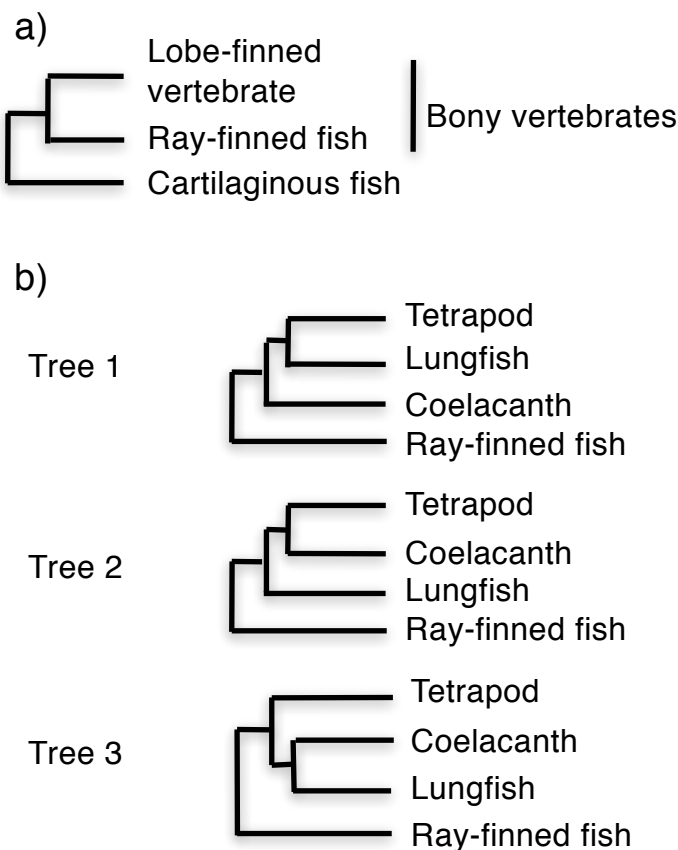


# Resolving the Phylogenetic Position of Coelacanth: The Closest Relative Is Not Always the Most Appropriate Outgroup

Takezaki N, Nishihara H. 2016. *Genome Biol Evol* 8: 1208-1221.

## Abstract

Determining the phylogenetic relationship of two extant lineages of lobe-finned fish, coelacanths and lungfishes, and tetrapods is important for understanding the origin of tetrapods. We analyzed datasets from two previous studies along with a newly collected dataset, each of which had varying numbers of species and genes and varying extent of missing sites. We found that in all the datasets the sister relationship of lungfish and tetrapods was constructed with the use of cartilaginous fish as the outgroup with a high degree of statistical support. In contrast, when ray-finned fish were used as the outgroup, which is taxonomically an immediate outgroup of lobe-finned fish and tetrapods, the sister relationship of coelacanth and tetrapods was supported most strongly, although the statistical support was weaker. Even though it is generally accepted that the closest relative is an appropriate outgroup, our analysis suggested that the large divergence of the ray-finned fish as indicated by their long branch lengths and different amino acid frequencies made them less suitable as an outgroup than cartilaginous fish.



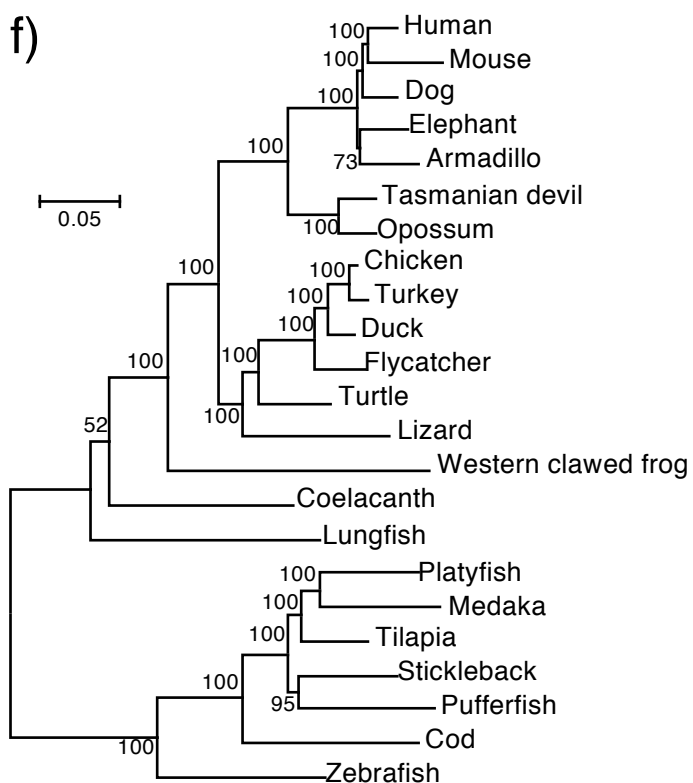
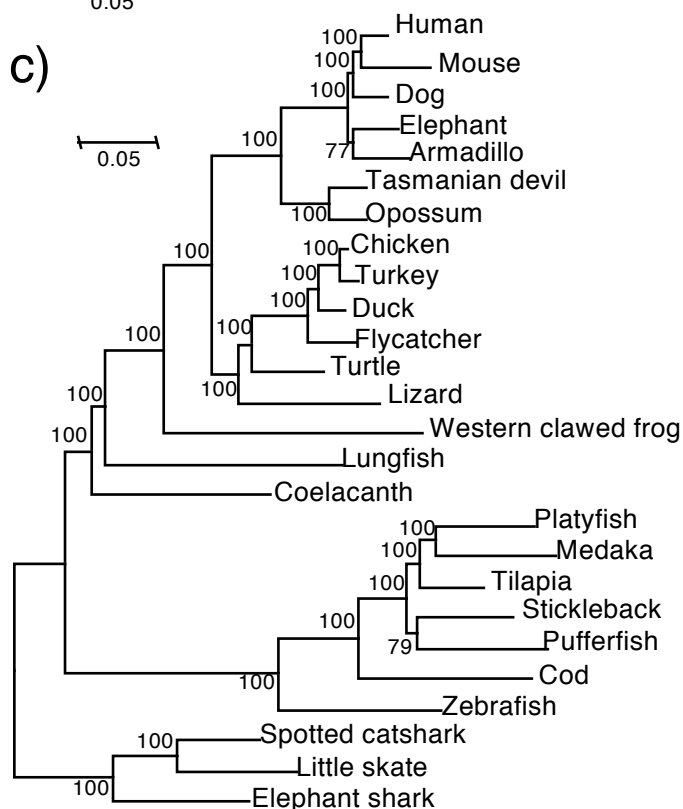
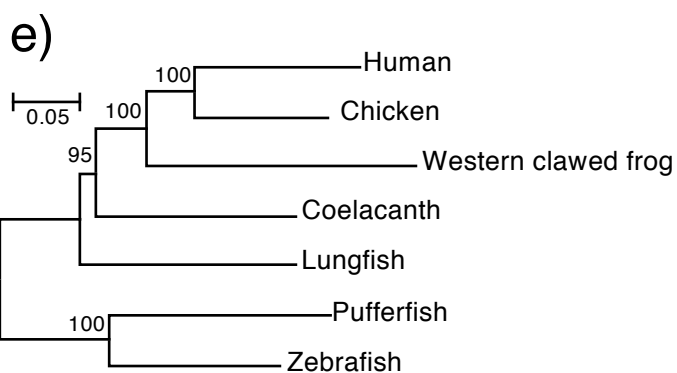
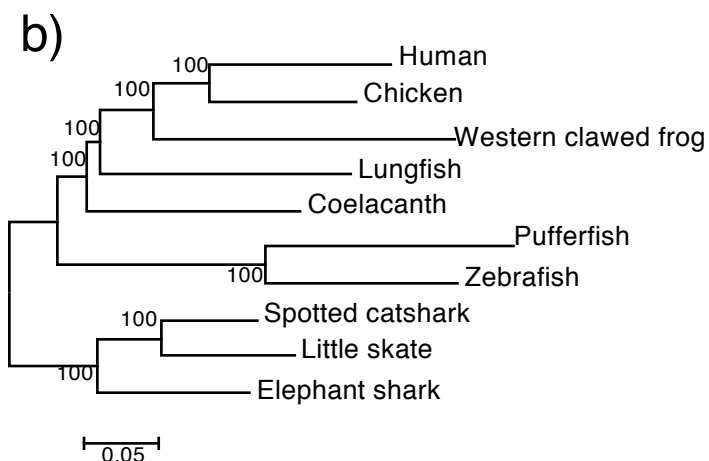
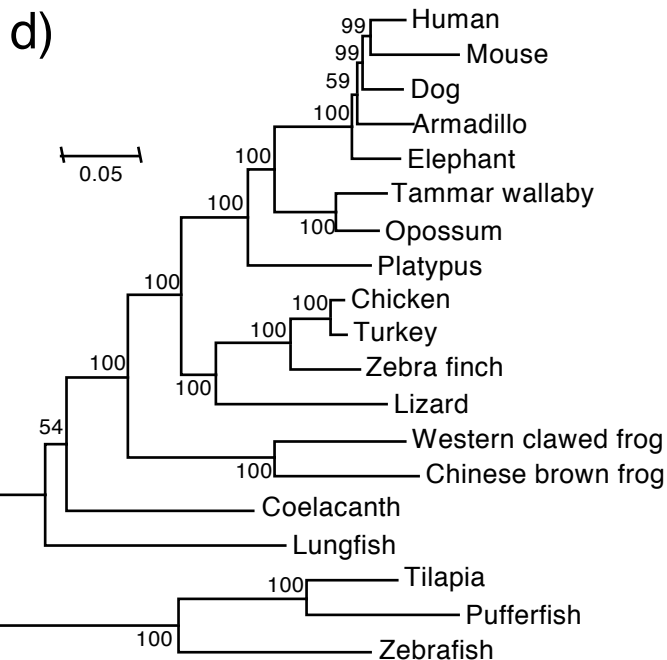
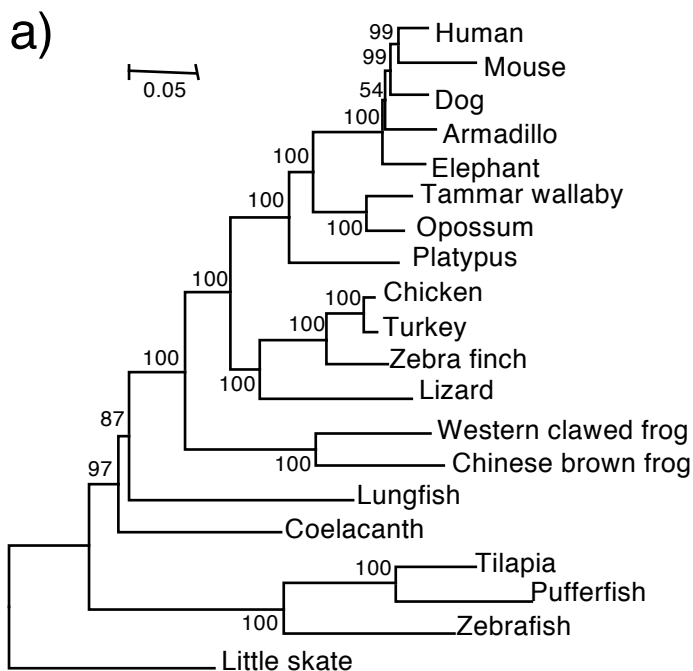
**Fig. 1.**—The phylogenetic relationship of the major lineages in jawed vertebrates and lobe-finned vertebrates. (a) The relationship of major lineages in jawed vertebrates. (b) Three possible relationships for the three extant lineages of lobe-finned vertebrates: sister relationships of lungfishes and tetrapods (Tree 1), coelacanths and tetrapods (Tree 2) and lungfishes and coelacanths (Tree 3).

**Table 1**

Datasets Analyzed in This Study

Note.—In dataset I, only concatenated sequence was available, and two shark species were missing. In dataset II genes with <50 amino acid sites were excluded.

Dataset	Source	Genes	Amino acid sites	Species	Missing sites (%)
I	Amemiya et al.	251	112,212	20	14.2
II	Liang et al.	1,288	618,946	10	6.5
III	This study	831	242,475	25	0



**Fig. 2.**—Maximum likelihood trees constructed for concatenated sequences of the three datasets. (a)–(c): CF and RF were used as the outgroup. (d)–(f): RF were used as the outgroup. (a), (d): Dataset I from Amemiya et al. (2013). (b), (e): Dataset II from Liang et al. (2013). (c), (f): Dataset III collected in this study. The numbers on the branches are BPs from 500 replications. The trees were constructed with the JTTFG4 setting by PhyML.