
Genome-scale approach proves that the lungfish-coelacanth sister group is the closest living relative of tetrapods with the BEST program

Yunfeng Shan[1] & Robin Gras[1,2]

The origin of tetrapods has not been resolved for decades. Three principal hypotheses (lungfish-tetrapod, coelacanth-tetrapod, or lungfish-coelacanth sister group) have been proposed. We used the Bayesian method under the coalescence model with the latest program (BEST) to perform a phylogenetic analysis for seven relevant taxa and 43 nuclear genes encoding amino acid residues with the jackknife method for taxon sub-sampling. The results, combined with those of other three genome-scale approaches, successfully prove the hypothesis that lungfishes and coelacanths form a monophyletic sister group and are equally related to tetrapods supported by high Bayesian posterior probabilities of the branch (a lungfish-coelacanth clade) and high taxon jackknife supports.

The origin of land vertebrates (tetrapods) has not been fully resolved until today after debating for many decades. Since the discovery of the "living fossil" in 1938, *Latimeria chalumnae*[1,2],

---

[1]School of Computer Science and [2]Department of Biology, University of Windsor, 401 Sunset Avenue, Windsor, ON N9B 3P4, Canada

the last discovered surviving species of a lineage of lobe-finned fish, was generally considered to be the closest living relative of the land vertebrates, the missing link between aquatic and terrestrial vertebrates. This is still the prevailing opinion in most general biology textbooks[3]. The tetrapod origin always commands considerable popular interest in public and academic fields because of the legendary fish discovery. Three hypotheses have been proposed for the phylogenetic relationship: e.g., lungfish-tetrapod (hypothesis 1, Fig. 1a), coelacanth-tetrapod (hypothesis 2, Fig. 1b), or, lungfish-coelacanth sister group (hypothesis 3, Fig. 1c). The coelacanth-lungfish-tetrapod trichotomy (Fig. 1d) is not generally considered as a hypothesis.

Fig. 1 |

Based on comparative morphological and paleontological studies, the lungfish were historically thought to be the closest living relatives of tetrapods[6,7], but the coelacanths were purported to have that claim[1,4-5] since its discovery in 1938, whereas the coelacanths and lungfish sister group (Tree III) was also proposed [8-10].

For the last two decades, single genes and whole mitochondrial genomes have been sequenced to infer phylogenetic relationships. Lungfishes as the closest relatives of tetrapods were supported by single genes[11-15] and mitochondrial whole genomes[16-20]; the coelacanth as the closest living sister group of tetrapods was preferred by single genes[21], and coelacanth-lungfish sister group relationship was suggested by the single gene[13] and the mitochondrial whole genome[17,19-20], while an unresolved coelacanth-lungfish-tetrapod trichotomy was shown by the 12S rRNA gene[12].

Recently, Takezaki *et al.*[22] re-investigated this question and showed an unresolved trichotomy using 44 genes with the concatenation genome-scale approach. Two recombination activation genes supported lungfish and not the coelacanth as the closet living relative of the tetrapods[15]. Our previous study provided strong evidence that hypothesis 2 is rejected, but hypothesis 1 or 3 could not be determined based on 43 genes with three common phylogenetic methods and three genome-scale approaches[23-24].

Although many morphological, paleontological and molecular phylogenetic studies have attempted to resolve this question, the results so far do not discover unequivocal evidence as to whether the lungfishes are the closet living relatives of tetrapods or that both lineages are equally closely related to tetrapods. Therefore, the origin of tetrapods continued to be debated and still is one of the longest standing major questions in vertebrate evolution.

To resolve the origin question of tetrapods, we here used the Bayesian method under coalescence model with a newly published program (BEST) for genome-scale phylogenetic analysis[25] and the jackknife method for taxon sub-sampling to analyze all 43 nuclear genes encoding amino acid residues that are currently available in Genbank, having considered the results of our previous study using three other genome-scale approaches with all three commonly used phylogenetic methods together. We sampled 7 taxa: Mammal (M), Bird (B), Amphibian (A), Coelacanth (C), Lungfish (L), Ray-finned Fish (R), and Shark (S).

Fig. 2 |

Table 1 |

Tree III was inferred with 88% Bayesian posterior probability of the branch of lungfish-coelacanth for 7 taxon set (Fig.2 and Table 1). Four taxon sets of five 6-taxon sets recovered tree III with Bayesian posterior probabilities of the branch of lungfish-coelacanth from 77 to 93% except for MBCLRS that recovered an alternative tree (Table 1). Seven of all nine 5-taxon sets inferred tree III, but BACLS recovered tree I and ACLRS reconstructed an alternative tree. While BCLS recovered tree I, all other five 4-taxon sets inferred tree III (Table 1).

The taxon jackknife support was 81.0% for tree III, 9.5% for tree I and an alternative tree, and 0 for tree II and tree IV, respectively, with Bayesian method under coalescence model for all 21 taxon sets (Table 1).

Tree III is consistently reconstructed with the Bayesian method under coalescence model in 17 of 21 taxon sets with the highest Bayesian posterior probability as 99%, except that tree I is inferred from BCLS and BACLS, and two alternative trees are recovered from ACLRS and MBCLRS (Table 1). Therefore, we provide strong evidence to support the hypothesis 3, e.g., that coelacanths and lungfish form a monophyletic group that is equally closely related to the tetrapods (Tree III). So, both should be considered as the phylogenetically closest living relatives of tetrapods. Our results agree with the other studies in morphological, palaeontological and molecular analyses below. The coelacanth and lungfish sister group relationship was supported by the single gene[13] and the whole mitochondrial genome[19], and by the nuclear 28S ribosomal RNA gene[17]. This relationship was also proposed [8-10] with comparative morphological and paleontological studies. In our previous study[24], we had observed that gene supports and taxon jackknife supports of tree II were significant lower. Therefore, we rejected hypothesis 2 that the coelacanth is the closest living relative of tetrapods, but we could not determine which of the lungfish (hypothesis 1) or the lungfish-coelacanth sister group (hypothesis 3) is the closest living relative of tetrapods based on phylogenetic analysis of 43 genes with those three common methods and those three genome-scale approaches at that time. This study provides further evidence for rejection of hypothesis 2 because none of 21 taxon sets recovers tree II. Recently published major palaeontological studies during the last decade proposed that lungfish are the closest living relatives of the tetrapods or alternatively, that coelacanths and lungfish form a monophyletic group that is equally closely related to the tetrapods[26,27]. The cause of this puzzle is that the divergence of coelacanth and lungfish happened in a relatively short time within a small (20-30 millions years) window in time around 400 million years ago in paleontological date[3,28]. This results in little time and chance for lineage-specific molecular changes to happen, but much time and opportunity for multiple and parallel changes and their accumulation since the origin of these two lineages[3]. So, for this tough phylogenetic question, it is most difficult to get a high resolution using ad hoc molecular phylogenetic methods and algorithms when the available sequence data of genes are currently limited before the BEST program with Bayesian method under coalescence model can be published[25]. However, we would like to point out that the species tree inferred from gene trees using the BEST program is not always correct for all cases. The wrong species trees such as those of ACLRS and MBCLRS may be recovered from gene trees (Table 1). Therefore, jackknife method for taxon sub-sampling is recommended to obtain statistical confidence with jackknife supports.

In conclusion, we successfully provide strong evidence to accept hypothesis 3 that the lungfish and coelacanth are equally related to tetrapods, and should both be the phylogenetically closest living relatives of tetrapods. These conclusions are supported by high Bayesian posterior probabilities of the branch (a lungfish-coelacanth clade) and high taxon jackknife supports based on the genome-scale phylogenetic analysis of 43 genes using the latest program (BEST)[25] with the Bayesian method under the coalescence model and the jackknife method for taxon sub-sampling, having considered with the results of our previous study with other three genome-scale approaches together.

METHODS SUMMARY

The sequences of encoding amino acid residues of 43 genes were mined from the GenBank using BLAST. Having been compared with the supplementary materials[22], these sequences of 43 genes were previously analyzed using the genome-scale approach of concatenated genes; however, the sequence length of some genes are different (Supplementary Table 1). One gene (FSCN1) is absent because some taxa lack its sequences in GenBank. In order to compare the results with the genome-scale concatenated multiple gene approach[22], the same 7 taxa were included: Mammal, Bird, Amphibian, Coelacanth, Lungfish, Ray-finned Fish, and Shark. Sequences of an individual gene were aligned using ClustalX with default settings. All alignments of single genes were manually edited to exclude insertions or deletions and uncertain positions from further analysis. The phylogenetic analysis software BEST (Version 1.1) with the Bayesian method under the coalescence[25] was used for tree inference under the GTR + Γ + I model and 4 simultaneous Markov chains for 20 million generations, starting with random initial trees and sampling every 2000 generations. Burnin value was 100. The majority rule consensus tree was generated using the remaining trees with posterior probability plotted on each node. We used a jackknife approach to sub-sample 6, 5 and 4 taxa from 7 taxa with permutation and combination. The debate over taxon sampling has not terminated. On the one hand, the accuracy was enhanced dramatically with the addition of taxa[29]. On the other hand, adding taxa can reduce accuracy and increase the probability of distorting the tree topology[30]. Adding characters can always increase the accuracy[29,30]. So, as many genes as possible should be included. The sequence data of 43 genes that are all currently available in GenBank were used in this study. Sequence data sets are available upon request.

1 Fritzsch, B. Inner ear of the coelacanth fish *Latimeria* has tetrapod affinities. *Nature* **327**, 153-154 (1987).

2 Thomson, K. S. *Living Fossil* (W. W. Norton & Company, New York ,1991).

3 Meyer, A. Molecular evidence on the origin of tetrapods and the relationships of the coelacanth. *Trends Ecol. Evol.* **10**, 111–116 (1995).

4 Long, J. A. A new rhizodontiform fish from the Early Carboniferous of Victoria, Australia, with remarks on the phylogenetic position of the group. *J. Vert. Paleontol.* **9**, 1–17 (1989).

5 Zhu, M. & Schultze, H. P. The oldest sarcopterygian fish. *Lethaia* **30**, 293–304 (1997).

6 Panchen, A. L. & Smithson, T. S. Character diagnosis, fossils and the origin of tetrapods. *Biol. Rev.* **62**, 341–438 (1987).

7 Ahlberg, P. E. Postcranial stem tetrapod remains from the Devonian Scat Craig, Morayshire. *Scotland. Zool. J. Linn. Soc.* **103**, 241–287 (1991).

8 Northcutt, R. G. in *The biology and evolution of lungfishes* (eds. Bemis, W. E. , Burggren, W. W. & Kemp, N. E.) 277–297 (Alan R. Liss, New York, 1986).

9 Chang, M. M. in *Origins of the higher groups of tetrapods: controversy and consensus* (eds. Schultze, H.-P. & Trueb, L.) 3–28 (Cornell University Press, Ithaca, New York, 1991).

10 Forey, P. L., Gardiner, B. G. & Patterson, C. in *Origins of the higher groups of tetrapods: controversy and consensus* (eds. Schultze, H.-P. & Trueb, L.) 145–172 (Cornell University Press, Ithaca, New York, 1991).

11 Meyer, A. & Wilson, A. C. Origin of tetrapods inferred from their mitochondrial DNA affiliation to lungfish. *J. Mol. Evol.* **31**, 359–364 (1990).

12 Meyer, A. & Dolven, S. I. Molecules, fossils, and the origin of tetrapods. *J. Mol. Evol.* **35**, 102–113 (1992).

13 Yokobori, A. I., Hasegawa, M., Ueda, T., Okada, N., Nishikawa, K. & Watanabe, K. Relationship among coelacanths, lungfishes, and tetrapods: a phylogenetic analysis based on mitochondrial cytochrome oxidase I gene sequences. *J. Mol. Evol.* **38**, 602–609 (1994).

14 Venkatesh, B., Erdmann, M. V. & Brenner, S. Molecular synapomorphies resolve evolutionary relationships of extant jawed vertebrates. *Proc. Natl. Acad. Sci. USA* **98**, 11382– 11387 (2001).

15 Brinkmann, H., Venkatesh, B., Brenner, S. & Meyer, A. Nuclear protein-coding genes support lungfish and not the coelacanth as the closest living relatives of land vertebrates. *Proc. Natl. Acad. Sci.* **101**, 4900-4905 (2004).

16 Hedges, S. B., Hass, C. A. & Maxson, L. R. Relations of fish and tetrapods. *Nature* **363**, 501–502 (1993).

17 Zardoya, R. & Meyer, A. Evolutionary relationships of the coelacanth, lungfish, and tetrapods based on the 28S ribosomal RNA gene. *Proc. Natl. Acad. Sci. USA* **93**, 5449–5454 (1996).

18 Zardoya, R. & Meyer, A. The complete DNA sequence of the mitochondrial genome of a 'living fossil,' the coelacanth (Latimeria chalumnae). *Genetics* **146**, 995-1010 (1997).

19 Zardoya, R., Cao, Y., Hasegawa, M. & Meyer, A. Searching for the closest living relative(s) of tetrapods through evolutionary analyses of mitochondrial and nuclear data. *Mol. Biol. Evol.* **15**, 506-517 (1998).

20 Zardoya, R., Cao, Y., Hasegawa, M. & Meyer, A. Searching for the closest living relative(s) of tetrapods through evolutionary analyses of mitochondrial and nuclear data. *Mol. Biol. Evol.* **15**, 506–517 (1998).

21 Gorr, T., Kleinschmidt, T. & Fricke, H. Close tetrapod relationships of the coelacanth Latimeria indicated by haemoglobin sequences. *Nature* **351**, 394–397 (1991).

22 Takezaki, N., Figueroa, F. Zaleska-Rutczynska, Z., Takahata, N. & Klein, J. The Phylogenetic Relationship of Tetrapod, Coelacanth, and Lungfish Revealed by the Sequences of Forty-Four Nuclear Genes. *Mol. Biol. Evol.* **21**, 1512-1524 (2004).

23 Shan, Y. & Li, X. Maximum Gene-support Tree. *Evolotionary Bioinformatics* **4**, 181-191 (2008).

24 Shan, Y., Li, X. & Gras, R. The hypothesis that coelacanth is the closest living relative of tetrapods was rejected based on three genome-scale approaches. arXiv:0910.1949v1 [q-bio.PE] (2009).

25 Liu, L. BEST: Bayesian estimation of species trees under the coalescent model. *Bioinformatics* **24**, 2542-2543 (2008).

26 Zhu, M, Yu, X, B. & Ahlberg, P. E. A primitive sarcopterygian fish with an eyestalk. *Nature* **410**, 81-84(2001).

27 Zhu, M. & Yu, X. A primitive fish close to the common ancestor of tetrapods and lungfish. *Nature* **418**, 767-770 (2002).

28 Ahlberg, P. E. Postcranial stem tetrapod remains from the Devonian Scat Craig, Morayshire. *Scotland. Zool. J. Linn. Soc.* **103**, 241–287 (1991).

29 Poe, S. & Swofford, D. L. Taxon sampling revisited. *Nature* **398**, 299–300 (1999).

30 Graybeal, A. Is it better to add taxa or characters to a difficult phylogenetic problem? *Syst. Biol.* **47**, 9–17 (1998).

Supplementary Information Table S1

Acknowledgements We thank Richard Winterbottom for valuable suggestions, and Adam Aspinall for reading and helpful comments. This work is partially supported by the NSERC grant ORGPIN 341854, the CRC grant 950-2-3617, and the CFI grant 203617.

Fig. 1 | Four alternative phylogenetic trees among tetrapod, coelacanth and lungfish lineages.

Fig. 2 | The phylogenetic relationship (tree III) of 7 taxa.

Table 1 | Tree types, Bayesian posterior probability of the branch of lungfish-coelacanth (tree III) or lungfish-tetrapods (tree I) for 7, 6, 5 and 4-taxon sets, and taxon jackknife supports with the Bayesian method under the coalescence model

Table 1 | Tree types, Bayesian posterior probability of the branch of lungfish-coelacanth (tree III) or lungfish-tetrapods (tree I) for 7, 6, 5 and 4-taxon sets, and taxon jackknife supports with the Bayesian method under the coalescence model

Taxon Set	Tree Type	Probability
7 taxon set		
MBACLRS	III	88%
6 taxon sets		
BACLRS	III	90%
MACLRS	III	80%
MBACLR	III	93%
MBACLS	III	77%
MBCLRS	AT	n/a
5 taxon sets		
ACLRS	AT	n/a
BACLR	III	49%
BACLS	I	43%
BCLRS	III	49%
MACLR	III	97%

MACLS	III	64%
MBCLR	III	45%
MBCLS	III	82%
MCLRS	III	62%
4 taxon sets		
ACLR	III	99%
ACLS	III	67%
BCLR	III	40%
BCLS	I	89%
MCLR	III	73%
MCLS	III	73%
JKF:	III (17/21)	81.0%
	I (2/21)	9.5%
	AT (2/21)	9.5%
	II	0
	IV	0

Notes:

The taxa included: Mammal (M), Bird (B), Amphibian (A), Coelacanth (C), Lungfish (L), Ray-finned Fish (R), and Shark (S); JKF = Taxon jackknife supports (%); AT = alternative tree; n/a = not available.

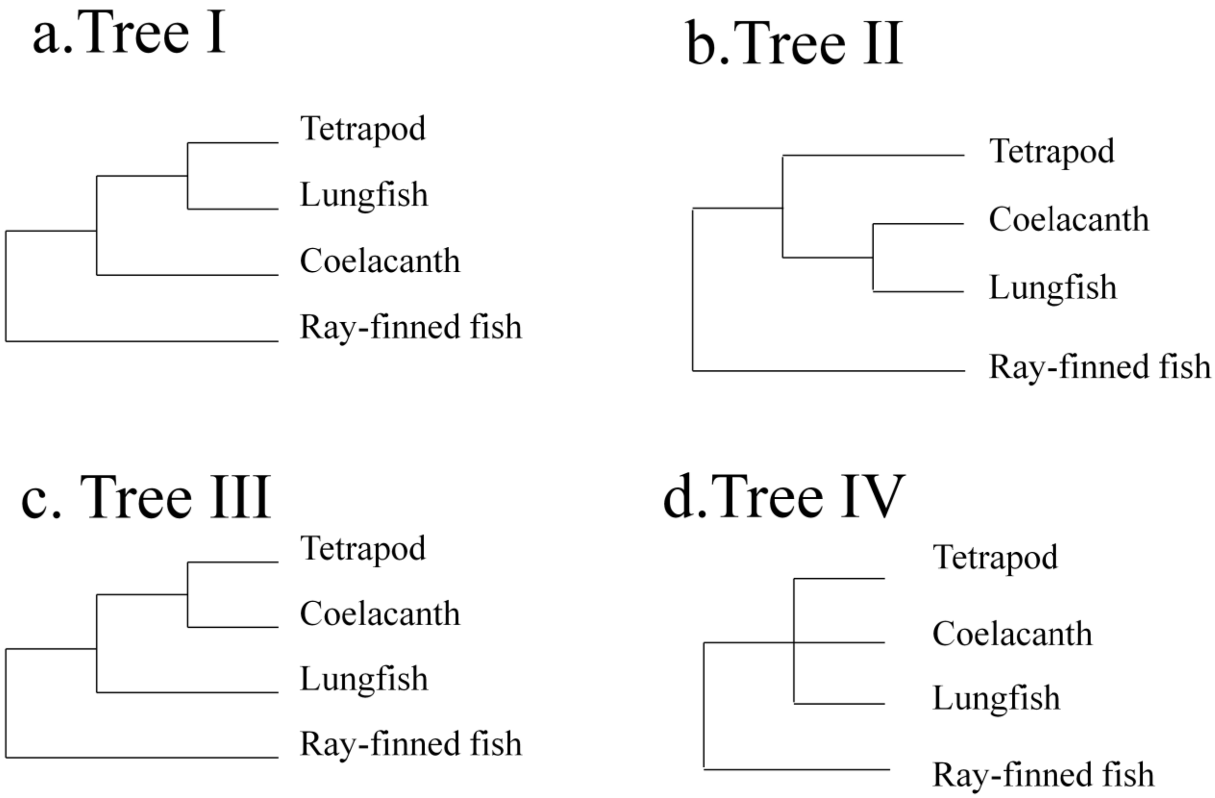
a.Tree I
Tetrapod
Lungfish
Coelacanth
Ray-finned fish
b.Tree II
Tetrapod
Coelacanth
Lungfish
Ray-finned fish
c. Tree III
Tetrapod
Coelacanth
Lungfish
Ray-finned fish
d.Tree IV
Tetrapod
Coelacanth
Lungfish
Ray-finned fish

Fig.1|

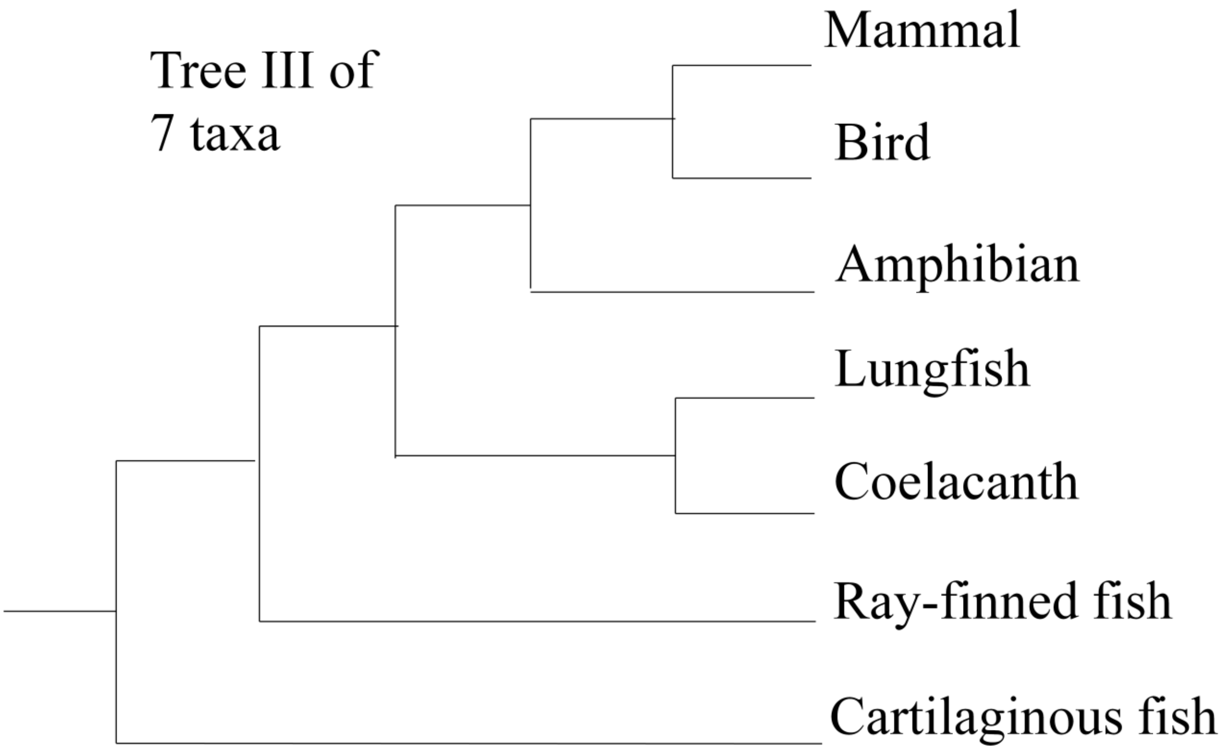
Tree III of
7 taxa
Mammal
Bird
Amphibian
Lungfish
Coelacanth
Ray-finned fish
Cartilaginous fish

Fig. 2|