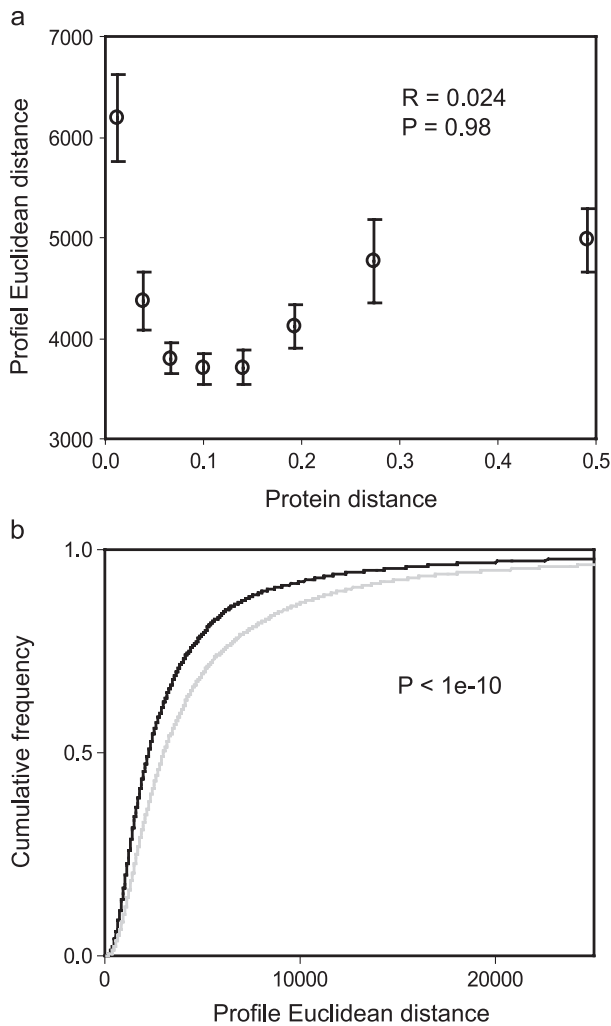


Fig. 2. Comparison between human and mouse orthologous gene expression profile divergence and gene sequence divergence. (a) Euclidean distances between human–mouse orthologous tissue-specific gene expression profiles plotted against human–mouse ortholog protein distances. Average expression values are shown for eight bins of ascending protein distances. The Spearman rank correlation ( $R$ ) and  $P$ -value for the correlation are shown. (b) Cumulative frequency distributions are shown for the gene expression profile Euclidean distances between human–mouse orthologs (black line) and between 10,000 randomly chosen human–mouse gene pairs (gray line).

purifying selection, too, and does not evolve in a purely neutral manner. However, there is also a notable difference in that changes in overall expression level clearly correlate with sequence divergence, whereas changes in expression profiles do not. It seems likely that expression level values are largely determined by one or a few tissues in which a given gene is highly expressed, functionally relevant, and subject to purifying selection which acts to retain the expression level, particularly for highly expressed genes with conserved sequences. In contrast, expression profiles are significantly affected by tissues with low and, perhaps, spurious expression of the gene in question; conceivably, in such tissues, the expression of a gene, indeed, evolves neutrally.

## 2.2. Neutral evolution accompanied by selective constraint

The model of purely neutral evolution of gene expression seems to represent an extreme and unrealistic view that is readily falsified by the data. This may not be particularly surprising because, after all, gene expression is surely an important aspect of gene function. A potentially more viable version of the neutral view of gene expression divergence holds that the functionally important component of gene expression is held constant by purifying selection, while the functionally irrelevant component evolves neutrally. Indeed, such a model is suggested by the differences observed between patterns of gene expression level divergence and gene expression profile divergence. Under this scenario, the vast majority of gene expression differences should reflect neutral (as opposed to adaptive) changes. As discussed previously here and elsewhere (Khaitovich et al., 2004), this model of neutral evolution with selective constraint predicts that gene expression divergence should accumulate in a clock-like manner, i.e., at a constant rate.

A relative rates test comparing the extent of gene expression divergence between human, mouse, and rat was used to test the prediction of rate constancy in expression divergence. This test is conceptually identical to a relative rates test previously employed with human, mouse, and rat gene sequences (Jordan et al., 2001). The idea is that human–mouse–rat evolutionary divergence can be partitioned into two components along the phylogenetic tree: 1—a within (W) rodent component and 2—a between (B) human–rodent component (Fig. 3). If evolutionary changes, in expression and/or sequences, accumulate linearly with time (i.e., in the clock-like manner), then there should be a constant ratio (W/B) of change. The clock-like model allows for different levels of selective constraint between genes, but holds the rate of change among lineages constant. Indeed, human–mouse–rat orthologous gene sequences do show such a (nearly) constant relative rate of change (Fig. 4a) as reported previously (Jordan et al., 2001). When differences in gene expression level for the same sets of human–mouse–rat orthologs were analyzed in this way, there was also a statistically significant, albeit

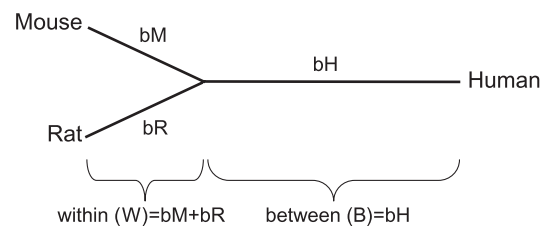


Fig. 3. Schematic of the relative rates test for gene sequence and gene expression divergence. The neutral model predicts a constant ratio of divergence within (W) and between (B) the two phylogenetic partitions. Expressed in terms of phylogenetic branch lengths, the null hypothesis of the test is constant  $(b_M + b_R)/b_H$ .

substantially weaker, correlation between the two phylogenetic components of variation (Fig. 4b). Furthermore, as noted above for human–mouse gene pairs (Fig. 1c), changes in gene expression level observed between human–mouse–rat orthologs were significantly lower than changes between random human–mouse–rat gene sets ( $t = 5.9$ ,  $P = 3.3 \times 10^{-9}$ ,

Student's  $t$ -test). These observations suggest that changes in expression level may be consistent with the model of neutral evolution accompanied by selective constraint.

When the divergence in tissue-specific gene expression patterns for the same sets of human–mouse–rat orthologs was partitioned along the phylogenetic tree, a different picture emerged. There is no evidence of a constant relative rate of tissue-specific gene expression profile divergence (Fig. 4c). Therefore, in the case of gene expression profiles, divergence does not accumulate in a regular clock-like manner as predicted by the model of neutral evolution accompanied by selective constraint. However, the gene expression profile divergence between human–mouse–rat orthologous sets is significantly lower than the divergence between random human–mouse–rat gene sets ( $t = 18.1$ ,  $P < 1 \times 10^{-10}$  Student's  $t$ -test) indicating that expression profile divergence is still constrained to some extent by purifying selection.

### 2.3. Conclusions

The results of the comparison of gene expression divergence and CDS divergence in human and mouse orthologs seem to falsify the recently proposed purely neutral model of the transcriptome evolution (Khaitovich et al., 2004; Yanai et al., 2004). Instead, we show that, although there may be a neutral component in the evolution of expression as evidenced by the lack of correlation between the divergence of expression profiles and gene sequence divergence (Jordan et al., 2004 and this work), much, if not most, of the change in expression is subject to purifying selection. The origin of the difference in these results and those of Khaitovich and coworkers and Yanai and coworkers remains to be investigated. One issue that could be pertinent is the use of pseudogenes as a proxy for completely neutral evolution of gene expression (Khaitovich et al., 2004). Conceivably, the expression of many pseudogenes could still be subject to purifying selection, perhaps, leading to low expression levels because high-level expression of a pseudogene is likely to be deleterious. This would result in low divergence between expression levels of orthologous pseudogenes and, if these employed as the neutral background, could lead to a false conclusion on purely neutral evolution of the expression of functional genes.

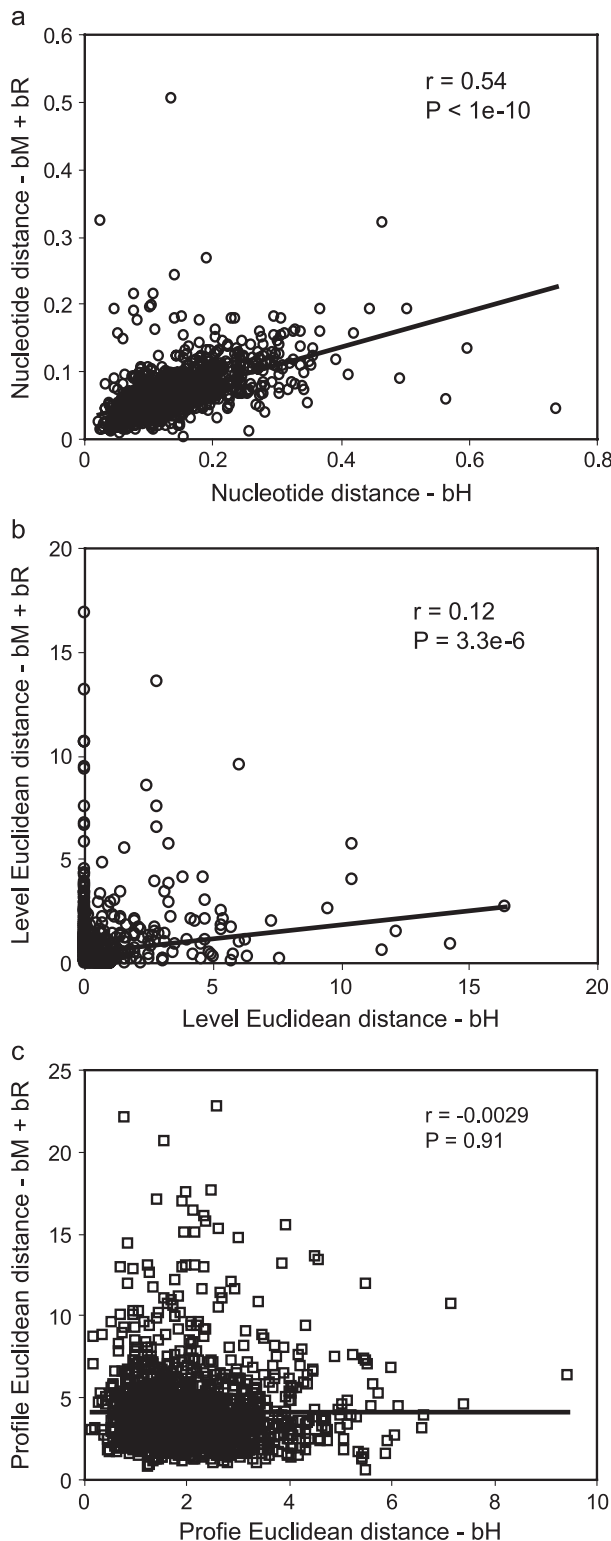


Fig. 4. Results of the relative rates test for gene sequence and gene expression divergence. 1427 sets of human–mouse–rat orthologs were compared. (a) Within (bM+bR) against between (bH) partition nucleotide divergence. Nucleotide divergence is measured as the Jukes–Cantor distance (Jukes and Cantor, 1969). (b) Within (bM+bR) against between (bH) partition expression level divergence. Expression level divergence is measured as the absolute difference between the species-specific relative gene expression levels. (c) Within (bM+bR) against between (bH) partition gene expression profile divergence. Tissue-specific gene expression profile divergence is measured as the Euclidean distance between orthologous vectors of tissue-specific expression values. For all plots, Pearson correlation co-efficients ( $r$ ) and  $P$ -values for the correlations are shown.

Whether or not there is a substantial correlation between gene expression divergence and sequence divergence had been an issue of much contest and contradiction. A lack of correlation between expression profile divergence and gene sequence divergence has been reported for yeast paralogs (Wagner, 2000) and, more recently, for worm paralogs (Conant and Wagner, 2004). However, another series of studies presents a significant correlation between the divergence of expression and sequence of duplicated genes for both yeast and humans (Gu et al., 2002; Makova and Li, 2003). Another recent study found that the divergence of putative promoter sequences among nematode genes was coupled to the divergence of the coding sequence for orthologs but not for paralogs, in line with the notion of rapid functional diversification of gene after duplication (Castillo-Davis et al., 2004). This lack of consensus on the relationship between evolution of coding sequences and evolution of expression might reflect the combination of neutral and selective forces affecting the latter as discussed in this paper; clearly, further analysis is required to resolve this issue.

The conclusion of this work that evolution of gene expression is, in large part, subject to purifying selection is hardly unexpected. The resulting view of gene expression evolution is not unlike the model of sequence evolution under the neutral theory which is compatible with the well-supported observation that highly expressed genes tend to evolve slowly (Pal et al., 2001; Jordan et al., 2004; Zhang and Li, 2004). However, another finding reported here does seem surprising, namely, that tissue-specific gene expression profiles, unlike gene expression levels, do not change at a constant rate (in a clock-like fashion) in the course of mammalian evolution. This observation does not seem to be compatible with any version of the neutral model for evolution of gene expression, even one that includes purifying selection. Instead, this result suggests that positive selection could be a substantial factor affecting changes in the patterns of gene expression during evolution. Clearly, much more analysis of the intragenomic and intergenomic patterns of expression divergence between homologous genes is required before a robust model of evolution of gene expression is developed.

### 3. Methods

#### 3.1. Gene sequence analysis

Human, mouse, and rat CDS and protein sequences were taken from the NCBI's RefSeq database (Pruitt and Maglott, 2001). The NCBI's LocusLink database (Pruitt and Maglott, 2001) was used to ensure that only one sequence per loci was retained for further analysis. For loci that encode multiple transcripts, the longest form was taken. Orthologous gene sets were identified as symmetrical best hits in

all-against-all between genome BLASTP searches as described previously (Jordan et al., 2001). The program ClustalW (Thompson et al., 1994) was used to align orthologous protein sequences. Nucleotide CDSs were aligned to correspond to protein sequence alignments. Nucleotide sequence distances were calculated using the Jukes–Cantor correction for multiple substitutions (Jukes and Cantor, 1969) and protein sequence distances were calculated using the correction based on the gamma distribution of site rate variation (Ota and Nei, 1994) with  $\alpha=2$ . Pairwise nucleotide distances ( $d$ ) were converted to branch lengths on the human–mouse–rat phylogeny as described previously (Jordan et al., 2001).

#### 3.2. Gene expression analysis

Gene expression data, based on Affymetrix microarray experiments, for human, mouse, and rat are from the mammalian gene expression atlas (Su et al., 2004). These expression data were retrieved from the UCSC Genome Browser (Karolchik et al., 2003). Affymetrix probe identifiers (ids) were mapped to human, mouse, and rat genomic loci using UCSC Genome Browser and NCBI annotations as shown below:

$$\begin{aligned} \text{Affymetrix probe id} &\rightarrow \text{GenBank accession} \\ &\rightarrow \text{RefSeq accession} \rightarrow \text{NCBI Locus id} \end{aligned} \quad (1)$$

Only affymetrix probes that map to unique genomic loci were considered for further analysis. When loci were found to be covered by multiple probes, the probe yielding the highest overall expression level was used in subsequent analyses.

To measure changes in gene expression levels and changes in tissue-specific expression patterns across species, tissue samples common to all species being compared were identified. There were 28 common tissues with expression data for both human and mouse and 10 common tissues with expression data for human, mouse, and rat. For the comparison between protein sequence divergence and expression level, expression levels were taken as the sum of all expression levels over 28 tissues shared between the human and mouse expression data sets. Using these same 28 tissues, changes in gene expression level between species were calculated by dividing the absolute value of the difference between expression levels by the sum of the expression levels being considered:

$$\frac{\text{Human expression level} - \text{Mouse expression level}}{\text{Human expression level} + \text{Mouse expression level}} \quad (2)$$

For the comparison between protein sequence divergence and expression breadth, expression breadth was taken as the number of tissues (out of 28 total possible) in which a gene showed an expression level  $\geq 350$  in the mammalian gene expression atlas. The relative differences

in expression breadth between species were calculated in the same way as shown above (Eq. (2)) for expression level differences. For the comparisons between human, mouse, and rat expression levels, each genes species-specific relative expression level was determined by taking the sum of all expression levels over the 10 shared tissues and dividing by the average of those sums for the species from which the gene is derived:

$$\frac{\sum_{\text{tissue}=1}^{10} \text{expression level}}{\frac{1}{1427} \left( \sum_{\text{gene}=1}^{1427} \sum_{\text{tissue}=1}^{10} \text{expression level} \right)} \quad (3)$$

The absolute distances between relative expression levels were taken and converted into branch lengths as described previously (Jordan et al., 2001).

In order to compare tissue-specific gene expression profiles, each gene (probe) is represented as a vector of tissue-specific expression levels. Orthologous gene expression patterns are then compared by calculating the Euclidean distances between the expression vectors ( $n=28$  for human–mouse and  $n=10$  for human–mouse–rat) of each species-specific gene in the orthologous set. Between species Euclidean distances were converted into branch lengths for the relative rates test as was done for the gene sequences (Jordan et al., 2001). Tissue-specific expression profiles between human and mouse were also compared using Pearson correlation co-efficients.

As a control for gene expression level and gene expression profile divergence, gene expression differences between 10,000 pairs, consisting of one randomly chosen human gene and one randomly chosen mouse gene, were compared as described above for orthologous gene pairs.

## References

- Adams, M.D., et al., 1991. Complementary DNA sequencing: expressed sequence tags and human genome project. *Science* 252, 1651–1656.
- Bloom, J.D., Adami, C., 2004. Evolutionary rate depends on number of protein–protein interactions independently of gene expression level: response. *BMC Evol. Biol.* 4, 14.
- Britten, R.J., Davidson, E.H., 1969. Gene regulation for higher cells: a theory. *Science* 165, 349–357.
- Britten, R.J., Davidson, E.H., 1971. Repetitive and non-repetitive DNA sequences and a speculation on the origins of evolutionary novelty. *Q. Rev. Biol.* 46, 111–138.
- Castillo-Davis, C.I., Hartl, D.L., Achaz, G., 2004. cis-Regulatory and protein evolution in orthologous and duplicate genes. *Genome Res.* 14, 1530–1536.
- Conant, G.C., Wagner, A., 2004. Duplicate genes and robustness to transient gene knock-downs in *Caenorhabditis elegans*. *Proc. R. Soc. Lond., B Biol. Sci.* 271, 89–96.
- Duret, L., Mouchiroud, D., 2000. Determinants of substitution rates in mammalian genes: expression pattern affects selection intensity but not mutation rate. *Mol. Biol. Evol.* 17, 68–74.
- Edgar, R., Domrachev, M., Lash, A.E., 2002. Gene expression omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res.* 30, 207–210.
- Fraser, H.B., Hirsh, A.E., Steinmetz, L.M., Scharfe, C., Feldman, M.W., 2002. Evolutionary rate in the protein interaction network. *Science* 296, 750–752.
- Fraser, H.B., Wall, D.P., Hirsh, A.E., 2003. A simple dependence between protein evolution rate and the number of protein–protein interactions. *BMC Evol. Biol.* 3, 11.
- Gollub, J., et al., 2003. The Stanford Microarray Database: data access and quality assessment tools. *Nucleic Acids Res.* 31, 94–96.
- Gu, Z., Nicolae, D., Lu, H.H., Li, W.H., 2002. Rapid divergence in expression between duplicate genes inferred from microarray data. *Trends Genet.* 18, 609–613.
- Hirsh, A.E., Fraser, H.B., 2001. Protein dispensability and rate of evolution. *Nature* 411, 1046–1049.
- Jordan, I.K., Kondrashov, F.A., Rogozin, I.B., Tatusov, R.L., Wolf, Y.I., Koonin, E.V., 2001. Constant relative rate of protein evolution and detection of functional diversification among bacterial, archaeal and eukaryotic proteins. *Genome Biol.* 2 (RESEARCH0053).
- Jordan, I.K., Rogozin, I.B., Wolf, Y.I., Koonin, E.V., 2002. Essential genes are more evolutionarily conserved than are nonessential genes in bacteria. *Genome Res.* 12, 962–968.
- Jordan, I.K., Wolf, Y.I., Koonin, E.V., 2003. No simple dependence between protein evolution rate and the number of protein–protein interactions: only the most prolific interactors tend to evolve slowly. *BMC Evol. Biol.* 3, 1.
- Jordan, I.K., Marino-Ramirez, L., Wolf, Y.I., Koonin, E.V., 2004. Conservation and co-evolution in the scale-free human gene coexpression network. *Mol. Biol. Evol.* 21, 2058–2070.
- Jukes, T.H., Cantor, C.R., 1969. Evolution of protein molecules. In: Munro, H.D. (Ed.), *Mammalian protein metabolism*. Academic Press, New York, pp. 21–132.
- Karolchik, D., et al., 2003. The UCSC genome browser database. *Nucleic Acids Res.* 31, 51–54.
- Khaitovich, P., et al., 2004. A neutral model of transcriptome evolution. *PLoS Biol.* 2, E132.
- Kimura, M., 1968. Evolutionary rate at the molecular level. *Nature* 217, 624–626.
- Kimura, M., 1983. *The Neutral Theory of Molecular Evolution*. Cambridge University Press, Cambridge.
- Kimura, M., Ohta, T., 1974. On some principles governing molecular evolution. *Proc. Natl. Acad. Sci. U. S. A.* 71, 2848–2852.
- King, J.L., Jukes, T.H., 1969. Non-Darwinian evolution. *Science* 164, 788–798.
- King, M.C., Wilson, A.C., 1975. Evolution at two levels in humans and chimpanzees. *Science* 188, 107–116.
- Krylov, D.M., Wolf, Y.I., Rogozin, I.B., Koonin, E.V., 2003. Gene loss, protein sequence divergence, gene dispensability, expression level, and interactivity are correlated in eukaryotic evolution. *Genome Res.* 13, 2229–2235.
- Li, W.H., 1997. *Molecular evolution*. Sinauer Associates, Sunderland.
- Li, W.H., Gojobori, T., Nei, M., 1981. Pseudogenes as a paradigm of neutral evolution. *Nature* 292, 237–239.
- Makova, K.D., Li, W.H., 2003. Divergence in the spatial pattern of gene expression between human duplicate genes. *Genome Res.* 13, 1638–1645.
- Ogurtsov, A.Y., Sunyaev, S., Kondrashov, A.S., 2004. Indel-based evolutionary distance and mouse–human divergence. *Genome Res.* 14, 1610–1616.
- Ohta, T., 1973. Slightly deleterious mutant substitutions in evolution. *Nature* 246, 96–98.
- Ota, T., Nei, M., 1994. Estimation of the number of amino acid substitutions per site when the substitution rate varies among sites. *J. Mol. Evol.* 38, 642–643.
- Pal, C., Papp, B., Hurst, L.D., 2001. Highly expressed genes in yeast evolve slowly. *Genetics* 158, 927–931.

- Pal, C., Papp, B., Hurst, L.D., 2003. Genomic function: rate of evolution and gene dispensability. *Nature* 421, 496–497.
- Pruitt, K.D., Maglott, D.R., 2001. RefSeq and LocusLink: NCBI gene-centered resources. *Nucleic Acids Res.* 29, 137–140.
- Rocha, E.P., Danchin, A., 2004. An analysis of determinants of amino acids substitution rates in bacterial proteins. *Mol. Biol. Evol.* 21, 108–116.
- Schena, M., Shalon, D., Davis, R.W., Brown, P.O., 1995. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* 270, 467–470.
- Shabalina, S.A., Ogurtsov, A.Y., Kondrashov, V.A., Kondrashov, A.S., 2001. Selective constraint in intergenic regions of human and mouse genomes. *Trends Genet.* 17, 373–376.
- Su, A.I., et al., 2004. A gene atlas of the mouse and human protein-encoding transcriptomes. *Proc. Natl. Acad. Sci. U. S. A.* 101, 6062–6067.
- Thompson, J.D., Higgins, D.G., Gibson, T.J., 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.
- Velculescu, V.E., Zhang, L., Vogelstein, B., Kinzler, K.W., 1995. Serial analysis of gene expression. *Science* 270, 484–487.
- Wagner, A., 2000. Decoupled evolution of coding region and mRNA expression patterns after gene duplication: implications for the neutralist–selectionist debate. *Proc. Natl. Acad. Sci. U. S. A.* 97, 6579–6584.
- Yanai, I., Graur, D., Ophir, R., 2004. Incongruent expression profiles between human and mouse orthologous genes suggest widespread neutral evolution of transcription control. *Omics* 8, 15–24.
- Zhang, L., Li, W.H., 2004. Mammalian housekeeping genes evolve more slowly than tissue-specific genes. *Mol. Biol. Evol.* 21, 236–239.