DNA Research **19**, 305–316, (2012)
Advance Access publication on 5 April 2012
doi:10.1093/dnares/dss012

Evolution of Vertebrate Tissues Driven by Differential Modes of Gene Duplication

Masanobu Satake[1], Masakado Kawata[2], Aoife McLysaght[3], and Takashi Makino[2,*]

Department of Molecular Immunology, Institute of Development, Aging and Cancer, Tohoku University, Seiryo-machi 4-1, Aoba-ku, Sendai 980-8575, Japan[1]; Department of Ecology and Evolutionary Biology, Graduate School of Life Sciences, Tohoku University, Aramaki-Aoba 6-3, Aoba-ku, Sendai 980-8578, Japan[2] and Smurfit Institute of Genetics, University of Dublin, Trinity College, Dublin 2, Ireland[3]

*To whom correspondence should be addressed. Tel. +81 22-795-6689. Fax. +81 22-795-6689.
E-mail: tamakino@m.tohoku.ac.jp

Edited by Toshihiko Shiroishi
(Received 29 December 2011; accepted 4 March 2012)

Abstract

In this study, we investigated the evolution of vertebrate tissues by examining the potential association among gene expression, duplication, and base substitution patterns. In particular, we compared whole-genome duplication (WGD) with small-scale duplication (SSD), as well as tissue restricted with ubiquitously expressed genes. All patterns were also analysed in the light of gene evolutionary rates. Among those genes characterized by rapid evolution and expressed in a restricted range of tissues, SSD was represented in a larger proportion than WGD. Conversely, genes with ubiquitous expression were associated with slower evolutionary rates and a larger proportion of WGD. The results also show that evolutionary rates were faster in genes expressed in endodermal tissues and slower in ectodermal genes. Accordingly, the proportion of the SSD and WGD genes was highest in the endoderm and ectoderm, respectively. Therefore, quickly evolving SSD genes might have contributed to the faster evolution of endodermal tissues, whereas the comparatively slowly evolving WGD genes might have functioned to maintain the basic characteristics of ectodermal tissues. Mesenchymal tissues occupied an intermediate position in this regard, whereas the patterns observed for haemocytes were unique. Rapid tissue evolution could be related to a specific gene duplication mode (SSD) and faster molecular evolution in response to exposure to the external environment. These findings reveal general patterns underlying the evolution of tissues and their corresponding genes.

Key words: gene duplication; ohnologue; gene expression; tissue evolution

1. Introduction

Mutations are the main factors driving genome evolution and may happen within genes through base substitutions or involve their entire duplication.[1] In the latter case, two mechanisms have been recognized: whole-genome duplication (WGD) and small-scale duplication (SSD), which occurs in relatively a small region of the genome during evolution. For instance, the early vertebrate ancestor is thought to have undergone two rounds of WGD,[2–5] as suggested by four vertebrate Hox gene clusters located in different chromosomes. Paralogous genes originated from WGD are referred to as ohnologues.[6] Singletons, in turn, are genes that did not undergo either WGD or SSD.

In general, duplicated genes are redundant and their functions overlap. Thus, in simple organisms such as yeasts and nematodes, the proportion of essential genes in duplicated genes is half as low as

that in singletons.[7,8] In mice, however, the proportion of essential genes is comparable between duplicated genes and singletons.[9,10] Furthermore, ohnologues are likely to contain a larger proportion of essential genes than SSD genes.[11] Ohnologues are indeed associated with development, the regulation of transcription, and protein complex formation.[11] For these genes to function properly, the relative amount of their products must be balanced.[12–14] WGD typically involves the simultaneous duplication of all genes, therefore preserving the relative dosages of each gene. Because either the loss or the gain of ohnologues may lead to dosage imbalance, ohnologues are expected to be retained intact in the genome.[15] Similarly, ancient WGD-derived ohnologues are expected to be more conservative than recently evolved SSD genes.

Gene expression can be tissue-specific, determining phenotypes such as the morphology and function of tissues, or ubiquitous. Previous research indicates that ubiquitously expressed genes are likely to evolve slowly.[16–18] For example, in humans and mice, the orthologous genes that are expressed in a limited number of tissues tend to evolve faster than ubiquitously expressed genes.[18] Little is known, however, about the extent to which the evolution of tissues is influenced by the differential modes of gene duplication and expression. Indeed, there are no reports exploring the effects of gene duplication events, such as WGD and SSD, on tissue-restricted or ubiquitous gene expression.

In this study, the potential associations between gene evolutionary rates, duplication (WGD, SSD, and singletons), and gene expression breadth in different tissues (restricted or ubiquitous) were investigated. In addition, these parameters were also analysed in relation to the developmental origin of tissues (endodermal, mesenchymal, or ectodermal). The results support the notion that both base substitutions within genes and gene duplication are associated with gene expression breadth and that the nature of duplication (WGD or SSD) differs substantially depending on the germ-layer origin of the tissue. Tissue evolution is therefore discussed here as the outcome of a process involving the gene evolutionary rate, duplication, and expression.

2. Materials and Methods

2.1. *Classification of human genes based on the gene duplication mode*

Protein-coding genes of human origin were obtained from Ensembl release 52 (http://www.ensembl.org). A total of 7294 ohnologues and 9027 SSD genes were defined as described in Makino and McLysaght.[15] Briefly, duplicated genes were so judged when the two aligned sequences showed homology in their $>30\%$ length with $e < 10^{-7}$ in BLAST search. Ohnologues were syntenic genes located on paralogous chromosomal regions and derived from WGD, whereas SSD genes were duplicated genes not experiencing WGD. Of the 9027 SSD genes, 1478 genes were classified as both ohnologues and SSD genes and, therefore, excluded from the analysis, resulting in 7549 pure SSD genes. An additional 6064 genes were classified as singletons. Thus, 20 907 genes ($7294 + 7549 + 6064 = 20\,907$) were considered.

To define the origin of SSDs, a sequence similarity search was performed within protein-coding human genes using the all-against-all BLASTP program. Synonymous substitution rates (K_S) were estimated for each close paralogue. There were 2510 and 5039 SSD genes for which K_S were <1 (recent SSD) and ≥ 1 (old SSD), respectively ($2510 + 5039 = 7549$).

2.2. *Human orthologous genes and gene evolutionary rates*

Human genes with orthologues in mice and other species were obtained from Ensembl release 52. Orthologous sequences were aligned using CLUSTALW,[19] and K_S and non-synonymous substitution rates (K_A) were deduced for each orthologous pair using the method of Yang and Nielsen[20] implemented in PAML.[21] Next, ω values (K_A/K_S) were calculated.

2.3. *Human EST data*

An expression sequence tag (EST) database[22] was used to determine the expression profile of human genes in various tissues. The data were registered at NCBI and included 3 199 559 reads from 47 different human tissues. Each tissue contained more than 10 000 ESTs (68 076 on an average). Using these 47 tissues, the breadth of expression of a gene was represented by the number of tissues in which the gene was identified, as based on the detection of ESTs.[22] Thus, breadth varied from 1 to 47. Among the 20 907 human genes considered for analysis, there were 3871 for which ESTs were not identified in the EST database. These genes were excluded from further analyses, therefore resulting in a total number of genes of 17 036, among which 6 952 were ohnologues, 5505 were SSDs, and 4579 were singletons.

2.4. *Classification of human tissues based on the developmental origin*

Forty-three tissues were classified into four subgroups based on their developmental origin, as

follows: endoderm (nasopharynx, thymus, stomach, colon, bladder, liver, trachea, lung, pancreas, uterus, cervix, prostate, and intestine), ectoderm (breast, skin, caudate nucleus, hypothalamus, eye, thalamus, subthalamic nucleus, cerebellum, hippocampus, corpus callosum, nervous system, amygdala, and substantia nigra), mesenchymal (synovium, kidney, adipose tissue, bone, adrenal gland, cartilage, muscle, pericardium, and heart), and haemocytes (B cells, bone marrow, germinal centre B cells, blood cells, lymph node, spleen, blood vessels, and T lymphocytes). The developmental origin of four tissues, such as ovary, testis, amnion, and placenta, was not categorized into any of the above four subgroups.

One caveat of this classification, however, is that the intestine is classified as an endodermal tissue based on the presence of endoderm-derived intestinal epithelium. However, as a macroscopic organ, the intestine includes not only epithelium but also mesoderm-derived tissues such as the submucosa and muscles. Thus, in the above classification, an endodermal tissue usually includes both endoderm- and mesoderm-derived tissues, whereas an ectodermal tissue includes both ectoderm- and mesoderm-derived tissues.

2.5. *Definition of tissue evolutionary rates*

The tissue evolutionary rate was originally calculated by Kuma *et al.*[23] and this method has been adopted in the succeeding studies.[18,24] Thus, we simply employed their definition in the present study. We assumed that a set of genes are expressed in a given tissue type and that the ω values of expressing genes are $X_1, X_2, \ldots, X_m$ (for ohnologues), $Y_1, Y_2, \ldots, Y_n$ (for SSD genes), and $Z_1, Z_2, \ldots, Z_o$ (for singletons). The evolutionary rate of this tissue is then given by: $[(X_1 + X_2 + \cdots + X_m) + (Y_1 + Y_2 + \cdots + Y_n) + (Z_1 + Z_2 + \cdots + Z_o)]/(m + n + o)$, where m, n, and o are the respective number of ohnologues, SSD genes, and singletons.

3. Results

3.1. *Association between expression breadth and duplication mode*

Figure 1 shows the gene number distribution over various expression breadths (note that the actual number of genes expressed in each tissue at each expression breadth is shown in Supplementary Table S1). Figure 1A–C displays the numbers of ohnologues, SSD genes, and singletons, respectively. The average number of tissues in which each gene type was expressed was 18.7 for ohnologues (blue), 16.6 for SSD genes (red), and 18.0 for singletons (green). Figure 1D shows that the proportions of each gene type in each expression breadth were noticeably different. Specifically, the proportion of SSD genes was relatively higher among breadth-restricted genes (see $n = 1$, $n \leq 3$, $n \leq 5$, and $n \leq 10$), whereas the proportion of ohnologues was increasingly higher among genes expressed in a larger number of tissues (see $n > 10$, $n > 20$, and $n > 40$).

3.2. *Association between gene evolutionary rates and duplication mode*

Ohnologues are believed to be more conservative than SSD genes with respect to functional essentiality and dosage-balance requirement.[11,15] To determine whether ohnologues are also conservative in terms of molecular evolution, non-synonymous nucleotide divergence in the coding region (between humans and mice) were examined. Table 1 shows that the average ω value of ohnologues (0.11) was 0.55–0.57-fold lower than that of SSD genes (0.19; $P < 2.2 \times 10^{-16}$, the Mann–Whitney U-test) and of singletons (0.20; $P < 2.2 \times 10^{-16}$). The result confirms the conservative nature of ohnologues in the evolution of coding regions.

3.3. *Tissue evolutionary rates in restricted expression breadths*

In Fig. 2A, tissue evolutionary rates are plotted in an increasing order, for each of 47 tissues, and under a condition of $n \leq 10$ (see 'Proportions of ohnologues and SSD genes in various expression breadths' as for a rationale of the use of $n \leq 10$). The ω values gradually increased from 0.13 in a slowly evolving, leftward tissue to 0.29 in a fast evolving, rightward tissue.

3.4. *Gene evolutionary rate in restricted expression breadths*

The average ω values were calculated separately for each type of gene as follows: $(X_1 + X_2 + \cdots + X_m)/m$ for ohnologues, $(Y_1 + Y_2 + \cdots + Y_n)/n$ for SSD genes, and $(Z_1 + Z_2 + \cdots + Z_o)/o$ for singletons. Figure 2B shows that the evolutionary rate (ω) of ohnologues (blue), SSD genes (red), and singletons (green) increased gradually in line with increases in tissue evolutionary rates, indicating that all gene types evolved in parallel with the tissue evolution. Therefore, gene evolutionary rates were higher in the genes that are expressed in fast evolving tissues irrespective of the gene type, whereas those were lower in slowly evolving tissues regardless of the gene type. It is worth noticing, however, that the average ω value of ohnologues was low (0.13), whereas that of SSD genes and singletons was high (0.26 and 0.29, respectively; Fig. 2B and see $n \leq 10$ in Table 1).

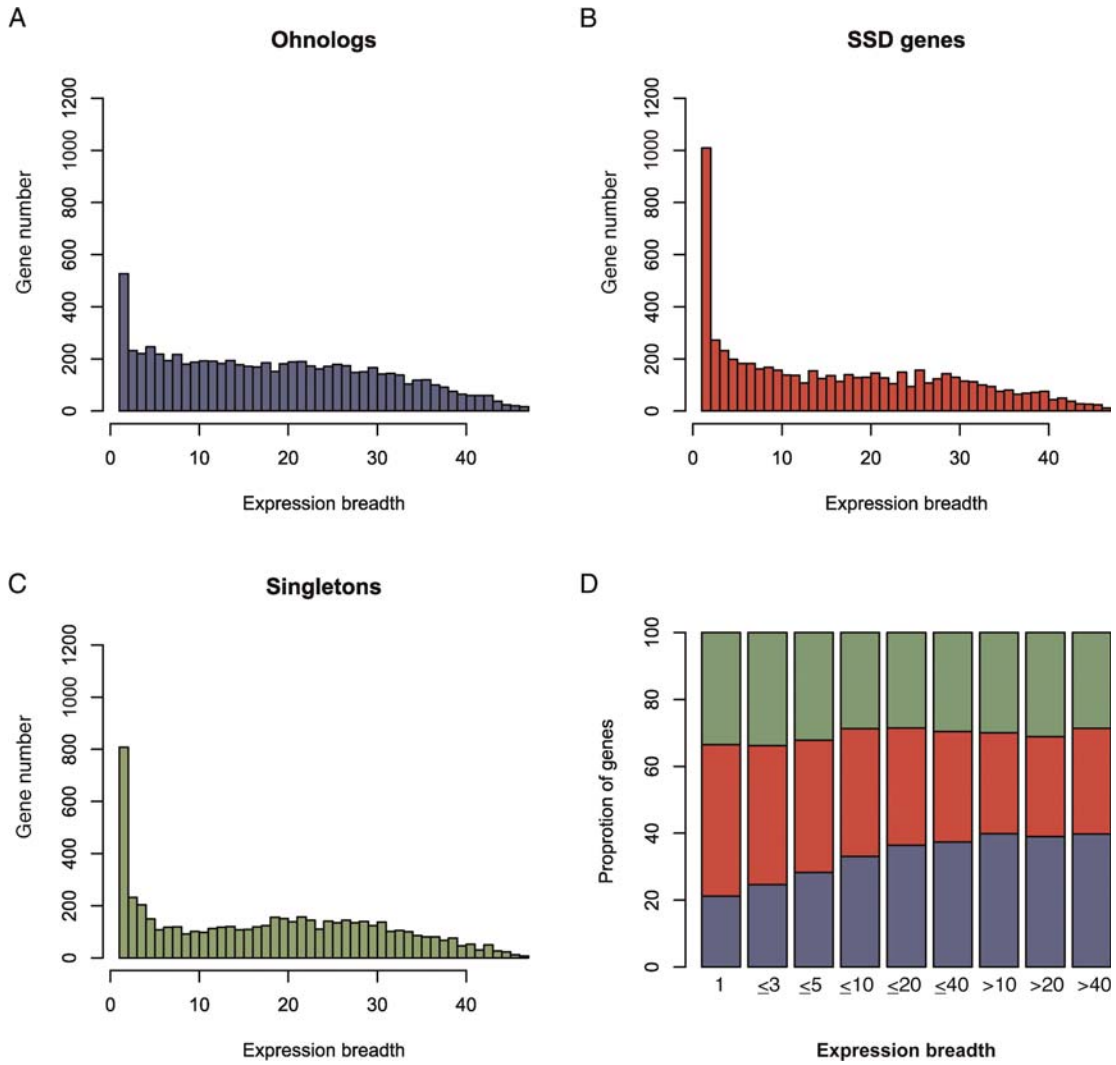

Figure 1. The number and proportion of ohnologues, SSD genes, and singletons at each tissue expression breadth (as determined by the number of tissues where a corresponding EST was detected for a respective gene). The number of genes at each expression breadth was counted for ohnologues (A, blue), SSD genes (B, red), and singletons (C, green). (D) The percentage of ohnologues, SSD genes, and singletons at each breadth.

Table 1. Average K_A/K_s values

Category	Subcategory	Number of compared orthologous gene pairs	Average K_A/K_s	SD
Expressed genes	Ohnologues	6952	0.11	0.11
	Total SSD genes	5505	0.19	0.16
	Old SSD genes ($K_s \geq 1$)	4356	0.17	0.14
	Recent SSD genes ($K_s < 1$)	1149	0.29	0.20
	Singletons	4579	0.20	0.22
Narrowly expressed genes ($n \leq 10$)	Ohnologues	2151	0.13	0.12
	Total SSD genes	2008	0.26	0.18
	Old SSD genes ($K_s \geq 1$)	1338	0.22	0.16
	Recent SSD genes ($K_s < 1$)	670	0.34	0.20
	Singletons	1101	0.29	0.19
Broadly expressed genes ($n > 10$)	Ohnologues	4801	0.10	0.11
	Total SSD genes	3497	0.15	0.14
	Old SSD genes ($K_s \geq 1$)	3018	0.14	0.12
	Recent SSD genes ($K_s < 1$)	479	0.23	0.19
	Singletons	3478	0.17	0.23

3.5. *Proportions of gene type in restricted expression breadths*

Figure 2C shows that an increase in tissue evolutionary rates was accompanied by a corresponding decrease in the proportion of ohnologues and an increase in the proportion of SSD genes. This result indicates that ohnologues and SSD genes tended to be expressed in slowly and fast evolving tissues, respectively. The proportion of singletons also appeared to increase in parallel with tissue evolutionary rates.

Figure 2B and C also indicates that ohnologues and SSD genes behave differently. In the case of SSD genes, ω values as well as proportions are positively associated with tissue evolutionary rates, suggesting that SSD genes contribute to faster tissue evolution. In the case of ohnologues, however, an increase in ω was associated with a decrease in proportion, accompanying the tissue evolutionary rate. Therefore, further analysis is needed to examine the relative contribution of ohnologues to tissue evolution.

3.6. *Differential contribution of ohnologues, SSD genes, and singletons among genes with restricted expression*

Both the evolutionary rate of each gene type and their proportion are incorporated in the following formula:

$$\begin{aligned}&\frac{(X_1+X_2+\cdots+X_m)+(Y_1+Y_2+\cdots+Y_n)+(Z_1+Z_2+\cdots+Z_o)}{m+n+o}\\&=\frac{X_1+X_2+\cdots+X_m}{m}\times\frac{m}{m+n+o}+\frac{Y_1+Y_2+\cdots+Y_n}{n}\times\frac{n}{m+n+o}\\&+\frac{Z_1+Z_2+\cdots+Z_o}{o}\times\frac{o}{m+n+o},\end{aligned}$$

where $[(X_1+X_2+\cdots+X_m)/m]\times[m/(m+n+o)]$ represents ohnologues, $[(Y_1+Y_2+\cdots+Y_n)/n]\times[n/(m+n+o)]$ represents SSD genes, and $[(Z_1+Z_2+\cdots+Z_o)/o]\times[o/(m+n+o)\}$ represents singletons. Figure 2D shows each ohnologue, SSD, and singleton component in relation to tissue evolutionary rates. For SSD genes and singletons, there was a positive association between tissue and gene evolutionary rates. In contrast, ohnologues were almost flat irrespective of tissue evolution. This flat line for ohnologues suggests that ohnologues did not play a major role in tissue evolution, in agreement with the notion that ohnologues are conservative in nature. However, this does not necessarily mean a lack of any role for ohnologues in tissue evolution. In fact, the proportions of ohnologues were substantially reduced in fast evolving tissues. This effect and the increase in the ω values cancelled out each other.

3.7. *Equal contribution of ohnologues, SSD genes, and singletons among genes with ubiquitous expression*

Figure 2E–H represents the results obtained under a condition of $n > 10$ tissues and show virtually no variation (flat lines) for all parameters measured and gene types among tissues. This is not surprising given the ubiquitous nature of the expression of each gene type in a broad range of tissues. The average ω values when $n > 10$ tissues were remarkably lower, particularly for SSD genes and singletons (0.15 and 0.17, respectively), than those in observed when $n \leq 10$ tissues (0.26 and 0.29, respectively; Fig. 2F and B and Table 1). In terms of the proportion of gene types (Fig. 2G), the proportion of ohnologues (40%) was substantially higher than that of SSD genes and singletons (30% in both cases). This result is in sharp contrast to that found for expression breadths of $n \leq 10$ tissues, indicating that among those genes that are expressed ubiquitously, there is a larger proportion of ohnologues and a lower proportion of SSD genes and singletons. Finally, the relative contribution of ohnologues, SSD genes, and singletons is nearly the same for all tissues (Fig. 2H), which is in high contrast to the situation where $n \leq 10$ tissues (Fig. 2D).

Overall, the results show that the relative proportion of ohnologues, SSD genes, and singletons, as well as their evolutionary rates, are substantially different between genes with ubiquitous and restricted expression breadths.

3.8. *Proportions of ohnologues and SSD genes in various expression breadths*

To confirm the previous findings, we analysed cases involving expression breadths other than $n \leq 10$ and $n > 10$ tissues. Similar results to those shown in Fig. 2A–D were obtained for $n = 1$, $n \leq 3$, $n \leq 5$, $n \leq 20$, and $n \leq 40$ tissues, whereas similar observations as those indicated in Fig. 2E–H were obtained for $n > 20$ and $n > 40$ tissues (data not shown). Figure 3 shows the proportion of ohnologues (blue) and SSD genes (red) for each tissue under various expression breadths ($n = 1$, $n \leq 3$, $n \leq 5$, $n \leq 10$, $n \leq 20$, $n \leq 40$, $n > 10$, $n > 20$, and $n > 40$). The x-axis represents the tissue evolutionary rate, whereas the y-axis represents the proportions of ohnologues and SSD genes. In tissue-restricted expression ($n = 1$, $n \leq 3$, $n \leq 5$, $n \leq 20$, and $n \leq 40$), the faster the tissue evolutionary rates, the higher the proportion of SSD genes and the lower the proportion of ohnologues expressed (correlation coefficients and P-values are shown in

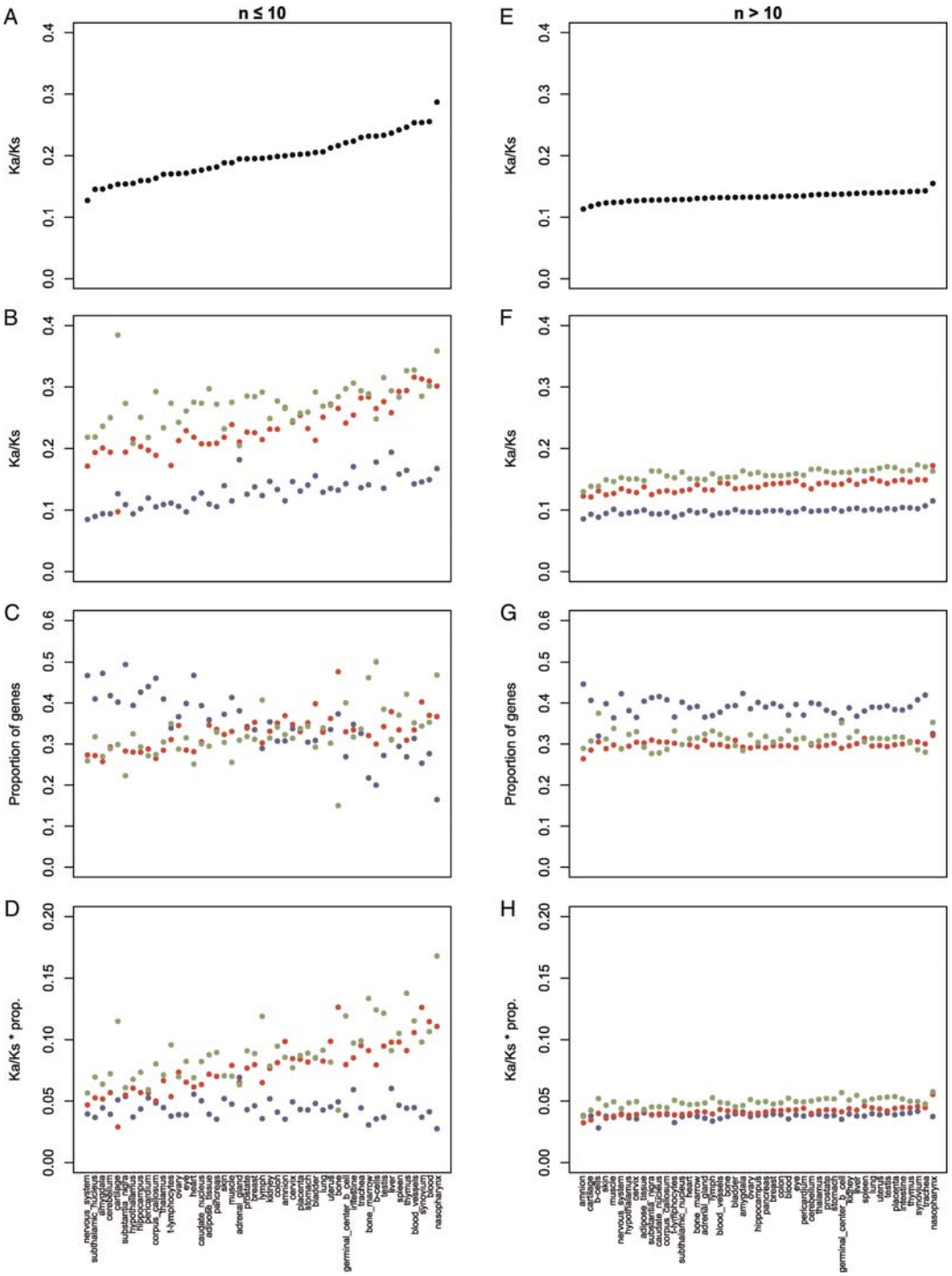

Figure 2. Contribution of ohnologues, SSD genes, and singletons among breadth-restricted ($n \leq 10$ tissues) and breadth-ubiquitous ($n > 10$ tissues) genes. Blue, red, and green represent the ohnologues, SSD genes, and singletons, respectively. Tissues are aligned on the *x*-axis in the order of magnitude of the K_A/K_S (ω) values of their expressed genes. Parameters used in the *y*-axis are as follows: (A and E) the average ω values of all genes expressed in a given tissue, which corresponds to the defined tissue evolutionary rate; (B and F) the ω values of ohnologues, SSD genes, and singletons; (C and G) the proportion of ohnologues, SSD genes, and singletons; and (D and H) the ω values and the proportions calculated for ohnologues, SSD genes, and singletons. The order of tissues aligned are as follows from the left to the right. In (A)–(D), the nervous system, subthalamic nucleus, amygdale, cerebellum, cartilage, substantia nigra, hypothalamus, hippocampus, pericardium, corpus callosum, thalamus, T lymphocytes, ovary, eye, heart, caudate nucleus, adipose tissue, pancreas, skin, muscle, adrenal gland, prostate, breast, lymph, kidney, colon, amnion, cervix, placenta, stomach, bladder, lung, uterus, bone, germinal centre B cell, intestine, trachea, bone marrow, B cells, testis, liver, spleen, thymus, blood vessels, synovium, blood, and nasopharynx. In (E)–(H), the amnion, cartilage, B cells, skin, muscle, nervous system, hypothalamus, cervix, adipose tissue, substantia nigra, caudate nucleus, corpus callosum, T lymphocytes, subthalamic nucleus, heart, bone marrow, adrenal gland, lymph, blood vessels, bone, bladder, amygdale, ovary, hippocampus, pancreas, breast, colon, blood, eye, pericardium, cerebellum, thalamus, prostate, stomach, germinal centre B cell, kidney, liver, spleen, lung, uterus, testis, placenta, intestine, thymus, synovium, trachea, and nasopharynx.

Supplementary Table S2). When the expression breadth was ≤ 10, a statistically significant correlation was detected for both the ohnologues and SSD genes ($P = 1.6 \times 10^{-13}$ and $P = 2.1 \times 10^{-5}$, respectively). A similar correlation was observed when the expression breadth was set at ≤ 20 and ≤ 40. However, the correlation became rather weak in these cases. Therefore, at expression breadth ≤ 10, correlation became maximum and most reliable.

On the other hand, in those cases where $n > 20$ and $n > 40$ tissues, the proportion of ohnologues was higher and the proportion of SSD genes was lower but there was no significant correlation between tissue evolutionary rates and the proportion of either gene type.

3.9. *Orthologous genes between human and other vertebrates*

The previous analyses involved genes that are orthologous in human and mice. However, the origin of ohnologues can be traced back to the emergence of vertebrates. Therefore, human orthologues in various vertebrates, including rat, cow, dog, opossum, and chicken, were extracted from Ensembl. Using an expression breadth of $n \leq 10$, analyses similar to those shown in Fig. 2C were performed. In the results shown in Supplementary Fig. S1, human genes were used as references for pairwise comparisons. Similar to the results shown in Fig. 2C, the comparison between humans and other species (Supplementary Table S3) shows that those tissues with higher ω values were associated with a larger proportion of SSD genes than with ohnologues.

3.10. *Recent SSD genes contribute to faster tissue evolution*

Approximate gene duplication times can be estimated based on K_S values, with lower and higher K_S values corresponding to recent and ancient duplication events, respectively. To examine the relative contribution of each SSD gene to tissue expression, a threshold value of K_S was set to 1.0 and the number and the percentage of each SSD gene at various expression breadths were counted (Fig. 4A–C). Narrower expression breadths were associated with a larger percentage of recent SSD genes ($K_S < 1.0$, red), whereas the percentage of ancient SSD genes ($K_S \geq 1.0$, pink) was nearly constant regardless of expression breadth. These results indicate a higher contribution of recent SSD genes to tissue-restricted gene expression.

An additional analysis similar to that shown in Fig. 2C was performed using recent and ancient SSD genes as criteria. The results are shown in Fig. 4D (expression breadth ≤ 10 tissues). Co-linearity between ω values and gene proportions was 0.58 for recent ($P = 1.7 \times 10^{-5}$) and 0.33 for ancient SSD genes ($P = 0.024$), indicating a larger contribution of recent SSD genes to the positive co-linearity among all SSD genes observed in Fig. 2C. Yet, the slightly positive co-linearity observed for ancient SSD genes indicates that ancient SSD genes are distinct from ohnologues, which showed negative co-linearity. These results indicate that the evolution of SSD genes of both recent and ancient origins, as well as of ohnologues, is distinct in nature.

3.11. *Association between gene evolutionary rates, duplication mode, and developmental origin of tissues*

In the analyses of the results shown in Fig. 2, fast and slow tissue evolutionary rates were based on ω values but we did not address tissue types. To examine how ω values and the proportion of SSD genes and ohnologues are associated with the developmental origin of tissues, ω values as well as the proportions of ohnologues, SSD genes, and singletons were calculated for each tissue type. Values calculated for the four subgroups with a breadth of $n \leq 10$ tissues are presented in Fig. 5 (Supplementary Table S4 shows *P*-values).

The average ω values were significantly higher in endodermal, intermediate in mesenchymal, and lower in ectodermal tissues (Fig. 5A). Conversely, the proportion of ohnologues was higher in ectodermal than in endodermal and mesenchymal tissues (Fig. 5B), whereas the proportion of SSD genes was highest in endodermal, intermediate in mesenchymal, and lowest in ectodermal tissues (Fig. 5C). The proportion of singletons did not vary among the three subgroups (Fig. 5D). Therefore, it is possible that SSD genes with higher ω values might have contributed to the faster evolution of endodermal tissues, whereas ohnologues with lower ω values might have functioned to maintain the essential characteristics of ectodermal tissues. Mesenchymal tissues occupied an intermediate position.

3.12. *The unique evolutionary position of haemocytes*

Haemocyte genes displayed unique features that were distinct from those of the three subgroups of tissues described previously. Specifically, the average ω value of haemocyte genes was the highest, whereas the proportion of haemocyte ohnologues was the lowest. Furthermore, the proportion of haemocyte singletons was higher than that of the other three subtypes of tissues.

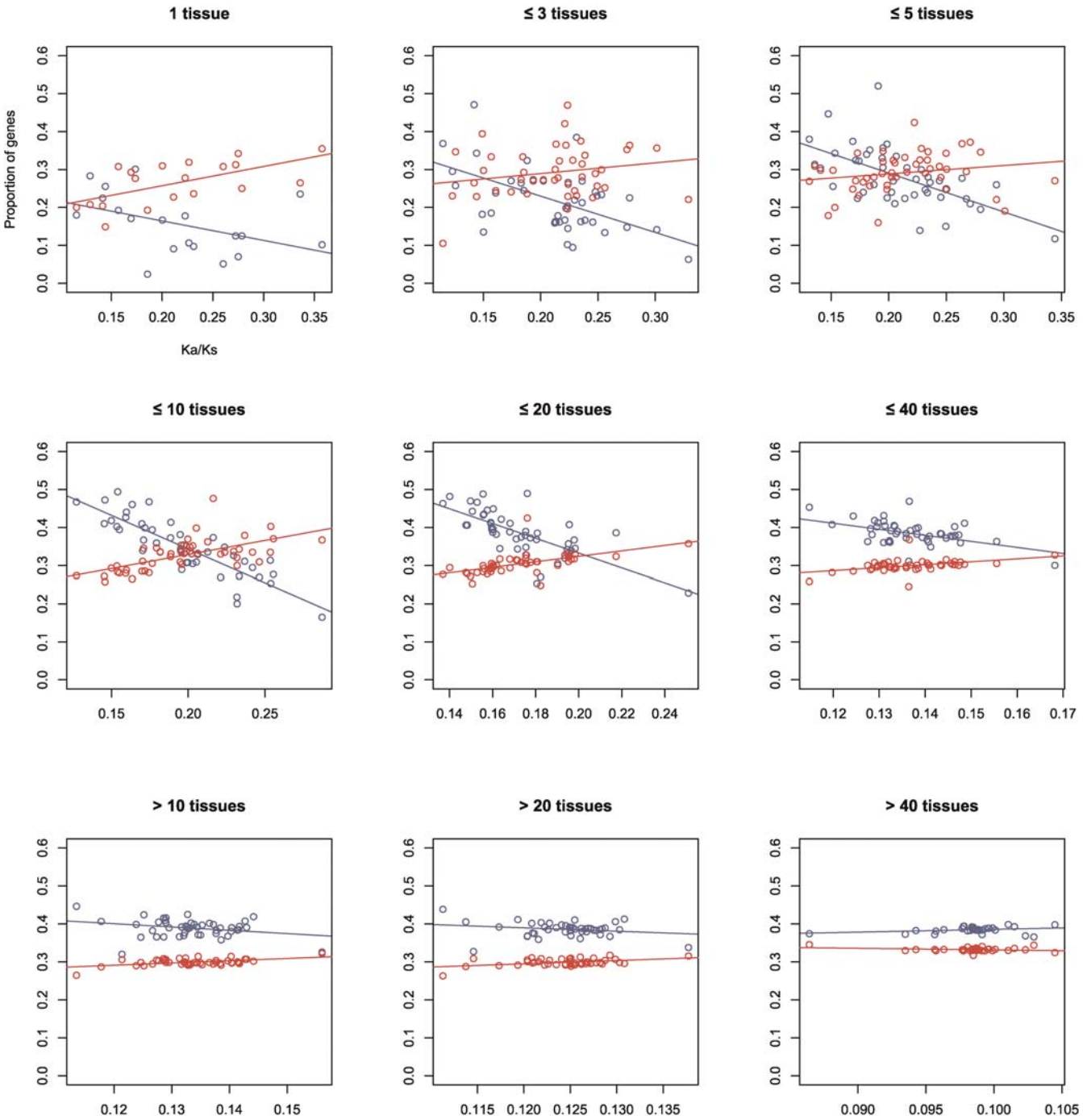

Figure 3. Proportion of ohnologues and SSD genes at various gene expression breadths. The *y*-axis represents the proportions of ohnologues (blue) and SSD genes (red), whereas the *x*-axis shows the tissue evolutionary rates [the average K_A/K_S (ω) value of all genes expressed in a given tissue]. *P*-values and linear correlation coefficients are shown in Supplementary Table S2.

4. Discussion

4.1. Gene evolutionary rate, duplication, and expression breadth

In the present study, the putative association between gene evolutionary rates, duplication, and their expression breadth in different tissues was examined. Gene evolutionary rates are affected by various factors. For example, duplicated genes are believed to evolve relatively fast, whereas singletons evolve more slowly.[1] Another important factor is the expression breadth of each gene. Genes expressed in a wide range of tissues tend to evolve more slowly, whereas those restricted to a narrow range of tissues evolve faster.[18,25] In addition, as the number of genes in a gene family increases, the number of tissues in

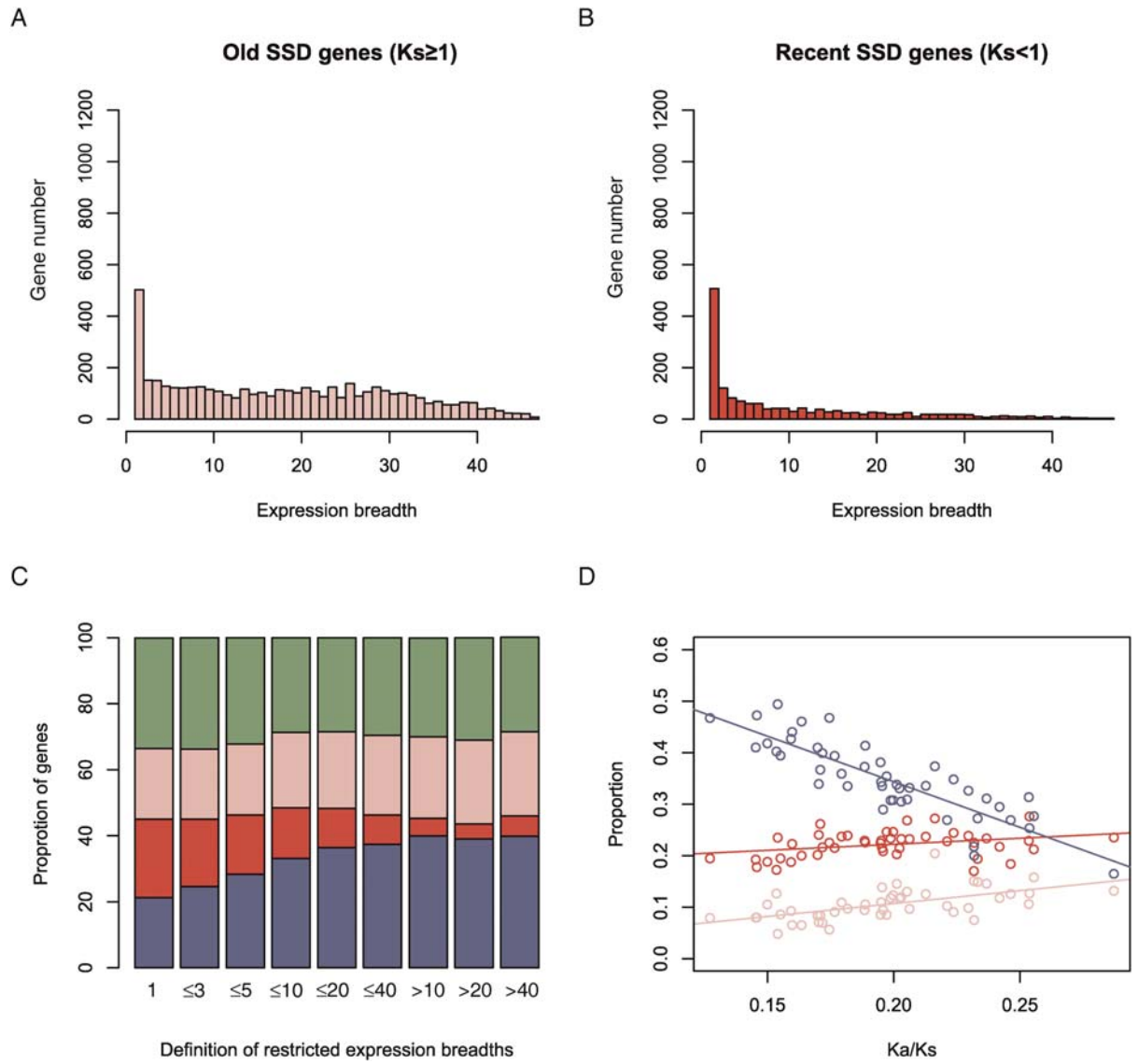

Figure 4. Recent and ancient SSD genes. (A–C) The number and the percentage of ancient (pink) and recent (red) SSD genes are shown for each tissue expression breadth. (D) The proportions of ancient and recent SSD genes among genes with tissue-restricted expression limited to $n \leq 10$ tissues.

which the corresponding gene is expressed decreases.[26] However, these findings were based on the analysis of only two parameters at a time, even though co-linearity among the three elements considered (gene evolutionary rate, gene duplication, and gene expression breadth) is also possible.

To analyse the involvement of duplicated genes on tissue evolution, we initially focused on the relationship between gene expression breadth and gene duplication. However, the percentages of singletons (Fig. 1D, green) and duplicated genes (Fig. 1D; red + blue) remained nearly constant at 30–35 and 65–70%, respectively, and were independent of gene expression breadth. This result suggests that the association between gene expression breadth and gene duplication is not as simple as reported previously.

The present study not only considered the analysis of gene evolutionary rates but also incorporated into it the distinction between SSD genes and ohnologues when considering gene duplication. Gene evolutionary rates were fast and slow for SSD genes and ohnologues, respectively (Table 1). Interestingly, SSD genes were represented in a larger proportion within narrow expression ranges, whereas the opposite was observed for ohnologues (Fig. 1D). The present study is therefore the first to show co-linearity among gene evolutionary rate, duplication, and expression breadth.

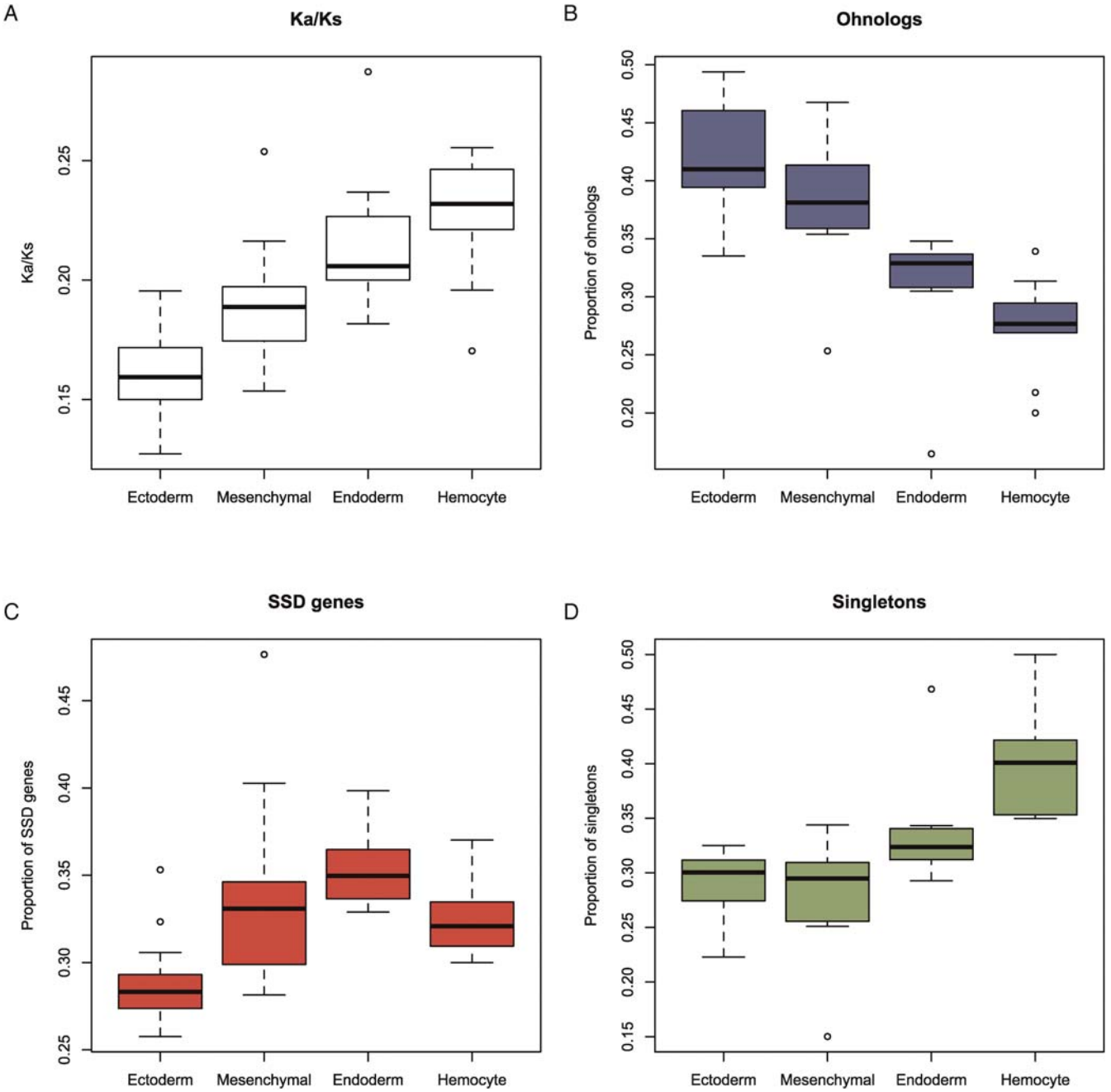

Figure 5. Contribution of ohnologues, SSD genes, and singletons to gene sets expressed in haemocytes, tissues of endodermal and ectodermal origin, and mesenchymal tissues. (A) The average K_A/K_S (ω) values associated with each gene set. (B)–(D) The proportion of ohnologues, SSD genes, and singletons expressed in each tissue type. In each panel, averages and standard deviations are shown. *P*-values for pairwise comparisons are shown in Supplementary Table S4.

4.2. *Contribution of SSD genes, ohnologues, and singletons to expression breadths in tissues*

Genes that are expressed in a narrow range of tissues are considered to play significant roles in the establishment of the characteristic features of the corresponding tissues. Prior studies have shown that those genes that are expressed specifically in neural tissues evolve more slowly, whereas those expressed in other tissue types evolve faster.[18,23] Alternatively, a substantial proportion of those genes with a tissue-specific expression tend to encode secreted polypeptides irrespective of tissue types, and there are few such genes in neural tissues.[24]

In this study, we also examined if the nature of SSD genes depended on the type of tissue in which the corresponding genes were expressed. To this end, the evolutionary rate of a tissue was defined as the average ω value of those genes expressed in that

tissue, whereas the nature of the expressed genes was evaluated by their ω values, proportions as well as by the product of these parameters. These parameters were calculated for each type of gene (SSD, ohnologues, and singletons) and plotted for each tissue (Fig. 2). Among genes with a narrow spectrum of expression, SSD genes (and singletons) played a minor role in slowly evolving tissues with low ω values. However, the evolutionary rates as well as the proportion of SSD genes were associated positively with the evolutionary rates of tissues. Thus, SSD genes (and singletons) seemed to play major roles in fast evolving tissues with high ω values, indicating that the relative contribution of SSD genes differs substantially depending on the type of tissue in which the gene is expressed. In contrast, genes with a wide spectrum of expression are presumed to maintain basic cellular functions irrespective of the tissue type. Among these genes, the relative contribution of SSD genes and ohnologues was similar.

Our results indicate that ohnologues with slower rates of evolution and wide expression ranges are associated with slowly evolving tissues, whereas SSD genes with faster rates of evolution and narrow range of expression are associated with fast evolving tissues. Specifically, there were less ohnologues and more SSD genes in endodermal tissues such as the digestive tract. These tissues face the outer environment directly and therefore they may differentiate functionally and pleiotropically to adjust to environmental changes. Rapid tissue evolutionary rates could be therefore related to a specific mode of duplication (SSD genes) and faster molecular evolution (high ω values) in response to this exposure to the outer environment.

Conversely, in ectoderm-derived tissues, such as the nervous system, there were more ohnologues and less SSD genes. Because these tissues are not exposed to the outer environment, their evolution may not be driven by functional differentiation but rather by the need to maintain basal functions. Ohnologues with a slow evolutionary rate and lower probability of further duplication were therefore predominant in slowly evolving tissues such as the ectoderm. Mesenchymal tissues were associated with an intermediate phenotype in terms of ω values and the proportion of ohnologues and SSD genes.

Those genes with expression restricted to haemocytes exhibited very unique features. The percentage of ohnologues was the smallest among all tissue types, whereas singletons were the most representative, a result not observed for the other three tissue types. A lower proportion of conservative ohnologues might allow the relaxation of those constraints required for the formation and functionality of tissues. This flexibility might be further accelerated by an increase in the ratio of singletons. Such findings might be expected in light of the dramatic transitions of haematopoietic organs and tissues from the aorta, gonads, and mesonephros regions to the liver, spleen, and bone marrow during mammalian development and vertebrate evolution. Furthermore, the haemocytes used in this study included immune-competent cells and tissues, and immune-related genes are known to evolve fast (immunoglobulin and T cell receptor gene families were excluded from the analysis here). In this sense, endodermal tissues and haemocytes appear to adopt distinct strategies to increase the proportion of SSD genes and singletons, respectively, when adapting to environmental changes.

Acknowledgements: We thank Koji Tamura for the valuable comments.

Supplementary Data: Supplementary Data are available at www.dnaresearch.oxfordjournals.org.

Funding

The work was supported by research grants from the Ministry of Education, Science, Sports, Culture and Technology of Japan and by the Creative Interdisciplinary Research Program of the Center for Interdisciplinary Research, Tohoku University, Japan. M.S. is supported by a GCOE program, 'Network Medicine', at Tohoku University.

References

1. Graur, D. and Li, W.-H. 1999, *Fundamentals of Molecular Evolution*, 2nd edition. Sinauer Associates: Sunderland, MA.
2. McLysaght, A., Hokamp, K. and Wolfe, K.H. 2002, Extensive genomic duplication during early chordate evolution, *Nat. Genet.*, **31**, 200–4.
3. Nakatani, Y., Takeda, H., Kohara, Y. and Morishita, S. 2007, Reconstruction of the vertebrate ancestral genome reveals dynamic genome reorganization in early vertebrates, *Genome Res.*, **17**, 1254–65.
4. Dehal, P. and Boore, J.L. 2005, Two rounds of whole genome duplication in the ancestral vertebrate, *PLoS Biol.*, **3**, e314.
5. Ohno, S. 1970, *Evolution by Gene Duplication*. Springer: Berlin.
6. Wolfe, K. 2000, Robustness—it's not where you think it is, *Nat. Genet.*, **25**, 3–4.
7. Gu, Z., Steinmetz, L.M., Gu, X., Scharfe, C., Davis, R.W. and Li, W.H. 2003, Role of duplicate genes in genetic robustness against null mutations, *Nature*, **421**, 63–6.
8. Conant, G.C. and Wagner, A. 2004, Duplicate genes and robustness to transient gene knock-downs in *Caenorhabditis elegans*, *Proc. Biol. Sci.*, **271**, 89–96.

9. Liang, H. and Li, W.H. 2007, Gene essentiality, gene duplicability and protein connectivity in human and mouse, *Trends Genet.*, **23**, 375–8.
10. Liao, B.Y. and Zhang, J. 2007, Mouse duplicate genes are as essential as singletons, *Trends Genet.*, **23**, 378–81.
11. Makino, T., Hokamp, K. and McLysaght, A. 2009, The complex relationship of gene duplication and essentiality, *Trends Genet.*, **25**, 152–5.
12. Veitia, R.A. 2002, Exploring the etiology of haploinsufficiency, *Bioessays*, **24**, 175–84.
13. Veitia, R.A. 2003, Nonlinear effects in macromolecular assembly and dosage sensitivity, *J. Theor. Biol.*, **220**, 19–25.
14. Veitia, R.A. 2004, Gene dosage balance in cellular pathways: implications for dominance and gene duplicability, *Genetics*, **168**, 569–74.
15. Makino, T. and McLysaght, A. 2010, Ohnologs in the human genome are dosage balanced and frequently associated with disease, *Proc. Natl Acad. Sci. USA*, **107**, 9270–4.
16. Liao, B.Y. and Zhang, J.Z. 2006, Low rates of expression profile divergence in highly expressed genes and tissue-specific genes during mammalian evolution, *Mol. Biol. Evol.*, **23**, 1119–28.
17. Gu, X. and Su, Z.X. 2007, Tissue-driven hypothesis of genomic evolution and sequence-expression correlations, *Proc. Natl Acad. Sci. USA*, **104**, 2779–84.
18. Duret, L. and 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.
19. Thompson, J.D., Higgins, D.G. and 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**, 4673–80.
20. Yang, Z. and Nielsen, R. 2000, Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models, *Mol. Biol. Evol.*, **17**, 32–43.
21. Yang, Z. 1997, PAML: a program package for phylogenetic analysis by maximum likelihood, *Comput. Appl. Biosci.*, **13**, 555–6.
22. Necsulea, A., Semon, M., Duret, L. and Hurst, L.D. 2009, Monoallelic expression and tissue specificity are associated with high crossover rates, *Trends Genet.*, **25**, 519–22.
23. Kuma, K., Iwabe, N. and Miyata, T. 1995, Functional constraints against variations on molecules from the tissue level: slowly evolving brain-specific genes demonstrated by protein kinase and immunoglobulin supergene families, *Mol. Biol. Evol.*, **12**, 123–30.
24. Winter, E.E., Goodstadt, L. and Ponting, C.P. 2004, Elevated rates of protein secretion, evolution, and disease among tissue-specific genes, *Genome Res.*, **14**, 54–61.
25. Hastings, K.E. 1996, Strong evolutionary conservation of broadly expressed protein isoforms in the troponin I gene family and other vertebrate gene families, *J. Mol. Evol.*, **42**, 631–40.
26. Huminiecki, L. and Wolfe, K.H. 2004, Divergence of spatial gene expression profiles following species-specific gene duplications in human and mouse, *Genome Res.*, **14**, 1870–9.