Article

Discoveries

Tissue-specificity of gene expression diverges slowly between orthologs, and rapidly between paralogs

Nadezda Kryuchkova-Mostacci[1,2], Marc Robinson-Rechavi[1,2,*]

[1]Department of Ecology and Evolution, University of Lausanne, Switzerland
[2]Swiss Institute of Bioinformatics, Lausanne, Switzerland

[*]Author of Correspondence: Marc Robinson-Rechavi, Department of Ecology and Evolution, University of Lausanne, Switzerland, +41 21 692 4220, marc.robinson-rechavi@unil.ch

Abstract

The ortholog conjecture implies that functional similarity between orthologous genes is higher than between paralogs. It has been supported using levels of expression and Gene Ontology term analysis previously, although the evidence was rather weak and there were also conflicting reports. In this study on 12 species we provide strong evidence of high conservation in tissue-specificity between orthologs, in contrast to low conservation between within-species paralogs. This allows us to shed a new light on the evolution of gene expression patterns. While there have been several studies of the correlation of expression between species, little is known about the evolution of tissue-specificity itself. Ortholog tissue-specificity is strongly conserved between all tetrapod species, with the lowest Pearson correlation between mouse and frog at r = 0.66. Tissue-specificity correlation decreases strongly with divergence time. Paralogs in human show much lower conservation, even for recent Primate-specific paralogs. Small-scale tissue-specific paralogs are mostly specific for the same tissue, while ancient whole genome duplication paralogs are often specific for different tissues. The same patterns are observed using human or mouse as focal species and are robust to choices of datasets and of thresholds.

Introduction

The ortholog conjecture is widely used to transfer annotation among genes, for example in newly sequenced genomes. But it is still debated how much orthologs share more similar functions than paralogs (Studer and Robinson-Rechavi 2009; Gabaldón and Koonin 2013). The most widely accepted model is that orthologs diverge slower, and that the generation of paralogs through duplication leads to strong divergence and even change of function. It is also expected that in general homologs diverge functionally with time. The test of these hypotheses poses fundamental questions of molecular evolution, about the rate of functional evolution and the role of duplications.

Surprisingly, there are several studies which have reported no difference between orthologs and paralogs, or even the opposite, that paralogs would be more functionally similar than orthologs. Tests of the ortholog conjecture using sequence evolution found no difference after speciation or duplication in positive selection (Studer et al. 2008), nor in amino acid shifts (Studer and Robinson-Rechavi 2010). The debate was truly launched by Nehrt et al. (Nehrt et al. 2011) who reported in a large scale study, based on expression levels similarity and Gene Ontology (GO) analysis, that paralogs are better predictors of function than orthologs. Of note, methodological aspects of the GO analysis of that study were criticized by several other authors (Chen and Zhang 2012; Thomas et al. 2012). Using a very similar GO analysis but correcting biases in the data, Altenhoff et al. (Altenhoff et al. 2012) found more functional similarity between orthologs than between paralogs based on GO annotation analysis, but the differences were very slight.

An early comparison of expression profiles of orthologs reported that they were very different, close to paralogs and even to random pairs (Yanai et al. 2004). Further studies, following Nehrt et al. (Nehrt et al. 2011), found little or no evidence for the ortholog conjecture in expression data. Rogozin et al. (Rogozin et al. 2014) reported that orthologs are more similar than between species paralogs but less similar than within-species paralogs based on correlations between RNA-seq expression profiles in mouse and human. Wu et al. (Wu et al. 2014) found only a small difference between orthologs and paralogs. Paralogs were significantly more functionally similar than orthologs, but by classifying in subtypes they reported that one-to-one orthologs are the most functionally similar. The analysis was done on the level of function by looking at expression network similarities.

On the other hand, the ortholog conjecture has been supported by several studies of gene expression. *Contra* Yanai et al. (Yanai et al. 2004), several studies have reported good correlations between expression levels of orthologs, between human and mouse (Liao and Zhang 2006), or among amniotes (Brawand et al. 2011). Moreover, some studies have

reported changes of expression following duplication, although without explicitly testing for the ortholog conjecture: duplicated genes are more likely to show changes in expression profiles than single-copy genes (Gu et al. 2004; Huminiecki and Wolfe 2004). Chung et al. (Chung et al. 2006) reported through network analysis that duplicated genes diverge rapidly in their expression profile. Recently Assis and Bachtrog (Assis and Bachtrog 2015) reported that paralog function diverges rapidly. They analysed among other things difference in tissue-specificity between a pair of paralogs and their single copy ortholog in closely related species. They conclude that divergence of paralogs results in increased tissue-specificity, and that there are differences between tissues. Finally, several explicit tests of the ortholog conjecture have also found support using expression data. Huerta-Cepas et al. (Huerta-Cepas et al. 2011) reported that paralogs have higher levels of expression divergence than orthologs of the similar age, using microarray data with calls of expressed/not expressed in human and mouse. They also claimed that a significant part of this divergence was acquired shortly after the duplication event. Chen and Zhang (Chen and Zhang 2012) re-analysed the RNA-seq dataset of (Brawand et al. 2011) and reported that expression profiles of orthologs are significantly more similar than within-species paralogs.

Thus while the balance of evidence appears to weight towards confirmation of the ortholog conjecture, functional data has failed so far to strongly support or invalidate it. Even results which support the ortholog conjecture often do so with quite slight differences between orthologs and paralogs (Altenhoff et al. 2012; Rogozin et al. 2014). Yet expression data especially should have the potential to solve this issue, since it provides functional evidence for many genes in the same way across species, without the ascertainment biases of GO annotations or other collections of small scale data. Part of the problem is that the relation of levels of expression to function is not direct, making it unclear what biological signal is being compared in correlations of these levels. Another problem is that the comparison of different transcriptome datasets in different species suffers from biases introduced by ubiquitous genes (Piasecka et al. 2012) or batch effects (Gilad and Mizrahi-Man 2015).

In our analysis we have concentrated on the tissue-specificity of expression. Tissue-specificity indicates in how many tissues a gene is expressed, and whether it has large differences of expression level between them. It reflects the functionality of the gene: if the gene is expressed in many tissues then it is "house keeping" and has a function needed in many organs and cell types; tissue-specific genes have more specific roles, and tissue adjusted functions. Previous results indicate that tissue-specificity is conserved between human and mouse orthologs, and that it is functionally informative (Kryuchkova-Mostacci and Robinson-Rechavi 2016). Moreover, tissue-specificity can be computed in a comparable manner in

different datasets without notable biases, as long as at least 6 tissues are represented, including preferably testis, nervous system, and proportionally not too many parts of the same organ (e.g. not many parts of the brain).

Are there major differences between the evolution of tissue-specificity after duplication (paralogs) or without duplication (orthologs)? We analyse the conservation of one-to-one orthologs and within-species paralogs with evolutionary time, using RNA-seq datasets from 12 species.

Results

We compared orthologs between 12 species: human, chimpanzee, gorilla, macaque, mouse, rat, cow, opossum, platypus, chicken, frog, and fruit fly. Overall 7 different RNA-seq datasets were used, including 6 to 27 tissues (see Materials and Methods). Three comparisons were performed with the largest sets as focal data: 27 human tissues from Fagerberg et al., 16 human tissues from Bodymap, and 22 tissues from mouse ENCODE (The ENCODE Project Consortium 2011; Fagerberg et al. 2014; Farrell et al. 2014).

The first notable result is that tissue-specificity is strongly correlated between one-to-one orthologs. The correlations between human and four other species are presented in Fig. 1a for illustration. This confirms and extends our previous observation (Kryuchkova-Mostacci and Robinson-Rechavi 2016), which was based on one human and one mouse datasets. Correlation of tissue-specificity varies between 0.74 and 0.89 among tetrapods, and is still 0.43 between human and fly, 0.38 between mouse and fly. The latter is despite the very large differences in anatomy and tissue sampling between the species compared, showing how conserved tissue-specificity can be in evolution.

The correlation between orthologs decreases with divergence time (Fig 2). The decline is linear. An exponential model is not significantly better: ANOVA was not significantly better for the model with $\log_{10}$ of time than for untransformed time for any dataset ($p > 0.0137$, $q > 1\%$). The trend is not caused by the outlier fly data point: removing it there is still a significant decrease of correlation for orthologs (see Supplementary Materials). Results are also robust to the use of Spearman instead of Pearson correlation between tissue-specificity values.

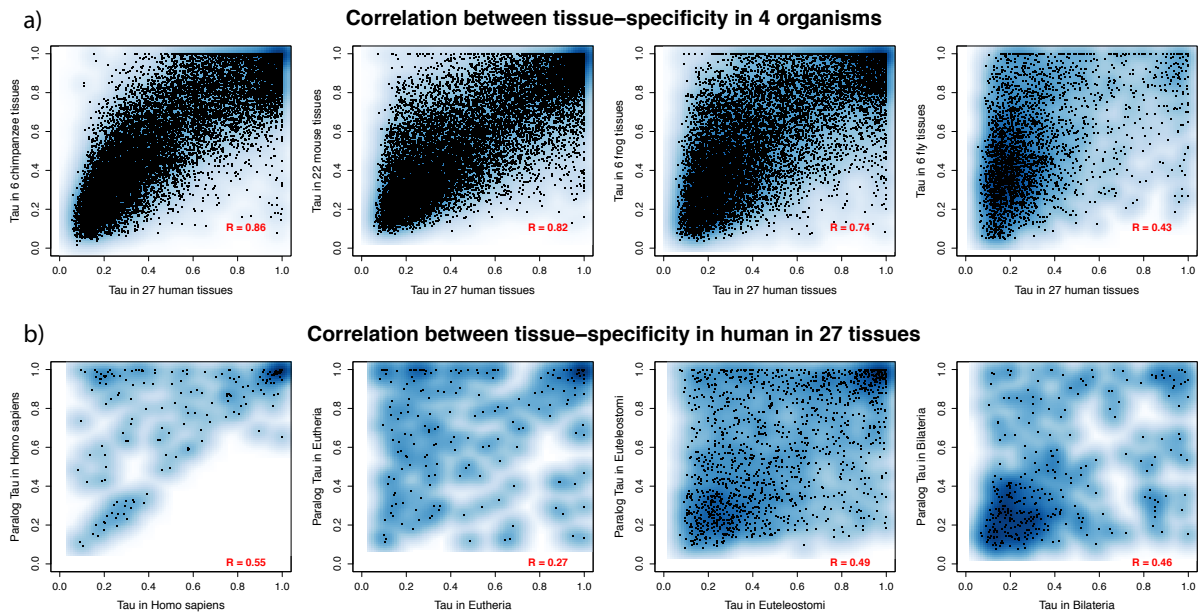

Fig. 1: Pearson correlation of tissue-specificity between a) orthologs and b) paralogs. a) Human ortholog vs. one-to-one ortholog in another species; b) highest expressed paralog vs. lowest expressed paralog in human, for different duplication dates.

The correlation between within-species paralogs is significantly lower than between orthologs (ANOVA $p<0.0137$, $q<1\%$ for all datasets) (Fig 2). Moreover, there is no significant decline in correlation with evolutionary time (neither linear nor exponential) for paralogs. This may indicate almost immediate divergence of paralogs upon duplication, although other scenarios are possible (see Discussion).

The results are consistent between human and mouse (Fig. 2a and b). Results are also consistent using a different human RNA-seq dataset (Fig S1).

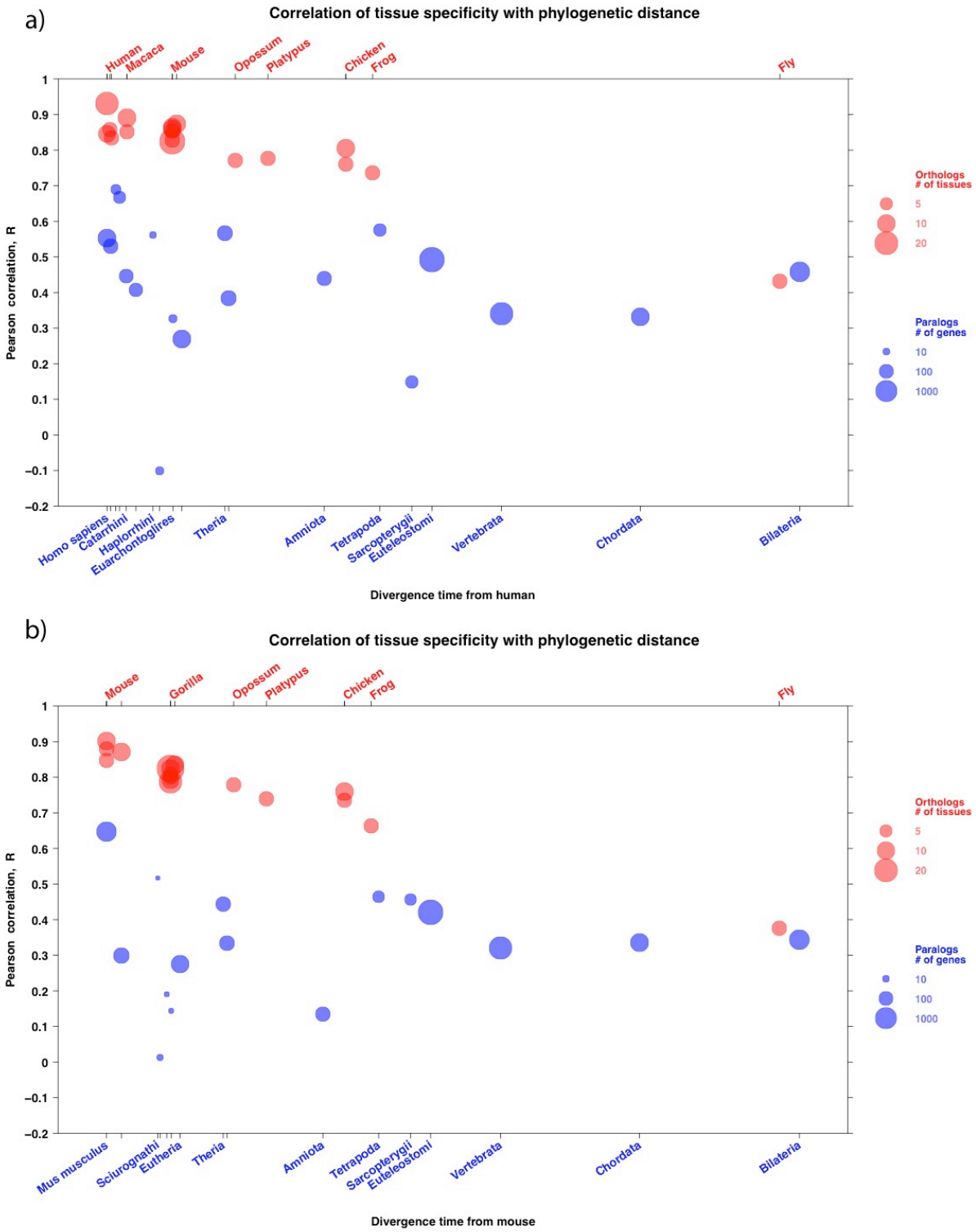

Fig. 2: Pearson correlation of tissue-specificity focusing on a) human and b) mouse. X-axis, divergence time in million years between the genes compared; Y-axis, Pearson correlation between values of τ over genes. In red, the correlation of orthologs between the focal species and other species; representative species are noted above the figure; there are several points when there are several datasets for a same species, e.g. four for mouse (Table 1); the size of red circles is proportional to the number of tissues used for calculation of tissue-specificity. In blue, the correlation of paralogs in the focal species, according to the date of duplication; representative taxonomic groups for this dating are noted under the figure; the size of blue circles is proportional to the number of genes in the paralog group.

This main analysis is based on the correlation of tissue-specificity on orthologs called pairwise between species. The number of orthologs used in the analysis is thus variable (available in Supplementary Materials). An additional analysis was also performed using the same orthologs for all tetrapods, 4785 genes (Fig. S2-S4). Correlations of these "conserved orthologs" are not significantly different from those observed over all orthologs.

The analysis was also performed on all the datasets excluding testis specific genes (Fig. S6-S8), defined as having their highest expression in testis (see Materials and Methods). The correlation between orthologs becomes significantly lower (ANOVA p=0.000178), while between paralogs it does not change significantly (ANOVA p=0.846). Even though correlation between orthologs becomes weaker there is still a significant difference between orthologs and paralogs (ANOVA p=1.299e-07).

We also performed the analysis removing genes on sex chromosomes (Fig. S9-S11). This analysis was done without frog, as sex chromosome information is not available. This does not change significantly the correlations between either orthologs (ANOVA p=0.856) or paralogs (ANOVA p=0.755).

In general paralogs have lower expression and are more tissue-specific than orthologs (Fig S12), which is consistent with the dosage-sharing model (Gout and Lynch 2015; Lan and Pritchard 2016). Young paralogs are very tissue-specific, and get more ubiquitous with divergence time (Fig 1b and Fig S13); this is true for all datasets, and for τ calculated without testis. We also observe that the higher expressed paralog has a stronger correlation with an ortholog outgroup, thus appears to keep more the ancestral tissue-specificity, while the lower expressed paralog has a lower correlation and appears to become more tissue-specific (Fig 3), which is consistent with a form of neo-functionalization.

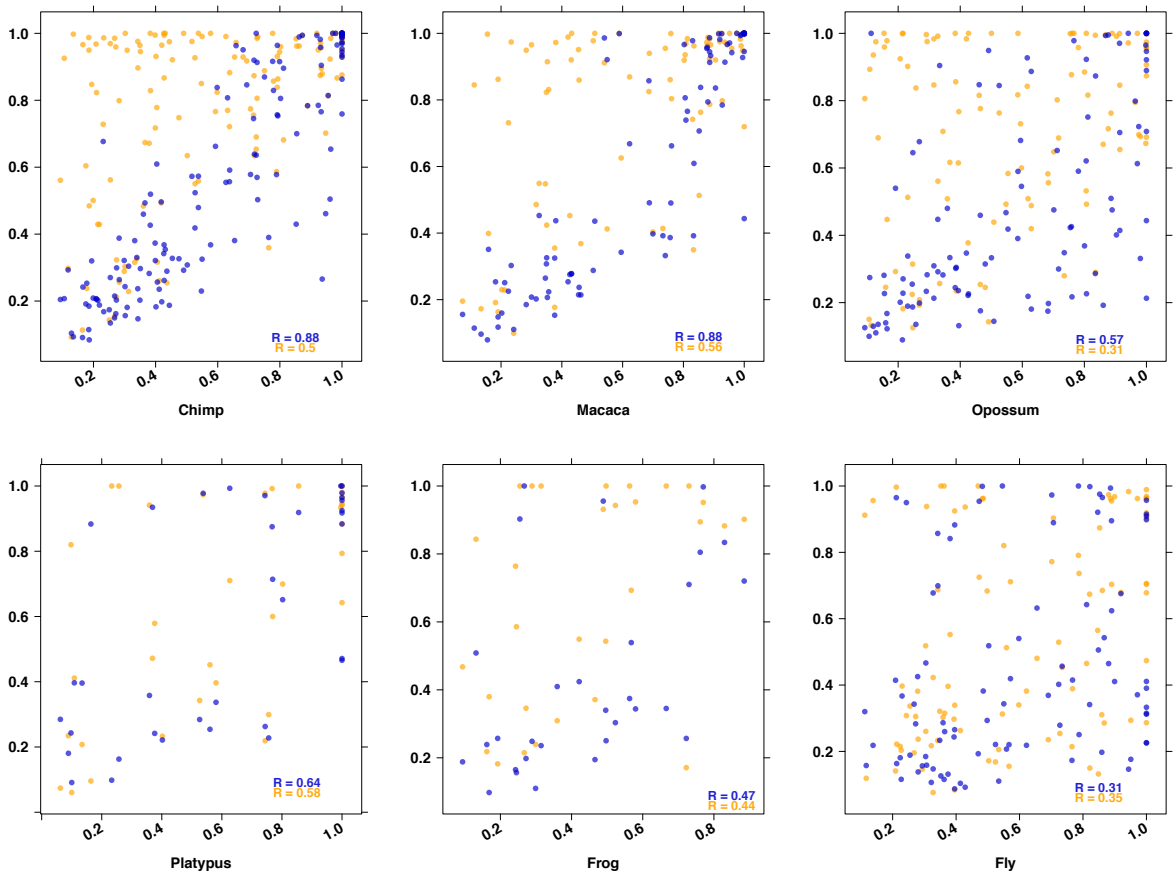

Fig. 3: Distribution of tissue-specificity in paralogs compared to an outgroup ortholog. For each graph, paralogs of a given phylogenetic age are compared to the closest outgroup un-duplicated ortholog; thus these paralogs are "in-paralogs" relative to the speciation node, and are both "co-orthologs" to the outgroup. X-axis, τ of unduplicated ortholog. Y-axis, τ of paralogs. Blue points are values for the paralog with highest maximal expression of the pair of paralogs, orange points are values for the other.

When both orthologs of a pair are tissue-specific ($\tau > 0.8$), they are most often expressed in the same tissue (Fig. 4). The same is observed when both paralogs are tissue-specific and are younger than the divergence of tetrapods. But for Euteleostomi and Vertebrata paralogs, if both are tissue-specific they are as likely to be expressed in the same as in different tissues; most of these are expected to be ohnologs, i.e. due to whole genome duplication. This analysis was performed on the Brawand et al. (2011) dataset, because it has the most organisms with the same 6 tissues. This result does not change after removing testis (Fig. S14), nor changing the τ threshold from 0.8 to 0.3 (Fig S15-S16).

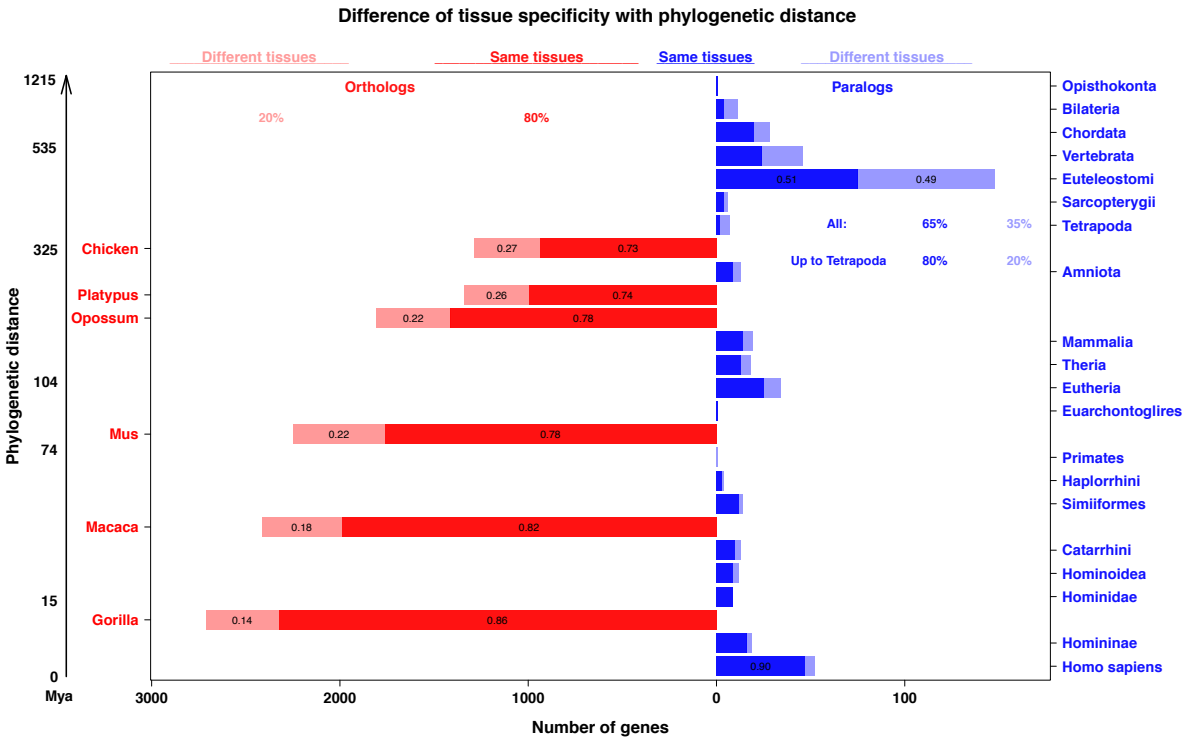

Fig. 4: Difference of tissue-specificity between orthologs and paralogs. Each bar represents the number of gene pairs of a given type for a given phylogenetic age, for which both genes of the pair are tissue-specific ($\tau > 0.8$). In dark colour, the number of gene pairs specific of the same tissue; in light colour, the number of gene pairs specific of different tissues. Orthologs are in red, in the left panel, paralogs are in blue, on the right panel; notice that the scales are different for orthologs and for paralogs. The overall proportions of pairs in the same or different tissues are indicated for orthologs and paralogs; in addition, for paralogs the proportion for pairs younger than the divergence of tetrapods is also indicated.

Discussion

Our results show that most genes have their tissue-specificity conserved between species. This provides strong new evidence for the evolutionary conservation of expression patterns. Using tissue-specificity instead of expression values allows easy comparison between species, as bias of normalisation or use of different datasets has little effect on results (Kryuchkova-Mostacci and Robinson-Rechavi 2016). All of our results were confirmed using three different focus datasets, from human or mouse, and thus appear to be quite robust.

The conservation of expression tissue-specificity of protein coding genes that we find is high even for quite distant one-to-one orthologs: the Pearson correlation between τ in human or mouse and τ in frog is $R = 0.74$ (respectively $R = 0.66$) over 361 My of divergence. Even between fly and mammals it is more than 0.38. Moreover, this tissue-specificity can be easily compared over large datasets without picking a restricted set of homologous tissues (e.g. in (Brawand et al. 2011; Chen and Zhang 2012)). The correlation between orthologs is strongest for recent speciations, and decreases linearly with divergence time. This decrease shows that

we are able to detect a strong evolutionary signal in tissue-specificity, which has not always been obvious in functional comparisons of orthologs (e.g. (Nehrt et al. 2011; Altenhoff et al. 2012)).

Correlation between within-species paralogs is much lower than between orthologs. Whereas the expression of young paralogs has been recently reported to be highly conserved (Assis and Bachtrog 2015), we find a large difference between even very young paralogs in tissue-specificity. In Assis and Bachtrog (2015), the measure of tissue-specificity is not clearly defined, but it seems to be TSI (Julien et al. 2012), which performed poorly as an evolutionarily relevant measure in our recent benchmark (Kryuchkova-Mostacci and Robinson-Rechavi 2016); they also treated female and male samples as different "tissues", confounding two potentially different effects. The low correlation that we observed for young paralogs does not decrease significantly with divergence time. It is possible that on the one hand paralogs do diverge in tissue-specificity with time, and that on the other hand this trend is compensated by biased loss of the most divergent paralogs. It is also possible that we lack statistical power to detect a slight decrease in correlation of paralogs, due to low numbers of paralogs for many branches of the phylogeny. The most likely interpretation is that for small-scale paralogs there is an asymmetry with a daughter gene which lacks regulatory elements of the parent gene upon birth; further independent changes in tissue-specificity in each paralog would preserve the original lack of correlation. In any case, we do not find support for a progressive divergence of tissue-specificity for paralogs.

The overall conservation of tissue-specificity could be due to a subset of genes, and most notably sex-related genes. Indeed, the largest set of tissue-specific genes are testis-specific (Kryuchkova-Mostacci and Robinson-Rechavi 2016). To verify the influence of sex-related genes, we performed all analyses without testis expression data, or without genes mapped to sex chromosomes. After removing testis expression from all datasets the correlation between paralogs does not change significantly, while between orthologs is gets significantly weaker. The lower correlation of orthologs suggests that testis specific genes are conserved between species, and as they constitute a high proportion of tissue-specific genes, they contribute strongly to the correlation. Removing sex chromosome located genes does not change results significantly. After removing testis expression the differences of conservation of tissue-specificity between orthologs and paralogs stay significant. Overall, it appears that tissue-specificity calculated with testis represents a true biological signal, and given its large effect it is important to include this tissue in analyses.

In general paralogs are more tissue-specific and have lower expression levels. This could be explained if ubiquitous genes are less prone to duplication or duplicate retention. Yet we do

not observe any bias in the orthologs of duplicates towards more tissue-specific genes (Fig 3; see also Supplementary Materials). With time both paralogs get more broadly expressed (Fig 1 and Fig S13). In the rare case where both paralogs are tissue-specific, small-scale young paralogs are expressed in the same tissue, while genome-wide old paralogs (ohnologs) are expressed in different tissues (Fig 4). With the data available, we cannot distinguish the effects of paralog age and of duplication mechanism, since many old paralogs are due to whole genome duplication in vertebrates, whereas that is not the case for the young paralogs. In many cases the higher expressed paralog has a similar tissue-specificity to the ancestral state, while the lower expressed paralog is more tissue-specific (Fig 3).

The overall picture that we obtain for the evolution of tissue-specificity is the following. In the absence of duplication, tissue-specificity evolves slowly, thus is mostly conserved, and tissue-specific genes do not change their main tissue of expression (Fig 2 and 4). After small-scale duplication (i.e., not whole genome) paralogs diverge rapidly in tissue-specificity, or already differ at birth. This difference is mostly due to the less expressed paralog losing the ancestral specificity, while the most expressed paralog keeps at first closer to the ancestral state, as estimated from a non duplicated outgroup ortholog (Fig 3). But over time, even the most expressed paralog diverges much more strongly than a non duplicated ortholog. While paralog divergence is rapid, in the small number of genes which stay tissue-specific for both paralogs the main tissue of expression is mostly conserved, for several hundred million years (i.e. origin of tetrapods, Fig 4). With increasing age of the paralogs, they both tend to become more broadly expressed (Fig 1 and S13) while keeping a low correlation. For whole genome duplicates we have less information, because of the age of the event in vertebrates and the lack of good outgroup data. The main difference is that when two genome duplication paralogs are both tissue-specific, they are often expressed in different tissues (Fig 4).

We have studied gene specificity without taking in account alternative splicing, or the possibility that different transcripts are expressed in different tissues, because it is still difficult to call transcript level expression reliably (Pelechano et al. 2014). This would probably not change our main observations, that tissue-specificity is conserved among orthologs, diverges with evolutionary time, and follows the ortholog conjecture. Of note, recent results have not supported an important role of alternative splicing for differences in transcription between tissues (Ezkurdia et al. 2015).

We have used tissue-specificity to estimate the conservation of function, rather than Gene Ontology annotations or expression levels. We believe that this metric is less prone to systematic errors, whether annotation biases for the Gene Ontology, or proper normalisation between datasets and choice of few tissues for expression levels.

Our results confirm the Ortholog Conjecture on data which is genome-wide and functionally relevant: orthologs are more similar than within-species paralogs. Moreover, orthologs diverge monotonically with time, as expected. On the contrary, even young paralogs show high divergence.

Material and Methods

RNA-seq data from 12 species (human, gorilla, chimpanzee, macaque, mouse, platypus, opossum, chicken, gorilla, cow, frog, rat and fruit fly) were used for the analysis. For human, mouse and chicken we used several datasets. All the datasets with the corresponding number of tissues are summarized in Table 1. The numbers of genes used for the analysis are in Table S1 and S2.

The orthology and paralogy calls and their phylogenetic dating were taken from Ensembl Compara (Vilella et al. 2009).

Table 1: Datasets used in the paper.

Organisms/ datasets	Fagerberg	Brawand	Bodymap	ENCODE	Necsulea	Merkin	Keane
Dataset ID	E-MTAB-1733	GSE30352	GSE30611	GSE36025 (mouse)	GSE43520	GSE41637	GSE30617
RPKM/FPKM source	Supp. mat.	Bgee	Bgee	(1)	Bgee	Bgee	Bgee
Human *Homo sapiens*	27	8	16				
Gorilla *Gorilla gorilla*		6					
Chimpanzee *Pan troglodytes*		6					
Macaque *Macaca mulatta*		6				9	
Mouse *Mus musculus*		6		22		9	6
Rat *Rattus norvegicus*						9	
Cow *Bos taurus*						9	
Opossum *Monodelphis domestica*		6					
Platypus *Ornithorhynchus anatinus*		6					
Chicken *Gallus gallus*		6				9	
Frog *Xenopus tropicalis*					6		
Fly *Drosophila melanogaster*				6			
Citations	(Fagerberg et al. 2014)	(Brawand et al. 2011)	(Farrell et al. 2014)	(The ENCODE Project Consortium 2011; Li et al. 2014)	(Necsulea and Kaessmann 2014)	(Merkin et al. 2012)	(Keane et al. 2011)

1. Supp. mat.; (Kryuchkova-Mostacci and Robinson-Rechavi 2015)

For the human dataset from Fagerberg et al. (Fagerberg et al. 2014) and the fly dataset (Li et al. 2014), FPKM values were downloaded from the respective papers Supplementary Materials; the mouse ENCODE project dataset was processed by an in house script (TopHat and Cufflinks (Trapnell et al. 2012)); all other data were processed by the Bgee pipeline (Bastian et al. 2008) pipeline. For all analyses gene models from Ensembl version 75 were used (Flicek et al. 2013). Only protein-coding genes were used for analysis. For the analysis of paralogs the youngest couple was taken, and sorted according to the maximal expression, i.e. the reference paralog (called "gene" in our R scripts) is always the one with the highest maximal expression.

Analyses were performed in R version 3.2.1 (R Core Team 2015) using Lattice (Sarcar 2008), plyr (Wickham 2011), gplots (Warnes et al. 2016) and qvalue (Storey and Tibshirani 2003; Storey 2015) libraries.

As a measure for tissue specificity τ (Tau) was used (Yanai et al. 2005). We have recently shown that τ is the best choice for calculating tissue specificity among existing methods (Kryuchkova-Mostacci and Robinson-Rechavi 2016). For comparing tissue-specific genes, they were called with $\tau \geq 0.8$, and assigned to the tissue with the highest expression.

A special case is testis-specificity, as many more genes are expressed in testis than other tissues. For control analysis, all genes with maximal expression in testis were called "testis specific", independently of τ value.

Over all ANOVA tests performed (81 tests), we used a q-value threshold of 1% of false positives, corresponding to a p-value threshold of 0.0137.

Acknowledgements

We thank Julien Roux and Jialin Liu for their helpful comments and suggestions.

Supplementary Materials

Supplementary Materials are available online.

References

Altenhoff AM, Studer RA, Robinson-Rechavi M, Dessimoz C. 2012. Resolving the ortholog conjecture: orthologs tend to be weakly, but significantly, more similar in function than paralogs. PLoS Comput. Biol. 8:e1002514.

Assis R, Bachtrog D. 2015. Rapid divergence and diversification of mammalian duplicate gene functions. BMC Evol. Biol. 15:1–7.

Bastian F, Parmentier G, Roux J, Moretti S, Lauder V, Robinson-Rechavi M. 2008. Bgee: integrating and comparing heterogeneous transcriptome data among species. In: Data Integration in the Life Sciences. Springer Berlin Heidelberg. p. 124–131.

Brawand D, Soumillon M, Necsulea A, Julien P, Csárdi G, Harrigan P, Weier M, Liechti A, Aximu-Petri A, Kircher M, et al. 2011. The evolution of gene expression levels in mammalian organs. Nature 478:343–348.

Chen X, Zhang J. 2012. The ortholog conjecture is untestable by the current gene ontology but is supported by RNA sequencing data. PLoS Comput. Biol. 8:e1002784.

Chung W-Y, Albert R, Albert I, Nekrutenko A, Makova KD. 2006. Rapid and asymmetric divergence of duplicate genes in the human gene coexpression network. BMC Bioinformatics 7:1–14.

Ezkurdia I, Rodriguez JM, Carrillo-de Santa Pau E, Vázquez J, Valencia A, Tress ML. 2015. Most highly expressed protein-coding genes have a single dominant isoform. J. Proteome Res. 14:1880–1887.

Fagerberg L, Hallstrom BM, Oksvold P, Kampf C, Djureinovic D, Odeberg J, Habuka M, Tahmasebpoor S, Danielsson A, Edlund K, et al. 2014. Analysis of the human tissue-specific expression by genome-wide integration of transcriptomics and antibody-based proteomics. Mol. Cell. Proteomics 13:397–406.

Farrell CM, O'Leary NA, Harte RA, Loveland JE, Wilming LG, Wallin C, Diekhans M, Barrell D, Searle SMJ, Aken B, et al. 2014. Current status and new features of the Consensus Coding Sequence database. Nucleic Acids Res. 42:D865–D872.

Flicek P, Ahmed I, Amode MR, Barrell D, Beal K, Brent S, Carvalho-Silva D, Clapham P, Coates G, Fairley S, et al. 2013. Ensembl 2013. Nucleic Acids Res. 41:D48–D55.

Gabaldón T, Koonin E V. 2013. Functional and evolutionary implications of gene orthology. Nat. Rev. Genet. 14:360–366.

Gilad Y, Mizrahi-Man O. 2015. A reanalysis of mouse ENCODE comparative gene expression data. F1000Research 4:121.

Gout J-F, Lynch M. 2015. Maintenance and loss of duplicated genes by dosage subfunctionalization. Mol. Biol. Evol. 32:2141–2148.

Gu Z, Rifkin SA, White KP, Li W-H. 2004. Duplicate genes increase gene expression diversity within and between species. Nat. Genet. 36:577–579.

Huerta-Cepas J, Dopazo J, Huynen MA, Gabaldón T. 2011. Evidence for short-time divergence and long-time conservation of tissue-specific expression after gene duplication. Brief. Bioinform. 12:442–448.

Huminiecki L, Wolfe KH. 2004. Divergence of spatial gene expression profiles following species-specific gene duplications in human and mouse. Genome Res. 14:1870–1879.

Julien P, Brawand D, Soumillon M, Necsulea A, Liechti A, Schütz F, Daish T, Grützner F, Kaessmann H. 2012. Mechanisms and evolutionary patterns of mammalian and avian dosage compensation. PLoS Biol. 10:e1001328.

Keane TM, Goodstadt L, Danecek P, White M a, Wong K, Yalcin B, Heger A, Agam A, Slater G, Goodson M, et al. 2011. Mouse genomic variation and its effect on phenotypes and gene regulation. Nature 477:289–294.

Kryuchkova-Mostacci N, Robinson-Rechavi M. 2015. Tissue-specific evolution of protein coding genes in human and mouse. PLoS One 10:e0131673.

Kryuchkova-Mostacci N, Robinson-Rechavi M. 2016. A benchmark of gene expression tissue-specificity metrics. Brief. Bioinform. forthcomin.

Lan X, Pritchard JK. 2016. Coregulation of tandem duplicate genes slows evolution of subfunctionalization in mammals. Science. 352:1009–1013.

Li JJ, Huang H, Bickel PJ, Brenner SE. 2014. Comparison of D. melanogaster and C. elegans

developmental stages, tissues, and cells by modENCODE RNA-seq data. Genome Res. 24:1086–1101.

Liao B-Y, Zhang J. 2006. Evolutionary conservation of expression profiles between human and mouse orthologous genes. Mol. Biol. Evol. 23:530–540.

Merkin J, Russell C, Chen P, Burge CB. 2012. Evolutionary dynamics of gene and isoform regulation in mammalian tissues. Science. 338:1593–1599.

Necsulea A, Kaessmann H. 2014. Evolutionary dynamics of coding and non-coding transcriptomes. Nat. Rev. Genet. 15:734–748.

Nehrt NL, Clark WT, Radivojac P, Hahn MW. 2011. Testing the ortholog conjecture with comparative functional genomic data from mammals. PLoS Comput. Biol. 7:e1002073.

Pelechano V, Wei W, Jakob P, Steinmetz LM. 2014. Genome-wide identification of transcript start and end sites by transcript isoform sequencing. Nat. Protoc. 9:1740–1759.

Piasecka B, Robinson-Rechavi M, Bergmann S. 2012. Correcting for the bias due to expression specificity improves the estimation of constrained evolution of expression between mouse and human. Bioinformatics 28:1865–1872.

R Core Team. 2015. R: A language and environment for statistical computing. :R Foundation for Statistical Computing, Vienna.

Rogozin IB, Managadze D, Shabalina SA, Koonin E V. 2014. Gene family level comparative analysis of gene expression n mammals validates the ortholog conjecture. Genome Biol. Evol. 6:754–762.

Sarcar D. 2008. Lattice: Multivariate data visualization with R. New York: Springer

Storey J, Tibshirani R. 2003. Statistical significance for genomewide studies. Proc Natl Acad Sci U S A 2003.

Storey JD. 2015. Qvalue: Q-value estimation for false discovery rate control.

Studer R, Penel S, Duret L, Robinson-Rechavi M. 2008. Pervasive positive selection on duplicated and nonduplicated vertebrate protein coding genes. Genome Res. 18:1393–1402.

Studer RA, Robinson-Rechavi M. 2009. How confident can we be that orthologs are similar, but paralogs differ? Trends Genet. 25:210–216.

Studer RA, Robinson-Rechavi M. 2010. Large-scale analysis of orthologs and paralogs under covarion-like and constant-but-different models of amino acid evolution. Mol. Biol. Evol. 27:2618–2627.

The ENCODE Project Consortium. 2011. A user's guide to the encyclopedia of DNA elements (ENCODE). PLoS Biol. 9:e1001046.

Thomas PD, Wood V, Mungall CJ, Lewis SE, Blake JA. 2012. On the use of gene ontology annotations to assess functional similarity among orthologs and paralogs: A short report. PLoS Comput. Biol. 8:1–7.

Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, Pimentel H, Salzberg SL, Rinn JL, Pachter L. 2012. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nat. Protoc. 7:562–578.

Vilella AJ, Severin J, Ureta-Vidal A, Heng L, Durbin R, Birney E. 2009. EnsemblCompara GeneTrees: Complete, duplication-aware phylogenetic trees in vertebrates. Genome Res. 19:327–335.

Warnes G, Bolker B, Bonebakker L, Gentleman R, Huber W, Liaw A, Lumley T, Maechler M, Magnusson A, Moeller S, et al. 2016. Gplots: Various R programming tools for

plotting data.

Wickham H. 2011. The Split-Apply-Combine strategy for data analysis. J. Stat. Softw. 40:1–29.

Wu Y-C, Bansal MS, Rasmussen MD, Herrero J, Kellis M. 2014. Phylogenetic identification and functional characterization of orthologs and paralogs across human, mouse, fly, and worm. bioRxiv.

Yanai I, Benjamin H, Shmoish M, Chalifa-Caspi V, Shklar M, Ophir R, Bar-Even A, Horn-Saban S, Safran M, Domany E, et al. 2005. Genome-wide midrange transcription profiles reveal expression level relationships in human tissue specification. Bioinformatics 21:650–659.

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.