

**Comparative genomics within and across Bilaterians
illuminates the evolutionary history of ALK and LTK proto-oncogene
origination and diversification**

Alex Dornburg^{a, 1,2}, Zheng Wang^{b,c, 1}, Junrui Wang^{b, d}, Elizabeth S. Mo^e, Francesc Lopez-Giraldez^f, and
Jeffrey P. Townsend^{b, c, g}

^a Department of Bioinformatics and Genomics, University of North Carolina Charlotte, Charlotte, North Carolina 28223.

^b Department of Ecology and Evolutionary Biology, Yale University, New Haven, CT 06520, USA.

^c Department of Biostatistics, Yale School of Public Health, New Haven, CT 06520, USA.

^d Crop Sciences, Chinese Academy of Agricultural Sciences, Beijing, China.

^e Yale Combined Program in the Biological and Biomedical Sciences, Yale School of Medicine, Yale University, New Haven, CT 06516

^f Yale Center for Genome Analysis (YCGA), Department of Genetics, Yale University, West Haven, CT 06516.

^g Program in Microbiology, Yale University, New Haven, CT 06510, USA.

¹ A.D. and Z.W. contributed equally to this work.

² To whom correspondence may be addressed. Email: dornburgalex@gmail.com

The authors declare no potential conflicts of interest.

Abstract

Comparative genomic analyses have enormous potential for identifying key genes central to human health phenotypes, including those that promote cancers. In particular, the successful development of novel therapeutics using model species requires phylogenetic analyses to determine molecular homology. Accordingly, we investigate the evolutionary histories of anaplastic lymphoma kinase (ALK)—which can underlie tumorigenesis in neuroblastoma, non-small cell lung cancer, and anaplastic large-cell lymphoma—its close relative leukocyte tyrosine kinase (LTK) and their candidate ligands. Homology of ligands identified in model organisms to those functioning in humans remains unclear. Therefore, we searched for homologs of the human genes across metazoan genomes, finding that the candidate ligands Jeb and Hen-1 were restricted to non-vertebrate species. In contrast, the ligand AUG was only identified in vertebrates. We found two ALK-like and four AUG-like protein-coding genes in lamprey. Of these six genes, only one ALK-like and two AUG-like genes exhibited early embryonic expression that parallels model mammal systems. Two copies of AUG are present in nearly all jawed vertebrates. Our phylogenetic analysis strongly supports the presence of previously unrecognized functional convergences of ALK and LTK between actinopterygians and sarcopterygians—despite contemporaneous, highly conserved synteny of ALK and LTK. These findings provide critical guidance regarding the propriety of fish and mammal models with regard to model-organism-based investigation of these medically important genes. In sum, our results provide the phylogenetic context necessary for effective investigations of the functional roles and biology of these critically important receptors.

Keywords: Cancer evolution, vertebrates, genome duplication, phylogenomics, functional divergence, gene evolution

Significance

Model organisms have the potential to provide vital research breakthroughs that revolutionize human biology and medicine. However, inferences from model organisms often fail to translate to humans because of evolved differences in gene function and interaction. Such evolved differences currently stymie translational research on the cancer driver gene *ALK*, its sister *LTK*, and the genes they interact with. Our comparative analysis revealed their evolutionary history, demonstrating that *LTK* and *ALK* have switched and subsequently duplicated gene partners between invertebrates and vertebrates, and that these genes have exchanged functions between humans and fish model species. These results illuminate the model organisms with which future research will effectively translate to human cancer biology—specifically for guiding the development of novel therapeutics.

Introduction

A diversity of non-human model species distributed across the Tree of Life have been essential to investigations of the biology of key genes responsible for trait evolution and human health phenotypes. Correspondingly, comparative investigations of the genes that prevent or promote the origin and spread of cancer have been fruitful for understanding human cancer biology ((Kocher & Piwnica-Worms 2013; O'Hagan et al. 2005; Stoletov & Klemke 2008; Rowell et al. 2011; Paoloni & Khanna 2008). However, the challenge to such investigations is that gene function and regulation evolves. Underlying molecular divergences can become obscured by convergences in phenotype that can confound our ability to infer the homology of receptors and their interacting partners between species. Accordingly, accurate homology predictions have the potential to accelerate the development of novel therapeutics. Therefore, estimating the evolutionary history of oncogenes provides a critical reference for translation of fundamental findings in model organisms (Reshetnyak et al. 2015; Gudernova et al. 2017; Mo et al. 2017; Lemke 2015).

This problem of distinguishing homology from convergence has grown particularly acute within anaplastic lymphoma kinase (ALK) and leukocyte tyrosine kinase (LTK), two well-known RTK proto-oncogenes whose roles in oncogenesis and potential as therapeutic targets have been increasingly investigated (Chiarle et al. 2008; Bresler et al. 2011; Fujioka et al. 2006; Marzec et al. 2006; Lin et al. 2017; Soda et al. 2007; Hallberg & Palmer 2013; Wellstein 2012; Roll & Reuther 2012; Holla et al. 2017). In particular, structural conservation of ALK between invertebrate models as diverse as fruit flies, nematodes, and humans, model organisms has been central to illuminating the biology of oncogenic alterations such as mutation type prevalences in cancer tissues (Ogawa et al. 2011), over-expression triggering abnormal activation of ALK (Zhu et al. 2012; De Brouwer et al. 2010), and ligand-dependent mutations (Mosse et al. 2008). However, disentangling potentially complex patterns of divergent and convergent evolution in ALK and LTK requires investigating three core aspects of evolutionary history, each establishing the extent to which model systems provide functional parallels of humans.

The first aspect of evolutionary history with translational relevance regards the homology of ALK and LTK. A gene that by a simple BLAST search appears to be ALK or LTK in one model organism

could in fact be the other with a very divergent regulatory apparatus and functional repertoire. Second, key domains in ALK and LTK have been gained and lost and may have undergone convergent as well as divergent evolution. ALK and LTK have a tripartite structure including an intracellular kinase domain (KD), a single transmembrane domain (TMD), and an extracellular domain (ECD; Inoue & Thomas 2000; Englund et al. 2003; Bilsland et al. 2008; Weiss et al. 2012). Within the ECD, a low-density lipoprotein receptor class A (LDLa) repeat and two protein tyrosine phosphatase Mu (MAM) domains are also conserved, the latter playing a role in RTK homo-dimerization, which leads to rapid activation of the kinase domains across metazoans (Cismasiu et al. 2004). Knowing the history of gain, loss, and sequence evolution of these domains is essential to knowing functional parallels between model systems and humans. Third, three ligands of Alk have been identified: Jelly belly (Jeb) in *Drosophila melanogaster*, Hesitation behavior-1 (Hen-1) in *Caenorhabditis elegans* (Rohrbough & Broadie 2010; Lee et al. 2003; Ishihara et al. 2002), and Augmentor (FAM150 or AUG- α in *Homo sapiens*; Guan et al 2015, Reshetnyak et al. 2015). Understanding the identity and functional interactions of these ligands with their cognate receptors has been argued to be vital to the development of inhibitors and other small-molecule pharmaceuticals (Slavish et al. 2011; Gambacorti-Passerini 2016; Tartari et al. 2011; Grande et al. 2011; Fadeev et al. 2018), yet the homology of these ligands identified in diverse organisms has not been established (Guan et al. 2015).

In this study, we performed an exhaustive search for homologs of vertebrate ALK, LTK, and AUG against genomes of organisms that include all major vertebrate lineages, additional chordates, hemichordates, and protostomes. Across these genomes, we identified genes homologous to those known to encode ALK ligands, providing evidence for the origins of AUG ligands as a vertebrate-specific innovation. We further reconstructed ancestral sequences of vertebrate AUGs and performed phylogenetic analysis of the evolution of vertebrate ALKs, LTKs and AUGs to reveal the history of major events in their evolution, including the gains and losses of genes and the evolution of functional domains. Using the gene phylogenies obtained, we further identified amino acids that likely play essential roles in the functional divergence between gene paralogs. By determining the origins and evolution of the proto-

oncogenic tyrosine kinases ALK and LTK and their ligands across the history of Metazoans, our results provide the necessary foundation for effective, phylogenetically informed investigations of the functional roles of these genes.

Results

The evolutionary history of *ALK*, *LTK*, *JEB*, *HEN-1*, and *AUG* homologs

BLAST searches (including *blastp* and *tblastn*) revealed a diversity of sequences that are potentially homologous to sequences of ALKs and associated ligands from model and non-model organisms across the genomes available at Ensembl (Kersey et al. 2017) and NCBI (NCBI Resource Coordinators 2017; **Table S1**). ALKs identified in nematode and fruit fly genomes exhibited sufficient conservation of sequence for homologous alignment in the GR and kinase domains. One homolog of vertebrate ALK (NCBI XP_032818754) was found in the sea lamprey (Smith et al. 2013, 2018b) and Japanese lamprey genomes (Mehta et al. 2013), that included identifiable MAM2, LDLa, and tyrosine kinase domains. In addition, a sequence with potential ALK similarity that was annotated as LTK was found (NCBI XP_032804141). This annotated LTK protein exhibited 65% sequence similarity with the predicted lamprey ALK protein along with conservation of a glycine-rich tyrosine kinase domain. However, no MAM and LDLa domains were present in the annotated LTK protein. BLASTp searches using this protein yielded hits on ALK and LTK proteins in birds and mammals; with the highest-scoring hit against the human genome being the anaplastic lymphoma kinase (Ki-1) receptor, followed by ALK, then LTK. Consequently, sequence similarity alone was not sufficient to assign this gene to either ALK or LTK. Functional or structural domains of ALK, LTK, and ALK-like proteins were predicted with sequence comparison and by InterPro (Mitchell et al. 2019). We find orthologs of *ALK* are present in almost all vertebrate genomes as well as most protostomes. However, *ALK* appears to have been lost in the early diverging chordate lineage *Ciona*, as well as in hagfish—the

sister lineage to lamprey. In contrast, we detected ALK-like and Jeb-like proteins within the genome of *lancelets*.

In a parallel to our finding of *ALK* and a candidate LTK homolog in the lamprey genomes, multiple AUG-like proteins (XP_032817706, XP_032809021, XP_032826340, XP_032810391) were also found in the sea lamprey and Japanese lamprey genomes. In contrast, no orthologs of *Jeb* or *Hen-1* were found in any vertebrates, with *Hen-1* restricted entirely to the nematodes *C. elegans* and *Loa*. *Jeb* was present in most protostomes: our identification of Jeb-like proteins in *Aplysia* and *Strongylocentrotus* respectively extends the presence of this ligand to mollusca and represents the first identification of this protein in a deuterostome. However, we found no evidence of *Jeb* in other chordate lineages, indicating that *Jeb* may have been lost prior to the diversification of vertebrates (**Fig. 1**). As lampreys are members of the earliest-diverging lineage of living vertebrates, our finding of an *ALK*, a possible *LTK*, and multiple *AUG* genes is suggestive of two potential scenarios. One is that *ALK*, *LTK*, and two copies of *AUG* arose at the dawn of vertebrates. A second possibility is that a duplication of the *ALK-AUG* ligand-receptor gene pair gave rise to *LTK* and a gene for a second *AUG* ligand after the divergence of the lamprey lineage from the vertebrate ancestor and prior to the diversification of jawed vertebrates (Smith et al. 2018b; Dehal & Boore 2005; **Fig. 1**). The latter hypothesis is supported by conserved synteny: genes adjacent to *ALK* and *LTK* in humans are found adjacent to *ALK* and *LTK* within lineages as divergent as *Anolis*, elephant sharks, and gar, with evidence for conserved synteny between lamprey and other jawed vertebrates restricted to genes surrounding *ALK* (**Fig. 2**), thereby suggesting other lamprey genes to be lineage-specific paralogs.

Phylogenetic analysis of vertebrate ALK and LTK provided further strong support for a history in which the duplication of *ALK* gave rise to *LTK* prior to the most recent common ancestor (MRCA) of jawed vertebrates. The origin of *LTK* mapped to the MRCA of chondrichthyans and Osteichthyes (sarcopterygians + actinopterygians; **Fig. 1**). Synteny of some genes neighboring *ALK* and *LTK* was conserved between and within major clades of vertebrates (**Fig. 2**). Moreover, there was strong support for the reciprocal monophyly of jawed-vertebrate ALK and LTK [Bayesian posterior probability

(BPP) = 1.0; **Fig. 3**) and strong support for a clade comprising lamprey ALK & jawed vertebrate ALK + LTK that excluded the annotated lamprey LTK protein (**Fig. S1**). Comparisons of the evolutionary history of ALK and LTK demonstrate divergent rates of molecular evolution as well as divergent patterns of domain acquisitions and losses (**Fig. 3**). For example, mammals exhibit a signature of decelerated evolution of nonsynonymous substitution in *ALK*, a deceleration that contrasts with a significantly faster rate of molecular divergence of mammal *LTKs* (**Fig. 3; Tables S4 and S6**). We found that both ALK (Iss = 1.08, Iss.c = 0.83, $P < 0.01$) and LTK (Iss = 1.21, Iss.c = 0.83, $P < 0.01$) have experienced severe amino-acid substitution saturation (Xia & Lemey 2009; DAMBE; Xia 2018) and exhibit a sharp decline of phylogenetic informativeness (PhyInformR; Dornburg et al. 2016) within mammals (**Fig. S6**). Rates of evolution of ALK and LTK between “fish” (ray-finned fishes, sharks, and Coelacanth) and mammals contrasted: mammal LTK exhibited significant saturation (Iss = 1.07, Iss.c = 0.84, $P < 0.01$) relative to “fish” (Iss = 0.67, Iss.c = 0.84); and ALK for “fish” exhibited significant saturation (Iss = 1.07, Iss.c = 0.84, $P < 0.01$) relative to mammals (Iss = 0.65, Iss.c = 0.84). We found evidence for functional divergence in different sites between ALK and LTK, with six sites (745-Lys, 760-Leu, 767-Lys, 795-Ile, 808-Asn, and 863-Asn positions of human ALK) identified as important to the evolution of differential function of ALK and LTK ($P < 0.05$; **Fig. S3**). These sites were all located between the MAM2 domain and the GR region, with the exception of 863-Asn, located in the GlyR domain. Functional divergence analyses further predicted 11 amino acids in human LTK to be of significant ($P < 0.01$) importance in the differential function of LTK between mammals and non-mammals groups (98-Thr, 120-Leu, 152-Leu, 171-Gly, 200-Gly located before the GlyR domain and 216-Tyr, 226-Glu, 245-Arg, 261-Ala, 262-Pro, 267-Arg located within the GlyR domain), five of which are conserved across ALK and LTK in jawed vertebrates (**Fig. S4**).

The evolution of augmentor (AUG) in vertebrates

Investigation of transcriptomes and sequenced genomes revealed that AUG is an innovation shared by all vertebrates. Using the annotated lamprey AUG sequence and the maximum-likelihood ancestral sequence for the most recent common ancestor of vertebrate AUG and lamprey AUG revealed no potential homologs in searches of any invertebrate deuterostome, protostome, or non Metazoan genomes. Our reconstruction of the evolutionary history of AUG strongly supports its duplication into AUG- α and AUG- β prior to the most recent common ancestor of jawed vertebrates, suggesting that the other three AUG genes in lamprey are lamprey-specific paralogs (**Fig. 1 & S5**). This hypothesis is further supported by the loss of signaling peptides in two of the lamprey AUG genes (XP_032826340, XP_032810391) and the absence of conserved amino-acid motifs shared by vertebrates in the N-terminus of AUG- α or AUG- β in XP_032809021. In parallel, conserved synteny supports duplication and divergence: genes adjacent to AUG- α or AUG- β in humans are also found adjacent to AUG- α or AUG- β across representative chondrichthyan, actinopterygian, and sarcopterygian lineages (**Fig. 4**). Collectively these lineages span the most recent common ancestor of all jawed vertebrates (**Fig. 4** and—assuming complete and accurate genome annotation—provide evidence that absences of AUG- α (**Fig. 5a**) or AUG- β (**Fig. 5b**) are a consequence of heterogeneous lineage-specific losses. Further, sequence comparisons of jawed vertebrate AUG- α and AUG- β reveal these ligands to share structural conservation with lamprey AUG: all encode four cysteines near the C-terminus (**Fig. 5C**). Functional divergence analysis further identified three sites in human AUG- α (81-Glu, 91-Leu, and 146-Val) as being significant ($P < 0.01$) to the differential function of the *AUG* paralogs.

Orthologs of AUG- α exhibited rapid evolution (high non-synonymous substitution) across jawed vertebrates, suggesting a comparatively conserved function of AUG- β (**Table 1**). This hypothesis is supported by a relative ratio test demonstrating significantly greater numbers of amino acid substitutions in AUG- α than in AUG- β ($P < 0.05$). Both AUG- α (Iss = 1.39; Iss.c = 0.76) and AUG- β (Iss = 1.45; Iss.c = 0.76) have experienced severe substitution saturation (DAMBE; Xia & Lemey 2009; Xia 2018) and by the sharp decline of phylogenetic informativeness (PhyInformR; Dornburg et al. 2016) of the protein sequence at timescales corresponding with the divergence of mammals (**Fig. S6**).

Tissue-specific expression profiles of ALK and AUG in lamprey

Quantifying expression levels of *ALK* and *AUG* transcripts in lampreys identified similarities in expression that are consistent with their operation as a ligand-receptor pair and that align with function in humans. We performed PCR-based (RACE) strategies in muscle, brain, liver, and eye tissues from adult and ammocoete lampreys (**Table S2**) and analyzed RNA sequencing data from 76 experiments in eight transcriptomic projects of *P. marinus* in NCBI Sequence Read Archive (SRA). Short reads were aligned to the recently updated annotation of *P. marinus* genome from NCBI, and we quantified expression as raw read counts that mapped to the two *ALK*-like genes and four *AUG*-like genes (**Table S3**). For *ALK* and the annotated LTK gene, expression of the annotated *LTK* (XP_032818754) was detected in 67 out of the 76 experiments with an average 55 counts (maximum 483 counts), while *ALK* (XP_032804141) exhibited very low expression (1–12 counts) in 20 out of the 76 experiments as well as our RACE results. This suggests a restricted role for XP_032804141 in lamprey development. Expression of *AUG* was not detected from our lamprey specimens in either the ammocoete and adult life stage. Among the four *AUG*-like genes, expression of XP_032826340 and XP_032810391—which have no predicted signal peptides—was either undetectable or very low in all 76 experiments, while *AUG*-XP_032817706 and *AUG*-XP_032809021 were respectively detected in 55 and 61 of the 76 experiments, always at high expression levels (**Fig. S7**). We found expression levels of the two *AUG* homologs—summed, as would be consistent with subfunctionalization—were correlated with expression of *ALK*, supporting the hypothesis that *AUG* and *ALK* are a ligand-receptor pair in lamprey (Pearson's correlation coefficient $r(84) = 0.59$, $P < 10^{-5}$; **Table S3**, **Fig. S7**). These findings are consistent with the hypothesis that they were a ligand-receptor pair in the common ancestor of lamprey and jawed vertebrates. *AUG* transcripts were found in the neurula stage, olfactory tissues after exposure to copper, meiotic testes and brains, and samples with post-injury spinal cord and brain tissue (**Table S3**). No transcripts of *ALK*, *AUG*-XP_032817706, or *AUG*-XP_032809021 were detected in 8 samples, including parasitic olfactory

epithelium, adult olfactory epithelium, adult brain, parasitic larval brain, laval liver, parasitic liver, parasitic kidney and parasitic liver samples (**Table S3**).

Discussion

Here we have reported our discovery that the augmentor ligands of ALK and LTK are a vertebrate innovation, and demonstrated that invertebrate genes *JEB* and *HEN-1* are not homologs of *AUG*. This lineage-specific evolution of ALK and LTK function across Metazoa likely underlies a diversity of hypotheses regarding how human functions and phenotypes of ALK homologues relate to the functions and phenotypes observed in protostome models such as nematodes and fruit flies (Hallberg & Palmer 2016; Ishihara et al. 2002; Reshetnyak et al. 2018; Hallberg & Palmer 2013). The split between ALK-bearing vertebrates and ALK-bearing protostomes corresponds not only to a divergence in receptor-ligand association, but also to a greater functional divergence. For instance, ALK plays a central role in the visceral gut formation, growth, and neurogenesis in protostomes (Wolfstetter et al. 2017; Gouzi et al. 2011), and a role in neuronal proliferation, differentiation, and survival in vertebrates (Yao et al. 2013; Weiss et al. 2012). Similarly, we have demonstrated that after the genesis of *AUG*, an additional duplication likely occurred prior to the most recent common ancestor of jawed vertebrates, giving rise to the ligands *AUG- α* and *AUG- β* . This duplication of *AUG* coincided with a duplication of anaplastic lymphoma kinase (*ALK*) creating leukocyte tyrosine kinase (*LTK*).

The duplication of ALK into *ALK* and *LTK* reveals striking functional similarities between these two lineages of tyrosine kinases. These functional similarities within ALK, within LTK, and between ALK and LTK encourage the use of a wide range of candidate vertebrate model species for investigation of the receptors, ligands, and their interactions. For instance, the LTKs of non-tetrapods, including models such as zebrafish, exhibit a strong signature of shared domain structure with mammal ALK. The only notable difference is the sequence divergence between the N-termini of ALK and LTK within the first MAM domain. In addition, we also found evidence for the conservation of 11 key amino acids between

non-tetrapod and mammal LTK. The amino-acid identity of five of these eleven amino acids is—remarkably—also shared between non-tetrapod LTK and mammal ALK. This conservation is encouraging and consistent with previous research emphasizing the importance of the nonhuman models in ALK tumorigenesis (Hallberg & Palmer 2013). Experimental research investigating the effects of induced point mutations in the ALK and LTK coding sequences at the sites identified here would be especially likely to reveal functionally divergent aspects of ALK and LTK signaling among humans and relevant model species.

Structurally, we found lamprey ALK is similar to mammalian LTK: both feature the PTK and GR domains and lack MAM1 and LDL domains. Our rapid-amplification-of-cDNA-ends experimental results and our analysis of public transcriptomic data mapping to lamprey *ALK* and *AUG* demonstrated that *ALK* and *AUG* are coexpressed post-fertilization. Post-fertilization expression and maximal expression during early embryonic development and after injury to nerve or brain tissues suggests a role of ALK and AUG in nerve and brain development. These results are consistent with expression of *ALK* in mice and humans that is also highest during embryonic development, quickly drops after birth, and is subsequently maintained at a low level (Iwahara et al. 1997; Vernersson et al. 2006). Together, our results and prior studies suggest that the functional roles of ALK and AUG may be conserved across vertebrates. Future research examining spatiotemporal changes in *ALK* and *AUG* expression between species—in particular in response to stressors—presents an exciting and potentially fruitful avenue toward an increasingly thorough understanding of the general role of these genes in humans and as well as their role in human cancers.

Our phylogenetic analyses of AUG- α revealed an accelerated evolutionary rate that is unexpected for proteins executing critical biological functions. Lower rates of sequence evolution are typically expected for proteins believed to have a colocalized dual specificity of interaction between genes. Biophysical binding data suggests AUG- α is a dual-specific ligand for both ALK and LTK (Reshetnyak et al. 2015). Our results demonstrate that this dual specificity has not constrained the evolution of AUG- α to a slower substitution rate within mammals. Instead, molecular rates of AUG- α exceed those estimated for

AUG- β , which is a monospecific ligand of only LTK (Reshetnyak et al. 2015). The high substitution rates observed in mammal LTK, non-tetrapod ALK, and jawed vertebrate AUG- α could indicate increased functional specificity and lower promiscuity of interaction in these genes. In contrast, we found AUG- β homologs to be more conserved than AUG- α homologs, a signature consistent with expectations of co-evolution between signaling and receiving molecules (Goh et al. 2000; Monte et al. 2018; Laisney et al. 2010). In mammals, ALK is activated via both ligand-dependent (Takita 2017; Reshetnyak et al. 2015) and ligand-independent (Deuel 2013; Perez-Pinera et al. 2007) processes, implying multiple functions of ALK and high interaction specificity between ALK and its ligand(s). These lower substitution rates potentially indicate promiscuous interactions among AUG- β and its receptors. Biological relevance of these interactions has been indicated by research on both zebrafish development and human cancers (Reshetnyak et al. 2015; Mo et al. 2017; Guan et al. 2015). The differences in evolutionary rates between AUG- α & AUG- β , mammal LTK & non-tetrapod ALK, and mammal ALK & non-tetrapod LTK represent evolutionary trade-offs between functional specificity and the number of interactors a protein can achieve (Zhang & Yang 2015).

Rapidly evolving proteins have been shown to exhibit greater functional specificity—for example, higher tissue specificity or higher promoter methylation in mammals (Zhang & Yang 2015). If the different rates of evolution of these receptors and their ligands are the outcomes of evolutionary trade-offs, we might expect a higher complexity of the protein networks associated with mammal LTK and non-tetrapod ALK, and expect more functional generality in networks associated with mammal ALK and non-tetrapod LTK. Consequent hypotheses that mammal LTK and non-tetrapod ALK regulate nerve development in conjunction with many highly specialized partner proteins, and that mammal ALK and non-tetrapod LTK play broadly important and general roles in internal developmental signaling, warrant further molecular biological investigation. Additional collection of data on genome-wide or gene-specific spatial and temporal co-expression in vertebrates would provide additional insight into regulatory gene interaction networks, narrowing the scope of viable hypotheses for protein-protein interaction experiments as well as revealing evolutionary change in the structure and function of the ALK, LTK, and

AUG signaling networks. Building on the evolutionary history we have here revealed, comparative analyses of inter-specific protein interaction networks will reveal how these genes and gene domains are co-opted in tumorigenesis and cancer progression. Such insights will enable translational research toward interventions that successfully target the cellular function of ALK and LTK in human cancers.

Materials and methods

Identification homologs of ALK and its possible ligand in non-models—To investigate homologs of vertebrate ALKs, LTKs, and their associated ligands across model organisms spanning the protostomes and deuterostomes, amino-acid alignments of ALK homologs as well alignments of JEB, HEN-1 and AUGs from the genomes of zebrafish (*Danio rerio*), fruit fly (*Drosophila melanogaster*), and nematode (*Caenorhabditis elegans*) were used to perform HMMER ortholog searches (Wheeler & Eddy 2013) against available genomes from genome databases in NCBI (www.ncbi.nlm.nih.gov) and Ensembl (www.ensembl.org; **Table S1**). For genomes whose annotation did not report homologs of these genes, additional BLASTp and tBLASTn searches were conducted with protein sequences derived from the three representative genomes listed above. Sequences with the mutually best matches between two sequences in genome pairs via BLAST search (Moreno-Hagelsieb & Latimer 2008) were subjected to further phylogenetic analyses to confirm their homology with annotated ALK, LTK, or ligand proteins. To investigate the evolution of ALK/LTK and associated AUG in vertebrates, sequences from model species that included zebrafish, frog (*Xenopus tropicalis*), chicken (*Gallus gallus*), zebra finch (*Taeniopygia guttata*), mouse (*Mus musculus*) and human (*Homo sapiens*) were queried against non-model vertebrate genomes (**Table S1**) using HMMER and best-hit reciprocal-BLAST searches. To illuminate the origin of vertebrate ALK/LTK and AUG, special attention was devoted to thoroughly ascertain the presence of ALK and possible ligand(s) in the genomes of the jawless vertebrates (hagfish and lampreys), as these animals represent the earliest-diverging extant vertebrate lineage (Shimeld & Donoghue 2012). No sequence in the hagfish genome exhibited any similarity to vertebrate ALK, LTK, or AUG. BLAST searches of multiple ALK and AUG sequences from non-tetrapod genomes recovered highly conserved

regions between the two lamprey genomes indicating the presence of homologs of jawed-vertebrate ALK and AUG. To analyze synteny of ALK, LTK and AUG orthologs among vertebrate representatives, predicted genes around targets were extracted from genome annotations of the sea lamprey, elephant shark, zebrafish, *Xenopus*, chicken, armadillo, pig, dog, mouse and human genomes available from the UCSC genome browser, as well as from the NCBI genome browser for gar, and from the Ensemble genome browser for the opossum.

Searches for *ALK*, *LTK*, and paired ligands were additionally conducted against transcriptomes of sea lampreys (*Petromyzon marinus*), including 86 publicly available transcriptomes (**Table S3**). RNAs were additionally sampled from tissues of an ammocoete and an adult sea lamprey. The ammocoete lamprey was flash-frozen in liquid nitrogen before tissues of head, muscle, and viscera were dissected for RNA extraction. Tissues of the large adult lamprey were dissected from muscle, eyes, liver, brain, and heart. All tissues were preserved in *RNAlater*, then maintained at -76 C prior to RNA extraction. Total RNA was extracted from homogenized tissue with TRI REAGENT (Molecular Research Center). Messenger RNA was purified using Dynabeads oligo(dT) magnetic separation (Invitrogen). A cDNA library was generated using a SMARTer 5'/3' RACE Kit (Takara cat no. 634860) as per the manufacturer's instructions. First-Strand cDNA synthesis was performed using 11 uL of RNA extract and 1 uL of 3'-CDS Primer A. Rapid amplification of cDNA ends (RACE) was also performed as per the manufacturer's instructions with custom gene-specific primers (GSPs, **Table S2**).

The genome sequence of the lamprey *P. marinus* was downloaded from Ensembl (Aken et al. 2017) and was used as the reference sequence for HISAT2 (Kim et al. 2015) to build the index and perform read alignment. Transcripts were assembled and gene-expression levels were quantified using *StringTie* (Pertea et al. 2015). Sequence read data totaled 323 Gb, and the largest single dataset (based on paired-end sequencing with the Illumina HiSeq 4000 platform) amounted to 21.7 Gb. We used the most recently annotated *P. marinus* genome at NCBI as a reference for mapping reads, specifying the *HHEX* gene as a control (Sharman & Holland 1998). Due to its high efficiency, HISAT2 (Kim et al.

2015) was chosen to map the reads, which were subsequently extracted with SAMtools (Li et al. 2009) using parameter setting -F 4 to filter unmapped reads. All RNAseq datasets (Table S3) generated by previous studies (Libants et al. 2009; Ravi et al. 2016; Bryant et al. 2016; Smith et al. 2018a; Goetz et al. 2016; Chang et al. 2013; Jones et al. 2019; Herman et al. 2018; Chung-Davidson et al. 2020; Ren et al. 2015; Ravi et al. 2019; Zhang et al. 2017) and downloaded from NCBI. Reads were derived from samples of whole embryos 1–2.5 days post-fertilization; neural crest, kidney, brain, liver and olfactory tissue after 24 h exposure to 5 g/mL, 10 g/mL, and 30 g/mL of copper; and brains (whole brains without the olfactory lobes) and spinal cords (1 cm surrounding the lesion), harvested from 6 h to 12 weeks after injury of lamprey specimens (Table S3).

Molecular Phylogeny

We obtained amino acid sequences and nucleotide sequences of ALK / LTK and AUG-□ / AUG-□ genes from NCBI and Ensemble respectively (Table S1). The amino-acid sequences of ALK and Jeb in *Drosophila melanogaster*, and SCD-2 (homologous with ALK in *Drosophila*) and Hen-1 in *Caenorhabditis elegans*, were accessed for sequence comparison to lamprey ALK and AUG. We subsequently used lamprey sequences as outgroups for phylogenetic inference of ALK/LTK and AUG evolution in jawed vertebrates. Amino-acid sequences were aligned using MAFFT in Saté-II (Katoh & Standley 2013; Liu et al. 2011), while nucleotide sequences were aligned based on amino-acid sequences by the TranslatorX online server (Abascal et al. 2010). We inferred phylogenies of ALK and AUG using Markov chain Monte Carlo (MCMC) methods implemented in MrBayes 3.2 (Ronquist et al. 2011). Our Bayesian phylogenetic analyses were executed for 10,000,000 generations, sampling every 1,000 generations with four chains. We set lamprey ALK and AUG as outgroups for each analysis and discarded 2500 (25%) of the 10,000 trees as burn-in. We assessed convergence of the chains by quantifying potential scale reduction factors (PSF = 1.0). We visually compared computed log likelihoods across chains to confirm stationarity. For nucleotide sequences, the GTR + I + Γ substitution model was

specified. For amino-acid sequences, a mixed model of amino-acid substitution was specified that allowed each model of amino-acid substitution to contribute in relation to its posterior probability. We deemed branches that exhibited a posterior probability (PP) higher than 0.98 to be strongly supported (**Figs. 2–3, Fig. S1**). We conducted analyses on both the aligned amino-acid and nucleotide sequences separately.

Ancestral reconstruction

Using phylogeny reconstructed for ALK and AUG homologs in vertebrates, ancestral sequence reconstructions were performed using both likelihood and Bayesian approaches as implemented in PAML 4 (Yang 2007). We used `codeml` to conduct codon-based ancestral sequence reconstruction of the common ancestor of mammals as well as all vertebrates. Reconstructed sequences of ALK and AUG for the common ancestor of lampreys and jawed vertebrates were also used to search against non-vertebrate genomes for possible homologs. Ancestral states of MAM domains in vertebrate genomes were estimated using maximum likelihood (ML) criteria in `BayesTraits` (Pagel & Meade 2004; Pagel 1999). We coded the presence or absence of MAM domains for ALK and LTK, and used the `multiState` method of discrete character evolution to reconstruct gains or losses of MAM domains (**Fig. S2**).

Selection tests

To test for positive selection along specific branches in vertebrate ALK/LTK and AUG evolution, we used branch models implemented in PAML (Yang 2007, 1997), in which the ratio of nonsynonymous to synonymous substitution (ω) was allowed to vary among branches in the phylogeny (**Table S4, S5**). The ratio ω was estimated for the branch of interest (the ‘foreground’ branch) and the rest of the tree (the ‘background’) in the phylogeny reconstructed from a multiple sequence alignment. To evaluate whether there was a statistically significant difference between the branch model and the null model, a likelihood ratio test (LRT) was applied. To search for positively selected sites, site models permitting ω to vary among sites were used. We set `NSsites` to equal 0, 1, 2, 7, and 8, then conducted likelihood ratio tests

between pairs of the models to identify the best fitting model comparing M_{1a} (nearly neutral) against M_{2a} (positive selection), and M_7 (beta) against M_8 (beta & ω), each with two degrees of freedom. A Bayes-Empirical Bayes analysis was performed to identify the sites evolving under significant positive selection (Yang et al. 2005). Clade-model C was fit to the data to evaluate whether ω differed between ALK and LTK and between AUG- α and AUG- β in mammals (Bielawski & Yang 2004; Yang et al. 2005). The improved model M_{2a_rel} was used as the null model for the likelihood ratio test on clade-model C results (Weadick & Chang 2011).

The degree of saturation of substitutions for ALK, LTK, and AUG proteins was assessed by DAMBE (Xia & Lemey 2009) and by inspection of phylogenetic informativeness profiles (Townsend 2007) visualized with the R package PhyInformR (Dornburg et al. 2016). To estimate site rates, we used HyPhy (Cummings 2004) in the PhyDesign web interface (López-Giráldez & Townsend 2011). Profiles of phylogenetic informativeness were depicted along a relative ultrametric guide topology generated in BEAST v. 2.4.7 (Drummond et al. 2012) with a prior root height of 1.0. As the depths of divergence examined exhibited some evidence of saturation with regard to substitutions, we compared their maximum-likelihood rates of sequence evolution using a likelihood-ratio test conducted in HyPhy (Cummings 2004), enabling meaningful comparisons of relative-rate differences between AUG paralogs.

Functional divergence analysis

We used DIVERGE 3.0 (Gu et al. 2013) to test for functional divergence of the gene pairs. DIVERGE tests for site-specific shifts in evolutionary rates after gene duplication or speciation. The coefficient of divergence (θ_D) was calculated to test against a null hypothesis of no functional divergence between ALK and LTK, between mammal ALK and fish ALK, between mammal LTK and fish LTK, and between AUG- α and AUG- β . We employed the default posterior probability cutoff of 0.5 for detection of site-specific shifted evolutionary rates (Gu et al. 2013). Amino acids with significant ($P < 0.05$) roles in functional divergence between gene paralogs were predicted.

Availability of Data and Materials

All data is available on NCBI. Alignments of sequences accessed for this study have been deposited on Zenodo under DOI: 10.5281/zenodo.3972039.

Supplementary Material

Supplementary materials are available at *Genome Biology and Evolution* online.

Acknowledgments. We thank J. Morkert at Lamprey Service and G. J. Watkins-Colwell at the Peabody Museum of Natural History for help collecting lamprey specimens and tissue samples, and J. Yoder, K. Zapf, L. Abrams, as well as J. Smith and an anonymous reviewer for valuable discussions and critical comments. This research was supported by National Science Foundation grant IOS-1755242 to AD, and by National Science Foundation grant GCR-1934860 and a gift from the Notsew Orm Sands Foundation to JPT. ZW and JPT designed the research; ZW and EM performed the research; ZW, AD, JW, and FL analyzed the data; and AD, ZW, JW, EM, and JPT wrote the paper. The authors declare no conflict of interest.

References

- Abascal F, Zardoya R, Telford MJ. 2010. TranslatorX: multiple alignment of nucleotide sequences guided by amino acid translations. *Nucleic Acids Res.* 38:W7–13.
- Aken BL et al. 2017. Ensembl 2017. *Nucleic Acids Res.* 45:D635–D642.
- Bielawski JP, Yang Z. 2004. A maximum likelihood method for detecting functional divergence at individual codon sites, with application to gene family evolution. *J. Mol. Evol.* 59:121–132.
- Bilsland JG et al. 2008. Behavioral and neurochemical alterations in mice deficient in anaplastic lymphoma kinase suggest therapeutic potential for psychiatric indications. *Neuropsychopharmacology.* 33:685–700.
- Bresler SC et al. 2011. Differential inhibitor sensitivity of anaplastic lymphoma kinase variants found in neuroblastoma. *Sci. Transl. Med.* 3:108ra114.
- Bryant SA, Herdy JR, Amemiya CT, Smith JJ. 2016. Characterization of Somatic-ly-Eliminated Genes During Development of the Sea Lamprey (*Petromyzon marinus*). *Molecular Biology and Evolution.* 33:2337–2344. doi: 10.1093/molbev/msw104.
- Chang S et al. 2013. The sea lamprey has a primordial accessory olfactory system. *BMC Evolutionary Biology.* 13:172. doi: 10.1186/1471-2148-13-172.
- Chiarle R, Voena C, Ambrogio C, Piva R, Inghirami G. 2008. The anaplastic lymphoma kinase in the pathogenesis of cancer. *Nat. Rev. Cancer.* 8:11–23.
- Chung-Davidson Y et al. 2020. TGF- β Signaling Plays a Pivotal Role During Developmental Biliary Atresia in Sea Lamprey (*Petromyzon marinus*). *Hepatology Communications.* 4:219–234. doi: 10.1002/hep4.1461.
- Cismasiu VB, Denes SA, Reiländer H, Michel H, Szedlacsek SE. 2004. The MAM (Meprin/A5-protein/PTPmu) Domain Is a Homophilic Binding Site Promoting the Lateral Dimerization of Receptor-like Protein-tyrosine Phosphatase μ . *J. Biol. Chem.* 279:26922–26931.
- Cummings MP. 2004. HyPhy (Hypothesis Testing Using Phylogenies). In: *Dictionary of Bioinformatics and Computational Biology.*
- De Brouwer S et al. 2010. Meta-analysis of neuroblastomas reveals a skewed ALK mutation spectrum in tumors with MYCN amplification. *Clin. Cancer Res.* 16:4353–4362.
- Dehal P, Boore JL. 2005. Two Rounds of Whole Genome Duplication in the Ancestral Vertebrate. *PLoS Biol.* 3:e314.
- Deuel TF. 2013. Anaplastic lymphoma kinase: ‘Ligand Independent Activation’ mediated by the PTN/RPTP β/ζ signaling pathway. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics.* 1834:2219–2223.
- Dornburg A, Fisk JN, Tamagnan J, Townsend JP. 2016. PhyInformR: phylogenetic experimental design and phylogenomic data exploration in R. *BMC Evol. Biol.* 16:262.
- Drummond AJ, Suchard MA, Xie D, Rambaut A. 2012. Bayesian Phylogenetics with BEAUti and the

BEAST 1.7. *Mol. Biol. Evol.* 29:1969–1973.

Englund C et al. 2003. Jeb signals through the Alk receptor tyrosine kinase to drive visceral muscle fusion. *Nature*. 425:512–516.

Fadeev A et al. 2018. ALKALs are in vivo ligands for ALK family receptor tyrosine kinases in the neural crest and derived cells. *Proc. Natl. Acad. Sci. U. S. A.* 115:E630–E638.

Fujioka A et al. 2006. Dynamics of the Ras/ERK MAPK cascade as monitored by fluorescent probes. *J. Biol. Chem.* 281:8917–8926.

Gambacorti-Passerini C. 2016. ALK inhibitors for clinical use in cancer therapy. *Front. Biosci.* 8:46–60.

Goetz F, Smith SE, Goetz G, Murphy CA. 2016. Sea lampreys elicit strong transcriptomic responses in the lake trout liver during parasitism. *BMC Genomics*. 17:675.

Goh C-S, Bogan AA, Joachimiak M, Walther D, Cohen FE. 2000. Co-evolution of proteins with their interaction partners 1 Edited by B. Honig. *J. Mol. Biol.* 299:283–293.

Gouzi JY et al. 2011. The receptor tyrosine kinase Alk controls neurofibromin functions in *Drosophila* growth and learning. *PLoS Genet.* 7:e1002281.

Grande E, Bolós M-V, Arriola E. 2011. Targeting oncogenic ALK: a promising strategy for cancer treatment. *Mol. Cancer Ther.* 10:569–579.

Guan J et al. 2015. FAM150A and FAM150B are activating ligands for anaplastic lymphoma kinase. *Elife*. 4:e09811.

Gudernova I et al. 2017. Inhibitor repurposing reveals ALK, LTK, FGFR, RET and TRK kinases as the targets of AZD1480. *Oncotarget*. 8:109319–109331.

Gu X et al. 2013. An update of DIVERGE software for functional divergence analysis of protein family. *Mol. Biol. Evol.* 30:1713–1719.

Hallberg B, Palmer RH. 2013. Mechanistic insight into ALK receptor tyrosine kinase in human cancer biology. *Nat. Rev. Cancer*. 13:685–700.

Hallberg B, Palmer RH. 2016. The role of the ALK receptor in cancer biology. *Annals of Oncology*. 27:iii4–iii15. doi: 10.1093/annonc/mdw301.

Herman PE et al. 2018. Highly conserved molecular pathways, including Wnt signaling, promote functional recovery from spinal cord injury in lampreys. *Sci. Rep.* 8:742.

Holla VR et al. 2017. ALK: a tyrosine kinase target for cancer therapy. *Molecular Case Studies*. 3:a001115. doi: 10.1101/mcs.a001115.

Inoue T, Thomas JH. 2000. Suppressors of transforming growth factor-beta pathway mutants in the *Caenorhabditis elegans* dauer formation pathway. *Genetics*. 156:1035–1046.

Ishihara T et al. 2002. HEN-1, a secretory protein with an LDL receptor motif, regulates sensory integration and learning in *Caenorhabditis elegans*. *Cell*. 109:639–649.

Iwahara T et al. 1997. Molecular characterization of ALK, a receptor tyrosine kinase expressed

specifically in the nervous system. *Oncogene*. 14:439–449.

Jones J, Wellband K, Zielinski B, Heath DD. 2019. Transcriptional Basis of Copper-Induced Olfactory Impairment in the Sea Lamprey, a Primitive Invasive Fish. *G3*. 9:933–941.

Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* 30:772–780.

Kersey PJ et al. 2017. Ensembl Genomes 2018: an integrated omics infrastructure for non-vertebrate species. *Nucleic Acids Res.* doi: 10.1093/nar/gkx1011.

Kim D, Langmead B, Salzberg SL. 2015. HISAT: a fast spliced aligner with low memory requirements. *Nat. Methods*. 12:357–360.

Kocher B, Piwnica-Worms D. 2013. Illuminating cancer systems with genetically engineered mouse models and coupled luciferase reporters in vivo. *Cancer Discov.* 3:616–629.

Laisney JAGC, Braasch I, Walter RB, Meierjohann S, Scharf M. 2010. Lineage-specific co-evolution of the Egf receptor/ligand signaling system. *BMC Evol. Biol.* 10:27.

Lee H-H, Norris A, Weiss JB, Frasch M. 2003. Jelly belly protein activates the receptor tyrosine kinase Alk to specify visceral muscle pioneers. *Nature*. 425:507–512.

Lemke G. 2015. Adopting ALK and LTK. *Proc. Natl. Acad. Sci. U. S. A.* 112:15783–15784.

Libants S et al. 2009. The sea lamprey *Petromyzon marinus* genome reveals the early origin of several chemosensory receptor families in the vertebrate lineage. *BMC Evol. Biol.* 9:180.

Li H et al. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. 25:2078–2079.

Lin JJ, Riely GJ, Shaw AT. 2017. Targeting ALK: Precision Medicine Takes on Drug Resistance. *Cancer Discov.* 7:137–155.

Liu K et al. 2011. SATé-II: Very Fast and Accurate Simultaneous Estimation of Multiple Sequence Alignments and Phylogenetic Trees. *Syst. Biol.* 61:90.

López-Giráldez F, Townsend JP. 2011. PhyDesign: an online application for profiling phylogenetic informativeness. *BMC Evol. Biol.* 11:152.

Marzec M et al. 2006. Oncogenic tyrosine kinase NPM/ALK induces activation of the MEK/ERK signaling pathway independently of c-Raf. *Oncogene*. 26:813–821.

Mehta TK et al. 2013. Evidence for at least six Hox clusters in the Japanese lamprey (*Lethenteron japonicum*). *Proc. Natl. Acad. Sci. U. S. A.* 110:16044–16049.

Mitchell AL et al. 2019. InterPro in 2019: improving coverage, classification and access to protein sequence annotations. *Nucleic Acids Res.* 47:D351–D360.

Mo ES, Cheng Q, Reshetnyak AV, Schlessinger J, Nicoli S. 2017. Alk and Ltk ligands are essential for iridophore development in zebrafish mediated by the receptor tyrosine kinase Ltk. *Proceedings of the National Academy of Sciences*. doi: 10.1073/pnas.1710254114.

Monte I et al. 2018. Ligand-receptor co-evolution shaped the jasmonate pathway in land plants. *Nat.*

Chem. Biol. doi: 10.1038/s41589-018-0033-4.

Moreno-Hagelsieb G, Latimer K. 2008. Choosing BLAST options for better detection of orthologs as reciprocal best hits. *Bioinformatics*. 24:319–324.

Mosse YP et al. 2008. Identification of ALK as a major familial neuroblastoma predisposition gene. *Nature*. 455:930–U22.

NCBI Resource Coordinators. 2017. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res*. doi: 10.1093/nar/gkx1095.

Ogawa S, Takita J, Sanada M, Hayashi Y. 2011. Oncogenic mutations of ALK in neuroblastoma. *Cancer Sci*. 102:302–308.

O'Hagan RC, Wu M, Rideout WM, Zhou Y, Heyer J. 2005. Genetically Engineered Mouse Models of Human Cancer for Drug Discovery and Development. *Cancer Drug Discovery and Development*. 247–261. doi: 10.1385/1-59259-893-5:247.

Pagel M. 1999. The Maximum Likelihood Approach to Reconstructing Ancestral Character States of Discrete Characters on Phylogenies. *Syst. Biol*. 48:612–622.

Pagel M, Meade A. 2004. A Phylogenetic Mixture Model for Detecting Pattern-Heterogeneity in Gene Sequence or Character-State Data. *Syst. Biol*. 53:571–581.

Paoloni M, Khanna C. 2008. Translation of new cancer treatments from pet dogs to humans. *Nat. Rev. Cancer*. 8:147–156.

Perez-Pinera P, Zhang W, Chang Y, Vega JA, Deuel TF. 2007. Anaplastic Lymphoma Kinase Is Activated Through the Pleiotrophin/Receptor Protein-tyrosine Phosphatase β/ζ Signaling Pathway. *J. Biol. Chem*. 282:28683–28690.

Pertea M et al. 2015. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat. Biotechnol*. 33:290–295.

Ravi V et al. 2016. Cyclostomes Lack Clustered Protocadherins. *Mol. Biol. Evol*. 33:311–315.

Ravi V et al. 2019. Lampreys, the jawless vertebrates, contain three Pax6 genes with distinct expression in eye, brain and pancreas. *Sci. Rep*. 9:19559.

Ren J et al. 2015. Genome-wide analysis of the ATP-binding cassette (ABC) transporter gene family in sea lamprey and Japanese lamprey. *BMC Genomics*. 16:436.

Reshetnyak AV et al. 2015. Augmentor alpha and beta (FAM150) are ligands of the receptor tyrosine kinases ALK and LTK: Hierarchy and specificity of ligand-receptor interactions. *Proc. Natl. Acad. Sci. U. S. A*. 112:15862–15867.

Reshetnyak AV et al. 2018. Identification of a biologically active fragment of ALK and LTK-Ligand 2 (augmentor- α). *Proceedings of the National Academy of Sciences*. 115:8340–8345. doi: 10.1073/pnas.1807881115.

Rohrbough J, Broadie K. 2010. Anterograde Jelly belly ligand to Alk receptor signaling at developing synapses is regulated by Mind the gap. *Development*. 137:3523–3533.

- Roll JD, Reuther GW. 2012. ALK-activating homologous mutations in LTK induce cellular transformation. *PLoS One*. 7:e31733.
- Rowell JL, McCarthy DO, Alvarez CE. 2011. Dog models of naturally occurring cancer. *Trends Mol. Med.* 17:380–388.
- Sharman AC, Holland PW. 1998. Estimation of Hox gene cluster number in lampreys. *Int. J. Dev. Biol.* 42:617–620.
- Shimeld SM, Donoghue PCJ. 2012. Evolutionary crossroads in developmental biology: cyclostomes (lamprey and hagfish). *Development*. 139:2091–2099.
- Slavish PJ et al. 2011. Synthesis of an aryloxy oxo pyrimidinone library that displays ALK-selective inhibition. *Bioorg. Med. Chem. Lett.* 21:4592–4596.
- Smith JJ et al. 2018a. Publisher Correction: The sea lamprey germline genome provides insights into programmed genome rearrangement and vertebrate evolution. *Nat. Genet.* 50:1617.
- Smith JJ et al. 2013. Sequencing of the sea lamprey (*Petromyzon marinus*) genome provides insights into vertebrate evolution. *Nat. Genet.* 45:415–21, 421e1–2.
- Smith JJ et al. 2018b. The sea lamprey germline genome provides insights into programmed genome rearrangement and vertebrate evolution. *Nat. Genet.* 50:270–277.
- Soda M et al. 2007. Identification of the transforming EML4-ALK fusion gene in non-small-cell lung cancer. *Nature*. 448:561–U3.
- Stoletov K, Klemke R. 2008. Catch of the day: zebrafish as a human cancer model. *Oncogene*. 27:4509–4520.
- Takita J. 2017. The role of anaplastic lymphoma kinase in pediatric cancers. *Cancer Sci.* 108:1913–1920.
- Tartari CJ, Scapozza L, Gambacorti-Passerini C. 2011. The ALK gene, an attractive target for inhibitor development. *Curr. Top. Med. Chem.* 11:1406–1419.
- Townsend JP. 2007. Profiling phylogenetic informativeness. *Syst. Biol.* 56:222–231.
- Vernersson E et al. 2006. Characterization of the expression of the ALK receptor tyrosine kinase in mice. *Gene Expr. Patterns*. 6:448–461.
- Weadick CJ, Chang BSW. 2011. An Improved Likelihood Ratio Test for Detecting Site-Specific Functional Divergence among Clades of Protein-Coding Genes. *Mol. Biol. Evol.* 29:1297–1300.
- Weiss JB et al. 2012. Anaplastic lymphoma kinase and leukocyte tyrosine kinase: functions and genetic interactions in learning, memory and adult neurogenesis. *Pharmacol. Biochem. Behav.* 100:566–574.
- Wellstein A. 2012. ALK receptor activation, ligands and therapeutic targeting in glioblastoma and in other cancers. *Front. Oncol.* 2:192.
- Wheeler TJ, Eddy SR. 2013. nhmmer: DNA homology search with profile HMMs. *Bioinformatics*. 29:2487–2489.
- Wolfstetter G et al. 2017. The scaffolding protein Cnk binds to the receptor tyrosine kinase Alk to

promote visceral founder cell specification in. *Sci. Signal.* 10. doi: 10.1126/scisignal.aan0804.

Xia X. 2018. DAMBE7: New and improved tools for data analysis in molecular biology and evolution. *Mol. Biol. Evol.* doi: 10.1093/molbev/msy073.

Xia X, Lemey P. 2009. Assessing substitution saturation with DAMBE. In: *The Phylogenetic Handbook*. pp. 615–630.

Yang Z. 2007. PAML 4: Phylogenetic Analysis by Maximum Likelihood. *Mol. Biol. Evol.* 24:1586–1591.

Yang Z. 1997. PAML: a program package for phylogenetic analysis by maximum likelihood. *Bioinformatics.* 13:555–556.

Yang Z, Wong WSW, Nielsen R. 2005. Bayes empirical bayes inference of amino acid sites under positive selection. *Mol. Biol. Evol.* 22:1107–1118.

Yao S et al. 2013. Anaplastic lymphoma kinase is required for neurogenesis in the developing central nervous system of zebrafish. *PLoS One.* 8:e63757.

Zhang H et al. 2017. Lampreys, the jawless vertebrates, contain only two ParaHox gene clusters. *Proc. Natl. Acad. Sci. U. S. A.* 114:9146–9151.

Zhang J, Yang J-R. 2015. Determinants of the rate of protein sequence evolution. *Nat. Rev. Genet.* 16:409–420.

Zhu S et al. 2012. Activated ALK collaborates with MYCN in neuroblastoma pathogenesis. *Cancer Cell.* 21:362–373.

Figure Legends

Figure 1. Phylogenetic distribution of *HEN-1* (red), *JEB* (blue), *ALK* (brown), *LTK* (yellow), *AUG-α* (green), and *AUG-β* (teal) orthologs in metazoan genomes (arcs span taxa with each gene; light blue taxon names correspond to model organisms discussed in the text). Identification of homologous genes was confirmed with phylogenetic analyses, and gene accession numbers are provided (**Table S1**; Photo credits Dan Warren [koala], Katerina Zapfe [squirrel], Bronwyn Williams [urchin], Matt Bertone [*Drosophila*], Lynn Ketchum [zebrafish: creative commons, cropped from the original], and Alex Dornburg [all others]).

Figure 2. Synteny (dotted lines and curves) of up to ten genes (syntenic with human *ALK*: black rectangles; non-syntenic with human *ALK*: light grey rectangles) that are annotated as present in genomes on either side of *ALK* (light blue bars; orange trace) and *LTK* (dark grey bars; blue trace) across major clades of vertebrates.

Figure 3. Phylogeny of vertebrate *ALKs* and *LTKs*. *ALK* in Lamprey (*Petromyzon*) features only GR and PTK domains (sand and onyx blocks). *M*₂, *L*, and possibly *M*₁ domains (teal, light brown, and light gray blocks) are duplicated in jawed vertebrates, giving rise to *ALK* and *LTK* (black dotted lines from simple *ALK*, indicated with gray lines from inset text). Dashed lines from MAM domains indicate the presence of *M*_{1a} and *M*_{1b} in *ALK* and *LTK*, respectively. Some domains in some lineages exhibited low levels of sequence similarity to domains observed in other vertebrates (“?”s on GR block on the lineage to *Tursiops*, and *M*₂ on the lineage to both *Latimeria* and *Taeniopygia*). Domain losses (-) are indicated at internodes where they were reconstructed to occur (e.g. loss of *M*₁, *L*, and *M*₂ in *LTK* in mammals—highlighted with a gold gradient, and loss of *M*₁, *L*, and *M*₂ in *Tursiops* in *ALK*). Unlabeled internodes all exhibited strong statistical support (Bayesian posterior probability > 0.98), labels on internodes indicate other BPP values. Photos: AD.

Figure 4. (A) Synteny (dotted lines or curves—blue: synteny to a non-tetrapod, brown: synteny within tetrapods) of up to ten genes (syntenic with human *AUG-α* or *AUG-β*: black rectangles; not syntenic with human *AUG-α* or *AUG-β*: light grey rectangles) that are annotated in genomes as present on either side of *AUG-α* (light blue bars; orange trace) and *AUG-β* (dark grey bars; blue trace) across major clades of vertebrates. **(B)** In zebrafish, an *AUG-α* paralog is located within 15 genes of *LTK*.

Figure 5. Phylogeny of vertebrate **(A)** *AUG-β* and **(B)** *AUG-α*, in which labels at branches indicate Bayesian posterior probabilities < 0.98, illustrated with **(C)** an amino-acid alignment of selected AUGs demonstrating sequence conservation between lamprey *AUG* and its homologs in mammals and zebrafish including key cysteine positions near the C-terminus (aquamarine columns). The lamprey alignment possessed unique sequence insertions (vertical lines) that are not shown here. An asterisk (*) indicates the *Petromyzon* sequence that is most similar to jawed-vertebrate *AUG*. Some lamprey *AUG* sequences have lost their signalling peptide (grey shading). Images: AD (Possum at dinner courtesy of Dan Warren).

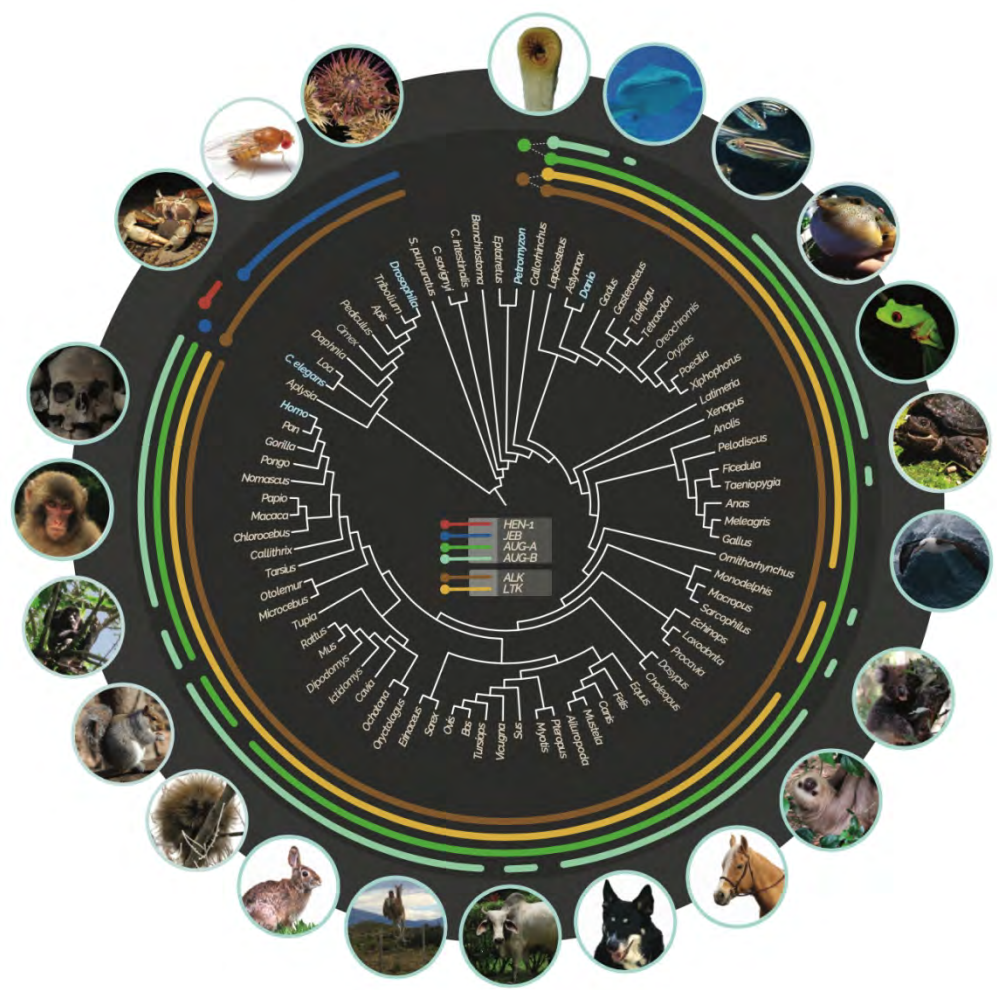


Figure 1. Phylogenetic distribution of HEN-1 (red), JEB (blue), ALK (brown), LTK (yellow), AUG-a (green), and AUG-β (teal) orthologs in metazoan genomes (arcs span taxa with each gene; light blue taxon names correspond to model organisms discussed in the text). Identification of homologous genes was confirmed with phylogenetic analyses, and gene accession numbers are provided (Table S1; Photo credits Dan Warren [koala], Katerina Zapfe [squirrel], Bronwyn Williams [urchin], Matt Bertone [Drosophila], Lynn Ketchum [zebrafish: creative commons, cropped from the original], and Alex Dornburg [all others]).

464x464mm (72 x 72 DPI)



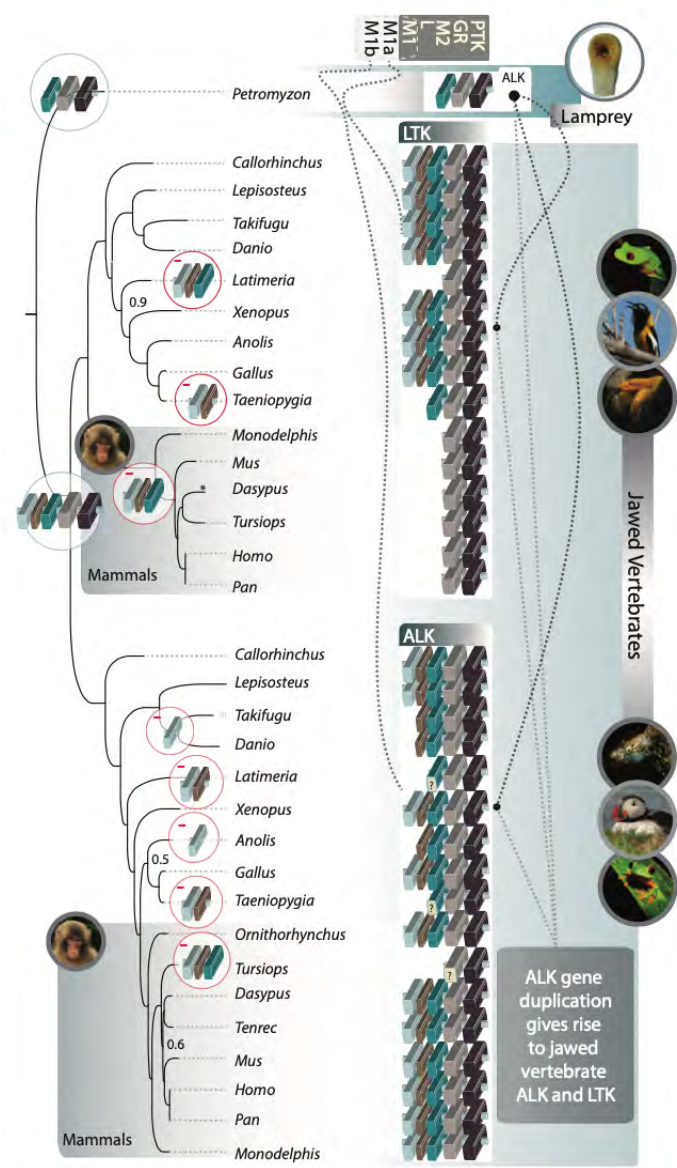
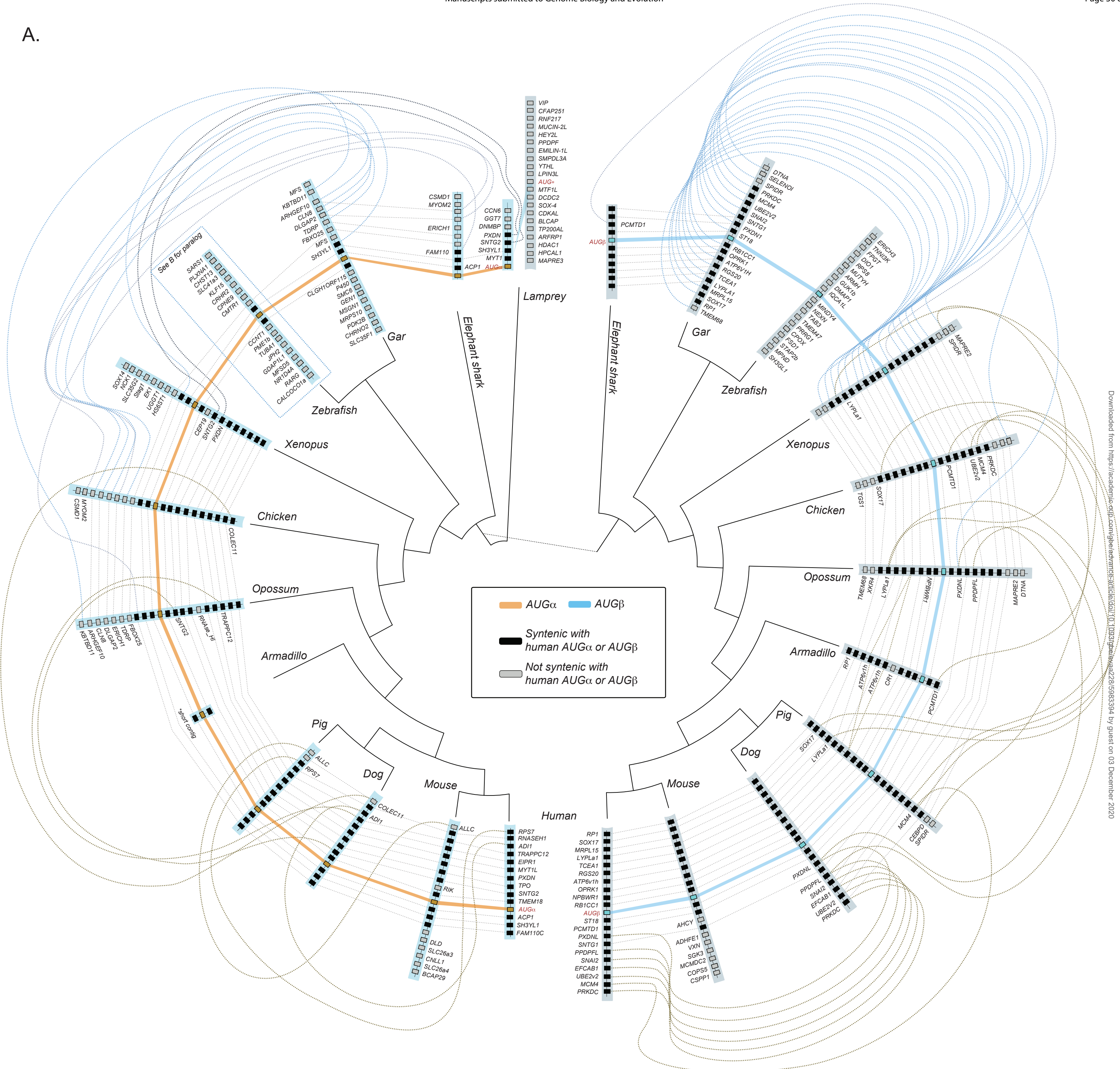


Figure 3. Phylogeny of vertebrate ALKs and LTKs. ALK in Lamprey (Petromyzon) features only GR and PTK domains (sand and onyx blocks). M2, L, and possibly M1 domains (teal, light brown, and light gray blocks) are duplicated in jawed vertebrates, giving rise to ALK and LTK (black dotted lines from simple ALK, indicated with gray lines from inset text). Dashed lines from MAM domains indicate the presence of M1a and M1b in ALK and LTK, respectively. Some domains in some lineages exhibited low levels of sequence similarity to domains observed in other vertebrates ("?"s on GR block on the lineage to Tursiops, and M2 on the lineage to both Latimeria and Taeniopygia). Domain losses (-) are indicated at internodes where they were reconstructed to occur (e.g. loss of M1, L, and M2 in LTK in mammals—highlighted with a gold gradient, and loss of M1, L, and M2 in Tursiops in ALK). Unlabeled internodes all exhibited strong statistical support (Bayesian posterior probability > 0.98), labels on internodes indicate other BPP values. Photos: AD.

228x396mm (72 x 72 DPI)

A.



B.



Zebrafish *AUGα* paralog

