SG363: Elucidating shark evolution via comparative transcriptomics

Small Grant Report – Anthony Redmond (University of Aberdeen)

Background

The cartilaginous fishes diverged from other jawed vertebrates ~450 million years ago (mya) [1]. Despite this key evolutionary position, the only high-quality cartilaginous fish genome available is for the elephant shark (*Callorhinchus milii*) [2], a chimaera that split from the elasmobranch (sharks and batoids [skates and rays]) lineage ~420 mya [1].

1. *Immunity*

Initial analysis of the elephant shark genome resource led to proposals that components of the cartilaginous fish adaptive immune system, notably the repertoire of CD4+ T cell subsets, were primitive compared to mammals [2]. This was surprising, as robust immune responses are reported in elasmobranchs after immunization [3,4], and the importance of CD4+ T cells to vertebrate immune responses is exemplified by the prognosis of HIV patients. This indicates a need for deeper investigation of cartilaginous fish immunogenetics.

2. *Phylogeny*

The earliest split in the cartilaginous fish phylogeny is generally considered to fall between sharks and batoids, with this scenario being supported by mitogenome analyses [1] and some morphological studies [5]. However, more recent morphological studies, and the largest dataset of nuclear genes [6] used to date do not support this, rather supporting links between batoids and either of two major shark groups, Galeomorphii or Squalomorphii. Further mitochondrial data is not always appropriate for resolving deep relationships in vertebrate phylogenetic trees [7,8]. While, a draft assembly for the little skate (*Leucoraja erinacea*) was available at the outset of this project [9], the lack of genomic data for these two shark groups prohibits the application of phylogenomics to this question.

The project

The goal of this project was to generate multi-tissue transcriptome data for two distantly related, model shark species, the small-spotted catshark (*Scyliorhinus canicula*;

Galeomorphii) and the spiny dogfish (*Squalus acanthias*; Squalomorphii) to provide genomic insight into this poorly studied vertebrate lineage. And then use this data in phylogenomic analyses, with that for elephant shark and little skate to resolve the root of the elasmobranch phylogeny (i.e. whether sharks are likely to be mono- or para-phyletic), and explore the origins of vertebrate immunity, particularly of the CD4+ T cell gene repertoire.

Use of Funds

The funds awarded in this grant, in combination with a Royal Society grant to Helen Dooley (University of Maryland) and a PhD studentship to myself from University of Aberdeen Centre for Genome-Enabled Biology and Medicine, enabled the procurement of equipment and reagents required for animal sampling, through to RNA and cDNA preparation, and RNA-seq of small-spotted catshark and spiny dogfish. This grant funded the computational requirements of de novo transcriptome assembly and phylogenomic analyses.

Materials & Methods

A full suite of organs and other major tissues was harvested from a single mature female each for small-spotted catshark (captive-bred) and spiny dogfish. High quality RNA samples were pooled to create a single species multi-tissue RNA sample which was used for subsequent cDNA synthesis and normalization (to maximize representation of lowly expressed transcripts) using the Evrogen Mint-2 and Trimmer-2 kits, respectively. Small-spotted catshark sequencing was performed on the Ion Proton (Life Technologies) using 2x Ion PI v2 BC Chips (Life Technologies). Spiny dogfish sequencing was intended for the Illumina NextSeq 500, however an initial chip failed, and we then discovered from collaborators also attempting to sequence Evrogen normalized cDNA, that such libraries appear to produce poor yield on the NextSeq. However, we have obtained preliminary spiny dogfish RNA-seq data from a Roche 454ti run.

Following assembly, bioinformatic searches of these datasets were used to identify putative immune genes of interest, and to identify one-to-one orthologues for phylogenomic analyses. Multiple sequence alignments were generated using MAFFT v7 [10]. Maximum likelihood phylogenetic analyses were performed in IQ-tree (omp-1.5.4) [11]. Bayesian phylogenetics were carried out in BEASTv1.8.3 and/or Phylobayes 4.1b [12].

Results

1. *Immune genes in cartilaginous fish transcriptomes*

Bioinformatic searches of the new transcriptome data for elasmobranchs resulted in the discovery of putative orthologues for many previously undiscovered immune genes in cartilaginous fishes. Most interestingly, I found sequences showing similarity to those of the CD4+ T cell pathways thought to be restricted to bony vertebrates. Phylogenetic analyses employing complex modelling strategies and a variety of tree rooting strategies, revealed that IL-6R (Fig. 1A), IL-9 (Fig. 1B), IL4/13 (Fig. 1C), IL-23R (Fig. 1D), IL-23, IL-27, IL-11 (Fig. 1E), and RORC (Fig. 2), all existed in the ancestor of jawed vertebrates, contrary to previous reports. This data is consistent with the presence of a sophisticated, mammalian-like, T cell repertoire in cartilaginous fishes, in line with the evidence for immunological memory and antibody affinity maturation in this lineage. In summary, a mammalian-like repertoire of CD4+ helper and regulatory T cell subsets evolved in the jawed vertebrate ancestor and still exist in cartilaginous fishes (Fig. 3).

2. *The root of the elasmobranch phylogeny*

143 one-to-one orthologues were identified to be present in each of elephant shark, little skate, small-spotted catshark, and spiny dogfish datasets (i.e. no missing data in cartilaginous fishes), as well as most other vertebrates. These were aligned and concatenated into a supermatrix to perform phylogenomic analyses to investigate whether sharks are paraphyletic. Results under partitioned analyses, and under the mixture model CAT [13], which permits across site rate heterogeneity and thereby helps to mitigate phylogenetic artefacts [14], provide maximal support for a sister group relationship between small-spotted catshark and spiny dogfish, to the exclusion of little skate (Fig. 4). This best fits a placement of the elasmobranch root such that batoids are sister to a monophyletic shark group, rather than derived from within, in line with mitogenome data, and early morphological work [1,5].

Publication plan & impact of new datasets to community

I intend to publish this work in two parts, the first of which is currently in preparation:

1. Phyloimmunogenomics of the ancestral jawed vertebrate

2. Phylogenomic resolution of the elasmobranch root

Because of the broad utility of genomic datasets, the new datasets generated here have already provided benefit to other research:

1. Pettinello, R., Redmond, A.K., Secombes, C.J., Macqueen, D.J. and Dooley, H., 2017. Evolutionary history of the T cell receptor complex as revealed by small-spotted catshark (Scyliorhinus canicula). *Developmental & Comparative Immunology*, *74*, pp.125-135.

2. Gillespie, K.M., Bachvaroff, T.R. and Jagus, R., 2016. Expansion of eIF4E and 4E-BP Family Members in Deuterostomes. In *Evolution of the Protein Synthesis Machinery and Its Regulation* (pp. 165-185). Springer International Publishing.

Future Work

-> Obtaining further spiny dogfish RNA-seq data.

-> Sequencing multi-tissue transcriptomes and/or genomes for more shark species.

-> Further exploring the origins of vertebrate immunity and other vertebrate novelties using the current datasets.

References:

1. Inoue JG, Miya M, Lam K, Tay BH, Danks JA, Bell J, et al. Evolutionary Origin and Phylogeny of the Modern Holocephalans (Chondrichthyes: Chimaeriformes): A mitogenomic perspective. Mol. Biol. Evol. 2010;27:2576–86.

2. Venkatesh B, Lee AP, Ravi V, Maurya AK, Lian MM, Swann JB, et al. Elephant shark genome provides unique insights into gnathostome evolution. Nature. 2014;505:174–9.

3. Dooley H, Stanfield RL, Brady RA, Flajnik MF. First molecular and biochemical analysis of in vivo affinity maturation in an ectothermic vertebrate. Proc. Natl. Acad. Sci. U. S. A. 2006;103:1846–51.

4. Dooley H, Flajnik MF. Shark immunity bites back: Affinity maturation and memory response in the nurse shark, Ginglymostoma cirratum. Eur. J. Immunol. 2005;35:936–45.

5. Maisey JG. An Evaluation of Jaw Suspension in Sharks. Am. Museum Novit. 1980;1–17.

6. Li C, Matthes-Rosana KA, Garcia M, Naylor GJP. Phylogenetics of Chondrichthyes and the problem of rooting phylogenies with distant outgroups. Mol. Phylogenet. Evol. 2012;63:365–73.

7. Morgan CC, Creevey CJ, O'Connell MJ. Mitochondrial data are not suitable for resolving placental mammal phylogeny. Mamm. Genome. 2014;25:636–47.

8. Rasmussen AS, Arnason U. Phylogenetic studies of complete mitochondrial DNA molecules place cartilaginous fishes within the tree of bony fishes. J. Mol. Evol. 1999;48:118–23.

9. Wyffels J, L. King B, Vincent J, Chen C, Wu CH, Polson SW. SkateBase, an elasmobranch genome project and collection of molecular resources for chondrichthyan fishes. F1000Research. 2014;3.

10. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. Mol. Biol. Evol. 2013;30:772–80.

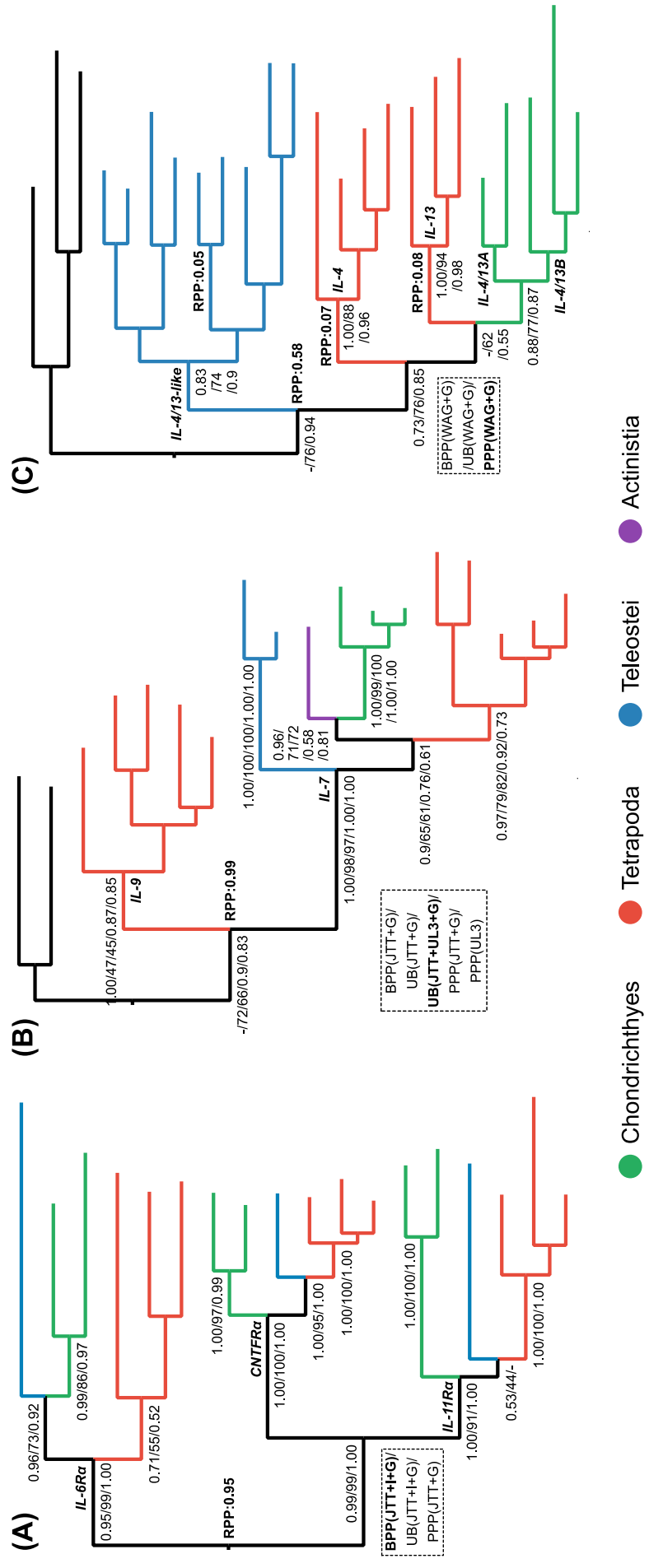
11. Nguyen LT, Schmidt HA, Von Haeseler A, Minh BQ. IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol. Biol. Evol. 2015;32:268–74.

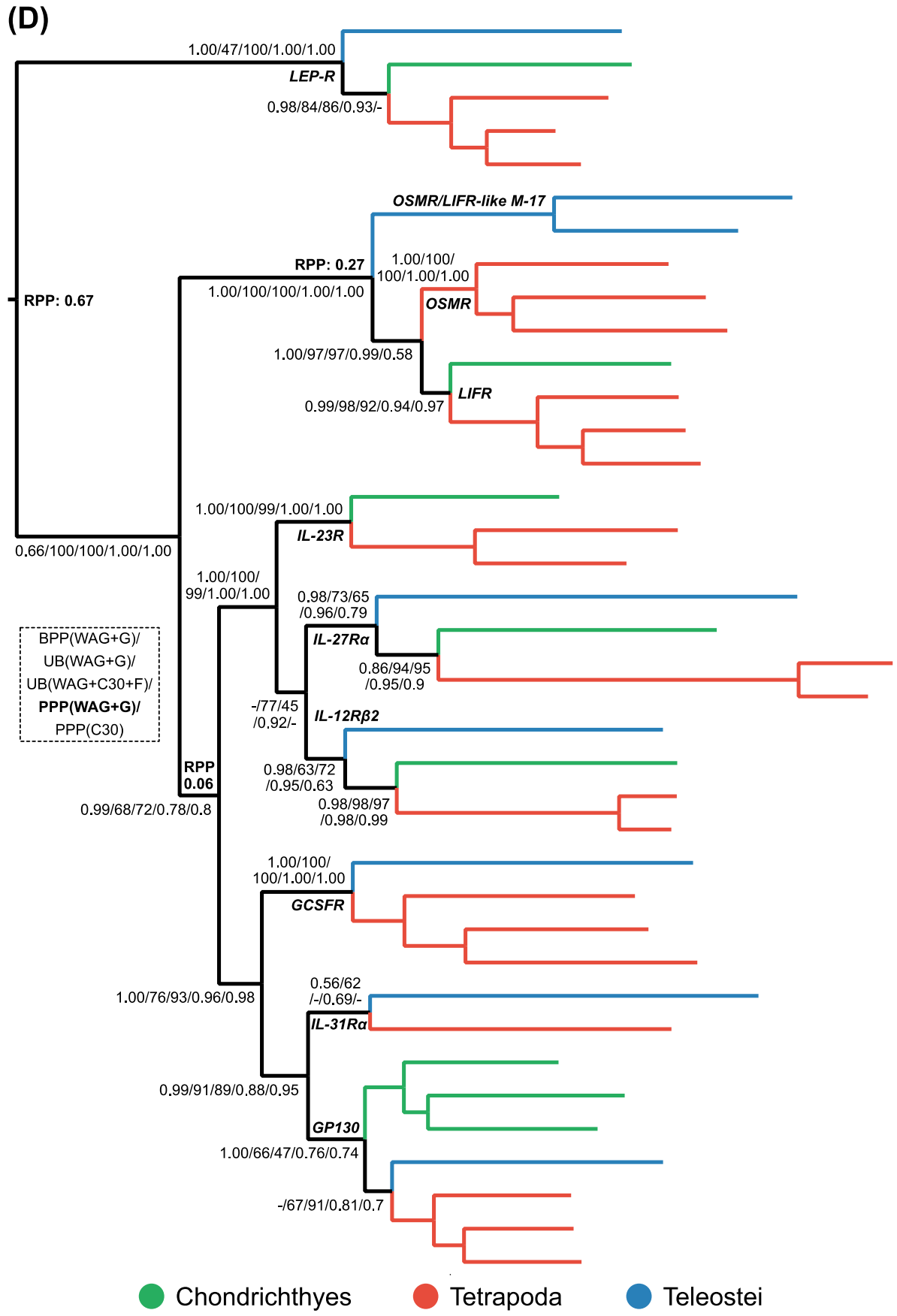
12. Lartillot N, Lepage T, Blanquart S. PhyloBayes 3: A Bayesian software package for phylogenetic reconstruction and molecular dating. Bioinformatics. 2009;25:2286–8.

13. Lartillot N, Philippe H. A Bayesian mixture model for across-site heterogeneities in the amino-acid replacement process. Mol. Biol. Evol. 2004;21:1095–109.

14. Lartillot N, Brinkmann H, Philippe H. Suppression of long-branch attraction artefacts in the animal phylogeny using a site-heterogeneous model. BMC Evol. Biol. 2007;7 Suppl 1:S4.

15. Dijkstra JM. T H 2 and T reg candidate genes in elephant shark. Nature. Nature Publishing Group; 2014;511:E7–9.


(A)
IL-6Rα
0.96/73/0.92
0.99/86/0.97
0.95/99/1.00
0.71/55/0.52
RPP:0.95
BPP(JTT+I+G)/
UB(JTT+I+G)/
PPP(JTT+G)
0.99/99/1.00
CNTFRα
1.00/97/0.99
1.00/100/1.00
1.00/95/1.00
1.00/100/1.00
IL-11Rα
1.00/100/1.00
1.00/91/1.00
0.53/44/-
1.00/100/1.00
(B)
1.00/47/45/0.87/0.85
IL-9
RPP:0.99
-/72/66/0.9/0.83
1.00/100/100/1.00/1.00
0.96/
71/72
/0.58
/0.81
IL-7
1.00/98/97/1.00/1.00
1.00/99/100
/1.00/1.00
BPP(JTT+G)/
UB(JTT+G)/
UB(JTT+UL3+G)/
PPP(JTT+G)/
PPP(UL3)
0.9/65/61/0.76/0.61
0.97/79/82/0.92/0.73
(C)
IL-4/13-like
0.83
/74
/0.9
RPP:0.05
RPP:0.58
-/76/0.94
RPP:0.07
IL-4
1.00/88
/0.96
RPP:0.08
IL-13
1.00/94
/0.98
0.73/76/0.85
BPP(WAG+G)/
/UB(WAG+G)/
PPP(WAG+G)
IL-4/13A
-/62
/0.55
0.88/77/0.87
IL-4/13B
Chondrichthyes
Tetrapoda
Teleostei
Actinistia


(D)
1.00/47/100/1.00/1.00
LEP-R
0.98/84/86/0.93/-
OSMR/LIFR-like M-17
RPP: 0.27
1.00/100/
100/1.00/1.00
1.00/100/100/1.00/1.00
OSMR
RPP: 0.67
1.00/97/97/0.99/0.58
LIFR
0.99/98/92/0.94/0.97
1.00/100/99/1.00/1.00
IL-23R
0.66/100/100/1.00/1.00
1.00/100/
99/1.00/1.00
0.98/73/65
/0.96/0.79
IL-27Rα
BPP(WAG+G)/
UB(WAG+G)/
UB(WAG+C30+F)/
PPP(WAG+G)/
PPP(C30)
0.86/94/95
/0.95/0.9
-/77/45
/0.92/-
IL-12Rβ2
RPP
0.06
0.98/63/72
/0.95/0.63
0.99/68/72/0.78/0.8
0.98/98/97
/0.98/0.99
1.00/100/
100/1.00/1.00
GCSFR
0.56/62
/-/0.69/-
1.00/76/93/0.96/0.98
IL-31Rα
0.99/91/89/0.88/0.95
GP130
1.00/66/47/0.76/0.74
-/67/91/0.81/0.7
Chondrichthyes
Tetrapoda
Teleostei

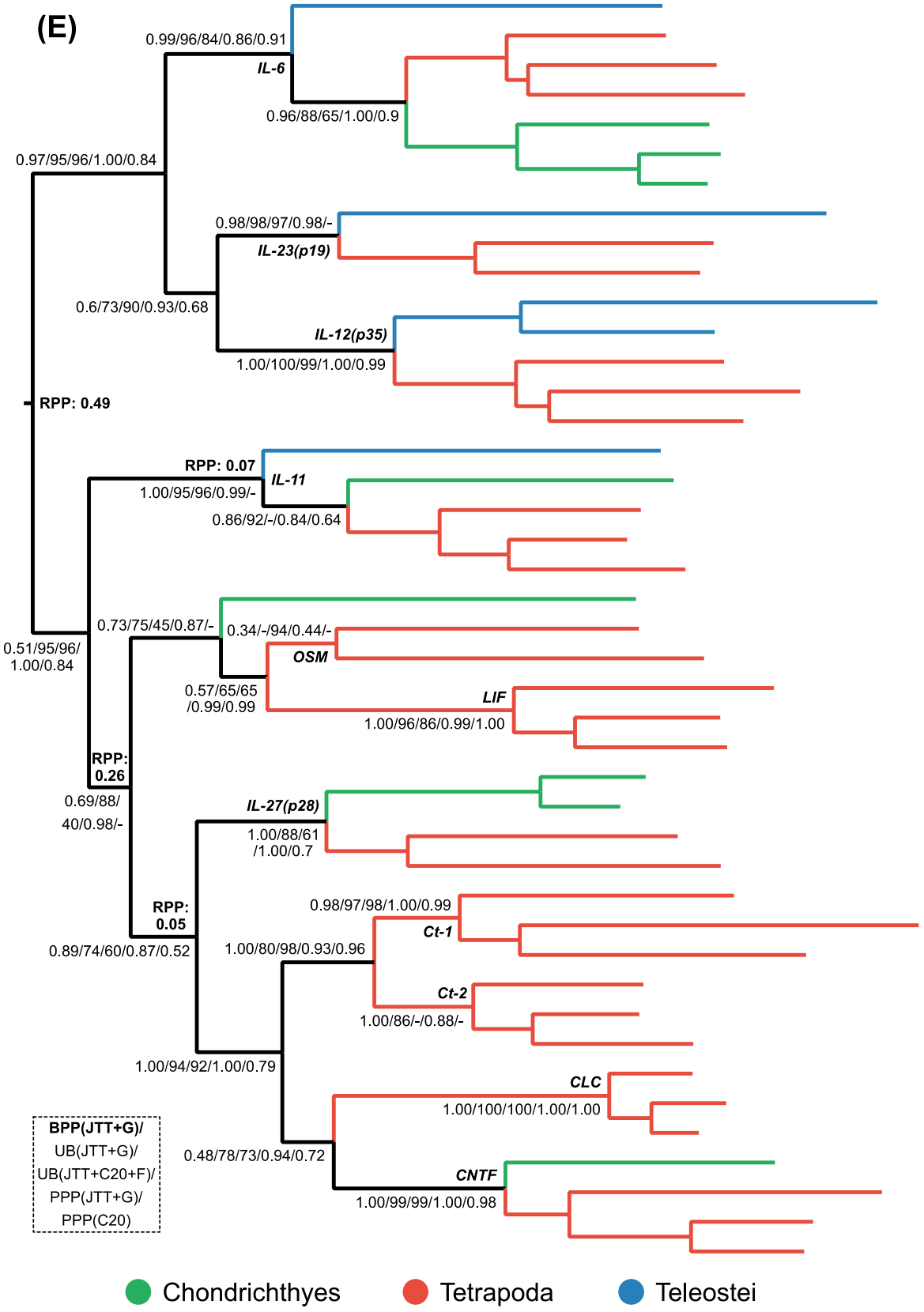

(E)
0.99/96/84/0.86/0.91
IL-6
0.96/88/65/1.00/0.9
0.97/95/96/1.00/0.84
0.98/98/97/0.98/-
IL-23(p19)
0.6/73/90/0.93/0.68
IL-12(p35)
1.00/100/99/1.00/0.99
RPP: 0.49
RPP: 0.07
IL-11
1.00/95/96/0.99/-
0.86/92/-/0.84/0.64
0.73/75/45/0.87/-
0.34/-/94/0.44/-
OSM
0.51/95/96/
1.00/0.84
0.57/65/65
/0.99/0.99
LIF
1.00/96/86/0.99/1.00
RPP:
0.26
0.69/88/
40/0.98/-
IL-27(p28)
1.00/88/61
/1.00/0.7
RPP:
0.05
0.98/97/98/1.00/0.99
Ct-1
0.89/74/60/0.87/0.52
1.00/80/98/0.93/0.96
Ct-2
1.00/86/-/0.88/-
1.00/94/92/1.00/0.79
CLC
1.00/100/100/1.00/1.00
BPP(JTT+G)/
UB(JTT+G)/
UB(JTT+C20+F)/
PPP(JTT+G)/
PPP(C20)
0.48/78/73/0.94/0.72
CNTF
1.00/99/99/1.00/0.98
Chondrichthyes
Tetrapoda
Teleostei

FIGURE 1. Phylogenetic analyses reveal orthologues of 'missing' genes in cartilaginous fishes. Trees represent the **(A)** IL-6Rα family, **(B)** IL-7/9 family, **(C)** IL-4/IL-13 family, **(D)** Class-1 Group-2 cytokine receptors, and **(E)** IL-6 superfamily. Branches are coloured according to the taxonomic key in the figure. Statistical support is shown for key nodes as per boxes accompanying each tree, wherein the analysis shown in bold is the topology shown. Root Posterior Probability <0.5 not shown. Posterior probabilities from Phylobayes analyses using poorly fitting mixture models are shown for completeness where applicable (fifth support value).

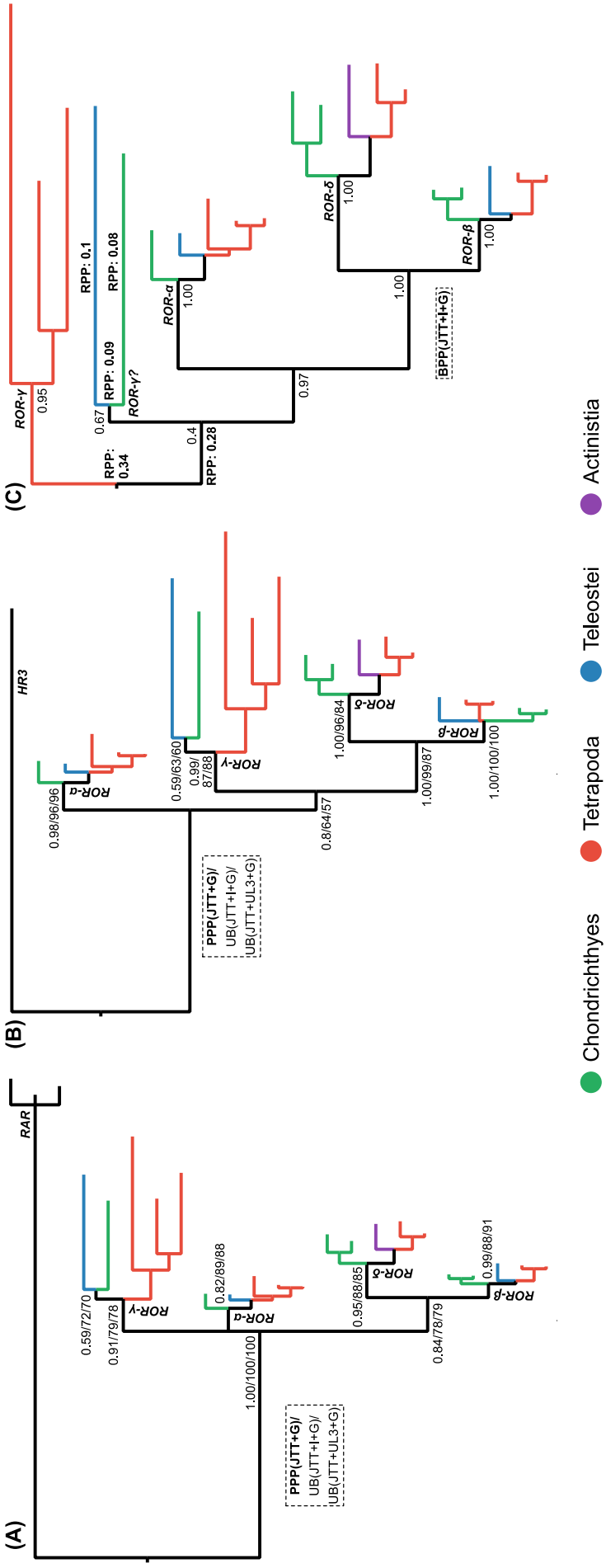

(A)
RAR
ROR-γ
ROR-α
ROR-δ
ROR-β
0.59/72/70
0.91/79/78
0.82/89/88
1.00/100/100
0.95/88/85
0.84/78/79
0.99/88/91
PPP(JTT+G)/
UB(JTT+I+G)/
UB(JTT+UL3+G)
(B)
HR3
ROR-α
ROR-γ
ROR-δ
ROR-β
0.98/96/96
0.59/63/60
0.99/
87/88
0.8/64/57
1.00/96/84
1.00/99/87
1.00/100/100
PPP(JTT+G)/
UB(JTT+I+G)/
UB(JTT+UL3+G)
(C)
ROR-γ
RPP: 0.1
RPP: 0.08
RPP: 0.09
ROR-γ?
RPP:
0.34
RPP: 0.28
ROR-α
ROR-δ
ROR-β
BPP(JTT+I+G)
0.95
0.67
0.4
0.97
1.00
1.00
1.00
Chondrichthyes
Tetrapoda
Teleostei
Actinistia

FIGURE 2. Phylogenetic analyses of the vertebrate ROR family shows that ROR-γ existed in the jawed vertebrate ancestor, and reveals a new vertebrate ROR-β paralog not found in mammals (which we dub ROR-δ). Alternative rooting strategies, using (A) RARs as outgroup, (B) HR3 as outgroup, or (C) a relaxed clock model, show that the root of the ROR phylogeny cannot be congruently placed. Other details as per Fig. 1.

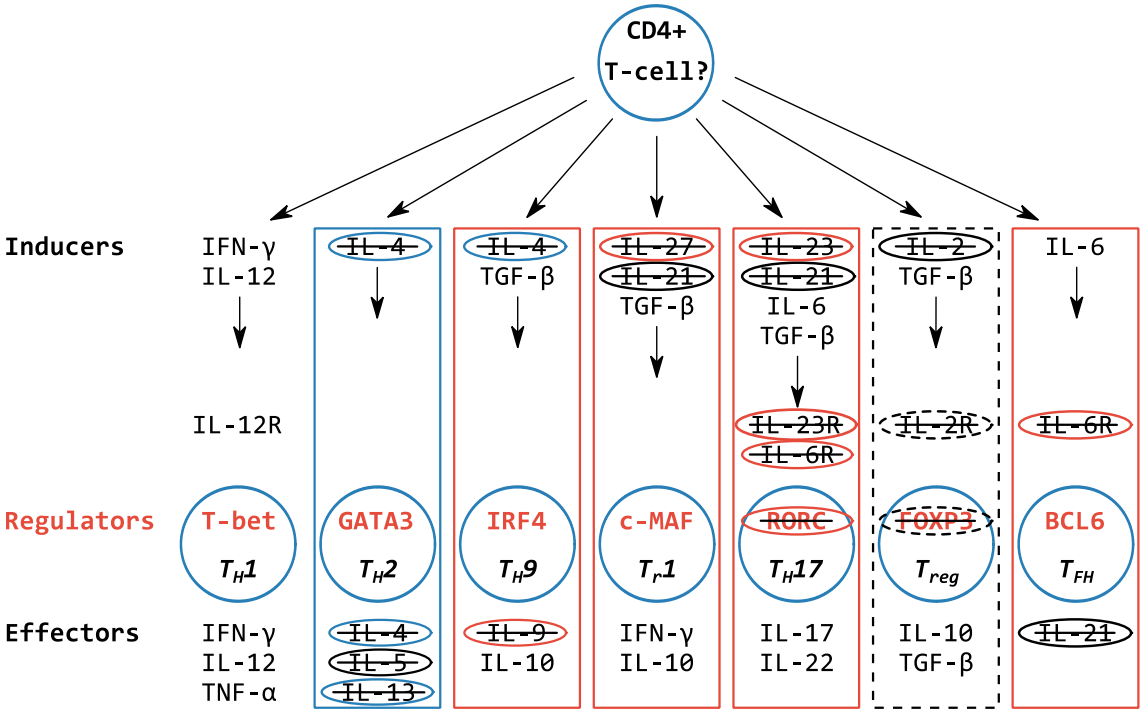


FIGURE 3. Revised summary of the presence/absence of major CD4+ T-cell lineages and associated genes in the jawed vertebrate ancestor. The figure and gene selection are based on Fig. 5 from Venkatesh et al. [2]. Boxed lineages or encircled genes represent genes/lineages that were predicted to have emerged in the ancestor of bony vertebrates due to absence from the elephant shark genome [2], but for which evidence has been presented here (red circles/boxes), by Dijkstra [15] (black circles/boxes), or both (blue circles/boxes). Dotted edges indicate uncertainty of presence (e.g. IL-2R) may not have duplicated until the evolution of bony vertebrates), or function (e.g. FOXP3).

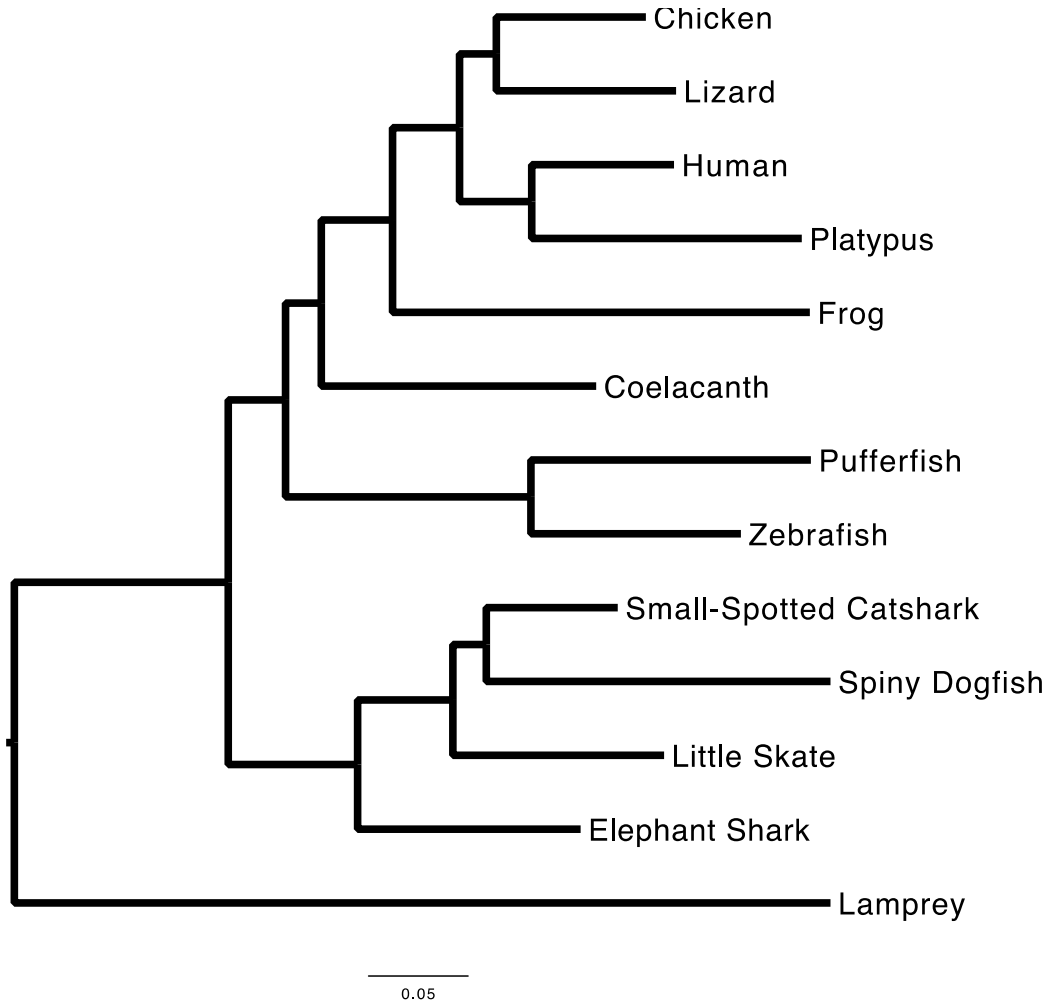


FIGURE 4. Consensus topology from IQ-tree analysis of 143 gene vertebrate phylogenomic dataset. IQ-tree (partitioned by gene) and Phylobayes (CAT) support values are maximal for all nodes.