Chapter 9
Quartet Partitioning Reveals Hybrid Origins of the Vertebrate

Michael Syvanen, Bryan Ericksen, Simone Linz and Jonathan Ducore

Abstract It is generally accepted that humans and sea urchins are deuterostomes and that fruit flies and jelly fish are outgroups. However, when we analyzed proteins from the genomes of these four species and submitted them to 4 taxa phylogenetic analysis, we found that, while as expected, most of the proteins (563) supported the notion of human and sea urchin in one clade and jelly fish and fruit flies in the other clade (Tree1), a large number of proteins (353) showed human and fruit fly in one clade with the sea urchin and jelly fish in the other (Tree3). Homologs were found in the genomes from 5 other metazoa. Tree1 proteins resulted in the expected 9 taxa tree, while the Tree3 proteins show vertebrates, to the exclusion of the other chordates, in the protostome clade. The two 9 taxa trees were fused into a single most parsimonious net that supports an introgression event between a vertebrate ancestor and a primitive protostome.

M. Syvanen (✉) · B. Ericksen
Department of Microbiology, University of California at Davis School of Medicine, Davis, CA 95617, USA
e-mail: syvanen@ucdavis.edu

B. Ericksen
e-mail: ericksen.b@gmail.com

S. Linz
Department of Computer Science, Center for Bioinformatics (ZBIT), University of Tübingen, Sand 14, 72076 Tübingen, Germany
e-mail: simone_linz@yahoo.de

J. Ducore
Department of Pediatrics, University of California at Davis School of Medicine, Davis, CA 95617, USA
e-mail: jmducore@ucdavis.edu

P. Pontarotti (ed.), *Evolutionary Biology: Genome Evolution, Speciation, Coevolution and Origin of Life*, DOI: 10.1007/978-3-319-07623-2_9,

9.1 Introduction

Fossil metazoan phyla appeared over a relatively short period of time 540 million years ago (MYA), an event called the Cambrian explosion or the metazoan radiation. Figure 9.1 shows that major metazoan assemblages appeared in the fossil record 540 MYA. If modern metazoan phyla radiated from a single point, then it would not be possible to assemble the various phyla in higher taxonomic assemblages. However, with the advent of phylogenomics which enabled hundreds if not thousands of genes to be analyzed, it became clear that the radiation occurred over a longer period of time than had been appreciated and that diversification began well before any recognizable metazoans could be seen in fossil record (Wray et al. 1996; Douzery et al. 2004; Blair and Hedges 2005; Philippe et al. 2009; Osigus et al. 2013) as is shown in Fig. 9.1. Thus, it is a realistic goal to reconstruct the pre-Cambrian evolutionary relationships of those taxa that gave rise to the modern metazoa.

In 1985, one of us offered the conjecture that horizontal gene transfer events were a major factor during the emergence of the metazoan phyla as indicated by the widespread occurrence of parallelism in the fossil record (Syvanen 1985).

Since then, a number of horizontal gene transfer events have been documented in metazoans as reviewed by Syvanen (2012). Most of the published examples involve transfers from bacteria or fungi into animals. Documenting gene transfer between metazoan phyla, a type of transfer required to explain parallel evolution implicit in Fig. 9.1, is a much more difficult problem, especially if the transfer events occurred deep in time. Earlier, we presented evidence for a possible major gene transfer event. It was found that the genome of the tunicate *Ciona intestinalis* (Sea squirt) consists of two sets of genes that support two different phylogenies (Syvanen and Ducore 2010). The simplest explanation for this result is that *C. intestinalis* descended from a hybrid, one donor being a chordate and the other belonging to an extinct phyla – likely a sister to primitive protostomes. Quartet partitioning was used to identify the proteins that fell into one or the other of these two groups; we use this method and extend it in the current study. Quartet partitioning has found application in analyzing reticulate evolution (Huson and Bryant 2006; Gauthier and Lapointe 2007), most prominently in identifying relatively recent introgressions between plant species.

9.2 Result

9.2.1 Quartet Partitions

The number of variables required to test alternative trees can be minimized by analyzing four taxa since there are only three competing unrooted trees and a single internal branch (Gaut and Lewis 1995; Hillis and Huelsenbeck 1992;

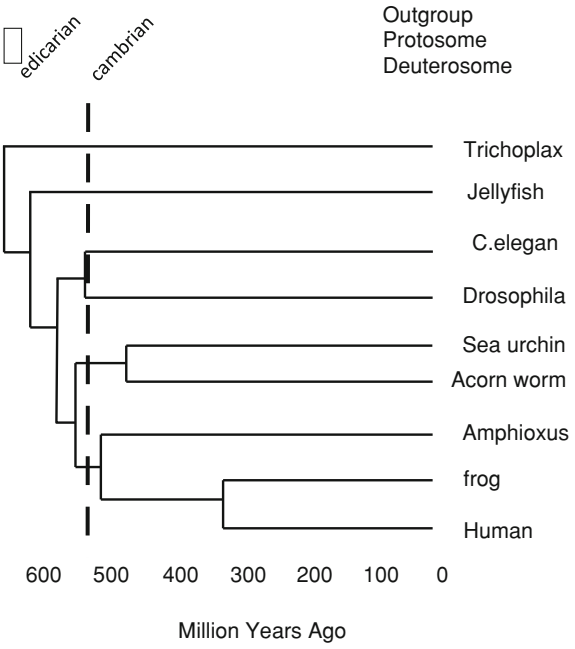

Fig. 9.1 Generally recognized relationships among metazoan phyla. The chronology is based on the geological record and the displayed taxa are the ones used in the current study

Syvanen 2002). We have chosen the jelly fish, *Nematostella vectensis*, *H. sapiens*, the Purple sea urchin *Strongylocentrotus purpuratus* and the fruit fly *Drosophila melanogaster*. Figure 9.2 shows the expected rooted four taxa tree derived from Fig. 9.1 and Fig. 9.3 shows the topologies of the three possible unrooted 4 taxa trees. By convention, Tree1 shows the generally accepted relationship among the four taxa while and Tree2 and 3 show the two alternatives. According to simple parsimony, the best tree is the tree that has the most phylogenetically informative characters (PIC) in its support. (This principle applies as well to weighted parsimony, maximum likelihoo, Bayesian, and protein distance methods though there are quantitative differences between these different approaches.) Let us assume that Tree1 represents the evolutionary history of the four taxa. Tree1 can then be supported by single changes that occur on the internal branch (refer to Fig. 9.3). Tree1 can also be supported by multiple (parallel or convergent) changes that occur on the distal branches that are homoplastic replacements. Thus, if Tree1 represents the actual history, then the number of phylogenetic informative characters (PIC) in its support (defined as N_1) will be those in which Sea urchin and human share one character and the fruit fly and jelly fish share another. N_1 will be determined by the sum of changes on the central branch and the homoplastic changes. There will also be PIC where the other two pairs of taxa share characters that can only arise by means of homoplastic changes on the distal branches. If the distal branches are relatively equal in length and the occurrence of homoplastic changes is randomly distributed, then we would expect to see the number of PIC due to homoplasy to be approximately equal, in which case $N_1 > N_2 = N_3$.

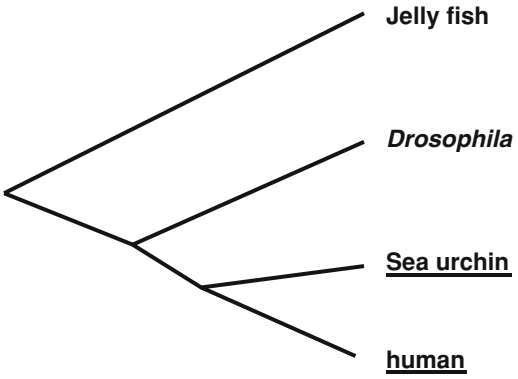

Fig. 9.2 Rooted 4 taxa tree. Shown are the four taxa that are used in the quartet partitioning. This is the generally accepted relationship among the four taxa showing the two deuterostomes—human and sea urchin to the exclusion of the protostome (drosophila) and jelly fish (cnidarian) outgroups

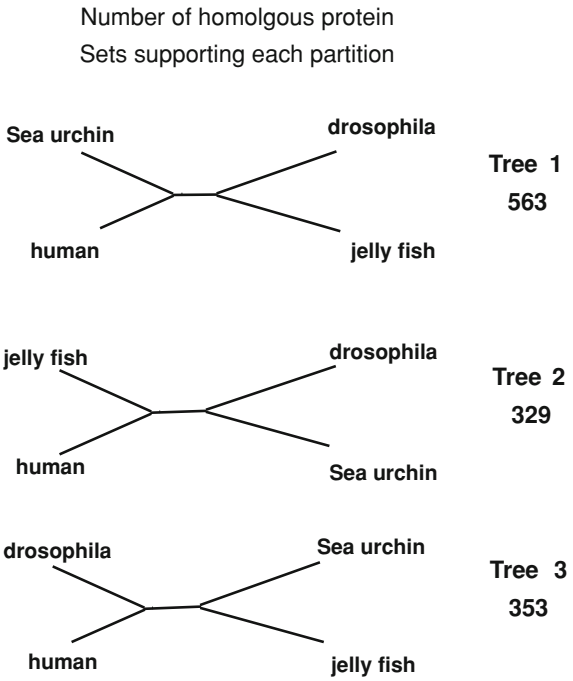

Fig. 9.3 Three unrooted four taxa topologies. Tree1 is the unrooted version of Fig. 9.2. Tree2 and Tree3 are the two remaining topologies. This defines the three topologies and the taxa used in quartet partitioning. The numbers correspond to the number of homologous protein sets that support each tree. The Tree3 seen here captures a relationship seen in Tree3 in our earlier paper (Syvanen and Ducore 2010)

Or more generally

$$N_i > N_j = N_k \tag{9.1}$$

where Tree i is the most parsimonious or the preferred tree.

A priori we can consider N_1 as putative support for Tree1, N_2 as support for Tree2 and N_3 as support for Tree3. By the principles of parsimony, the tree with the largest PIC in its support is the preferred one. Thus, the empirical finding of,

Table 9.1 Four taxa analysis of the protein sets in common between *H. sapiens*, *N. vectensis*, *S. purpuratus* and *D. melanogaster*

	N_1	N_2	N_3
(i) There 563 protein sets supporting Tree1			
PIC (total)	4,835	2,766	2,953
PIC (average)	8.5	4.9	5.2
Ratio	1.7	1.0	1.1
Chi sq (P) = 6 (0.1)			
(ii) There are 329 protein sets supporting Tree2			
PIC (total)	1,557	2,776	1,638
PIC (average)	4.7	8.4	4.9
Ratio	1.0	1.8	1.05
Chi sq (P) = 2 (0.5)			
(iii) There are 353 protein sets supporting Tree3			
Total	2,310	2,321	4,100
PIC (average)	6.5	6.6	11.6
Ratio	1.0	1.0	1.8
Chi sq (P) = 0.1 (0.95)			

The three sets were identified by the bootstrap partition. If more than 70 % of the bootstrap replicates supported Tree1, Tree2, or Tree3, then that protein set was assigned to that particular partition. Tree1 = (hu, su)(dr, cn), Tree2 = (hu, cn)(su, dr) and Tree3 = (hu, dr)(su, cn) as in Fig. 9.3. Chi square and (probability) give the results of the chi square that tests the distribution of N_1, N_2, and N_3 does not significantly deviate from the model $N_i > N_j = N_k$ (Eq. 9.1)

for example, $N_1 > N_2$ and N_3 is taken as evidence that Tree1 reflects the evolutionary history of the four taxa. In our approach, we applied the further restriction that Eq. 9.1 describes the PIC distribution and that deviation from this inequality raises questions about the consistency of the data. In four taxa analysis controls, we have shown that Eq. 9.1 holds reasonably well for those taxa that have undisputed relationships.

The current work with the four species shown in Fig. 9.2 begins by identifying a common set of proteins using Blast. Those proteins that are members of large gene families (i.e., copy numbers in excess of 10 in any of the four taxa) were excluded. This process identified about 2800, quartets that were aligned, submitted to parsimony analysis and the number of PICs in support of each tree determined as described in Methods. These sets will be referred to as "protein sets." The protein sets supporting alternative phylogenies are identified by determining the phylogeny for each of the approximately 2,800 protein sets and assessing bootstrap support for each tree. Each protein was submitted to a bootstrap analysis (200 replicates). Only those protein sets that had bootstrap support >70 % were included. There were only about 1,200 protein sets that significantly supported one of the three trees, the remaining 1,600 protein sets were excluded from further analysis. As summarized in Table 9.1, more protein sets support Tree1 (563

protein sets) than Tree2 (329 protein sets) or Tree3 (353 protein sets), nevertheless a large number of protein sets supported Tree2 and Tree3. The total number of PICs that support Tree1, Tree2, and Tree3, respectively, were combined and these combined sets support Eq. 9.1 (see Sect. 9.2.5 for further discussion of Table 9.1).

9.2.2 Nine Taxa Analysis

The four taxa partitions suggest potentially different phylogenies. However, there is not enough information to explain why. We therefore used the human sequence from each of the three partitions in separate Blast searches against a data base of nine taxa. This larger group of taxa includes two outgroups to the main metazoan cluster (the cnidarian, *N. vectensis* and the placozoan, *Trichoplax adherens*). The resulting trees can be rooted to these two outgroups so that the ancestral node for the Deuterostome/Protostome bifurcation can be identified. Increasing the number of taxa also allows one to assess whether or not "the taxon sampling artifact" is responsible for the incongruent trees revealed by the four taxa analysis (Lecointre et al. 1993; Matus et al. 2006; Dunn et al. 2008).

Let us first focus on the 9 taxa trees produced by the Tree1 and Tree3 supporting partitions. Figure 9.4 shows the topologies based on maximum likelihood analysis of the respective concatenated protein sets. There are a few salient points. Both trees preserve the 4 taxa topology that was found in Fig. 9.3. Hence, the phylogenetic information defining those two groups is not lost upon increasing the number of taxa. We can see that the two preselected outgroup taxa occupy an appropriate position in the tree, allowing us to infer a root. The Tree1 and Tree3 partitions produce clearly incongruent trees. Similar topologies were observed with parsimony and both Fitch and neighbor-joining analyses of protein distance matrices.

The two trees in Fig. 9.4 were submitted to a maximum likelihood analysis, and the log likelihood scores and standard deviations were recorded according to the Shimodaira and Hasegawa (1999) test to assess their differences. Table 9.2 shows the number of standard deviations separating the two datasets. The Tree1 partition significantly supports the Tree1 topology over Tree3 and vice versa.

The Tree2 partition gave conflicting results. This set seems to experience the taxon sampling artifact. Namely, the relative relationship of the four taxa defining the Tree2 partition changed as more taxa were added. When Tree2 partitions were submitted to the 9-taxa analysis, the four key taxa assumed topologies different from that which would be predicted by the 4 taxa analysis; some of these assumed a Tree1–like appearance and others a Tree3-like appearance (data not shown). This was not seen with the Tree1 and Tree3 partitions. These results indicate that there is considerable homoplasy in the character states for the Tree2 partition when additional taxa are added. No further effort to unravel this puzzle was made.

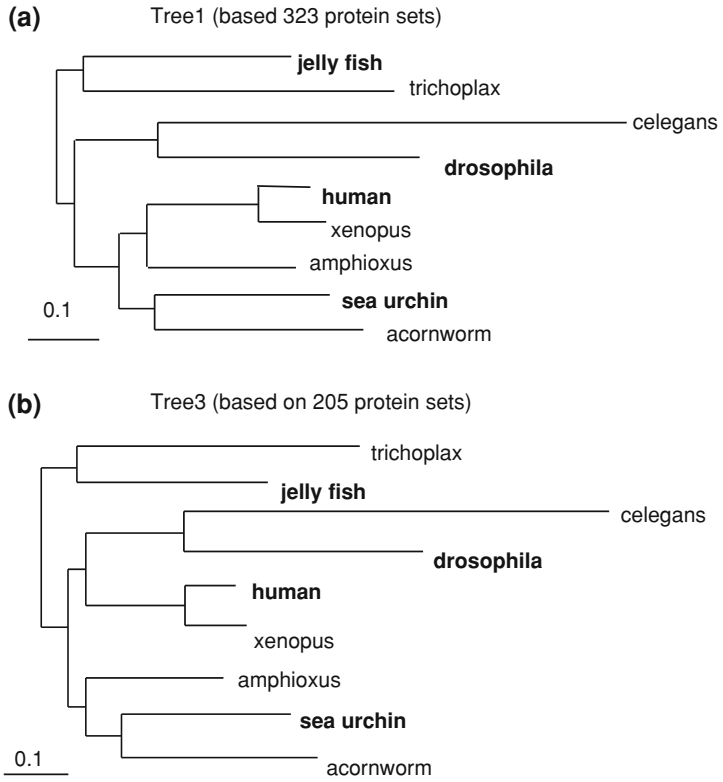

Fig. 9.4 **a** displays the tree using the Tree1 partition and **b** displays the tree using the Tree3 partition. All of the protein sets for each partition were concatenated into their respective file and the maximum likelihood trees were computed. The bold face taxa makes are those from Fig. 9.3

Table 9.2 Tree1 and Tree3 9-taxa topologies compared to the Tree1 and Tree3 character sets

Characters for:	Number std dev	
	From topology for:	
	Tree1	Tree3
Tree1	<0.1	15
Tree3	17	<0.1

Topology1 is from Fig. 9.4a and topology3 is from Fig. 9.4b and were submitted as user defined trees and analyzed by maximum likelihood. The number of standard deviations was determined using the Shimodaira and Hasegawa (1999) and Templeton test provided in Phylip

9.2.3 A Single Network Reconciling Tree1 and Tree3

Recently, phylogenetic networks have attracted attention as a useful analytical tool. If the underlying dataset contains conflicting signals that are due to reticulation (e.g., horizontal gene transfer or hybridization) a phylogenetic network may be more appropriate than a phylogenetic tree (reviewed in Huson and Scornavacca 2011).

To reconcile Tree1 and Tree3 from Fig 9.4 into a network, we used the algorithm HybridInterleave (Collin et al. 2013). This algorithm decomposes each of two input trees into a set of subtrees by deleting a minimum number of edges such that the resulting two sets are identical. In a subsequent step, the set of subtrees can then be used to reconstruct a network that explains all ancestral relationships given by the two input trees, and whose number of reticulation vertices (i.e., vertices with two incoming edges) is one less than the number of subtrees resulting from the decomposition step (Theorem 2 in Baroni et al. 2005).

Applying HybridInterleave to Tree1 and Tree3 from Fig. 9.4 results in the phylogenetic network depicted in Fig. 9.5. This network identifies the vertebrate as the hybrid clade while all other taxa have descended from the last common ancestor by vertical inheritance. In other words, if the vertebrate clade is removed from Tree1 and Tree3, the resulting 7-taxa trees are identical. Note that Fig. 9.5 depicts the unique phylogenetic network that results from applying HybridInterleave to Tree1 and Tree3. No other network with only one reticulation vertex can simultaneously explain Tree1 and Tree3. Based on this parsimony principle, the ancestor that gave rise to the vertebrate was a hybrid between an early protostome (or a sister group thereof) and a vertebrate ancestor that excludes the chordate amphioxus and the other two deuterostomes, the acorn worm and the sea urchin.

9.2.4 Temporal Patterns of Change

We made an effort to determine an age for the protostome deuterostome bifurcation shown in Fig. 9.4. These efforts were not fruitful because not only are the rates of change variable in the terminal branches of those trees, but variation in the rates of change in the internal branches seemed even larger. However, there is a feature seen in Fig. 9.4 that is noteworthy. We can see that the Human/xenopus clade is much closer to the root in Tree3 than it is in Tree1. Maximum likelihood distances on user defined trees can be in error if the trees are wrong. However, this truncated xenopus/human branch seen in the Tree3 protein set was also inferred using direct distances in pair-wise distance matrices, i.e., it is supported by the relative rate test. Vertebrate distances calculated from the Tree3 partition set are 18 % closer to the root of the tree than from the Tree1 partition set. This implies the presence of an ancestral lineage that is sister to the protostomes. Figure 9.6 shows the network that incorporates this inferred lineage.

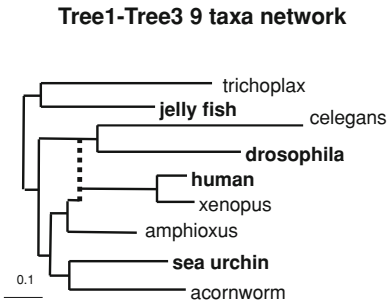

Fig. 9.5 The nine taxa network. The two trees in Fig. 9.2 were submitted to the HybridInterleave algorithm to find the most parsimonious network to resolve the conflicting trees

9.2.5 Alternative Explanations

Statistical Artifact

Let us consider the possibility that there is only a single class of protein sets supporting Tree1, but that the variance of N1, N2, and N3 is very high. According to this scenario, the partitioning method would have selected incorrect trees by chance. If there were a single distribution and it was based upon Tree1, we would expect to see three things not seen in the data.

First, if high variance were responsible for the Tree2 and Tree3 assignments, we would expect the number of PICs supporting the incorrect tree to be lower than those supporting Tree1 since small sample size is less reliable than large sample size. The data in Table 9.1 do not show this; in fact the number of PICs per protein set supporting Tree3 is greater than those supporting Tree1. Second protein sets from the Tree2 partition should show a distribution $N2 > N1 > N3$ and those from the Tree3 partition should show a distribution $N3 > N1 > N2$. This follows because in the first case we are selecting for $N2 > N1$ AND N3. While N1 may be a low outlier in some samples, N2 would be a high outlier in others. Since in either case N3 should remain normally distributed, we would expect $N1 > N3$. The same argument applies to the Tree3 protein set where we would expect $N1 > N2$. The data in Table 9.1 rules this out. Third, if a high variance were causing an incorrect assignment, we would expect that the Tree1 partition, as seen in the N1, N2, N3 distribution, would more robustly support its tree than would the Tree2 and Tree3 partitions. This is also not the case. As shown, the Ni:Nj:Nk distribution roughly equals 1.7:1:1 for each of the three partitions in Table 9.1.

Confusing Parology with Orthology

In our initial screen that identified protein sets, we selected for homology and did not distinguish orthology from parology. We tried to minimize this problem by eliminating large protein families in the original blast search. However, we should expect some paralogous families in the final dataset. These would arise if the last common ancestor of the four taxa contained multigene families but orthologues

were differentially lost in one of the taxa. In the original 4 taxa Blast searches for homologs, we used *H. sapiens* proteins as the query. The top score for each of the other three taxa were assembled as a protein set. We repeated the entire Blast search, once using the *N. vectensis* as the query and then again using *D. melanogaster* as the query. Basically, different Blast searches resulted in datasets (after the bootstrap selection for Tree1, Tree2, and Tree3 supporting protein sets) that yielded Tree1 and Tree3 9-taxa topologies similar to that seen in Fig. 9.4 (data not shown). To be sure, the number of Tree1 and Tree3 supporting protein sets resulting from these different searches are not the same. This is probably due to the fact that there are some sets composed of paralog/ortholog mixtures.

A second effort was made to minimize the parology/ortholog confusion. We edited the entire 1,200 protein set for duplicate unique identifying proteins and then removed those sets that had a protein found in more than one file. This truncated the number of protein sets but it did not change the relative support for Tree1, Tree2, and Tree3. Thus, there is no extreme bias toward selective loss in one taxa that could explain the tree incongruity in Fig. 9.4. The simplest explanation for the tree incongruity is a preexistence of two groups of protein sets with different evolutionary histories.

9.3 Discussion

The most straightforward explanation for the results presented here is that a primitive vertebrate ancestor, appearing after the split from the cephalochordates and tunicates, received an influx of genes from some unknown ancestor that is likely a sister group to modern protostomes. The size of the influx can only be approximated, but given the relative size of the partitions in Table 9.1, it appears that at least 20–30 % (if not more) of the modern vertebrate genes moved into the vertebrate lineage by this mechanism. If a single event is responsible, it is probably simplest to invoke a major hybridization between taxa that likely belonged to different phyla. Though entertaining such a big genetic upheaval may seem like a radical concept, there has been acceptance of the idea that a major genetic rearrangement occurred in an ancestor of the vertebrates that occurred after the cephalochordates and tunicates had diverged. This theory posits that there were two complete genome duplications during this period or that vertebrates evolved from a polyploid ancestor. Hughes and Friedman (2003) employed phylogenetic analysis of many duplicated genes to test this hypothesis. They found very little support for even a single duplication event, i.e., phylogenetic analysis revealed that duplicated regions of the chromosome diverged earlier than would be predicted by a simple genome duplication. If at least one of the major duplication events was the result of remote species hybridization, this pattern would be expected since that "duplication" event would be timed to the protostome-deuterostome speciation event, not to the hybridization event.

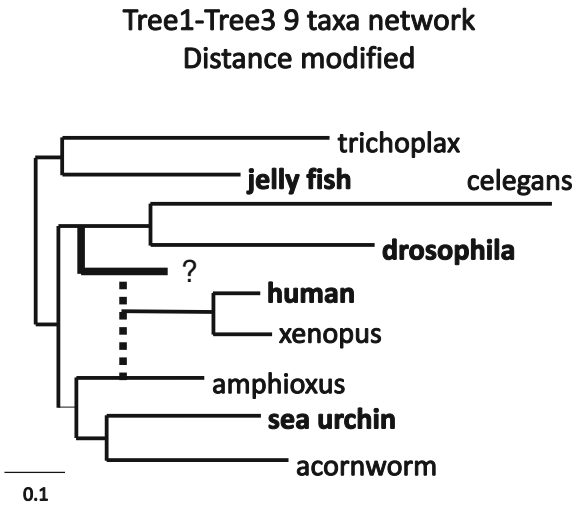

Fig. 9.6 Modified nine taxa network. A hypothesized internal branch was added to accommodate the molecular distance discrepancy seen in Fig. 9.4. The time to the last common vertebrate ancestor to in Tree1 (Fig. 9.4a) and Tree3 (Fig. 9.4b) must be the same. However, the molecular distance of the last common ancestor to is much shorter in Tree3 than in Tree1. This means the rate of evolution in an ancestral lineage for the Tree3 partition is much slower than is the rate for the Tree1 partition. Since the rate of evolution of the other protostome taxa seem to be even faster than the chordate lineages, a new unknown ancestral lineage is postulated

We have made some efforts to use molecular clock considerations to estimate the time of the hybridization event. This is not possible with any reasonable degree of precision, but we can see that there is considerable distance between the hypothesized influx of genes and the last common ancestor to the xenopus/human bifurcation; hence it appears quite possible that the event occurred early, probably before the Cambrian. We found that the 9-taxa Tree1 displays a much larger distance between the vertebrate LCA and the outgroup when compared to the LCA and outgroup distance in 9-taxa Tree3 (Fig. 9.4). This serendipitous result unexpectedly revealed properties of the donor ancestor. Namely, the donor ancestor experienced a relatively long period of evolution with an unusually slow molecular clock as compared to the extant taxa. This, we believe, reflects large differences in the rate of protein evolution among the lineages, including lineages in the internal branches. Further, these large differences in rates between ancestral lineages provide us with evidence that the evolutionary history of Tree3 partition proteins found in vertebrates followed a significantly different path than did the evolution of these proteins in the extant protostomes. Application of the HybridInterleave algorithm to the nine taxa Tree1 and Tree3 topologies identified the vertebrate as the hybrid clade. This unexpected pattern in rate also provides a second line of evidence supporting the hypothesis that the hybrid clade is the vertebrate.

The result shown in Fig. 9.5 is qualitatively similar to the result we published showing that the *C. intestinalis* evolved from a chordate-protostome hybrid

ancestor (Syvanen and Ducore 2010). Our earlier result cannot explain the current result. The four taxa partition sets that resulted from the *C. intestinalis*, vertebrate, *S. purpuratus, D. melanogaster* quartet behaved independently from the four taxa partitions produced in this study with the jelly fish, *H. sapiens*, *S. purpuratus, D. melanogaster* quartet. That is, the Tree1 partition from this study does not overlap with the Tree1 partition from the earlier study.

One can reasonably ask why such a major evolutionary event would have gone undetected until now. It should be stated that the differences between the protein sets found in partition 1 and partition 3 are not that large. For example, the average protein distances between these two partitions are very small compared to the variance between individual protein sets. Thus, for example, a distance distribution for the Tree1 and Tree3 partition sets, at first appearance, look the same. Additionally, there has not been much interest in looking for deep branch networks given that most work is devoted to finding a single tree (Fuchs et al. 2009; Delsuc 2009; Philip et al. 2005; Blair et al. 2002; DeSalle and Schierwater 2008) even when multiple trees are uncovered (Eitel et al. 2013; Nosenko et al. 2013).

A theory that posits major horizontal gene transfer early in metazoan history can explain two major observations: the taxon sampling paradox and leaf instability, which are two related phenomena that are a reflection of underlying homoplasy in the character data set. The classical "one true tree" theory deals with phenomena of this kind by assuming them to be unexplained noise. A theory that incorporates horizontal gene transfer can provide a mechanistic explanation. The results in this chapter also shed light on what has long been considered a paradox. The fossil record supports the notion that the modern metazoan phyla radiated from a single point in time. However, modern genomics has established that multiple and varied ancestral animals preceded and contributed to the post Cambrian explosion, and considerable parallelism in morphological evolution is evident. A theory of evolution incorporating horizontal gene transfer can also easily explain that apparent paradox.

9.4 Materials and Methods

A group of 3,800 protein sequences from the Human genome sequence were used as query sequences in Blast searches. These sequences were selected from ca 25,000 human proteins on the basis of having homologs in a variety of other metazoa and also belonging to gene families with a copy number less than 10. Searches were made against a database consisting of the protein sequences obtained from the genome projects for the following metazoans: *H. sapiens* (Human Genome Resources 2010), *Xenopus laevis* (JGI 2009) the sea urchin *Strongylocentrotus purpuratus* (Sea Urchin et al. 2006) the fruit fly *D. melanogaster* (Celniker et al. 2002) and the round worm *Caenorhabditis elegans* (*C. elegans* 1998) the amphioxus *Branchiostoma floridae* (Nicholas 2008) the cnidarian *Nematostella vectensis* (Sullivan et al. 2006), and the placozoan

Trichoplax adherens (Srivastava et al. 2008) and the acorn worm *Saccoglossus kowalevskii* (acorn worm). An expectation score of less than 10^{-13} was used in all cases. These Blast results were screened such that each contained at least one homolog for the particular 4 taxa or 9 taxa analysis. For those proteins that had multiple listings for the same taxa, the protein with the smallest expectation value was used.

Sequences from each output file were recovered, and multisequence alignments were performed using Clustal (Thompson et al. 1994) and then gaps were deleted with the sequence editor Gblocks (Castresana 2000). Phylogenetic analysis was performed using the Phylip suite of programs (Felsenstein 2005). Four different types of trees were determined. Simple parsimony, maximum likelihood, nearest neighbor, and Fitch distance trees were determined as noted. For tree and molecular clock estimations protein distances were calculated after concatenating the protein sets for each partition. The Jones, Thornton, Taylor distance matrix (Jones et al. 1992) was used in the distance and maximum likelihood methods. In preliminary screens of the protein sets it was shown that distances up to 2.5 changes per residue were linear with time of divergence (data not shown), and those protein sets containing distances in excess of 2.5 were removed from further consideration. The phylogenetic maximum likelihood program proml was used to calculate log likelihood scores that uses the Shimodaira-Hasegawa test (1999). Programs within Phylip were also used to perform the bootstrap procedure. Treeview (Page 1996) was used for tree visualization. All computations were performed on a standard pc with a Linux OS and data was processed using shell script files, Perl scripts and standard spread sheets. The HybridInterleave algorithm was used transform two incongruent phylogenetic trees into a single phylogenetic network (Collins et al. 2011).

The number of phylogenetic informative characters (N) that supports tree i is $N_i = (\text{pic} - 2T_i + T_j + T_k)/3$ where PIC is the total of number of PICs and T is the total length of the parsimony tree in units of unweighted amino acid differences. In a four taxa tree the only PIC are those in which two taxa share one amino acid and the other two share another.

References

Acorn worm site. ftp://ftp.hgsc.bcm.tmc.edu/pub/data/Skowalevskii/fasta

Baroni M, Grünewald S, Moulton V, Semple C (2005) Bounding the number of hybridisation events for a consistent evolutionary history. J Math Biol 51:171–182

Blair JE, Blair Hedges S (2005) Molecular phylogeny and divergence times of deuterostome animals. Mol Biol Evol 22:2275–2284

Blair JE, Ikeo K, Gojobori T, Hedges SB (2002) The evolutionary position of nematodes. BMC Evol Biol 8(2):7

C. elegans Sequencing Consortium (1998) Genome sequence of the nematode *C. elegans*: a platform for investigating biology. Science 11:2012–2018

Castresana J (2000) Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol Biol Evol 17:540–552
Celniker et al (2002) Finishing a whole-genome shotgun: release 3 of the Drosophila melanogaster euchromatic genome sequence. Genome Biol 3:1–0079
Collins J, Linz S, Semple C (2011) Quantifying hybridization in realistic time. J Comput Biol 18:1305–1318
Delsuc F, Brinkmann H, Chourrout D, Philippe H (2006) Tunicates and not cephalochordates are the closest living relatives of vertebrates. Nature 439:965–968
DeSalle R, Schierwater B (2008) An even newer animal phylogeny. BioEssays 30:1043–1047
Douzery EJ, Snell EA, Bapteste E, Delsuc F, Philippe H (2004) The timing of eukaryotic evolution. Proc Natl Acad Sci USA 101:15386–15391
Dunn CW, Hejnol A, Matus DQ, Pang K, Browne WE, Smith SA, Seaver E, Rouse GW, Obst M, Edgecombe GD, Sørensen MV, Haddock SH, Schmidt-Rhaesa A, Okusu A, Kristensen RM, Wheeler WC, Martindale MQ, Giribet G (2008) Broad phylogenomic sampling improves resolution of the animal tree of life. Nature 452:745–749
Eitel M, Osigus HJ, DeSalle R, Schierwater B (2013) Global diversity of the Placozoa. PLoS One 8(4):e57131
Felsenstein J (2005) PHYLIP (Phylogeny inference package) version 3.6. Distributed by the author. Department of Genome Sciences, University of Washington, Seattle
Fuchs J, Obst M, Sundberg P (2009) The first comprehensive molecular phylogeny of Bryozoa (Ectoprocta) based on combined analyses of nuclear and mitochondrial genes. Mol Phylogenet Evol 52:225–233
Gaut BS, Lewis PO (1995) Success of maximum likelihood phylogeny inference in the four-taxon case. Mol Biol Evol 12:152–162
Gauthier O, Lapointe F (2007) Hybrids and phylogenetics revisited: a statistical test of hybridization using quartets. Syst Bot 32:8–15
Hillis DM, Huelsenbeck JP (1992) Signal, noise, and reliability in molecular phylogenetic analyses. J Hered 83:189–195
Hughes AL, Friedman R (2003) 2R or not 2R: testing hypotheses of genome duplication in early vertebrates. J Struct Funct Genomics 3:85–93
Human Genome Resources 2010. http://www.ncbi.nlm.nih.gov/projects/genome/guide/human/
Huson DH, Bryant D (2006) Application of phylogenetic networks in evolutionary studies. Mol Biol Evol 23:254–267
Huson DH, Scornavacca C (2011) A survey of combinatorial methods for phylogenetic networks. Genome Biol Evol 3:23–35
JGI *X. tropicalis* genome assembly (2009). http://genome.jgi-psf.org/Xentr4/Xentr4.home.html
Jones DT, Taylor WR, Thornton JM (1992) The rapid generation of mutation data matrices from protein sequences. Comput Appl Biosci 8:275–282
Lecointre G, Philippe H, Vân Lê HL, Le Guyader H (1993) Species sampling has a major impact on phylogenetic inference. Mol Phylogenet Evol 2:205–224
Matus DQ, Copley RR, Dunn CW, Hejnol A, Eccleston H, Halanych KM, Martindale MQ, Telford MJ (2006) Broad taxon and gene sampling indicate that chaetognaths are protostomes. Curr Biol 8:R575–R576
Nosenko T, Schreiber F, Adamska M, Adamski M, Eitel M, Hammel J, Maldonado M, Müller WE, Nickel M, Schierwater B, Vacelet J, Wiens M, Wörheide G (2013) Deep metazoan phylogeny: when different genes tell different stories. Phylogenet Evol 67:223–233
Osigus HJ, Eitel M, Bernt M, Donath A, Schierwater B (2013) Mitogenomics at the base of Metazoa. Mol Phylogenet Evol 69:339–351
Page RDM (1996) TREEVIEW: an application to display phylogenetic trees on personal computers. Comput Appl Biosci 12:357–358
Peterson KJ, Cotton JA, Gehling JG, Pisani D (2008) The Ediacaran emergence of bilaterians: congruence between the genetic and the geological fossil records. Philos Trans R Soc Lond B Biol Sci 363:1435–1443

Philip GK, Creevey CJ, McInerney JO (2005) The Opisthokonta and the Ecdysozoa may not be clades: stronger support for the grouping of plant and animal than for animal and fungi and stronger support for the Coelomata than Ecdysozoa. Mol Biol Evol 22:1175–1184

Philippe H, Derelle R, Lopez P, Pick K, Borchiellini C, Boury-Esnault N, Vacelet J, Renard E, Houliston E, Quéinnec E, Da Silva C, Wincker P, Le Guyader H, Leys S, Jackson DJ, Schreiber F, Erpenbeck D, Morgenstern B, Wörheide G, Manuel M (2009) Phylogenomics revives traditional views on deep animal relationships. Curr Biol 19:706–712

Putnam NH, T Butts, Ferrier DEK, Furlong RF, Hellsten U, Kawashima T, Robinson-Rechavi M, Shoguch E, Terry A et al (2008) The amphioxus genome and the evolution of the chordate karyotype. Nature 453:1064–1071

Sea Urchin Genome Sequencing Consortium, Sodergren E et al (2006) The Genome of the sea urchin Strongylocentrotus purpuratus. Science 314:941–952

Shimodaira H, Hasegawa M (1999) Multiple comparisons of log-likelihoods with applications to phylogenetic inference. Mol Biol Evol 16:1114–1116

Srivastava, M, Begovic E, Chapman J, Putnam NH, Hellsten U, Kawashima T, Kuo A, Mitros T, Salamov A, Carpenter ML, Signorovitch AY, Moreno MA, Kamm K, Grimwood J (2008) The Trichoplax genome and the nature of placozoans. Nature 454:955–960

Sullivan JC, Ryan JF, Watson JA, Webb J, Mullikin JC, Rokhsar D, Finnerty JR (2006) StellaBase: the Nematostella vectensis genomics database. Nucleic Acids Res 1:34

Syvanen M (1985) Cross-species gene transfer: implications for a new theory of evolution. J Theor Biol 112:333–343

Syvanen M (2002) On the occurence of horizontal gene transfer among an arbitrarily chosen group of 26 Genes. J Mol Evol 54:258–266

Syvanen M (2012) Evolutionary implications of horizontal gene transfer. Ann Rev Genet 46:341–358

Syvanen M, Ducore J (2010) Whole genome comparisons reveals a possible chimeric origin for a major metazoan assemblage source. J Biol Syst 18:261–275

Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 22:4673–4680

Wray GA, Levinton JS, Shapiro LH (1996) Molecular evidence for deep Precambrian divergences among metazoan phyla. Science 274:568–573