


# Distinguishing Evolutionary Conservation from Derivedness

Jason Cheok Kuan Leong <sup>1,\*</sup>, Masahiro Uesaka <sup>2</sup> and Naoki Irie <sup>1,3,\*</sup>

<sup>1</sup> Department of Biological Sciences, Graduate School of Science, The University of Tokyo, Tokyo 113-0033, Japan

<sup>2</sup> Laboratory for Evolutionary Morphology, RIKEN Center for Biosystems Dynamics Research (BDR), Kobe 650-0047, Japan; masahiro.uesaka@riken.jp

<sup>3</sup> Universal Biology Institute, The University of Tokyo, Tokyo 113-0033, Japan

\* Correspondence: jasonleong@bs.s.u-tokyo.ac.jp (J.C.K.L.); irie@bs.s.u-tokyo.ac.jp (N.I.)

**Abstract:** While the concept of “evolutionary conservation” has enabled biologists to explain many ancestral features and traits, it has also frequently been misused to evaluate the degree of changes from a common ancestor, or “derivedness”. We propose that the distinction of these two concepts allows us to properly understand phenotypic and organismal evolution. From a methodological aspect, “conservation” mainly considers genes or traits which species have in common, while “derivedness” additionally covers those that are not commonly shared, such as novel or lost traits and genes to evaluate changes from the time of divergence from a common ancestor. Due to these differences, while conservation-oriented methods are effective in identifying ancestral features, they may be prone to underestimating the overall changes accumulated during the evolution of certain lineages. Herein, we describe our recently developed method, “transcriptomic derivedness index”, for estimating the phenotypic derivedness of embryos with a molecular approach using the whole-embryonic transcriptome as a phenotype. Although echinoderms are often considered as highly derived species, our analyses with this method showed that their embryos, at least at the transcriptomic level, may not be much more derived than those of chordates. We anticipate that the future development of derivedness-oriented methods could provide quantitative indicators for finding highly/lowly evolvable traits.

**Keywords:** derivedness; evolutionary conservation; evo-devo; phenotypic evolution



**Citation:** Leong, J.C.K.; Uesaka, M.; Irie, N. Distinguishing Evolutionary Conservation from Derivedness. *Life* **2022**, *12*, 440. <https://doi.org/10.3390/life12030440>

Academic Editors: Koichiro Tamura, Guojie Zhang and Daiqin Li

Received: 22 February 2022

Accepted: 14 March 2022

Published: 17 March 2022

**Publisher’s Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Searching for shared features, or conserved features, among different species allows biologists to estimate a variety of ancestral features of organisms, including possible phenotypes of common ancestors, such as Urbilateria [1], their signaling pathways [2] and biomolecules [3,4]. However, when it comes to evaluating more evolved, or highly derived features or organisms, focusing only on the conserved nature may cause various inconsistencies and confusion among studies. To be specific, when trying to determine which species has phenotypically more derived features than others, comparing only conserved or commonly shared traits may underestimate how much phenotypic change has occurred since their common ancestors. This is because novel traits or lost traits are often overlooked or even ignored. For example, the pentamerous symmetric body plan evolved in echinoderms is a novel trait which should not be excluded when evaluating how much the phenotypes of these species have changed since their common ancestor with other bilaterians.

This also applies to molecular-level studies. With recent increasing interest in using genome-scale omics data to study phenotypic evolution, such as the evolution of embryonic phenotypes [5–11], novel traits [12–17], loss of anatomical features [18–20], and adaptive or convergent evolution [21–24], the ambiguous use of conservation and derivedness could cause significant inconsistencies. For instance, comparisons of 1:1 orthologs tend

to consider commonly shared genes among species of interest; this can be regarded as a conservation-oriented analysis. With this approach, genes that were gained or lost during evolution are often excluded. In other words, evolutionary changes achieved by genes other than 1:1 orthologs might be overlooked by these 1:1 ortholog comparisons, leading to underevaluations of how much the phenotype or the organism has derived since the common ancestor.

In essence, the concept of conservation, as it has often been used in previous studies [5,11,12,25], represents information (including phenotypes and genotypes) retained during evolution. For this purpose, conservation-oriented methods tend to compare commonly shared or homologous genes and traits among the species being compared (Figure 1a). These conservation-oriented methods have been especially powerful in explaining ancestral features and species. In contrast, the concept of derivedness represents changes that have accumulated in organisms during evolution, and thus, derivedness-oriented methods tend to cover not only conserved traits and genes, but also those that were newly acquired or lost since the split from the common ancestor (Figure 1b). Because of these essential differences, it also should be noted that a less conserved feature is not always equal to a more derived one. This also applies to the relationship between being highly conserved and less derived. A possible scenario would be that if a certain organism has lost a tremendous number of traits (or genes) during its evolution, it is possible that it would be identified as an organism with highly conserved traits (or gene expression levels) by only comparing the commonly shared traits (or 1:1 orthologs). However, in this case, evolutionary changes by the loss of traits (or genes) would be overlooked, and so when such loss of traits (or genes) is also considered, it is more reasonable to consider the organism as highly derived rather than less derived.

In this review, we suggest that the differences between the two concepts, evolutionary conservation and evolutionary derivedness, and their associated methods should be handled with care; each of them has its own advantage in addressing different biological questions. We will first discuss several examples of inconsistencies which could have arisen from confusion between conservation and derivedness. We will then discuss the technical limitations and challenges of current conservation-oriented and derivedness-oriented molecular methods. Finally, as an application of the concept of derivedness, we will briefly describe our recently developed method, “transcriptomic derivedness index” [26], to estimate the evolutionary derivedness of transcriptomes of various echinoderm and chordate embryos from their common ancestor.

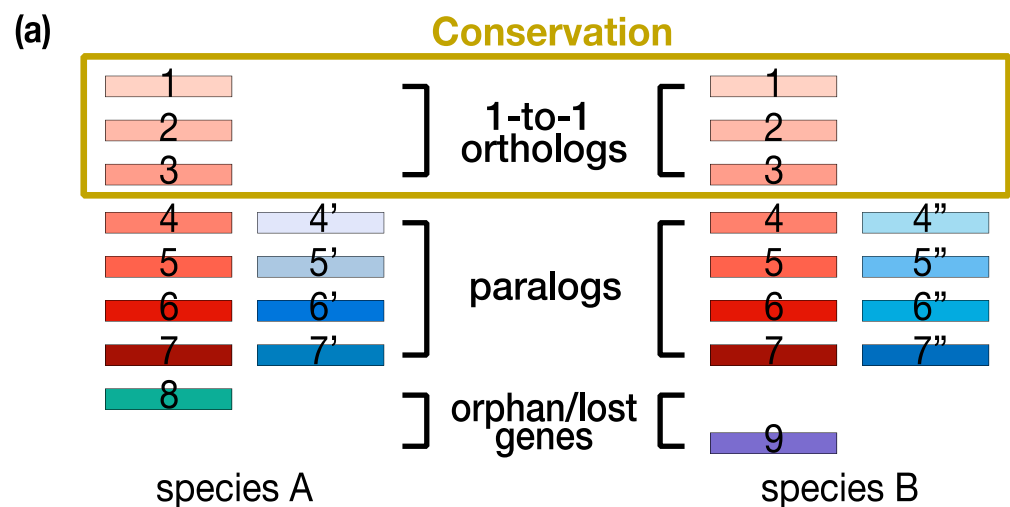
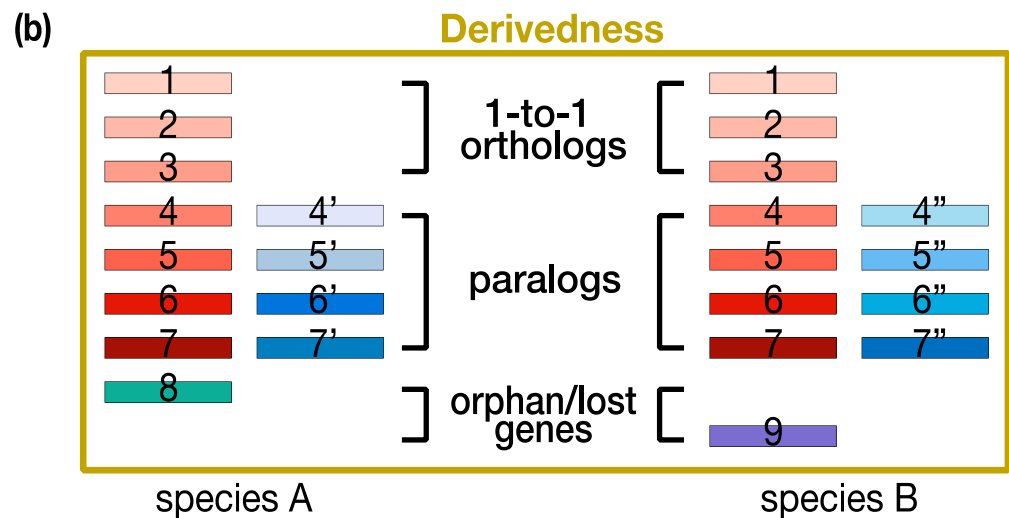


Figure 1. Cont.

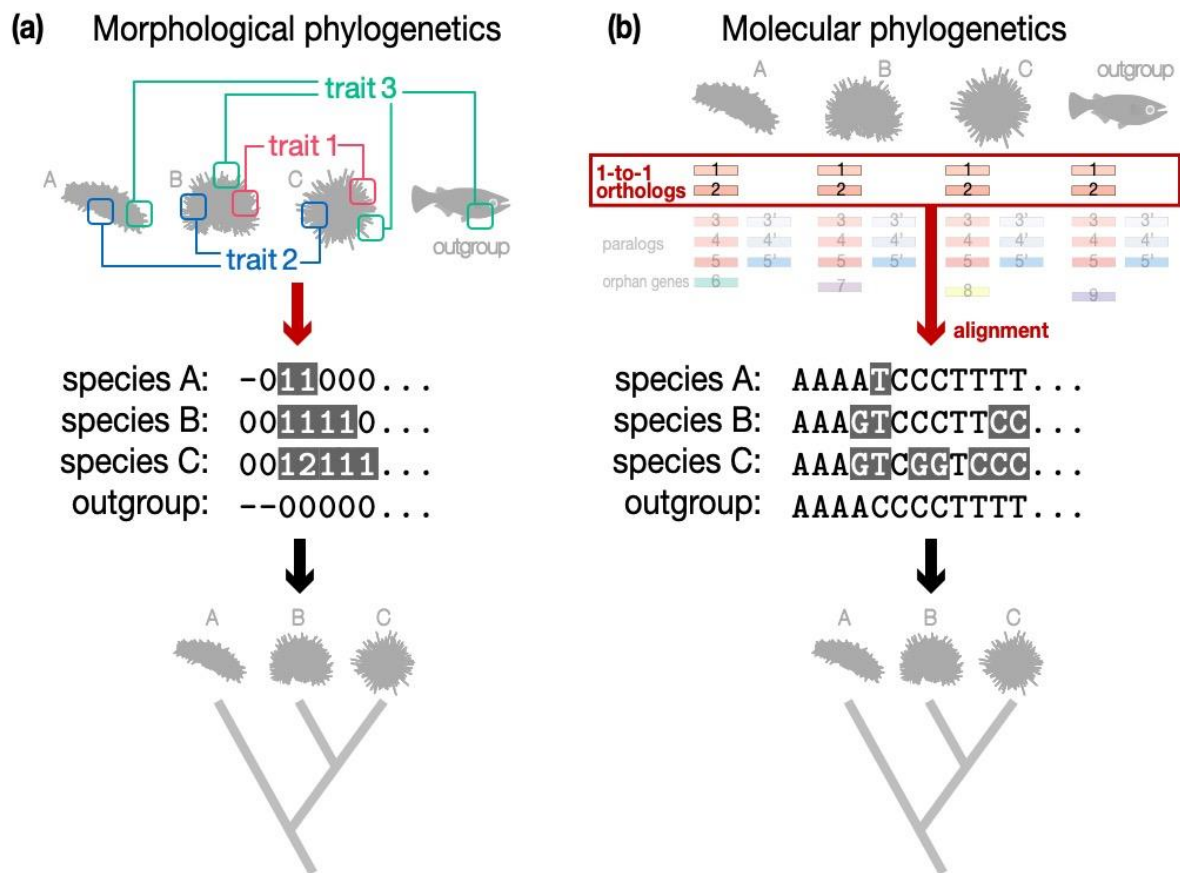


**Figure 1.** Conservation-oriented versus derivedness-oriented gene comparisons. (a) Conservation-oriented methods tend to compare commonly shared genes (e.g., 1:1 orthologs). (b) Derivedness-oriented methods additionally cover evolutionary changes of nonshared genes, such as 1-to-many orthologs, paralogs, and species-specific acquired or potentially lost genes. Rectangles: genes. Red: homologous genes inherited from the common ancestor of species A and B [27]. Blue: genes duplicated after the speciation event leading to species A and B (additionally marked by ' and '' signs). Green and purple: orphan genes (genes without recognizable homologs in the other species).

## 2. Confusion between Conservation and Derivedness Leads to Inconsistent Conclusions

### 2.1. Inconsistent Results Are Often Obtained from Molecular-Based Phylogenetic Estimation and Morphology-Based Estimation

Incongruence often appears between phylogenetic relationships estimated by a molecular and morphological approach. As a well-known example, the phylogenetic position of turtles among other vertebrates has been controversial. Many of the morphological comparisons tend to place turtles either as an early diverging lineage of reptiles or as the sister group of squamates (lizards and snakes), whereas molecular studies tend to suggest that turtles are the sister group of archosaurs (crocodiles and birds) [12,28–30]. This incongruence is most likely attributable to essential differences between these approaches. In short, the morphological approach tends to cover as many traits as possible (e.g., eggshell composition; Figure 2a), and evolutionary distance calculated with these traits tends to reflect the derivedness of the species, as it covers many novel and lost traits. Covering as many traits as possible is, of course, partially because there could be only a few strictly homologous morphological traits that are present in all analyzed organisms [31]. Analysis with fossil records adds further difficulty to identify commonly shared traits due to their incomplete preservation. Having said so, a major advantage of using morphological information for phylogenetics is the capability to unite extinct and extant taxa on the same tree [32], even though the approach tends to evaluate derivedness. For instance, to study early dinosaur evolution, Baron and colleagues curated the largest ever matrix comparing 457 morphological traits of dinosaurs and proposed a new hypothesis about the phylogenetic relationships of early dinosaurs [33–35]. This matrix included traits that are not commonly shared, as well as many missing data points due to, for example, partial fossil record.



**Figure 2.** Typical methods of molecular and morphological phylogenetics. **(a)** A derivedness-oriented method, which compares not only commonly shared traits (morphological trait 3) but also those that are not shared among all species (morphological traits 1 and 2), is often used in morphological phylogenetics. For phylogenetic reconstruction, the traits of different species are encoded as different states (for example, “-” represents the absence of a trait here, “0” represents ancestral state, and “1” and “2” represent derived states). **(b)** A conservation-oriented method, which compares only commonly shared genes, is often used in molecular phylogenetics. Only the alignable sequences of the commonly shared 1:1 orthologs tend to be used to infer phylogeny.

In contrast to the derivedness-oriented morphological approach, the molecular approach tends to focus on commonly shared genomic sequences, and distances calculated with this approach tend to reflect conservation rather than derivedness. The conservation-oriented molecular methods usually align sequences of orthologous genes or their encoded amino acid sequences, followed by the reconstruction of a phylogenetic tree based on the aligned sequences (reviewed in detail in [36–39]; Figure 2b). The alignment often requires strict orthology among the genes, while paralogous genes are excluded [38]. One reason for the wide use of 1:1 orthologs is that the evolutionary model or the probability of nucleotide substitutions during evolution [40–44] is understood better than that of gene loss and duplications. Therefore, distinguishing true orthologs from paralogs has been a major aim of many programs in molecular phylogenetics [45]. Practically, one widely used method to identify 1:1 orthologs is to extract genes through “reciprocal BLAST best-hits”, where orthologous genes are assumed to achieve the best scores in alignments against the whole genome of another species. With the recent accumulation of whole-genome sequencing data, evolutionary analyses have switched from using only a few available genes to genome-wide comparisons. However, the basic analytical logic has not changed drastically, and commonly shared 1:1 orthologs are still preferred. For example, current popular approaches include inferring a phylogenetic tree from concatenated 1:1 orthologous genes [38]. In other

words, conservation-oriented molecular methods tend to accommodate less information from paralogs and species-specific “orphan” genes (those that do not have recognizable homologous genes in other species). We note that this does not mean that these often excluded genes do not convey any phylogenetic information, as a recent study suggested that paralogous genes may also have the potential for phylogenetic reconstruction [46].

Therefore, in addition to potential impact from missing data [47] and evolutionary events such as convergent evolution [48] that could mask the true phylogeny, phylogenies estimated by the conservation-oriented and derivedness-oriented approach could possibly be inconsistent.

## 2.2. Conservation-Oriented Approach May Be Less Sufficient in Elucidating Derived Features

We next discuss the potential insufficiency in understanding derived traits and species if only conservation-oriented methods are deployed. Especially, arguments over the evolutionary rate often become confusing or seemingly contradictory when conservation and derivedness are mixed up. One good example would be the discussion on the evolutionary rate of species such as “living fossils”. These species, such as the coelacanth and the tuatara, retain many ancestral features and remain morphologically similar to their fossil relatives that existed millions of years ago. In other words, their overall rates of morphological changes tend to be slow, and it may be more suitable to describe them as less derived species rather than highly conserved species. Seemingly consistent with this, the genomic evolutionary rates of these species estimated on the basis of conserved 1:1 orthologous genes are reportedly slower than those of their sister groups [49,50]. However, occasional counterarguments to this view [51,52] point out that studies which demonstrated rapid molecular evolution in the coelacanth, especially those focused on specific gene sets, are often neglected. Such conflict could potentially arise because these molecular arguments tend to be based on commonly shared genes, or conservation-oriented evaluations while the former standpoint is focusing more on derivedness-oriented features. To be specific, an example is the analysis performed by Cavin and Guinot [53], which showed that the coelacanth acquired much fewer autapomorphies (new derived traits) per million years during its evolution than two other crown-group vertebrate species. Thus, living fossil species are expected to show low phenotypic derivedness as a whole (also considering the nonshared traits [53]), but this does not necessarily mean that the rate of sequence evolution (estimated with conservation-oriented methods for a limited set of genomic sequences) must also be slow. To add, this cannot be fully explained by a gap between genotype and phenotype. Even if we could measure evolutionary rate of all the conserved traits that are shared in species being compared, it would be insufficient to evaluate if the species acquired fewer novel traits or lost fewer traits to remain less derived.

## 3. Technical Limitations of Current Conservation-Oriented and Derivedness-Oriented Molecular Approaches

Although both conservation-oriented and derivedness-oriented approaches have different strengths in understanding evolution, each approach has various limitations, and thus methodology should be carefully selected depending on the purpose of research.

The conservation-oriented molecular comparisons tend to rely on commonly shared genes (e.g., 1:1 orthologs), and this approach has a limitation when a large number of species are analyzed. To be specific, due to the strict definition of 1:1 orthologs, these genes often cover only a small part of the entire gene repertoire in the genome, especially when a large number of species are compared. For example, in our recent analysis comparing 13 species of chordates and echinoderms [26], we identified only 271 1:1 orthologous genes, which covered only ~1.5% of all the genes in a typical deuterostome genome (~20,000 genes). While these 1:1 orthologs would still be sufficient for reconstructing phylogeny [38] or estimating the evolutionary rate of each species, their analysis may not be sufficient for evaluating derived features, or changes that accumulated during evolution, as it ignores changes that took place in the remaining ~98.5% of the gene repertoire. Even when a

more sophisticated search method was employed, only 1126 1:1 orthologs were found in 18 metazoan species [54]. As a result, the evolutionary changes within more than 90% of the genes will be excluded. Thus, the conservation-oriented molecular approach may significantly underestimate evolutionary changes, especially when derivedness is in focus. Consequently, the conservation-oriented approach may be less sufficient to elucidate the complete molecular mechanism of how derived phenotypic traits emerged or evolved. One good example would be a study done by Gildor and colleagues [11]. To understand how changes in developmental gene expressions alter morphogenesis of echinoderm species, they analyzed the gene expression profiles of 8735 1:1 orthologs among three echinoderm species (a sea star and two sea urchins), and these 1:1 orthologs correspond to only around half of the entire gene repertoire of each species. This conservation-oriented approach was sufficient to identify conserved developmental stages (such as the gastrula stage) among the three echinoderm species. However, as the authors argued, they pointed out that their analysis might have underestimated the differences of expressions of nonshared genes (such as genes that are important for skeletogenesis in sea urchins but are absent from the sea star genome). And they also argued that these overlooked genes might contribute to species-specific, or at least sea urchin-specific, morphological features. A similar discussion was made in the study trying to elucidate the evolution of an elaborated structure called the “helmet” in treehoppers. By comparing the transcriptomic profiles of 1:1 orthologs between treehoppers and leafhoppers (their sister group which retains an ancestral condition in the dorsal wall), Fisher and colleagues [15] found that the elaborated helmet may have evolved through coopting the wing-patterning network from the common ancestor of treehoppers and leafhoppers. Using this conservation-oriented approach, many commonly shared genes in the wing-patterning network could be identified as being expressed in the elaborated helmet, supporting the co-option hypothesis. However, this has yet to explain why the coopted gene expressions could transform the dorsal wall into a much more elaborated helmet morphology in treehoppers. Further studies are awaited to fully elucidate the molecular mechanism, and it is tempting to know if derivedness-oriented approach encompassing the nonshared genes would yield additional insights. As in the other studies, 7635 1:1 orthologs could be identified from the de novo assembled transcriptomes of the two species, which only corresponds to approximately half of the entire gene repertoire.

In contrast to the conservation-oriented approach, evolutionary changes achieved by duplicated, newly acquired and lost genes could play significant roles in phenotypic evolution because they are often considered as important drivers of phenotypic innovations [55–57]. However, despite the possibility to include these nonshared genes into comparison, current derivedness-oriented approaches also have a major limitation. There is still no widely accepted way to compare the attributes (such as expression levels) of paralogs, newly acquired genes, species-specific orphan genes, and potentially lost genes across species, although several trials have attempted to compare them using an ortholog-group approach [7,58]. In Levin et al. [58], the expression level of an ortholog-group was determined by the expression level of the gene with the largest expression fold change, although they found no significant difference when randomly selecting a paralogous gene as the representative expression level of the ortholog-group. Similarly, in Hu et al. [7], the authors found essentially the same results (i.e., persistent conservation of the mid-embryonic body plan-developing phase in vertebrates) by taking either the mean or the sum expression level of all predicted paralogous genes to define the expression level of an ortholog-group. These approaches are mostly based on the assumption that the functions of the putative ancestral gene became distributed among paralogs while neofunctionalization could occur, and yet further research is needed to assess to what extent these estimation methods are appropriate, or whether more sophisticated models of transcriptome evolution should be incorporated into the estimation (such as insights from studies that aimed to investigate the evolution of gene expressions involving duplicated genes [59,60]). For other

potential weak points of the derivedness-oriented molecular approach, we will further discuss them in the next section.

Lastly, we note that although we discussed comparing 1:1 orthologs in comparative transcriptomic studies tend to be a conservation-oriented method, it does not mean that comparing 1:1 orthologs must always be conservation-oriented. A possible exception is that when two organisms being compared have the exact same pairs of orthologs, comparison of 1:1 orthologs could be regarded as both conservation-oriented and derivedness-oriented method. Hence, methodologies are not always linked with either conservation-oriented or derivedness-oriented, and analytical methods and data should be designed to fit the purpose of research.

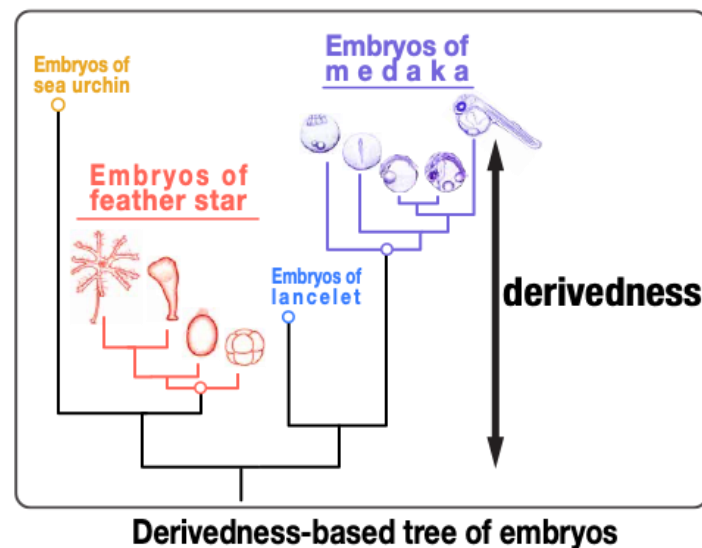
#### 4. “Transcriptomic Derivedness Index” for Quantifying Degree of Phenotypic Evolution of Embryos

In the last decade, the field of evolutionary biology field has shown increased interest in the use of molecular data to study phenotypic evolution, as in the examples discussed above. However, without further development of derivedness-oriented methods, the evolutionary changes that were achieved by genes other than 1:1 orthologs could be largely underestimated or even overlooked.

In this respect, we recently attempted to develop a method to quantify the degree of phenotypic evolution of embryos, or degree of their derivedness, by incorporating the expression profiles of paralogous genes (where we utilized the transcriptome as a phenotype of an embryo) [26]. We focused on echinoderms, a group of animals that are generally considered as highly derived species on the basis of their unique anatomical features, such as pentamer symmetry in the adult stage [61,62]. Meanwhile, it remains to be tested whether their molecular developmental systems, especially their pentamer developmental stages, are also much more derived than the bilateral embryonic stages. Similarly, although the evolutionary conservation of the mid-embryonic period has been clarified in a wide variety of animals (including vertebrates [5,6,12,63–65], *Drosophila* species [66,67], nematodes [68], and mollusks [69]), this does not necessarily mean that this embryonic period remains less derived than the earlier and the later stages because the nonshared genes have not been considered. In our recent study [26], we compared the expression profiles of not only 1:1 orthologs, but also paralogs and potentially lost genes of various echinoderm and chordate species to quantify the evolutionary derivedness of their developmental stages (Figure 3). Data were previously collected from wild-type embryos, with biological replicates from independent parents in order to represent the statistical population of interest by incorporating phenotypic variations among the wild-type embryos.

To mitigate the limitation imposed by conventional conservation-oriented methods, we proposed that transcriptomic derivedness could be better estimated by comparing the expression profiles of ortholog-groups instead of 1:1 orthologs, taking references from the methods used in some pioneering studies [7,58]. To identify the genes of ortholog-groups encompassing distantly related species, we compared protein-coding genes, and identified ortholog-groups using the PORTHOMCL software (which clusters genes on the basis of sequence alignment results from BLASTP) [70,71]. These ortholog-groups also include paralogs, acquired and potential lost genes, so their expression changes could also be evaluated. For the 13 species, 22,689 ortholog-groups were identified, covering an average of 76% of the entire gene repertoire in each species. Meanwhile, only 271 1:1 orthologs could be identified from reciprocal BLAST best-hits. We note that although this approach allowed the comparison using ortholog-groups, the ortholog-prediction software may have limited power to distinguish genes newly acquired in certain lineages and genes lost in other species. In comparing the expression levels of ortholog-groups, the expression level of a certain ortholog-group was calculated as the mean expression value of the constituent paralogous genes (taking the sum expression value yielded similar results). For those species that lack any homologs of the ortholog-group (i.e., potentially lost genes), an expression level of 0 was assigned. We note that the ortholog-group prediction approach

that we utilized may have a limited power to distinguish between newly acquired genes and genes potentially lost in other lineages, but it is reasonable to assign 0 expression levels to these potentially lost genes because they are not expressed. A method suitable for calculating evolutionary distance based on the expression profiles of ortholog-groups was defined by the criteria including: developmental stages cluster by species on the inferred tree of embryonic transcriptomes, and the topology of this tree is consistent with the phylogeny estimated from genomic sequences (as in Figure 3). However, the tree covering orphan genes (genes that have no homologs in the other species) violated the second criterion. Therefore, to be conservative, in further analyses in the study, we adhered to the method and the tree that showed a topology consistent with the phylogeny estimated from genomic sequences, where only 1:1 orthologs, paralogs and potentially lost genes were considered in the calculation (discussed in detail in [26]). We note that it is also possible that the derivedness tree does not strictly match the phylogenetic relationship estimated from genomic sequences, compromising the second criterion. For instance, if the phenotype of a species (including the developmental transcriptomes) evolves extraordinarily rapidly, the phenotypic derivedness tree topology may be inconsistent with that estimated from genomic sequences. Moreover, phylogeny estimated by genomic sequences mostly rely on alignable sequences across species (as explained briefly in Section 2.1) where species-specific orphan genes are excluded. Besides these, accurate estimation of evolutionary changes due to orphan genes could be challenging, because they could be misannotated as they have no homologous counterparts in other species. In *Xenopus laevis* (African clawed frog), ~40% (7879 genes) of its predicted orphan genes are expressed at low or undetectable levels in all developmental stages examined, which could imply misannotation [26]. Therefore, although our proposed method has greatly increased the number of genes that could be compared among species and allowed us to evaluate the transcriptomic derivedness of embryos, this method is still unable to cover evolutionary changes caused by species-specific orphan genes, and further research is required to better accommodate these genes into the estimation of the transcriptomic derivedness of embryos.



**Figure 3.** Quantification of phenotypic derivedness of embryos. The whole-embryonic transcriptome was utilized as a phenotype, and we developed a method to estimate the transcriptomic derivedness of each developmental stage from the common ancestor of echinoderms and chordates. In addition to 1:1 orthologs, expression levels of paralogs and potentially lost genes were also considered when calculating the evolutionary distances between embryonic transcriptomes. Transcriptomic derivedness of each developmental stage was then plotted as the total branch length from the common ancestral node on the inferred tree (modified from [26]).



Finally, using this ortholog-group approach, irrespective of whether the species-specific orphan genes were included or not, we surprisingly found that the embryonic transcriptomes of echinoderm species as a whole may not be much more derived than those of chordates since their common ancestor, but the pentamer phase of echinoderms indeed shows higher transcriptomic derivedness than the bilateral phase. In vertebrate species, we found that the conserved mid-embryonic period is also less derived compared to the earlier or the later developmental stages. We further sought to ask which orthologous gene groups could potentially explain the higher or lower derivedness of these developmental stages of interest. Our analysis showed that the higher derivedness of the pentamer phase of the green sea urchin appears to be attributable to genes being expressed at different levels in the pentamer phase than in the bilateral phase, rather than having different sets of genes being expressed in different phases. Meanwhile, the lower derivedness of the vertebrate mid-embryonic period is correlated with the expression of many genes involved in organogenesis such as the *Hox* ortholog-groups. Interestingly, different functional subsets of ortholog-groups may have different explanatory power in addressing the differences in transcriptomic derivedness index across species. For example, the embryonic transcriptomes of the green sea urchin tended to show higher derivedness than those of the purple sea urchin, but their overall difference was reduced in a tree drawn by ortholog-groups consisting only of transcription factors.

One surprising, or perplexing finding from our study [26] was that the embryos of the tunicate species, *Ciona intestinalis*, showed lower transcriptomic derivedness than those of the other chordate species despite the rapid evolutionary rate of the *C. intestinalis* genome [72,73]. A possible reason for this inconsistency would be that, similar to what we discussed in Section 2 of this review, estimation of evolutionary rate is mostly evaluated using commonly shared genes whereas the estimation of transcriptomic derivedness additionally covers nonshared genes. This could be further complicated by the inclusion of a large number of potentially lost genes in *C. intestinalis* as the effect of assigning 0-expression levels to potentially lost genes remains to be explored. Therefore, it may be interesting to further include developmental transcriptomic datasets of other species, such as the appendicularian tunicate *Oikopleura dioica* which is considered to have lost many essential genes, to investigate whether low transcriptomic derivedness is general in tunicates or not, and to study potential technical biases related to potentially lost genes. Importantly, while our method could be more suitable for evaluating derived features of embryos than merely comparing expression levels of 1:1 orthologs, further research is necessary to take into account technical concerns such as: (i) the bias from potential differences in read depth among RNA-seq samples, which may affect the measured expression levels, (ii) difference in the quality of genomes and annotations, which may affect ortholog-group prediction, and (iii) the effect of assigning 0-expression levels to potentially lost genes. We also found a moderately strong correlation between the transcriptomic derivedness index and the number of predicted ortholog-groups (i.e., more ortholog-groups could be identified for vertebrate and sea urchin species, and their median derivedness indices tend to be higher); however, the reason for this correlation and the extent to which the expression levels of ortholog-groups contributed to their higher derivedness remain unknown; however, the reason for this correlation and the extent to which the expression levels of ortholog-groups contributed to their higher derivedness remain unknown. Future studies are needed to assess the potential influence of the number of ortholog-groups being compared. Finally, when these technical issues are more thoroughly investigated, it would be intriguing to apply this method to study the derivedness of species in other phyla and other phenotypic traits as more data of other lineages of interest continue to accumulate. To add, to more comprehensively evaluate phenotypic derivedness of embryos or species, other phenotypic traits besides the transcriptome should ideally be considered as well.

## 5. Conclusions

To conclude, herein, we discussed the differences between the concepts of conservation and derivedness: conservation mainly represents information retained during evolution, whereas derivedness represents changes during evolution; this difference should be dealt with carefully. Accordingly, evaluations of conserved information are not equivalent to evaluations that include derived or nonshared traits, and confusion between the two could lead to various problems such as inconsistencies between phylogenies estimated by molecular and morphological approaches, and confusing arguments over the evolutionary rates of living fossil species.

In this respect, it is important to carefully select the appropriate method depending on the purpose of research. We discussed the notion that ancestral, shared features could be more easily identified by conservation-oriented methods, while derivedness-oriented methods could be more sufficient in addressing evolutionary changes that were achieved not only by modifications in conserved genes, but also by gene duplication, acquisition, or loss. Technically, when analyzing genome-scale data, 1:1 orthologous genes may cover only a small proportion (~1%) of the entire genome when many species are compared, so conservation-oriented methods may unintentionally discard much of the information which is important for the analysis of phenotypic evolution. However, derivedness-oriented methods are currently not as well developed as conservation-oriented ones. Although the ortholog-group approach has been used in several studies [7,26,58], the question of how to compare the expression levels of ortholog-groups by covering paralogs, acquired genes, and potentially lost genes may still require further investigation. Well-developed evolutionary models for novel and potentially lost traits and genes are also needed. We anticipate that further studies in these aspects will facilitate future development of derivedness-oriented methods.

Finally, considering that phenotypic evolvability addresses phenotypic changeability during evolution, it is expected that highly evolvable features would be, at least, highly derived. Given this, although derivedness itself does not directly reflect evolvability, a potential future application of derivedness-oriented methods would be to find candidate highly evolvable traits or biological processes (see [1] for a more detailed discussion related to the implications of our recently developed method, the “transcriptomic derivedness index”).

**Author Contributions:** Conceptualization, N.I.; methodology, J.C.K.L., M.U. and N.I.; writing, J.C.K.L., M.U. and N.I. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was funded by the National Natural Science Foundation of China (grant number No. 31561143016; No. 31621062) and the Strategic International Collaborative Research Program (SICORP) of JST.

**Institutional Review Board Statement:** The study was conducted according to the guidelines of the University of Tokyo (approval ID: 14-03, 16-2).

**Informed Consent Statement:** Not applicable.

**Acknowledgments:** We thank the 2nd AsiaEvo conference for selecting the poster presentation of Jason Cheok Kuan Leong as a winner, as this led to the opportunity of writing this review paper. We thank Alexandria Borges and Jinwoong Hwang for critically reading the manuscript. We thank one of the reviewers for the idea to include *Oikopleura* in the discussion.

**Conflicts of Interest:** The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

## References

1. Erwin, D.H. The Origin of Animal Body Plans: A View from Fossil Evidence and the Regulatory Genome. *Development* **2020**, *147*, dev182899. [[CrossRef](#)] [[PubMed](#)]
2. Carroll, S.B.; Grenier, J.K.; Weatherbee, S.D. *From DNA to Diversity: Molecular Genetics and the Evolution of Animal Design*, 2nd ed.; Blackwell Publishing: Malden, MA, USA, 2005.

3. Olsen, G.J.; Woese, C.R. Ribosomal RNA: A Key to Phylogeny. *FASEB J.* **1993**, *7*, 113–123. [[CrossRef](#)] [[PubMed](#)]
4. Scotland, R.W. Deep Homology: A View from Systematics. *Bioessays* **2010**, *32*, 438–449. [[CrossRef](#)] [[PubMed](#)]
5. Irie, N.; Kuratani, S. Comparative Transcriptome Analysis Reveals Vertebrate Phylotypic Period during Organogenesis. *Nat. Commun.* **2011**, *2*, 248. [[CrossRef](#)] [[PubMed](#)]
6. Yanai, I.; Peshkin, L.; Jorgensen, P.; Kirschner, M.W. Mapping Gene Expression in Two *Xenopus* Species: Evolutionary Constraints and Developmental Flexibility. *Dev. Cell* **2011**, *20*, 483–496. [[CrossRef](#)] [[PubMed](#)]
7. Hu, H.; Uesaka, M.; Guo, S.; Shimai, K.; Lu, T.-M.; Li, F.; Fujimoto, S.; Ishikawa, M.; Liu, S.; Sasagawa, Y.; et al. Constrained Vertebrate Evolution by Pleiotropic Genes. *Nat. Ecol. Evol.* **2017**, *1*, 1722–1730. [[CrossRef](#)] [[PubMed](#)]
8. Uesaka, M.; Kuratani, S.; Takeda, H.; Irie, N. Recapitulation-like Developmental Transitions of Chromatin Accessibility in Vertebrates. *Zool. Lett.* **2019**, *5*, 33. [[CrossRef](#)]
9. Li, Y.; Kikuchi, M.; Li, X.; Gao, Q.; Xiong, Z.; Ren, Y.; Zhao, R.; Mao, B.; Kondo, M.; Irie, N.; et al. Weighted Gene Co-Expression Network Analysis Reveals Potential Genes Involved in Early Metamorphosis Process in Sea Cucumber *Apostichopus japonicus*. *Biochem. Biophys. Res. Commun.* **2018**, *495*, 1395–1402. [[CrossRef](#)]
10. Hogan, J.D.; Keenan, J.L.; Luo, L.; Ibn-Salem, J.; Lamba, A.; Schatzberg, D.; Piacentino, M.L.; Zuch, D.T.; Core, A.B.; Blumberg, C.; et al. The Developmental Transcriptome for *Lytechinus variegatus* Exhibits Temporally Punctuated Gene Expression Changes. *Dev. Biol.* **2020**, *460*, 139–154. [[CrossRef](#)]
11. Gildor, T.; Cary, G.A.; Lalzar, M.; Hinman, V.F.; de-Leon, S.B.-T. Developmental Transcriptomes of the Sea Star, *Patiria miniata*, Illuminate How Gene Expression Changes with Evolutionary Distance. *Sci. Rep.* **2019**, *9*, 16201. [[CrossRef](#)]
12. Wang, Z.; Pascual-Anaya, J.; Zadissa, A.; Li, W.; Niimura, Y.; Huang, Z.; Li, C.; White, S.; Xiong, Z.; Fang, D.; et al. The Draft Genomes of Soft-Shell Turtle and Green Sea Turtle Yield Insights into the Development and Evolution of the Turtle-Specific Body Plan. *Nat. Genet.* **2013**, *45*, 701–706. [[CrossRef](#)] [[PubMed](#)]
13. Lin, Q.; Fan, S.; Zhang, Y.; Xu, M.; Zhang, H.; Yang, Y.; Lee, A.P.; Woltering, J.M.; Ravi, V.; Gunter, H.M.; et al. The Seahorse Genome and the Evolution of Its Specialized Morphology. *Nature* **2016**, *540*, 395–399. [[CrossRef](#)] [[PubMed](#)]
14. Seki, R.; Li, C.; Fang, Q.; Hayashi, S.; Egawa, S.; Hu, J.; Xu, L.; Pan, H.; Kondo, M.; Sato, T.; et al. Functional Roles of Aves Class-Specific Cis-Regulatory Elements on Macroevolution of Bird-Specific Features. *Nat. Commun.* **2017**, *8*, 14229. [[CrossRef](#)] [[PubMed](#)]
15. Fisher, C.R.; Wegrzyn, J.L.; Jockusch, E.L. Co-Option of Wing-Patterning Genes Underlies the Evolution of the Treehopper Helmet. *Nat. Ecol. Evol.* **2020**, *4*, 250–260. [[CrossRef](#)] [[PubMed](#)]
16. Tejada-Martinez, D.; de Magalhães, J.P.; Opazo, J.C. Positive Selection and Gene Duplications in Tumour Suppressor Genes Reveal Clues about How Cetaceans Resist Cancer. *Proc. R. Soc. B* **2021**, *288*, 20202592. [[CrossRef](#)]
17. Lü, Z.; Gong, L.; Ren, Y.; Chen, Y.; Wang, Z.; Liu, L.; Li, H.; Chen, X.; Li, Z.; Luo, H.; et al. Large-Scale Sequencing of Flatfish Genomes Provides Insights into the Polyphyletic Origin of Their Specialized Body Plan. *Nat. Genet.* **2021**, *53*, 742–751. [[CrossRef](#)]
18. Chan, Y.F.; Marks, M.E.; Jones, F.C.; Villarreal, G.; Shapiro, M.D.; Brady, S.D.; Southwick, A.M.; Absher, D.M.; Grimwood, J.; Schmutz, J.; et al. Adaptive Evolution of Pelvic Reduction in Sticklebacks by Recurrent Deletion of a Pitx1 Enhancer. *Science* **2010**, *327*, 302–305. [[CrossRef](#)]
19. McLean, C.Y.; Reno, P.L.; Pollen, A.A.; Bassan, A.I.; Capellini, T.D.; Guenther, C.; Indjeian, V.B.; Lim, X.; Menke, D.B.; Schaar, B.T.; et al. Human-Specific Loss of Regulatory DNA and the Evolution of Human-Specific Traits. *Nature* **2011**, *471*, 216–219. [[CrossRef](#)]
20. Sadier, A.; Sears, K.E.; Womack, M. Unraveling the Heritage of Lost Traits. *J. Exp. Zool. Part B Mol. Dev. Evol.* **2021**, *338*, 107–118. [[CrossRef](#)]
21. Parker, J.; Tsagkogeorga, G.; Cotton, J.A.; Liu, Y.; Provero, P.; Stupka, E.; Rossiter, S.J. Genome-Wide Signatures of Convergent Evolution in Echolocating Mammals. *Nature* **2013**, *502*, 228–231. [[CrossRef](#)]
22. Castoe, T.A.; de Koning, A.P.J.; Hall, K.T.; Card, D.C.; Schield, D.R.; Fujita, M.K.; Ruggiero, R.P.; Degner, J.F.; Daza, J.M.; Gu, W.; et al. The Burmese Python Genome Reveals the Molecular Basis for Extreme Adaptation in Snakes. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 20645–20650. [[CrossRef](#)] [[PubMed](#)]
23. Gallant, J.R.; Traeger, L.L.; Volkening, J.D.; Moffett, H.; Chen, P.-H.; Novina, C.D.; Phillips, G.N.; Anand, R.; Wells, G.B.; Pinch, M.; et al. Genomic Basis for the Convergent Evolution of Electric Organs. *Science* **2014**, *344*, 1522–1525. [[CrossRef](#)] [[PubMed](#)]
24. Foote, A.D.; Liu, Y.; Thomas, G.W.C.; Vinař, T.; Alföldi, J.; Deng, J.; Dugan, S.; van Elk, C.E.; Hunter, M.E.; Joshi, V.; et al. Convergent Evolution of the Genomes of Marine Mammals. *Nat. Genet.* **2015**, *47*, 272–275. [[CrossRef](#)] [[PubMed](#)]
25. Li, Y.; Omori, A.; Flores, R.L.; Satterfield, S.; Nguyen, C.; Ota, T.; Tsurugaya, T.; Ikuta, T.; Ikeo, K.; Kikuchi, M.; et al. Genomic Insights of Body Plan Transitions from Bilateral to Pentameral Symmetry in Echinoderms. *Commun. Biol.* **2020**, *3*, 371. [[CrossRef](#)]
26. Leong, J.C.K.; Li, Y.; Uesaka, M.; Uchida, Y.; Omori, A.; Hao, M.; Wan, W.; Dong, Y.; Ren, Y.; Zhang, S.; et al. Derivedness Index for Estimating Degree of Phenotypic Evolution of Embryos: A Study of Comparative Transcriptomic Analyses of Chordates and Echinoderms. *Front. Cell Dev. Biol.* **2021**, *9*, 749963. [[CrossRef](#)]
27. Koonin, E.V. Orthologs, Paralogs, and Evolutionary Genomics. *Annu. Rev. Genet.* **2005**, *39*, 309–338. [[CrossRef](#)]
28. Hedges, S.B. Amniote Phylogeny and the Position of Turtles. *BMC Biol.* **2012**, *10*, 64. [[CrossRef](#)]

29. Green, R.E.; Braun, E.L.; Armstrong, J.; Earl, D.; Nguyen, N.; Hickey, G.; Vandewege, M.W.; John, J.A.S.; Capella-Gutierrez, S.; Castoe, T.A.; et al. Three Crocodylian Genomes Reveal Ancestral Patterns of Evolution among Archosaurs. *Science* **2014**, *346*, 1254449. [[CrossRef](#)]
30. Lyson, T.R.; Bever, G.S. Origin and Evolution of the Turtle Body Plan. *Annu. Rev. Ecol. Evol. Syst.* **2020**, *51*, 1–24. [[CrossRef](#)]
31. Hillis, D.M. Molecular versus Morphological Approaches to Systematics. *Annu. Rev. Ecol. Syst.* **1987**, *18*, 23–42. [[CrossRef](#)]
32. Lee, M.S.Y.; Palci, A. Morphological Phylogenetics in the Genomic Age. *Curr. Biol.* **2015**, *25*, R922–R929. [[CrossRef](#)] [[PubMed](#)]
33. Baron, M.G.; Norman, D.B.; Barrett, P.M. A New Hypothesis of Dinosaur Relationships and Early Dinosaur Evolution. *Nature* **2017**, *543*, 501–506. [[CrossRef](#)] [[PubMed](#)]
34. Langer, M.C.; Ezcurra, M.D.; Rauhut, O.W.M.; Benton, M.J.; Knoll, F.; McPhee, B.W.; Novas, F.E.; Pol, D.; Brusatte, S.L. Untangling the Dinosaur Family Tree. *Nature* **2017**, *551*, E1–E3. [[CrossRef](#)] [[PubMed](#)]
35. Baron, M.G.; Norman, D.B.; Barrett, P.M. Baron et al. Reply. *Nature* **2017**, *551*, E4–E5. [[CrossRef](#)] [[PubMed](#)]
36. Yang, Z.; Rannala, B. Molecular Phylogenetics: Principles and Practice. *Nat. Rev. Genet.* **2012**, *13*, 303–314. [[CrossRef](#)] [[PubMed](#)]
37. Felsenstein, J. *Inferring Phylogenies*; Sinauer Associates: Sunderland, MA, USA, 2004; ISBN 0878931775.
38. Kapli, P.; Yang, Z.; Telford, M.J. Phylogenetic Tree Building in the Genomic Age. *Nat. Rev. Genet.* **2020**, *21*, 428–444. [[CrossRef](#)] [[PubMed](#)]
39. Delsuc, F.; Brinkmann, H.; Philippe, H. Phylogenomics and the Reconstruction of the Tree of Life. *Nat. Rev. Genet.* **2005**, *6*, 361–375. [[CrossRef](#)] [[PubMed](#)]
40. Kimura, M. A Simple Method for Estimating Evolutionary Rates of Base Substitutions through Comparative Studies of Nucleotide Sequences. *J. Mol. Evol.* **1980**, *16*, 111–120. [[CrossRef](#)]
41. Muse, S.V.; Gaut, B.S. A Likelihood Approach for Comparing Synonymous and Nonsynonymous Nucleotide Substitution Rates, with Application to the Chloroplast Genome. *Mol. Biol. Evol.* **1994**, *11*, 715–724. [[CrossRef](#)] [[PubMed](#)]
42. Goldman, N.; Yang, Z. A Codon-Based Model of Nucleotide Substitution for Protein-Coding DNA Sequences. *Mol. Biol. Evol.* **1994**, *11*, 725–736. [[CrossRef](#)]
43. Yang, Z. Maximum Likelihood Phylogenetic Estimation from DNA Sequences with Variable Rates over Sites: Approximate Methods. *J. Mol. Evol.* **1994**, *39*, 306–314. [[CrossRef](#)] [[PubMed](#)]
44. Arenas, M. Trends in Substitution Models of Molecular Evolution. *Front. Genet.* **2015**, *6*, 319. [[CrossRef](#)] [[PubMed](#)]
45. Nichio, B.T.L.; Marchaukoski, J.N.; Raittz, R.T. New Tools in Orthology Analysis: A Brief Review of Promising Perspectives. *Front. Genet.* **2017**, *8*, 165. [[CrossRef](#)] [[PubMed](#)]
46. Hellmuth, M.; Wieseke, N.; Lechner, M.; Lenhof, H.-P.; Middendorf, M.; Stadler, P.F. Phylogenomics with Paralogs. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, 2058–2063. [[CrossRef](#)] [[PubMed](#)]
47. Prevosti, F.J.; Chemisquy, M.A. The Impact of Missing Data on Real Morphological Phylogenies: Influence of the Number and Distribution of Missing Entries. *Cladistics* **2010**, *26*, 326–339. [[CrossRef](#)]
48. Zou, Z.; Zhang, J. Morphological and Molecular Convergences in Mammalian Phylogenetics. *Nat. Commun.* **2016**, *7*, 12758. [[CrossRef](#)]
49. Amemiya, C.T.; Alföldi, J.; Lee, A.P.; Fan, S.; Philippe, H.; MacCallum, I.; Braasch, I.; Manousaki, T.; Schneider, I.; Rohner, N.; et al. The African Coelacanth Genome Provides Insights into Tetrapod Evolution. *Nature* **2013**, *496*, 311–316. [[CrossRef](#)]
50. Gemmell, N.J.; Rutherford, K.; Prost, S.; Tollis, M.; Winter, D.; Macey, J.R.; Adelson, D.L.; Suh, A.; Bertozzi, T.; Grau, J.H.; et al. The Tuatara Genome Reveals Ancient Features of Amniote Evolution. *Nature* **2020**, *584*, 403–409. [[CrossRef](#)]
51. Hay, J.M.; Subramanian, S.; Millar, C.D.; Mohandesan, E.; Lambert, D.M. Rapid Molecular Evolution in a Living Fossil. *Trends Genet.* **2008**, *24*, 106–109. [[CrossRef](#)]
52. Casane, D.; Laurenti, P. Why Coelacanths Are Not ‘Living Fossils’. *Bioessays* **2013**, *35*, 332–338. [[CrossRef](#)]
53. Cavin, L.; Guinot, G. Coelacanths as “Almost Living Fossils”. *Front. Ecol. Evol.* **2014**, *2*, 49. [[CrossRef](#)]
54. Creevey, C.J.; Muller, J.; Doerks, T.; Thompson, J.D.; Arendt, D.; Bork, P. Identifying Single Copy Orthologs in Metazoa. *PLoS Comput. Biol.* **2011**, *7*, e1002269. [[CrossRef](#)] [[PubMed](#)]
55. Prince, V.E.; Pickett, F.B. Splitting Pairs: The Diverging Fates of Duplicated Genes. *Nat. Rev. Genet.* **2002**, *3*, 827–837. [[CrossRef](#)] [[PubMed](#)]
56. Chen, S.; Krinsky, B.H.; Long, M. New Genes as Drivers of Phenotypic Evolution. *Nat. Rev. Genet.* **2013**, *14*, 645–660. [[CrossRef](#)] [[PubMed](#)]
57. Albalat, R.; Cañestro, C. Evolution by Gene Loss. *Nat. Rev. Genet.* **2016**, *17*, 379–391. [[CrossRef](#)] [[PubMed](#)]
58. Levin, M.; Anavy, L.; Cole, A.G.; Winter, E.; Mostov, N.; Khair, S.; Senderovich, N.; Kovalev, E.; Silver, D.H.; Feder, M.; et al. The Mid-Developmental Transition and the Evolution of Animal Body Plans. *Nature* **2016**, *531*, 637–641. [[CrossRef](#)] [[PubMed](#)]
59. Guschanski, K.; Warnefors, M.; Kaessmann, H. The Evolution of Duplicate Gene Expression in Mammalian Organs. *Genome Res.* **2017**, *27*, 1461–1474. [[CrossRef](#)] [[PubMed](#)]
60. Fukushima, K.; Pollock, D.D. Amalgamated Cross-Species Transcriptomes Reveal Organ-Specific Propensity in Gene Expression Evolution. *Nat. Commun.* **2020**, *11*, 4459. [[CrossRef](#)] [[PubMed](#)]
61. Hyman, L.H. *The Invertebrates: Echinodermata: The Coelomate Bilateria*; McGraw-Hill: New York, NY, USA, 1955; Volume 4.
62. Brusca, R.C.; Moore, W.; Shuster, S.M. *Invertebrates*, 3rd ed.; Sinauer Associates: Sunderland, MA, USA, 2016; ISBN 1605353752.
63. 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. Zool. Part B Mol. Dev. Evol.* **2005**, *304B*, 150–158. [[CrossRef](#)]

64. Irie, N.; Sehara-Fujisawa, A. The Vertebrate Phylotypic Stage and an Early Bilaterian-Related Stage in Mouse Embryogenesis Defined by Genomic Information. *BMC Biol.* **2007**, *5*, 1. [[CrossRef](#)]
65. Domazet-Lošo, T.; Tautz, D. A Phylogenetically Based Transcriptome Age Index Mirrors Ontogenetic Divergence Patterns. *Nature* **2010**, *468*, 815–818. [[CrossRef](#)] [[PubMed](#)]
66. Kalinka, A.T.; Varga, K.M.; Gerrard, D.T.; Preibisch, S.; Corcoran, D.L.; Jarrells, J.; Ohler, U.; Bergman, C.M.; Tomancak, P. Gene Expression Divergence Recapitulates the Developmental Hourglass Model. *Nature* **2010**, *468*, 811–814. [[CrossRef](#)]
67. Liu, J.; Viales, R.R.; Khoueiry, P.; Reddington, J.P.; Girardot, C.; Furlong, E.; Robinson-Rechavi, M. The Hourglass Model of Evolutionary Conservation during Embryogenesis Extends to Developmental Enhancers with Signatures of Positive Selection. *Genome Res.* **2021**, *31*, gr.275212.121. [[CrossRef](#)]
68. Levin, M.; Hashimshony, T.; Wagner, F.; Yanai, I. Developmental Milestones Punctuate Gene Expression in the *Caenorhabditis* Embryo. *Dev. Cell* **2012**, *22*, 1101–1108. [[CrossRef](#)] [[PubMed](#)]
69. Xu, F.; Domazet-Lošo, T.; Fan, D.; Dunwell, T.L.; Li, L.; Fang, X.; Zhang, G. High Expression of New Genes in Trochophore Enlightening the Ontogeny and Evolution of Trochozoans. *Sci. Rep.* **2016**, *6*, 34664. [[CrossRef](#)] [[PubMed](#)]
70. Li, L.; Stoeckert, C.J., Jr.; Roos, D.S. OrthoMCL: Identification of Ortholog Groups for Eukaryotic Genomes. *Genome Res.* **2003**, *13*, 2178–2189. [[CrossRef](#)]
71. Tabari, E.; Su, Z. PorthoMCL: Parallel Orthology Prediction Using MCL for the Realm of Massive Genome Availability. *Big Data Anal.* **2017**, *2*, 4. [[CrossRef](#)]
72. Berná, L.; Alvarez-Valin, F. Evolutionary Genomics of Fast Evolving Tunicates. *Genome Biol. Evol.* **2014**, *6*, 1724–1738. [[CrossRef](#)]
73. Holland, L.Z. Tunicates. *Curr. Biol.* **2016**, *26*, R146–R152. [[CrossRef](#)]