

The RAF oncogenes of vertebrates are ohnologs that derive from the two rounds of whole genome duplications early in vertebrate evolution

Amanda Coward Black<sup>1</sup>, Mary Clay Bailey<sup>1</sup>, Marisa Ruane-Foster<sup>2</sup>, Juan C. Opazo<sup>3,4</sup>, Federico G. Hoffmann<sup>1,5</sup>

<sup>1</sup> Mississippi State University, Department of Biochemistry, Molecular Biology, Entomology, and Plant Pathology

<sup>2</sup> Clarkson University, Department of Chemistry and Biomolecular Science

<sup>3</sup> Instituto de Ciencias Ambientales y Evolutivas, Facultad de Ciencias, Universidad Austral de Chile, Valdivia, Chile

<sup>4</sup> Millennium Nucleus of Ion Channels-Associated Diseases (MiNICAD)

<sup>5</sup> Mississippi State University, Institute for Genomics, Biocomputing & Biotechnology

**Corresponding Author:**

Federico G. Hoffmann  
Associate Professor  
Department of Biochemistry, Molecular Biology,  
Entomology, and Plant Pathology  
Box 9655  
Mississippi State University  
Mississippi State, MS 39762

Lab website: <http://globins.bch.msstate.edu>

E-mail: [federico.g.hoffmann@gmail.com](mailto:federico.g.hoffmann@gmail.com), [fgh19@msstate.edu](mailto:fgh19@msstate.edu)

ORCID: 0000-0001-5056-4929

Department phone: 662-325-2640

Fax: 662-325-8664

## **ABSTRACT**

The Rapidly Accelerated Fibrosarcoma (RAF) kinases are part of large group of serine/threonine-specific protein kinases that play important roles in cell differentiation and organism development. Animal RAF kinases are key connectors in the signaling cascade that links the small G protein RAS and the Mitogen-activated protein kinase phosphorylation pathway. Mutations in the RAF genes have been linked to a number of cancers including melanoma, lung cancer, colorectal cancer, thyroid cancer, and ovarian cancer. Most animals possess a single RAF gene, but vertebrates have three RAF genes in their genomes, named as A-, B-, and C-RAF, the latter also known as RAF-1. The emergence of the multiple copies of vertebrate RAFs is not well resolved, on the one hand, because of sequence and functional similarities, some authors speculate that vertebrate B-Raf is most closely related to the RAF genes of the fruit fly and *Caenorhabditis elegans*, whereas a competing hypothesis is that the A-,B- and C-RAF paralogs emerged from the two rounds of whole genome duplications that occurred early in the evolution of vertebrates. We applied a comparative approach grounded in synteny and phylogenetic analyses to evaluate these two scenarios. Our results are consistent with the hypothesis that the RAF genes of vertebrates are paralogs generated by whole genome duplications. Thus, the functional similarities between vertebrate B-RAF and invertebrate RAF probably reflect the retention of ancestral characters. Interestingly, data from the literature indicate that B-RAF is the paralog that associated with cancer more strongly, and also yields the most severe phenotypes when knocked out.

## **INTRODUCTION**

The Rapidly Accelerated Fibrosarcoma (RAF) kinases are part of large group of serine/threonine-specific protein kinases with regulatory and signaling roles (Cseh et al. 2014). In animals, RAF kinases play important roles in cell differentiation and organism development (Desideri et al. 2015). Animal RAF kinases are key connectors in the signaling cascade that links the small G protein RAS and the Mitogen-activated protein kinase (MAPK) phosphorylation pathway (Cseh et al. 2014; Desideri et al. 2015). This pathway relays extracellular signals from receptors on the cell membrane to the cytosol to induce a variety of cellular responses, including cell proliferation, migration, differentiation, and survival (Matallanas et al. 2011; Desideri et al. 2015). Basically, RAS binds RAF to activate it, RAF then phosphorylates MEK, which then phosphorylates ERK, which can then phosphorylate a large number of downstream targets. The genes encoding for proteins in the MAPK cascade are a common source of oncogenic mutations, and the RAF genes are no exception: In fact, the first RAF gene was described as an oncogene transduced by a murine sarcoma retrovirus (Rapp et al. 1983). Later, mutations in the RAF genes have been linked to a number of cancers including melanoma, lung cancer, colorectal cancer, thyroid cancer, and ovarian cancer (Leicht et al. 2007; Desideri et al. 2015).

While most animals possess a single RAF gene, vertebrates have three RAF genes in their genomes, named as A-, B-, and C-RAF, the latter also known as RAF-1. The three vertebrate RAFs are involved in the MAPK pathway, and even though they exhibit some functional overlap, experimental evidence indicates they have independent functions and non-MAPK pathway roles as well (Wojnowski et al. 2000; Cseh et al. 2014). These three RAF isoforms can assemble into homo- or heterodimers, which expand the signaling output of the pathway in vertebrates relative to invertebrates. The emergence of the multiple copies of vertebrate RAFs is not well resolved from an evolutionary standpoint, with competing explanations about their origin. In some cases, vertebrate RAFs have been considered as putative ohnologs derived from the two rounds of whole genome duplications early in vertebrate evolution (Singh et al. 2012). On the other hand, because of sequence and functional similarities between vertebrate B-Raf and the RAF genes of the fruit fly *Drosophila melanogaster* (known as D-Raf) and the nematode

*Caenorhabditis elegans* (known as lin-45), some authors suggest that B-RAF is most closely related to invertebrate RAFs than to the A- and C- paralogs of vertebrates (Matallanas et al. 2011; Desideri et al. 2015). These two hypotheses make alternative phylogenetic predictions - with vertebrate RAF paralogs expected to fall in a monophyletic clade in the first hypothesis and vertebrate B-RAF expected to group with the D-RAF gene of fruit fly and lin-45 gene of *C. elegans* in the second one - but the corresponding studies did not use phylogenetic inference to test their assertions and based their inferences on sequence similarity. Accordingly, the goals of our study are to characterize the RAF repertoire in vertebrates by integrating phylogenetic and synteny analyses to contrast competing hypotheses about the duplicative history of these genes and to incorporate an evolutionary perspective into the study of cancer-associated mutations in these genes.

## **MATERIALS AND METHODS**

### Sequence data

We used bioinformatic protocols to collect the repertoire of RAF-like genes in a representative set of invertebrates and vertebrates to explore copy number variation in this gene family and resolve the duplicative history of vertebrate RAFs. In the case of vertebrates, our sampling includes cyclostomes (Arctic lamprey, *Lethenteron camtschaticum*; sea lamprey, *Petromyzon marinus*; and inshore hagfish, *Eptatretus burgeri*); cartilaginous fishes (elephant shark, *Callorhinchus milii*; and whale shark, *Rhincodon typus*); and a representative sample of bony vertebrates. Bony vertebrates (Euteleostomes) were represented by one holostean fish (spotted gar, *Lepisosteus oculatus*), three teleost fish (zebrafish, *Danio rerio*; medaka, *Oryzias latipes*; Mexican cave fish, *Astyanax mexicanus*), one lobe-finned fish (West Indian Ocean coelacanth, *Latimeria chalumnae*), one amphibian (western clawed frog, *Silurana tropicalis*), four sauropsids (chicken, *Gallus gallus*; Chinese softshell turtle, *Pelodiscus sinensis*; American alligator, *Alligator mississippiensis*; and green anole lizard, *Anolis carolinensis*), and two mammals (human, *Homo sapiens*; mouse, *Mus musculus*). The search protocol combined information from the Ensembl comparative genomics assignments of orthology (Zerbino et al.

2018) with results of BLAST searches against the genomes of the American alligator, Arctic lamprey, elephant shark and whale shark.

In the case of invertebrates, we verified that the RAF genes of invertebrates are single-copy genes by screening the 70 species available in release 40 of the Ensembl Metazoa database (available at <http://jul2018-metazoa.ensembl.org/index.html>, last accessed on January 2019). In addition, we screened the genomes of the additional invertebrate deuterostomes- acorn worm, sea urchin, starfish and amphioxus- which are available in the NCBI sequence repository but not in Ensembl. We then selected a small number of invertebrate sequences to be included in the phylogenetic analysis corresponding to a sponge (*Amphimedon queenslandica*), a sea anemone (*Nematostella vectensis*), a placozoan (*Trichoplax*), an arthropod (fruit fly), a nematode (*C. elegans*), two sea squirts (*Ciona intestinalis* and *Ciona savignyi*), an amphioxus (*Branchiostoma belcheri*), an acorn worm (*Saccoglossus kowalevskii*), sea urchin (*Strongylocentrotus purpuratus*), and a starfish (*Acanthaster planci*).

### Phylogenetic analyses

We inferred phylogenetic relationships among the RAF genes using the conceptual translation of the coding sequence of a representative set of the genes identified above. Sequences from the animal Kinase Suppressor of Ras (KSR) gene family, which is the animal gene family closest to RAF according to Ensembl Compara, plus a small set of plant RAF kinases were included as outgroup sequences. The full set of sequences used in the analyses are listed in supplementary table 1. We aligned amino acid sequences using Kalign (Lassmann and Sonnhammer 2006), the E-INS-i, L-INS-i, and G-INS-i strategies from MAFFT (Katoh et al. 2009; Katoh and Standley 2013), MUSCLE (Edgar 2004) and T-coffee (Notredame et al. 2000), and compared the resulting alignments using MUMSA (Lassmann and Sonnhammer 2005, 2006). Subsequently, we used the best-scoring alignment for all downstream analyses. Phylogenetic relationships were estimated using maximum likelihood (ML) and Bayesian analyses (BA). ML analyses were run using IQ-Tree ver 1.6.6 (Nguyen et al. 2015) in the implementation available from the IQ-Tree web server (Trifinopoulos et al. 2016) last accessed on September 2018, and support for the nodes

was evaluated with the Shimodaira-Hasegawa approximate likelihood-ratio test (SH-aLRT), the aBayes test (Anisimova et al. 2011) and 1,000 pseudoreplicates of the ultrafast bootstrap procedure (Hoang et al. 2018). The best-fitting model of substitution was selected using the ModelFinder subroutine from IQ-Tree (Kalyaanamoorthy et al. 2017). Bayesian Analyses were performed via the CIPRES Science Gateway (Miller et al. 2015) in MrBayes version 3.2 (Ronquist et al. 2012), under a mixed model of substitution, running four simultaneous chains for  $2 \times 10^7$  generations, sampling trees every 1000 generations, and using default priors. We assessed convergence by measuring the standard deviation of the split frequency among parallel chains. Chains were considered to have converged once the average split frequency was lower than 0.01. We discarded trees collected before the chains reached convergence, and we summarized results with a majority-rule consensus of trees collected after convergence was reached.

### Synteny comparisons

Whole-genome duplications are expected to affect all genes in a given genome. So, synteny analyses can help identify sets of co-duplicated genes that map to similar regions of the genome. To do so, we examined genes found upstream and downstream of each member of the RAF gene family on representative species. Initial ortholog predictions were derived from the EnsemblCompara database (Zerbino et al. 2018) and visualized using the Genomicus platform v92.01, last accessed in August 2018 (Nguyen et al. 2018). In the case of the elephant shark, synteny was visualized on NCBI and inferred by comparing flanking genes to other vertebrates using BLAST tools.

### Organismal phylogeny and divergence dates

In all cases, we assumed that relationships among the different animal lineages follow the arrangement reported by (Dunn et al. 2014), and estimates of divergence times among the lineages were obtained from the TimeTree server (Kumar et al. 2017).

## **RESULTS AND DISCUSSION**

## RAF repertoires

We found limited variation in the RAF repertoires of animals. In the case of invertebrates, our bioinformatic surveys confirm that most invertebrates include a single member of this gene family in their genome. Fifty five species in a total of 70 have a 1-to-1 ortholog of the *Drosophila melanogaster* D-Raf gene, 7 have 1-to-many orthologs, and 8 have no annotated ortholog. A more critical evaluation of the sequences of the 7 species that have putative duplicates indicates that all of these duplicates correspond to either very recent duplications or to assembly artifacts. Pairwise comparisons among the putative RAF homologs within a species are either identical; as in the stalk-eyed fly *Teleopsis dalmanni*, the nematodes *Caenorhabditis brenneri* and *Caenorhabditis remanei*, the rotifer *Adineta vaga*, the cnidarian *Nematostella vectensis*, and the brachiopod *Lingula anatina*; or correspond to two fragments of a single gene, as in the case of the African social velvet spider *Stegodyphus mimosarum*. Bioinformatic searches in other invertebrate genomes in other repositories were consistent with these results and identified a single putative RAF ortholog in most cases, either spanning the gene completely or partially, as in the amphioxus *Branchiostoma floridae*. Because we found RAF orthologs in a poriferan (the sponge *Amphimedon queenslandica*), a ctenophore (the warty comb jelly *Mnemiopsis leidyi*), the placozoan *Trichoplax adherens*, a cnidarian (the sea anemone *Nematostella vectensis*), protostomes, and deuterostomes, we can trace the presence of the animal RAF gene back to the common ancestor of animals, though we could not find candidates in choanoflagellates.

In the case of vertebrates, most species include three RAF-like genes in their genomes, A-, B- and C-Raf, with the exception of the two lampreys, which include one or two paralogs, and Archelosaurs (the group that includes birds, crocodilians, and testudines), where we could find copies of B- and C-RAF, but not of A-RAF. Two of the teleost fish surveyed, zebrafish and Mexican cave fish, possess duplicate copies of C-RAF, and there are 2 separate fragments of A-RAF in the JGI4.2/XenTro3 release of the Western clawed frog (*Silurana tropicalis*) genome, which are located on separate scaffolds and annotated as separate genes. This is probably an assembly artifact, as these two fragments map to non-overlapping portions of the human A-RAF

gene. We included one of the A-RAF homologs of the closely related African clawed frog (*Xenopus laevis*) genome in the analyses as an additional representative of amphibians. The current assembly of the elephant shark genome lacks an A-Raf gene, but a partial match is found in the whale shark genome, indicating the gene was present in the ancestor of cartilaginous fish.

From a sequence standpoint, pairwise distance comparisons within the A-, B-, and C-paralogs of jawed vertebrates reveal that the B-RAF paralog is the most conserved, the C-RAF paralog is strongly conserved as well, and the A-RAF paralog is the most variable (Table 1)

### Evolutionary history of the RAF oncogenes

Our phylogenetic analyses include a representative subset of animal RAFs, plus a small sample of plant RAF-like and animal KSR sequences as outgroup sequences (see Supplementary Table 1 for a full list of sequences used in phylogeny reconstructions). The resulting trees show that animal RAFs form a monophyletic group that is distantly related to plant RAFs or animal KSR sequences (Fig. 1). Relationships between plant and animal RAFs are poorly understood and their functional roles seem to be different: plant RAFs are mostly involved with signaling networks involved in responses to stress, whereas animal RAFs are involved with cell cycle progression, cell differentiation, migration and organismal development (Popescu and Popescu 2011; Lehti-Shiu and Shiu 2012). Sequence similarities between animal and plant RAFs are restricted to the kinase domain. In agreement with these observations, support for the node uniting animal and plant RAFs is very low. Animal RAFs are separated into two distinct groups, one including vertebrate RAFs and the other one including invertebrate RAFs (Fig. 1). This arrangement is not entirely consistent with the expected organismal relationships, as it does not group all deuterostomes in a monophyletic group, however, support for the relevant nodes is weak. We suspect that deep divergences among some invertebrates in our study, potential inconsistencies in the annotations of the gene models, and incomplete coverage for some genes in our study could account for this observation.



In both our ML and BA phylogenies, vertebrate RAFs are placed in a monophyletic group with strong support (93/0.99/96/1.0), and these analyses resolved orthology and duplicative history for jawed vertebrate RAFs well (Fig. 1). Each of the A-, B- and C-RAF genes of jawed vertebrates are recovered in a strongly supported monophyletic group, with the A- and C-RAF paralogs as sister groups, and the B-RAF paralog as the deepest split within vertebrate RAFs (Fig. 1). Relationships among the genes within each of these clades did not deviate significantly from the expected organismal tree, placing cartilaginous fish sequences as sister to all other jawed vertebrates, and grouping fish and tetrapod paralogs in monophyletic groups within each RAF paralog clade (Fig. 1). On the other hand, our analyses could not fully resolve orthology for the RAF sequences from cyclostomes. This is not surprising, as resolving orthology between gnathostome and cyclostome genes using phylogenies has had limited (Qiu et al. 2011; Kuraku 2013; Schwarze et al. 2014; Opazo et al. 2015; Campanini et al. 2015). In the case of B-RAF, our analyses identified one putative copy of this gene in the sea lamprey and two in the hagfish. The tree places the sea lamprey gene as sister to one of the hagfish B-RAF paralogs, a phylogenetic arrangement that would suggest that the last common ancestor of lamprey and hagfish had duplicate copies of B-RAF in its genome. Lampreys and hagfish include an additional RAF gene with affinities for the A/C-RAF lineage of jawed vertebrates, but they have no clear orthologs. The hagfish A/C-RAF gene is placed sister to the clade that includes the RAF A- and C- paralogs of jawed vertebrates, and the A/C-RAF genes of lampreys fall in a clade sister to the group that unites the hagfish A/C-RAF gene and the RAF-A and -C paralogs of jawed vertebrates (Fig. 1).

### RAF synteny

Synteny comparisons are consistent with the results of the pairwise sequence comparisons and phylogenetic analyses. We observed higher synteny conservation in the B- and C-RAF genes of jawed vertebrates relative to the A-RAF paralog (Supplementary Fig. 1). In the case of A-Raf, there are very few genes that are syntenic between gar and human, which diverged approximately 430 million years ago. By contrast, in the case of C-RAF, the 6 genes found downstream are conserved in comparisons all the way from human to elephant shark, which

diverged approximately 470 million years ago, and 4 genes upstream are conserved from anole to elephant shark. In addition, one of the two zebrafish C-RAF duplicates also shows some conservation with the genes upstream from human C-RAF. The B-RAF gene shows an intermediate pattern, with strong conservation in amniotes, as represented by comparisons between anole and human, which diverged approximately 318 million years ago, but more limited in more distant comparisons. Still, there are copies of MRPS33 and TMEM178B immediately upstream of B-RAF in most jawed vertebrates assessed, plus a copy of WEE2 that is located downstream in the elephant shark and upstream in amniotes (Fig. 2). Unfortunately, synteny provides limited information to help resolve orthology between cyclostome and gnathostome RAFs, as the RAF paralogs of the sea lamprey and hagfish have very limited information on flanking genes: the A/C-RAF of hagfish is flanked by orthologs of RAB43 and ISY1 which are flanking C-RAF in jawed vertebrates.

### Origin of vertebrate RAFs

The two competing hypotheses about the emergence of animal RAFs make alternative phylogenetic predictions. If the B-RAF of vertebrates was more closely related to the single copy RAFs of *C. elegans* and fruit fly, these genes would fall in a monophyletic group to the exclusion of the other RAF paralogs of vertebrates, and this is clearly not that case (Fig. 1). Alternatively, if vertebrate RAFs were ohnologs that expanded as a result of the two rounds of whole genome duplications in early vertebrate evolution, they would be monophyletic relative to invertebrate RAFs, in agreement with our phylogenies (Fig. 1). In addition, synteny comparisons reveal that the genomic context of the A-, B-, and C-RAF paralogs is similar in most species of jawed vertebrates analyzed, which would indicate that their location in the genome was established early in vertebrate evolution and has remained relatively conserved (Fig. 2). The human A-, B-, and C-RAF paralogs are located on chromosomal segments X.4, 7.6 and 3.1 respectively, which all map to ancestral Chordate Linkage Group 13 (Putnam et al. 2008). Thus, our results are consistent with the hypothesis that the RAF genes of jawed vertebrates are ohnologs derived from the two rounds of WGD early in vertebrate evolution, and the presence of co-duplicated genes MKRN, CECR5, VGLL4, synapsin, and TIMP gene families linking the different RAF

paralogons is also in line with this inference. Interestingly, the A-, B- and C-RAF genes appear to have reverted to their single-copy state in most teleosts after the whole genome duplication occurred in that lineage, with the exception of the C-RAF gene in zebrafish and Mexican cave fish.

The three RAFs of jawed vertebrates work by phosphorylating MEK, with some MEK-independent pathways involving apoptosis (Desideri et al. 2015). All three show some functional redundancy, as evidenced by RAF knockouts and a comparison of single and double B- and C-RAF mutants (Wojnowski et al. 2000; Galabova-Kovacs et al. 2006). In the case of B-RAF, transgenic mice knockouts died in utero by embryonic day 10-12.5 (Wojnowski et al. 1997). In the case of C-RAF, homozygous knockouts died by embryonic day 12.5, while heterozygous died shortly after birth (Wojnowski et al. 1998). In the case of A-RAF, knockout mice of the C57Bl/6J strain die 7-21 days after birth while knockouts in mice outbred with the 129/OLA strain survive to adulthood with some neurological defects (Pritchard et al. 1996). Thus, the loss of B- or C-RAF appears to have more severe consequences relative to A-RAF, an inference that is in line with the fact that archelosaurs live without a copy of this gene, but that B- and C-RAF are present in all vertebrates. This is also in line with known essential roles of both B- and C-RAF, while most characterized roles of A-RAF overlap with the other two (Cseh et al. 2014; Desideri et al. 2015).

#### Gene copy retention and cancer

Utilizing the OncoKB database (Chakravarty et al. 2017), we identified 138 annotated point mutations and exon deletions across all three RAFs that are cancer-associated mutations or sites examined within the context of carcinogenic properties (Table 2). We found that most disruptive mutations map to the kinase domains of the B-RAF paralog, where 76% of confirmed oncogenic mutations and 61% of likely oncogenic mutations are found. Additionally, of the 138 alterations, all but 20 have an effect that is gain-of-function or likely gain-of-function and only 4 have an effect that is loss-of-function or likely loss-of-function (Supplementary Table 2). Singh et al. (2012) classified putative ohnologs derived from the 2R of WGD of vertebrates as either

‘essential’, defined as genes where loss-of-function mutations have very severe phenotypic consequences, or ‘dangerous’, defined as genes where gain of function mutations lead to constitutive activity and are in turn linked to cancer (Singh et al. 2012). They then proposed a model where the retention of ohnologs was mostly driven by ‘their “dangerousness.”’ Our results align well with this prediction. Further, if we take a more expansive definition of essentiality, where in addition to loss-of-function we also consider the phenotypic consequences of the loss of a given gene, both essentiality and dangerousness would align. B-RAF is the one that exhibits more severe phenotypes when lost, closely followed by C-RAF. Similarly, most cancer mutations in the gene family map to the B-RAF paralog followed by C-RAF, which are the two most conserved paralogs in terms of sequence, synteny, and phyletic distribution.

Integrating results of our phylogenetic, synteny, and sequence conservation analyses with literature searches suggests the RAF paralogs of vertebrates derive from the 2R of WGD early in the evolution of vertebrates. Our results imply that the functional similarities between the single copy RAF gene of invertebrates and the B-RAF paralog of jawed vertebrates derive from the retention of shared ancestral states and do not reflect evolutionary origin. This is probably also true for the A-, B- and C-RAF paralogs of jawed vertebrates, where functional similarities between B- and C- are probably ancestral, and functional similarities between A- and C-RAF reflect derived functional differentiation (Matallanas et al. 2011; Desideri et al. 2015).

Integrating functional data into our phylogenetic analyses would suggest that all three vertebrate RAF paralogs have maintained some of their ancestral roles, with functional divergence among them: B-RAF is the most potent activator of ERK whereas A-RAF is the least potent one (Desideri et al. 2015). However, and in contrast to the single-copy invertebrate counterparts, the different RAF proteins of vertebrates can assemble into heterodimers, expanding the potential outputs of the RAF signaling cascade, and each of them plays a slightly different role in the pathway, which suggest that their functional divergence involves a combination of subfunctionalization and neofunctionalization processes.

## **ACKNOWLEDGEMENTS**

We thank Sorina C. Popescu for help with identifying relevant plant RAF kinases. This work was partially funded by the National Science Foundation through grants EPS-0903787, DBI-1659630, DEB-1354147 and EPSCoR RII Track-2 FEC 1736026 to FGH, Millennium Nucleus of Ion Channels Associated Diseases (MiNICAD), Iniciativa Científica Milenio, Ministry of Economy, Development and Tourism from Chile to JCO and Fondo Nacional de Desarrollo Científico y Tecnológico from Chile (FONDECYT 1160627) to JCO.

## **REFERENCES**

- Anisimova M, Gil M, Dufayard J-F, et al (2011) Survey of Branch Support Methods Demonstrates Accuracy, Power, and Robustness of Fast Likelihood-based Approximation Schemes. *Syst Biol* 60:685–699. doi: 10.1093/sysbio/syr041
- Campanini EB, Vandeweghe MW, Pillai NE, et al (2015) Early Evolution of Vertebrate Mybs: An Integrative Perspective Combining Synteny, Phylogenetic, and Gene Expression Analyses. *Genome Biol Evol* 7:3009–3021. doi: 10.1093/gbe/evv197
- Chakravarty D, Gao J, Phillips S, et al (2017) OncoKB: A Precision Oncology Knowledge Base. *JCO Precis Oncol* 1:1–16. doi: 10.1200/PO.17.00011
- Cseh B, Doma E, Baccarini M (2014) “RAF” neighborhood: protein-protein interaction in the Raf/Mek/Erk pathway. *FEBS Lett* 588:2398–406. doi: 10.1016/j.febslet.2014.06.025
- Desideri E, Cavallo AL, Baccarini M (2015) Alike but Different: RAF Paralogs and Their Signaling Outputs. *Cell* 161:967–970. doi: 10.1016/j.cell.2015.04.045
- Dunn CW, Giribet G, Edgecombe GD, Hejnol A (2014) Animal Phylogeny and Its Evolutionary Implications. *Annu Rev Ecol Evol Syst* 45:371–395. doi: 10.1146/annurev-ecolsys-120213-091627
- Edgar RC (2004) MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 32:1792–1797. doi: 10.1093/nar/gkh340
- Galabova-Kovacs G, Kolbus A, Matzen D, et al (2006) ERK and Beyond: Insights from B-Raf and Raf-1 Conditional Knockouts. *Cell Cycle* 5:1514–1518. doi: 10.4161/cc.5.14.2981
- Hoang DT, Chernomor O, von Haeseler A, et al (2018) UFBoot2: Improving the Ultrafast Bootstrap Approximation. *Mol Biol Evol* 35:518–522. doi: 10.1093/molbev/msx281
- Kalyaanamoorthy S, Minh BQ, Wong TKF, et al (2017) ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods* 14:587–589. doi: 10.1038/nmeth.4285
- Katoh K, Standley DM (2013) MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Mol Biol Evol* 30:772–780. doi: 10.1093/molbev/mst010
- Katoh K, Asimenos G, Toh H (2009) Multiple Alignment of DNA Sequences with MAFFT. *Methods Mol Biol* 537:39–64. doi: 10.1007/978-1-59745-251-9\_3

- Kuraku S (2013) Impact of asymmetric gene repertoire between cyclostomes and gnathostomes. *Semin Cell Dev Biol* 24:119–127. doi: 10.1016/j.semcdb.2012.12.009
- Lehti-Shiu MD, Shiu S-H (2012) Diversity, classification and function of the plant protein kinase superfamily. *Philos Trans R Soc B Biol Sci* 367:2619–2639. doi: 10.1098/rstb.2012.0003
- Leicht DT, Balan V, Kaplun A, et al (2007) Raf kinases: Function, regulation and role in human cancer. *Biochim Biophys Acta - Mol Cell Res* 1773:1196–1212. doi: 10.1016/j.bbamcr.2007.05.001
- Miller MA, Schwartz T, Pickett BE, et al (2015) A RESTful API for Access to Phylogenetic Tools via the CIPRES Science Gateway. *Evol Bioinforma* 11:EBO.S21501. doi: 10.4137/EBO.S21501
- Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ (2015) IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. *Mol Biol Evol* 32:268–274. doi: 10.1093/molbev/msu300
- Nguyen NTT, Vincens P, Roest Crollius H, Louis A (2018) Genomicus 2018: karyotype evolutionary trees and on-the-fly synteny computing. *Nucleic Acids Res* 46:D816–D822. doi: 10.1093/nar/gkx1003
- Notredame C, Higgins DG, Heringa J (2000) T-coffee: a novel method for fast and accurate multiple sequence alignment 1 Edited by J. Thornton. *J Mol Biol* 302:205–217. doi: 10.1006/jmbi.2000.4042
- Opazo JC, Lee AP, Hoffmann FG, et al (2015) Ancient Duplications and Expression Divergence in the Globin Gene Superfamily of Vertebrates: Insights from the Elephant Shark Genome and Transcriptome. *Mol Biol Evol* 32:1684–1694. doi: 10.1093/molbev/msv054
- Popescu G V., Popescu SC (2011) Complexity and Modularity of MAPK Signaling Networks. In: *Handbook of Research on Computational and Systems Biology*. IGI Global, pp 355–368
- Pritchard CA, Bolin L, Slattery R, et al (1996) Post-natal lethality and neurological and gastrointestinal defects in mice with targeted disruption of the A-Raf protein kinase gene. *Curr Biol* 6:614–7
- Putnam NH, Butts T, Ferrier DEK, et al (2008) The amphioxus genome and the evolution of the chordate karyotype. *Nature* 453:1064–1071. doi: 10.1038/nature06967

- Qiu H, Hildebrand F, Kuraku S, Meyer A (2011) Unresolved orthology and peculiar coding sequence properties of lamprey genes: the KCNA gene family as test case. *BMC Genomics* 12:325. doi: 10.1186/1471-2164-12-325
- Rapp UR, Goldsborough MD, Mark GE, et al (1983) Structure and biological activity of v-raf, a unique oncogene transduced by a retrovirus. *Proc Natl Acad Sci U S A* 80:4218–22
- Ronquist F, Teslenko M, van der Mark P, et al (2012) MrBayes 3.2: Efficient Bayesian Phylogenetic Inference and Model Choice Across a Large Model Space. *Syst Biol* 61:539–542. doi: 10.1093/sysbio/sys029
- Schwarze K, Campbell KL, Hankeln T, et al (2014) The Globin Gene Repertoire of Lampreys: Convergent Evolution of Hemoglobin and Myoglobin in Jawed and Jawless Vertebrates. *Mol Biol Evol* 31:2708–2721. doi: 10.1093/molbev/msu216
- Singh PP, Affeldt S, Cascone I, et al (2012) On the Expansion of “Dangerous” Gene Repertoires by Whole-Genome Duplications in Early Vertebrates. *Cell Rep* 2:1387–1398. doi: 10.1016/j.celrep.2012.09.034
- Trifinopoulos J, Nguyen L-T, von Haeseler A, Minh BQ (2016) W-IQ-TREE: a fast online phylogenetic tool for maximum likelihood analysis. *Nucleic Acids Res* 44:W232–W235. doi: 10.1093/nar/gkw256
- Wojnowski L, Stancato LF, Lerner AC, et al (2000) Overlapping and specific functions of Braf and Craf-1 proto-oncogenes during mouse embryogenesis. *Mech Dev* 91:97–104. doi: 10.1016/S0925-4773(99)00276-2
- Wojnowski L, Stancato LF, Zimmer AM, et al (1998) Craf-1 protein kinase is essential for mouse development. *Mech Dev* 76:141–149. doi: 10.1016/S0925-4773(98)00111-7
- Wojnowski L, Zimmer AM, Beck TW, et al (1997) Endothelial apoptosis in Braf-deficient mice. *Nat Genet* 16:293–297. doi: 10.1038/ng0797-293
- Zerbino DR, Achuthan P, Akanni W, et al (2018) Ensembl 2018. *Nucleic Acids Res* 46:D754–D761. doi: 10.1093/nar/gkx1098



## Figure Legends

Figure 1. Maximum-likelihood phylogram depicting evolutionary relationships among animal RAFs, with plant RAFs and animal KSRs as outgroup sequences. Support for the relevant nodes is shown next to the relevant nodes. Numbers above the nodes correspond to support from the Shimodaira-Hasegawa approximate likelihood-ratio test/ aBayes/ maximum likelihood ultrafast bootstrap support, and number below the nodes correspond to Bayesian posterior probability support values. The coloring of vertebrate RAFs reflects their orthology: B-RAFs in red, jawed vertebrates C-RAFs in green, and A-RAFs in blue, and the A/C RAFs of jawless vertebrates in light blue. Note that the A/C RAFs of lampreys and hagfish are not monophyletic and might represent paralogous genes.

Figure 2. Patterns of conserved synteny in genomic regions that harbor paralogous RAF genes in representative vertebrate taxa. Genes are color coded following orthology. Genes that co-duplicated along with the RAF paralogs are in bold. Diagonal lines connecting genes indicate a change of position.

Table 1. Pairwise comparisons between the A-,B-, and C-Raf paralogs of human and spotted gar, and those of human and elephant or whale shark.

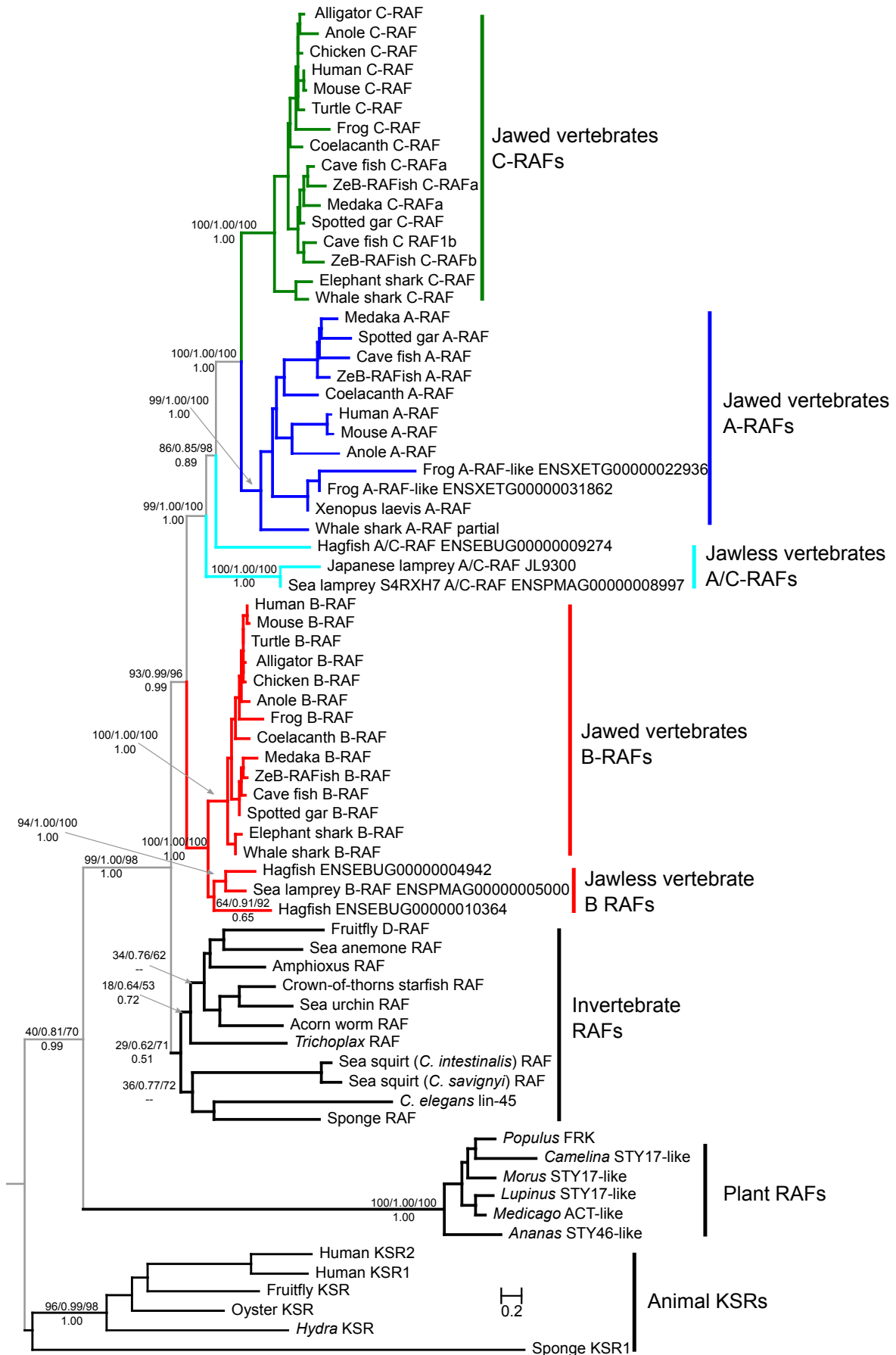
Table 2. Summary of OncoKB annotated mutations in RAF paralogs.

Supplementary Figure 1. Expanded synteny.

Supplementary Table 1. Sequence IDs.

Supplementary Table 2. Cancer mutations.

Figure 1



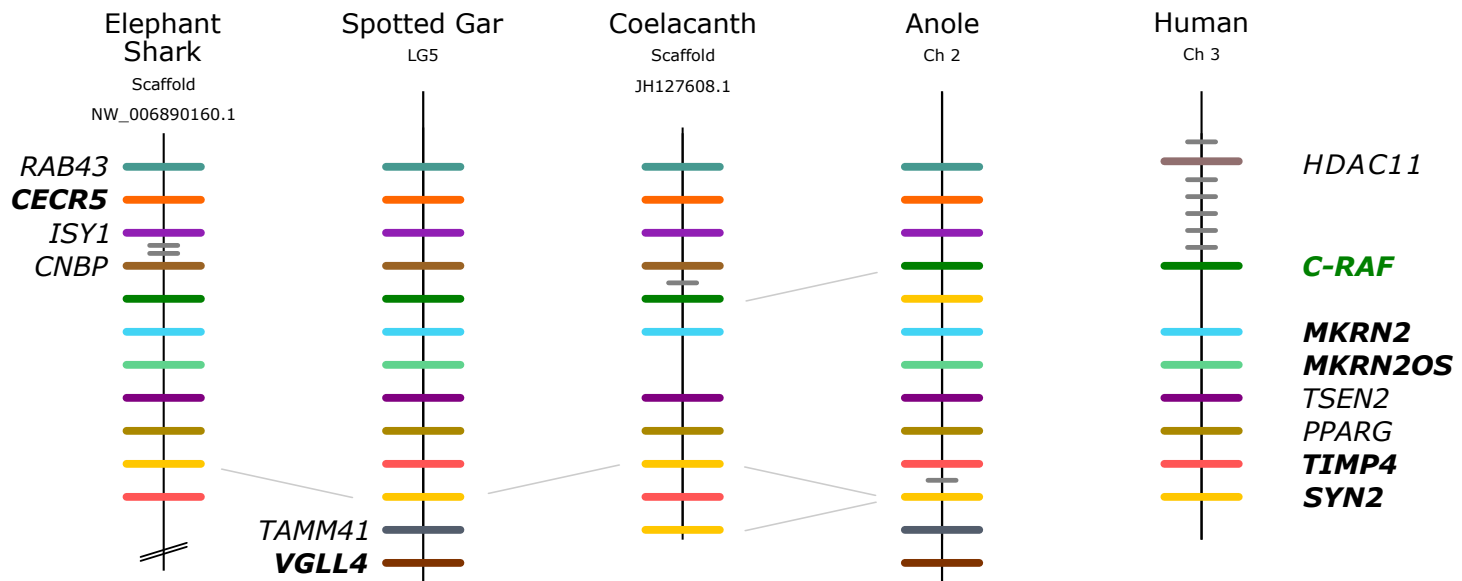
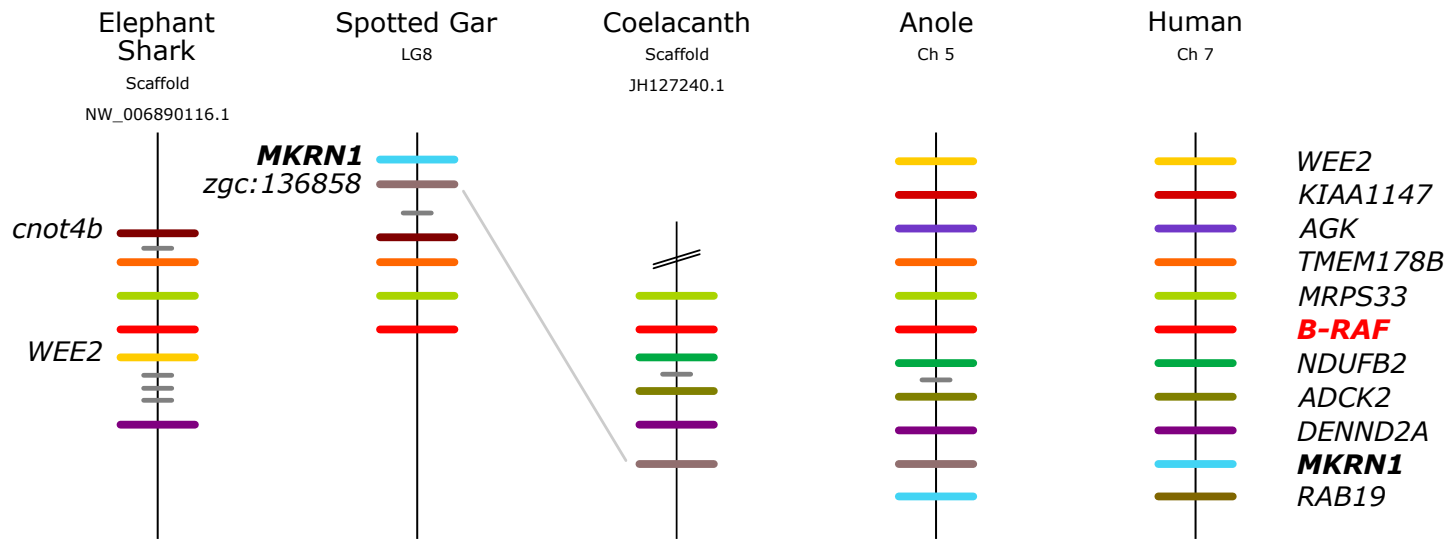
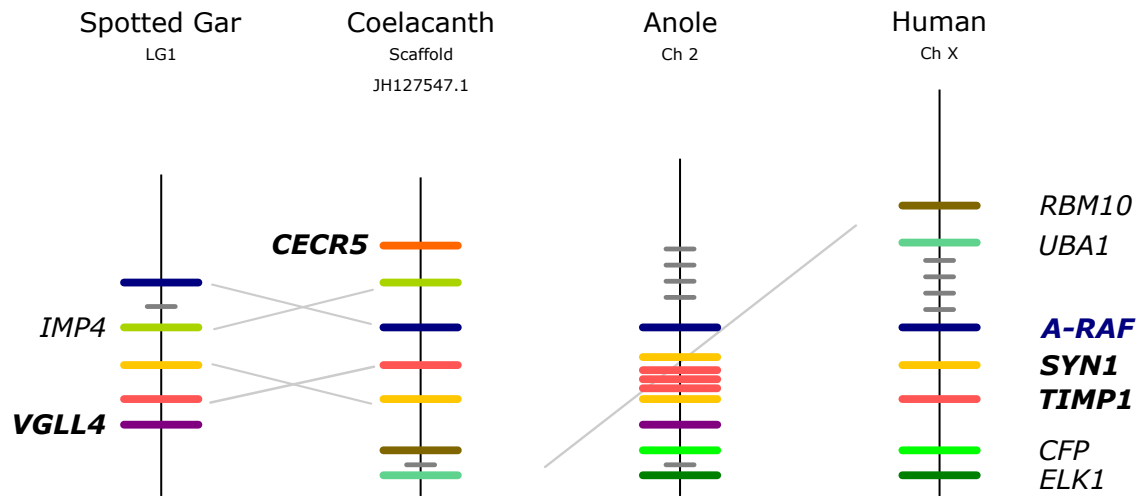


Table 1. Pairwise comparisons between the A-,B-, and C-Raf paralogs of human and spotted gar, and those of human and elephant or whale shark.

<b>Paralog</b>	<b>Comparison</b> (Protein length in amino acids)	<b>Score</b>	<b>Identities</b>	<b>Positives</b>	<b>Gaps</b>	<b>Pairwise distance</b> <b>(p-distance)</b>	<b>Pairwise distance</b> <b>(# of differences)</b>
A-Raf	Human (609) vs Spotted gar (627)	722	389/618	449/618	46/618	36.2%	217
	Human (609) vs Whale shark (358, partial)	320	179/329	222/329	28/329	42.7%	129
B-Raf	Human (766) vs Spotted gar (820)	1332	670/821	695/821	62/821	12.4%	94
	Human (766) vs Elephant shark (752)	1289	624/744	667/744	12/744	16.0%	119
	Human (766) vs Whale shark (266, partial)	530	248/257	254/257	0/257	4.6%	12
C-Raf	Human (668) vs Spotted gar (636)	1127	544/668	590/668	32/668	14.8%	94
	Human (668) vs Elephant shark (696)	957	467/657	527/657	26/657	26.7%	172
	Human (668) vs Whale shark (648)	965	472/666	536/666	27/666	26.4%	170

Table 2. Summary of OncoKB annotated mutations in RAF paralogs.

Paralog	Domain	Oncogenic Classification				Subdomain Total	Paralog total
		Inconclusive	Likely Neutral	Likely Oncogenic	Oncogenic		
A-RAF	RBD	0	0	0	0	0	9
	C1_1	0	0	0	0	0	
	Pkinase_Tyr	0	0	1	2	3	
	Between Domains	0	0	3	3	6	
B-RAF	RBD	1	0	0	1	2	105
	C1_1	0	0	5	0	5	
	Pkinase_Tyr	2	10	41	39	92	
	Between Domains	0	2	3	1	6	
C-RAF	RBD	0	1	1	0	2	24
	C1_1	0	0	0	0	0	
	Pkinase_Tyr	0	2	7	2	11	
	Between Domains	1	1	6	3	11	
Classification Totals		4	16	67	51	138	