LETTER

doi:10.1038/nature09632

A phylogenetically based transcriptome age index mirrors ontogenetic divergence patterns

Tomislav Domazet-Lošo[1,2] & Diethard Tautz[1]

Parallels between phylogeny and ontogeny have been discussed for almost two centuries, and a number of theories have been proposed to explain such patterns[1]. Especially elusive is the phylotypic stage, a phase during development where species within a phylum are particularly similar to each other[2–6]. Although this has formerly been interpreted as a recapitulation of phylogeny[1], it is now thought to reflect an ontogenetic progression phase[2], where strong constraints on developmental regulation and gene interactions exist[2,3]. Several studies have shown that genes expressed during this stage evolve at a slower rate, but it has so far not been possible to derive an unequivocal molecular signature associated with this stage[7–15]. Here we use a combination of phylostratigraphy[16] and stage-specific gene expression data to generate a cumulative index that reflects the evolutionary age of the transcriptome at given ontogenetic stages. Using zebrafish ontogeny and adult development as a model, we find that the phylotypic stage does indeed express the oldest transcriptome set and that younger sets are expressed during early and late development, thus faithfully mirroring the hourglass model of morphological divergence[2,3]. Reproductively active animals show the youngest transcriptome, with major differences between males and females. Notably, ageing animals express increasingly older genes. Comparisons with similar data sets from flies and nematodes show that this pattern occurs across phyla. Our results indicate that an old transcriptome marks the phylotypic phase and that phylogenetic differences at other ontogenetic stages correlate with the expression of newly evolved genes.

The evolutionary origin of genes can be traced by similarity searches in genomes representing the whole tree of life. We have called this approach 'phylostratigraphy' and have shown that meaningful comparisons can be derived from it[16–18] (see Supplementary Note 1). It is important to note that the procedure identifies specifically the origin of novel genes with no traceable relation to existing genes or protein domains (see Supplementary Note 2). Another important property of phylostratigraphy is that it establishes a phylogenetic scale where every gene within a genome has its phylogenetic rank. Here, using this phylogenetic hierarchy, we extend this approach by linking it to all expressed genes within the ontogenetic sequence. To link these two hierarchies quantitatively we developed a transcriptome age index (TAI), which integrates the age of a gene with its expression level at a given developmental stage and sums this over all genes expressed at the respective stage. The higher the TAI, the younger the transcriptome (see Methods).

To apply the TAI for a developmental model system, we have generated a fine-grained series of transcriptome data of zebrafish development, covering a total of 60 stages, from unfertilized eggs to ageing animals. Figure 1a shows the TAI profile, plotted along these stages. The comparatively oldest transcript sets are expressed during the late segmentation/early pharyngula stage, which is the developmental stage that is usually equated with the phylotypic stage in zebrafish[13]. The start of heart pulsations and blood circulation in the embryo (24 h)[19] is a morphological feature that approximately marks this period of lowest TAI values. Phylogenetically younger transcriptome sets are expressed before and after this stage. This correlates well with the observation that early and late stages of chordate development also show a higher morphological divergence between taxa[2,3]. During the mid-larval stage we see a second phase where older transcriptomes are expressed, which corresponds to metamorphosis[19]. Although metamorphosis in fish is not as overt as in some other chordates (for example, amphibians), it is nonetheless a phase with major changes in morphology and life-history strategy. It is particularly evident in the reshaping of the fins, which change from a basal pattern that is seen across all fish into the one that is more specific for zebrafish[20]. After this stage, the transcriptome becomes younger and peaks in young adults. Males and females show major differences in the overall age index, with females expressing the relatively youngest genes. Intriguingly, as animals become older, they express older genes again.

Analysing the contribution of the different phylostrata (ps) to the general profile shows that they contribute to different extents (Fig. 1b). Genes that have emerged before the evolution of metazoa (ps1 to ps5) are more equally expressed throughout ontogeny, whereas later-emerging genes contribute increasingly to the differential pattern. A more detailed contribution of the genes from the different phylostrata is summarized in Fig. 2. Here we have depicted the relative expression levels for each stage for several phylostrata. This representation is only partly comparable to that in Fig. 1b, as it disregards the actual number of genes within a phylostratum. But this analysis allows several more specific points to be made.

Most genes that have arisen in ps1 (cellular origin) are general enzymes and housekeeping genes, but their RNA is not highly expressed before gastrulation (Fig. 2a), indicating that the products of these genes are primarily stored as proteins in the egg. Intriguingly, this is very different for the genes from ps2 to ps4, which have their relatively highest expression levels at these early stages. This is also indicative of a correlation between phylogenetic age of a gene and ontogenetic use of its product.

The noticeable TAI peak during gastrulation (Fig. 1a) is mainly generated by the genes from ps5 (evolution of metazoa, Fig. 2b). Studies in sponges suggest that gastrulation is an embryological process present since the onset of the metazoan evolution[21], which is in agreement with the peak of ps5 genes. In addition, we have previously identified ps5 as the time of emergence of genes involved in cellular interactions[18], which are evidently of particular importance during gastrulation.

Genes that have evolved during chordate evolution (ps9) are particularly highly expressed at the end of the pharyngula stage and at the beginning of larval stages, before metamorphosis (Fig. 2b). This is again a very suggestive correlation, because during this phase the chordate body plan in zebrafish reaches, for the first time, a full functional differentiation that is reflected in chordate-specific undulatory swimming and the start of active feeding. Interestingly, ps7 genes (evolution of bilateria) start to be strongly expressed at the beginning

[1]Max-Planck-Institut für Evolutionsbiologie, August-Thienemannstrasse 2, 24306 Plön, Germany. [2]Laboratory of Evolutionary Genetics, Division of Molecular Biology, Ruđer Bošković Institute, Bijenička cesta 54, P.P. 180, 10002 Zagreb, Croatia.

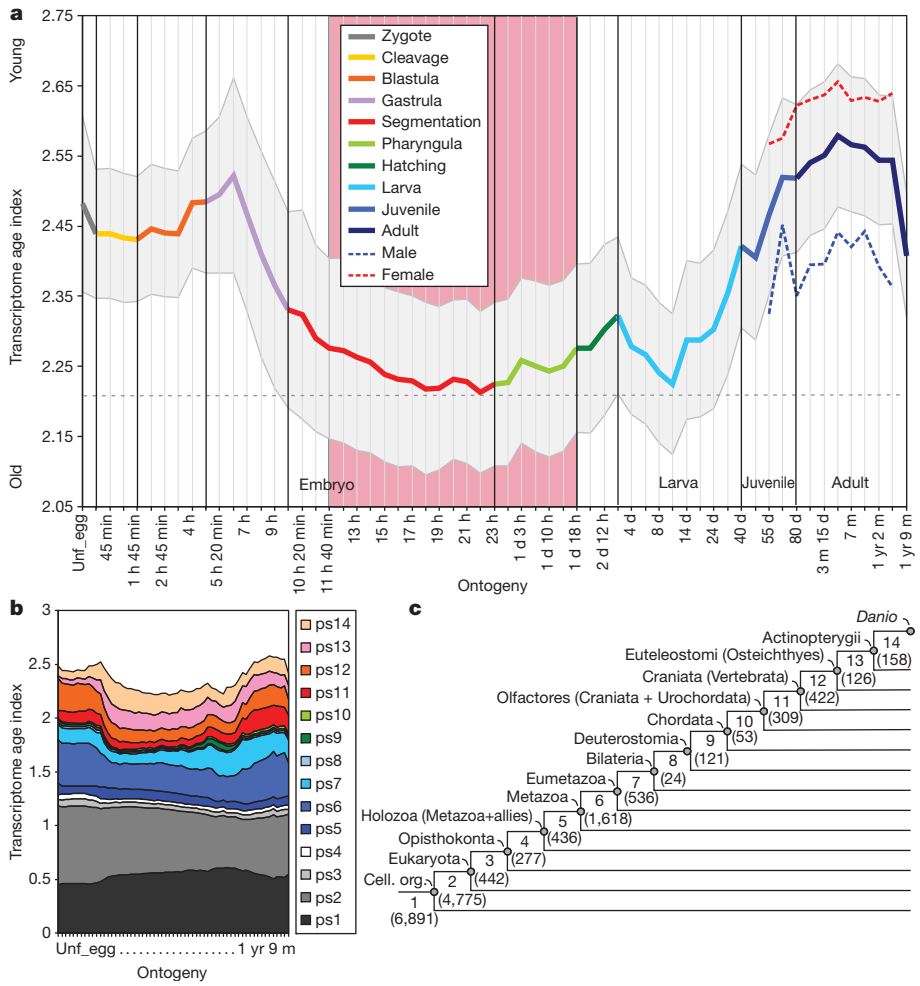

Figure 1 | Transcriptome age profiles for the zebrafish ontogeny. **a**, Cumulative transcriptome age index (TAI) for the different developmental stages. The pink shaded area represents the presumptive phylotypic phase in vertebrates. The overall pattern is significant by repeated measures ANOVA ($P = 2.4 \times 10^{-15}$, after Greenhouse–Geisser correction $P = 0.024$). Grey shaded areas represent ± the standard error of TAI estimated by bootstrap analysis. **b**, Transcriptome indices split according to the origin of the genes from the different phylostrata, based on the same developmental series as in **a**. **c**, Depiction of the phylostrata analysed; numbers in parentheses denote the number of array probes analysed for each phylostratum.

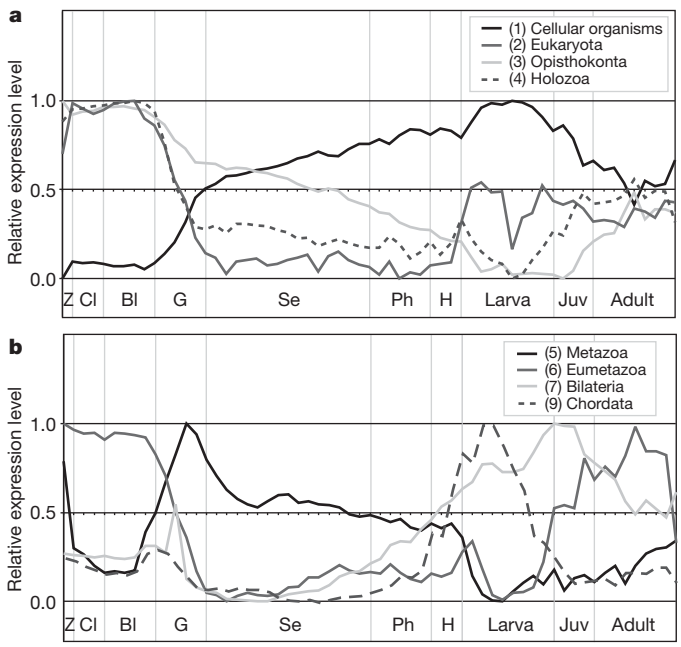

Figure 2 | Relative expression of the genes from each phylostratum across the zebrafish ontogeny (same stages as in Fig. 1) for selected phylostrata with significant differences. See Supplementary Fig. 3 for representation of all phylostrata and significance assessments. For easier comparisons, the relative expression calculated in relation to the highest (0) and lowest (1) expression values across developmental stages is shown (see Methods). Bl, blastula; Cl, cleavage; G, gastrula; H, hatching; Juv, juvenile; Ph, pharyngula; Se, segmentation; Z, zygote.

of metamorphosis, raising the possibility that metamorphosis-related genes or processes have already originated in parallel to the formation of bilateria. Ancient origins of hormonal signalling processes associated with metamorphosis have indeed been proposed[22].

Although comparable fine-grained data sets are currently not available for other model systems, one can still compare the trends based on available partial data sets. A good developmental transcriptome data series exists for *Drosophila*, although it covers only one-third of the expressed genes[23]. We have calculated the TAI for these data and find that the overall pattern is indeed comparable to zebrafish (Fig. 3). Most notably, the relatively oldest transcriptome is expressed during germband elongation, which can be equated to the phylotypic phase in arthropods[24]. Thus, this molecular signature is qualitatively comparable to the zebrafish data, but there are more novel genes among the post-embryonically expressed genes in *Drosophila* than in zebrafish, reflected in larger TAI values from differentiation stages onwards (Fig. 3a). Again, we see a major difference between males and females

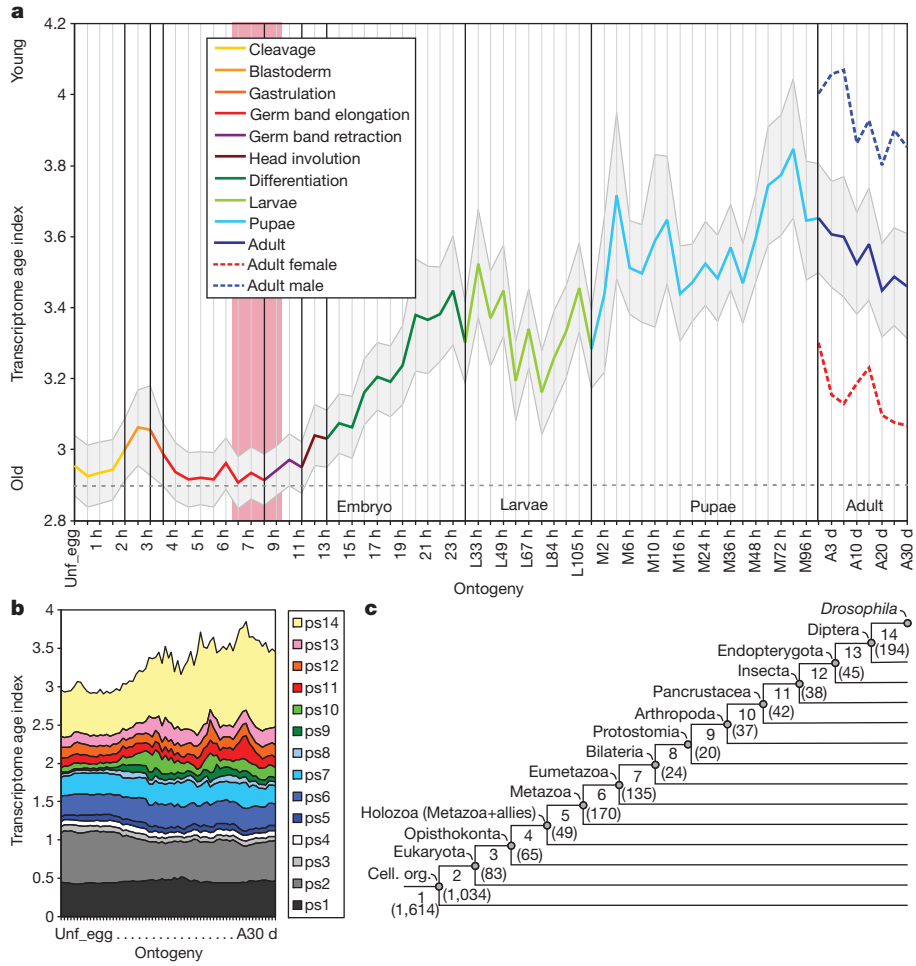

Figure 3 | Transcriptome age profiles for the *Drosophila* ontogeny, based on the data in ref. 23. **a**, Cumulative transcriptome index for the different developmental stages. The pink shaded area represents the presumptive phylotypic phase in insects. The overall pattern of differences in TAI is significant by repeated measures ANOVA ($P = 2.5 \times 10^{-93}$, after Greenhouse–Geisser correction $P = 1.22 \times 10^{-11}$). Grey shaded areas represent ± the standard error of TAI estimated by bootstrap analysis. **b**, **c** Same as for Fig. 1b, c.

after hatching, but in contrast to zebrafish, the males express the younger transcriptomes. Similar to the situation in zebrafish, ageing *Drosophila* express increasingly older transcriptomes (Fig. 3a). Breaking this pattern down to the contribution from the different phylostrata shows again that the oldest genes contribute little to the differential pattern, whereas genes that have emerged in ps9 (equivalent to the evolution of Arthropods) and later add increasingly to the final profile (Fig. 3b).

Comparable, but even more limited, ontogenetic transcriptome data are also available for the nematode *Caenorhabditis elegans*[25] and the mosquito *Anopheles*[26]. The same trends can be seen for those as well, namely the oldest genes expressed during the embryonic stages, the youngest towards adult stages and older genes in ageing animals (Supplementary Figs 1 and 2).

These consistent overall patterns across phyla, as well as the detailed analysis within zebrafish, suggest that there is a link between evolutionary innovations and the emergence of novel genes[16,27,28]. Adaptations are expected to occur primarily in response to altered ecological conditions. Juvenile and adults interact much more with ecological factors than embryos, which may even be a cause for fast postzygotic isolation[29]. Similarly, the zygote may also react to environmental constraints, for example, via the amount of yolk provided in the egg. In contrast, mid-embryonic stages around the phylotypic phase are normally not in direct contact with the environment and are therefore less likely to be subject to ecological adaptations and evolutionary change. As already suggested by Darwin (discussed in ref. 15), this alone could explain the lowered morphological divergence of early ontogenetic stages compared to adults, which would obviate the need to invoke particular constraints. Alternatively, the constraint hypothesis would suggest that it is difficult for newly evolved genes to become recruited to strongly connected regulatory networks[12,13,15].

The fact that ageing animals revert to older transcriptomes is in line with the notion that animals beyond the reproductive age are not 'visible' to natural selection and can therefore not be subject to specific adaptations any more. Also, the fact that major TAI differences can be seen between males and females could have been anticipated, because sexual selection is expected to continuously change phenotypic traits between them. However, the fact that the differences go in opposite directions in zebrafish and *Drosophila* is surprising. We have therefore studied in detail which phylostrata contribute most to these differences (Fig. 4). Both taxa show a female expression bias of ps2 genes, which may be correlated to egg production, as RNA from such genes is stored in the eggs (see above). But they strongly deviate at other phylostrata. Zebrafish shows a strong female bias of ps6 and ps12 genes, which is absent in *Drosophila* (Fig. 4). *Drosophila*, on the other hand, shows an extreme bias of ps14 genes in males (Fig. 4), which is caused by the many orphan genes involved in spermatogenesis[30]. Thus, in contrast to the ontogenetic similarities of the TAI trends between the two taxa, the sex differences are rather incongruent and indicate different evolutionary trajectories for male–female differences.

Our study provides strong molecular support for a correlate between phylogeny and ontogeny, as well as the hourglass model of

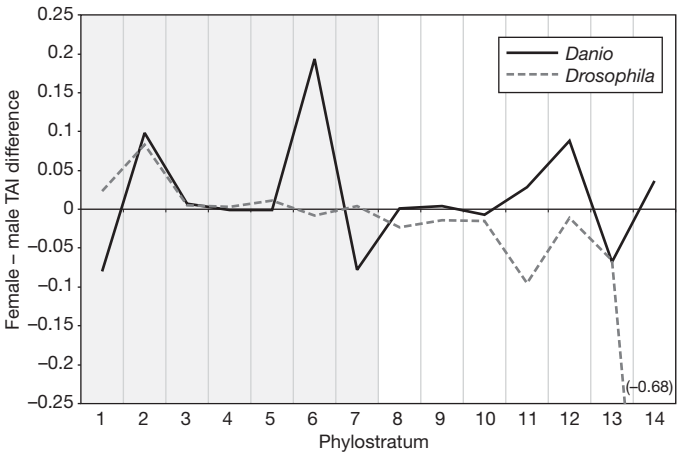

Figure 4 | Comparison of differences in TAI between females and males. Comparison across phylostrata in zebrafish (*Danio*) and *Drosophila* (see Supplementary Fig. 4 for a plot that includes the differences between stages). The grey shaded area designates the shared part of the phylogeny between the two species (origin of the first cell until the last common ancestor of Bilateria, ps1–ps7). Note that the ps14 value for *Drosophila* is off the scale (difference is given in parenthesis).

development. Under this scheme, the phylotypic phase can be defined as the ontogenetic progression during which the oldest gene set is expressed, either because this is the phase with the lowest opportunity for lineage-specific adaptations, or because it is internally so constrained that newly evolved genes cannot become integrated.

METHODS SUMMARY

The TAI is the weighted mean of phylogenetic ranks (phylostrata) and is calculated for every ontogenetic stage s as follows:

$$\mathrm{TAI}_S = \frac{\sum_{i=1}^{n} \mathrm{ps}_i e_i}{\sum_{i=1}^{n} e_i}$$

where ps_i is an integer that represents the phylostratum of the gene i (for example, 1, the oldest; 14, the youngest), e_i is the microarray signal intensity value (obtained from Agilent Zebrafish (V2) Gene Expression Microarrays) of the gene i that acts as weight factor and n is the total number of genes analysed. This way of calculating the index gives an increasingly stronger weight to younger phylostrata, thus compensating for the fact that the older phylostrata usually harbour the larger number of genes[16–18].

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 2 September; accepted 27 October 2010.

1. Gould, S. J. *Ontogeny and Phylogeny* (Harvard Univ. Press, 1977).
2. Duboule, D. Temporal colinearity and the phylotypic progression: a basis for the stability of a vertebrate Bauplan and the evolution of morphologies through heterochrony. *Dev. Suppl.* 135–142 (1994).
3. Raff, R. A. *The Shape of Life: Genes Development, and the Evolution of Animal Form* (Univ. Chicago Press, 1996).
4. Richardson, M. K. *et al.* There is no highly conserved embryonic stage in the vertebrates: implications for current theories of evolution and development. *Anat. Embryol. (Berl.)* **196,** 91–106 (1997).
5. Hall, B. K. Phylotypic stage or phantom: is there a highly conserved embryonic stage in vertebrates? *Trends Ecol. Evol.* **12,** 461–463 (1997).
6. Bininda-Emonds, O. R., Jeffery, J. E. & Richardson, M. K. Inverting the hourglass: quantitative evidence against the phylotypic stage in vertebrate development. *Proc. R. Soc. Lond. B* **270,** 341–346 (2003).
7. Hanada, K., Shiu, S. H. & Li, W. H. The Nonsynonymous/synonymous substitution rate ratio versus the radical/conservative replacement rate ratio in the evolution of mammalian genes. *Mol. Biol. Evol.* **24,** 2235–2241 (2007).
8. Comte, A., Roux, J. & Robinson-Rechavi, M. Molecular signaling in zebrafish development and the vertebrate phylotypic period. *Evol. Dev.* **12,** 144–156 (2010).
9. Yassin, A., Lienau, E. K., Narechania, A. & DeSalle, R. Catching the phylogenic history through the ontogenic hourglass: a phylogenomic analysis of *Drosophila* body segmentation genes. *Evol. Dev.* **12,** 288–295 (2010).
10. Davis, J. C., Brandman, O. & Petrov, D. A. Protein evolution in the context of *Drosophila* development. *J. Mol. Evol.* **60,** 774–785 (2005).
11. Hazkani-Covo, E., Wool, D. & Graur, D. In search of the vertebrate phylotypic stage: a molecular examination of the developmental hourglass model and von Baer's third law. *J. Exp. Zoolog. B* **304,** 150–158 (2005).
12. Irie, N. & Sehara-Fujisawa, A. The vertebrate phylotypic stage and an early bilaterian-related stage in mouse embryogenesis defined by genomic information. *BMC Biol.* **5,** 1 (2007).
13. Roux, J. & Robinson-Rechavi, M. Developmental constraints on vertebrate genome evolution. *PLoS Genet.* **4,** e1000311 (2008).
14. Cruickshank, T. & Wade, M. J. Microevolutionary support for a developmental hourglass: gene expression patterns shape sequence variation and divergence in *Drosophila*. *Evol. Dev.* **10,** 583–590 (2008).
15. Artieri, C. G., Haerty, W. & Singh, R. S. Ontogeny and phylogeny: molecular signatures of selection, constraint, and temporal pleiotropy in the development of *Drosophila*. *BMC Biol.* **7,** 42 (2009).
16. Domazet-Lošo, T., Brajković, J. & Tautz, D. A phylostratigraphy approach to uncover the genomic history of major adaptations in metazoan lineages. *Trends Genet.* **23,** 533–539 (2007).
17. Domazet-Lošo, T. & Tautz, D. An ancient evolutionary origin of genes associated with human genetic diseases. *Mol. Biol. Evol.* **25,** 2699–2707 (2008).
18. Domazet-Lošo, T. & Tautz, D. Phylostratigraphic tracking of cancer genes suggests a link to the emergence of multicellularity in metazoa. *BMC Biol.* **8,** 66 (2010).
19. Parichy, D. M., Elizondo, M. R., Mills, M. G., Gordon, T. N. & Engeszer, R. E. Normal table of postembryonic zebrafish development: staging by externally visible anatomy of the living fish. *Dev. Dyn.* **238,** 2975–3015 (2009).
20. Grandel, H. & Schulte-Merker, S. The development of the paired fins in the zebrafish (*Danio rerio*). *Mech. Dev.* **79,** 99–120 (1998).
21. Leys, S. P. & Eerkes-Medrano, D. Gastrulation in calcareous sponges: In search of Haeckel's Gastraea. *Integr. Comp. Biol.* **45,** 342–351 (2005).
22. Flatt, T., Moroz, L. L., Tatar, M. & Heyland, A. Comparing thyroid and insect hormone signaling. *Integr. Comp. Biol.* **46,** 777–794 (2006).
23. Arbeitman, M. N. *et al.* Gene expression during the life cycle of *Drosophila* melanogaster. *Science* **297,** 2270–2275 (2002).
24. Sander, K. In *Development and Evolution: the sixth Symposium of the British Society for Developmental Biology* (eds Goodwin, B. C., Holder, N. & Wylie, C. C.) 137–159 (Cambridge Univ. Press, 1983).
25. Hill, A. A., Hunter, C. P., Tsung, B. T., Tucker-Kellogg, G. & Brown, E. L. Genomic analysis of gene expression in *C. elegans*. *Science* **290,** 809–812 (2000).
26. Koutsos, A. C. *et al.* Life cycle transcriptome of the malaria mosquito *Anopheles gambiae* and comparison with the fruitfly *Drosophila melanogaster*. *Proc. Natl Acad. Sci. USA* **104,** 11304–11309 (2007).
27. Domazet-Lošo, T. & Tautz, D. An evolutionary analysis of orphan genes in *Drosophila*. *Genome Res.* **13,** 2213–2219 (2003).
28. Khalturin, K., Hemmrich, G., Fraune, S., Augustin, R. & Bosch, T. C. More than just orphans: are taxonomically-restricted genes important in evolution? *Trends Genet.* **25,** 404–413 (2009).
29. Nolte, A. W., Renaut, S. & Bernatchez, L. Divergence in gene regulation at young life history stages of whitefish (*Coregonus* sp.) and the emergence of genomic isolation. *BMC Evol. Biol.* **9,** 59 (2009).
30. Levine, M. T., Jones, C. D., Kern, A. D., Lindfors, H. A. & Begun, D. J. Novel genes derived from noncoding DNA in *Drosophila melanogaster* are frequently X-linked and show testis-biased expression. *Proc. Natl Acad. Sci. USA* **103,** 9935–9939 (2006).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank B. Walderich for providing zebrafish, A. Nolte, E. Blohm-Sievers, B. Kleinhenz, L. Turner and J. Bryk for laboratory support, R. Bakarić has provided the phylostratigraphic map of *C. elegans*, and M. Domazet-Lošo and V. Dunjko have helped with statistics. L. Boell, F. Chang and A. Pozhitkov have made suggestions on the manuscript. This work was supported by Unity Through Knowledge Fund (grant No. 49), Adris Foundation and funds of the Max-Planck Society. Computational resources were provided by CSTMB and RBI (Phylostrat Cluster).

Author Contributions T.D.-L. conceived the basic idea and conducted the experiments; D.T. contributed to the evaluation and interpretation of the results. Both authors wrote the manuscript.

Author Information The microarray data for zebrafish were deposited at the NCBI Gene Expression Omnibus (GEO) repository under the accession number GSE24616. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to T.D.-L. (tdomazet@irb.hr).

METHODS

Fish keeping and sampling conditions. Zebrafish (*Danio rerio*) were kept in 12 l flow-through tanks at 26.5 °C (around 60 animals per tank). For accurate staging, fertilized eggs were collected within 15-min intervals and incubated in Petri dishes at 28.5 °C with water changes every 2–6 h. After hatching, larvae were transferred to 1-l tanks and kept at 28.5 °C. We took, in total, 72 samples in two replicates that correspond to 60 stages across zebrafish ontogeny (50 samples before the sex could be clearly recognized plus 11 samples of males and females each). Staging was done according to post-fertilization time. Embryos were additionally staged under the dissecting microscope according to ref. 31 to check for the consistency of post-fertilization timing and morphological development at standard temperature (28.5 °C). Only healthy animals that showed the expected morphological features for a given post-fertilization time were sampled. Each sample contained around 50 individuals until the 1 day and 3 h embryo stage, 15 individuals until the 10 day larval stage, 10 individuals until the 18 day larval stage, 5 individuals until the 45 day juvenile stage, whereas in later juvenile and adult stages we sampled males and females separately and each sample contained 2 individuals. All samples were snap frozen in liquid nitrogen and stored at −80 °C until RNA extraction. To avoid severe biases owing to the excess of unfertilized eggs, we squeezed eggs from adult females before freezing them in liquid nitrogen.

Phylostratigraphy. A full account of phylostratigraphic analysis and theoretical underpinnings has been presented previously[16–18]. The zebrafish genes of the present study (28,546, ENSEMBL release 56) were mapped on the currently best supported phylogeny using BLAST searches against the cleaned up and additionally enriched NCBI NR database, which represents the most exhaustive set of known proteins across all organisms. Our choice of internodes (phylostrata) in the consensus phylogeny depended on the availability of complete annotated genomes, reliability of phylogenetic relationships and on the importance of evolutionary transitions. Similarly, the data of *Drosophila* (13,389 genes), *Anopheles* (12,457 genes) and *Caenorhabditis elegans* (19,077 genes) were mapped to the best supported phylogenies that represent their evolutionary lineages.

RNA isolation and microarray gene expression experiments. Total RNA was isolated using the TRIZOL plus protocol (Invitrogen). Four-hundred nanograms of total RNA per sample were Cy3 labelled according to the one-colour Quick Amp Labelling Kit protocol (Agilent). Labelled cRNAs were hybridized to Agilent Zebrafish (V2) Gene Expression Microarray slides (4 × 44k) for 17 h at 65 °C and washed according to the Agilent protocol. Hybridized microarray slides were scanned using an Agilent High-Resolution Microarray Scanner.

Microarray data extraction, filtering and analysis. Raw microarray image files were processed and quality checked by Agilent's Feature Extraction 10.7 Image Analysis Software. Background subtracted signal intensity values that contain correction for multiplicative surface trends (gProcessedSignal) generated by Feature Extraction Software were used for further data analysis. Using GeneSpring microarray data analysis software we filtered probes that were flagged as non-uniform or as population outlier. For every of the 72 samples we calculated average signal intensity values over the two biological replicates. Probes (60 bp) were mapped on the *Danio rerio* transcripts (ENSEMBL version 54) that passed the phylostratigraphic analysis (see below) using CD-hit software. This procedure yielded 16,188 unique probes that collapsed to 12,892 ENSEMBL predicted genes.

Phylostratigraphically mapped genes of *Drosophila* were linked to available microarray data[23]. This procedure yielded a data set of 3,550 genes. In a similar fashion, phylogenetically ranked microarray data sets were obtained for *C. elegans*[25] (16,832 genes) and *Anopheles*[26] (3,135 genes).

Transcriptome age index and statistical analysis. The TAI is the weighted mean of phylogenetic ranks (phylostrata) and is calculated for every ontogenetic stage s as follows:

$$\mathrm{TAI}_S = \frac{\sum_{i=1}^{n} \mathrm{ps}_i e_i}{\sum_{i=1}^{n} e_i}$$

where ps_i is an integer that represents the phylostratum of the gene i (for example, 1, the oldest; 14, the youngest), e_i is the microarray signal intensity value (obtained from Agilent Zebrafish (V2) Gene Expression Microarrays) of the gene i that acts as weight factor and n is the total number of genes analysed. This way of calculating the index gives an increasingly stronger weight to younger phylostrata, thus compensating for the fact that the older phylostrata usually harbour the larger number of genes[16–18].

We chose to calculate the TAI index based on the amount of expression per gene, rather than by simply adding up whether a gene is expressed or not. Although this latter approach would also seem feasible, it runs into a technical problem. To say that a given gene is expressed or not, one would have to impose a cutoff on the signals from the microarrays, which is more or less arbitrary, as a weak signal on a microarray could be derived from a gene with very low expression level, or from a highly expressed gene that is present in a few cells only. Also, absolute quantities are difficult to compare across microarrays and a single cutoff value would not be appropriate (see below). In balance, we have therefore opted for the expression level as a numerator, also because one could argue that genes that are broadly expressed at high levels should be more relevant than specialized genes.

The TAI formula can alternatively be written as:

$$\mathrm{TAI}_S = \mathrm{ps}_1 \frac{e_1}{e_1+e_2+\ldots+e_n} + \mathrm{ps}_2 \frac{e_2}{e_1+e_2+\ldots+e_n} + \ldots + \mathrm{ps}_n \frac{e_n}{e_1+e_2+\ldots+e_n}$$

The expression $e_1 + e_2 + \ldots + e_n$ represents the total signal of the analysed probes on the microarray, whereas the ratio $e_i/(e_1 + e_2 + \ldots + e_n)$, which can be denoted as f_i, represents the partial concentration (frequency) of probe i in the total microarray signal at a given stage; it is within a range between zero and one. It is important to note that the calculation of the partial concentration (f_i) inherently makes a global intensity normalization over the microarray experiment at a given stage and that at every stage the sum of partial concentrations will equal one. In many microarray studies it is common to assess the direction of expression change (over- or under-expression). This type of analysis requires that after the normalization procedure, which aims to remove noise from the experiment, expression signals that are measured across experiments still reflect absolute number of mRNA molecules per unit of biological material. In such situations, if global intensity normalization is applied, it must be assumed that the total number of mRNA copies for all genes on the array does not significantly differ between experiments. Contrary to this common application of microarrays, in our study we are not interested in the direction of expression change of particular genes. Instead, we are looking at how partial concentrations of RNAs contribute to the overall transcriptome across stages. For this purpose it is irrelevant which part of the transcriptome is responsible for change of the partial concentration. Therefore in our data treatment it is not necessary to assume that cumulative signals do not differ between experiments. This shift in perspective greatly simplifies the analysis on the scale of the complete ontogeny because abundance and distribution of transcripts is commonly very different between stages.

Thus, the TAI can be written as a sum of products between partial concentration and corresponding phylostratum:

$$\mathrm{TAI}_S = \sum_{i=1}^{n} \mathrm{ps}_i f_i = \mathrm{ps}_1 f_1 + \mathrm{ps}_2 f_2 + \ldots + \mathrm{ps}_n f_n$$

To asses the contributions of a specific phylostratum to the overall TAI (Figs 1b and 3b) we split the above total sum of $\mathrm{ps}_i f_i$ products to subsets of $\mathrm{ps}_i f_i$ sums where the value of ps_i (phylostratigraphic rank) was used as a grouping factor.

By applying repeated measures ANOVA on these $\mathrm{ps}_i f_i$ products we tested the significance of difference in TAI between stages. Repeated measures ANOVA was used because the same set of probes are measured at every stage, that is, there is dependence between the stages compared. Before means of these products across stages are compared by ANOVA we multiplied every $\mathrm{ps}_i f_i$ product with constant n (total number of analysed probes). This transformation does not influence the ANOVA analysis and its sole purpose is that means of $\mathrm{ps}_i f_i$ products compared in ANOVA are equal to the corresponding TAI values. Because the assumption of sphericity was violated in the data sets analysed by repeated measures ANOVA, we applied the Greenhouse–Geisser correction. Multivariate test statistics, an alternative approach that is not dependent on the assumption of sphericity, corroborated our statistical results of repeated measures ANOVA. We used the bootstrap approach (1,000 replicates) to asses the standard error of weighted mean (TAI)[32].

Relative expression of the genes for a given phylostratum (ps) and developmental stage (s) (Fig. 2) was calculated according to the equation:

$$\mathrm{RE}(\mathrm{ps})_s = \frac{\bar{f} - \bar{f}_{\min}}{\bar{f}_{\max} - \bar{f}_{\min}}$$

where $\bar{f}$ is the average partial concentration of RNAs from phylostratum ps for a given stage and $\bar{f}_{\max}, \bar{f}_{\min}$ are the maximal and minimal average partial concentration from phylostratum ps across all considered stages, respectively.

31. Kimmel, C. B., Ballard, W. W., Kimmel, S. R., Ullmann, B. & Schilling, T. F. Stages of embryonic development of the zebrafish. *Dev. Dyn.* **203,** 253–310 (1995).
32. Efron, B. & Tibshirani, R. Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. *Stat. Sci.* **1,** 54–75 (1986).