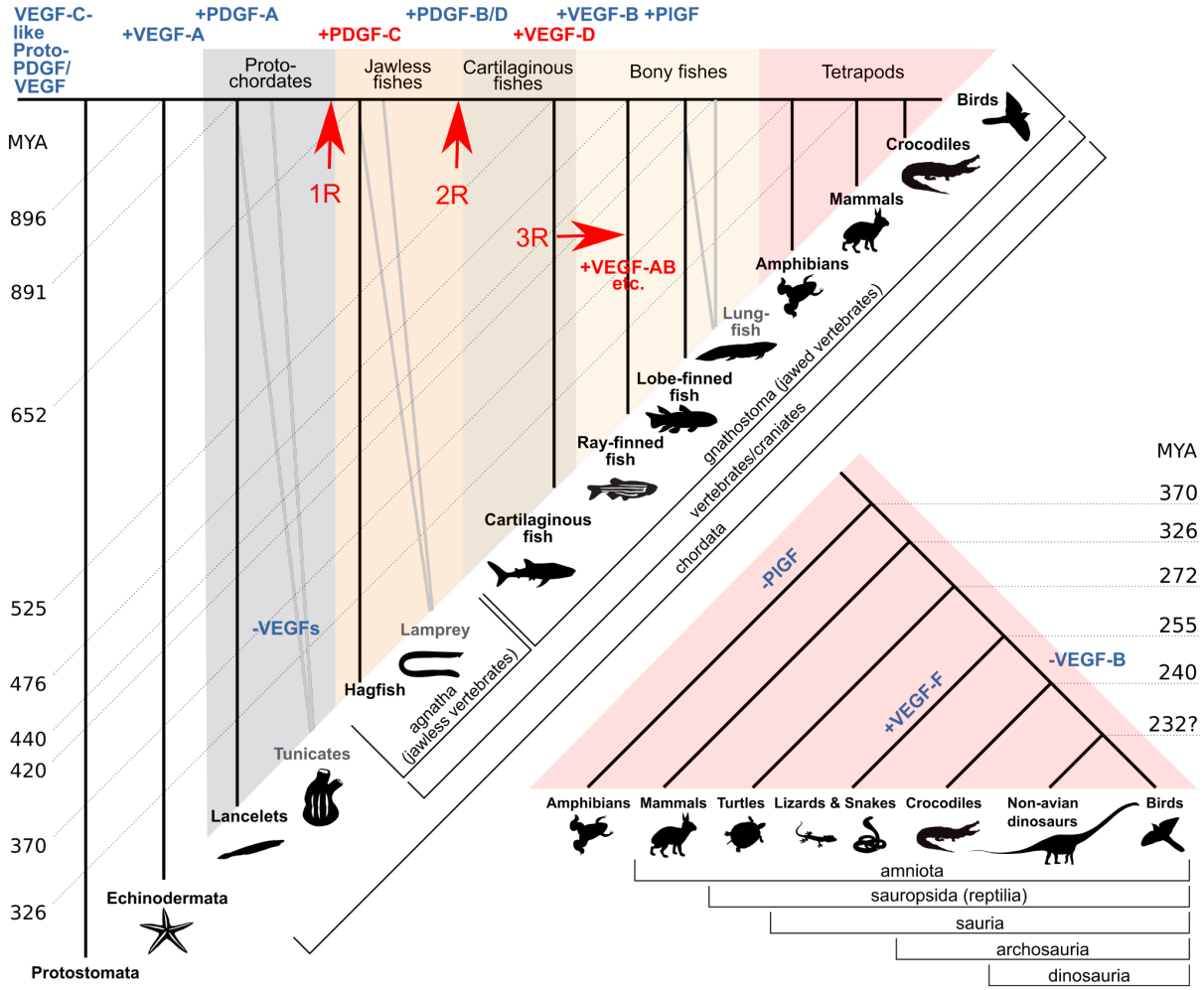
Expansion and collapse of VEGF diversity in major clades of the animal kingdom

Khushbu Rauniyar[1], ORCID 0000-0001-5485-7040; Honey Bokharaie[1], ORCID 0000-0002-3826-7582; Michael Jeltsch[1,2,3,4,*], ORCID 0000-0003-2890-7790

[1]Drug Research Program. Division of Pharmaceutical Biosciences, Faculty of Pharmacy, University of Helsinki, Finland; [2]Individualized Drug Therapy Research Program, Faculty of Medicine, University of Helsinki, Finland; [3]Wihuri Research Institute, Helsinki, Finland, [4]Helsinki One Health, University of Helsinki, Finland

*Author for correspondence: Dr. Michael Jeltsch, Drug Research Program, Faculty of Pharmacy, Biocenter 2, University of Helsinki, P.O.B. 56 (Viikinkaari 5E), 00790 Helsinki, Finland; Phone: +358-2941-25514; Fax: +358-2941-25510; E-mail: michael@jeltsch.org

Graphical abstract



Sources for the graphical abstract:

326 MYA and older (Blair and Hedges, 2005)

272-240 MYA (Benton, 1990)

235-65 MYA (Sereno, 1999)

ABSTRACT (max. 250)

The vascular endothelial growth factor (VEGF) family comprises in vertebrates five or six members: VEGF(-A), PlGF, VEGF-B, VEGF-C, VEGF-D, and – in venomous reptiles – VEGF-F. They fulfill mainly functions for the blood and lymphatic vascular systems. Together with the platelet-derived growth factors (PDGF-A to -D), they form the PDGF/VEGF subgroup among cystine-knot growth factors. Despite an absent vascular system in most invertebrates, PDGF/VEGF-like molecules have been found in, e.g., *Drosophila melanogaster* and *Caenorhabditis elegans*. The evolutionary relationship between PDGF and VEGF growth factors has only been addressed by older analyses, which were limited by the sparse sequencing data at the time. Here we perform a comprehensive analysis of the occurrence of PDGF/VEGF-like growth factors (PVFs) throughout all animal phyla and propose a likely phylogenetic tree. The three major vertebrate whole genome duplications play a role in the expansion of PDGF/VEGF diversity, but several limited duplications are necessary to account for the temporal pattern of emergence. The phylogenetically oldest PVFs likely featured a C-terminus with a BR3P signature, which is a hallmark of the modern-day lymphangiogenic growth factors, VEGF-C and VEGF-D. Some of the younger *VEGF* genes appeared completely absent in some clades, e.g., functional *VEGFB* genes in the clade Archosauria, which includes crocodiles, birds, and other dinosaurs, and *pgf* in amphibians. The lack of precise counterparts for human genes poses limitations, but also offers opportunities for research using organisms that diverge considerably from humans if the goal is to understand human physiology.

Key Words: VEGF, PDGF, phylogeny, VEGF-C, vascular biology, evolution, proto-PDGF/VEGF

Introduction

In biomedical research, model organisms are often used with the ultimate goal to understand the human organism. While the mouse is the most common model organism, other species can have specific advantages. Compared to the mouse, *Drosophila melanogaster* has for example a much shorter generation cycle, and in zebrafish or chicken, embryonic development can be directly and continuously visually observed.

Extrapolating molecular biomedical research results from a given model organism to humans is facilitated when the human proteins of interest have a corresponding counterpart (ortholog) in the model organism. Therefore, the choice of mouse as the most common model organism for biomedical research is understandable since, for many protein families, there is a 1:1 relationship of genes and proteins between *Mus musculus* and *Homo sapiens*. While there are stunning exceptions (e.g. among the kallikrein-like peptidases)(Lawrence et al., 2010), an early estimate found that less than 2% of mouse genes did not have a human counterpart (Mural et al., 2002). We were interested in whether this assumption of a 1:1 relationship holds true for the vascular endothelial growth factor (VEGF) family between humans and frequently used model organisms.

The VEGF protein family

The VEGF protein family is a highly conserved subgroup within the cystine-knot superfamily of growth factors, which share a conserved cystine knot structure, where six conserved cysteine residues are linked by three disulfide bridges such that two bridges form a ring through which the 3rd bridge passes (Vitt et al., 2001). In vertebrates, the VEGFs are primarily involved in angiogenesis and lymphangiogenesis, which are the two basic mechanisms of how blood and lymphatic vessels grow (Mattonet and Jeltsch, 2015; Risau,

1997). The VEGFs signal via the VEGF receptors (VEGFR-1, VEGFR-2, and VEGFR-3), which form a subgroup among the receptor tyrosine kinases (RTKs) which is characterized by seven extracellular Ig-like domains and an intracellular, split kinase domain (Simons et al., 2016). Of all VEGFs, VEGF-A was discovered first and soon shown to be of paramount importance for the development of blood vessels as it was the first-ever gene, for which a heterozygous deletion was found to be embryonically lethal (Carmeliet et al., 1996; Ferrara et al., 1996). Unlike VEGF-A, even the complete loss of the subsequently discovered placenta growth factor (PlGF) and VEGF-B were reasonably well tolerated in mice (Aase et al., 2001; Bellomo et al., 2000; Tayade et al., 2007), and thus VEGF-A has been regarded as the primary, most important VEGF. VEGF-C and VEGF-D, both first described in 1996, form a distinct subset within the VEGF family due to their unique structure and function (Joukov et al., 1996; Orlandini et al., 1996). They feature long, distinct N- and C-terminal propeptides, require multiple proteolytic cleavages for activation, and interact with VEGFR-3, which results in their exclusive ability to directly stimulate the growth of lymphatic vessels in vivo (Jeltsch et al., 1997, 2014; Joukov et al., 1997).

VEGF-E and VEGF-F

Two further VEGF family members have been described: VEGF-E and VEGF-F. VEGF-E is the collective name for VEGF-like molecules encoded by viruses (Ogawa et al., 1998) and VEGF-F denotes a group of VEGF-like molecules that have been identified from the venom of snakes, starting with the aspic viper in 1990 (Komori and Sugihara, 1990; Komori et al., 1999). In the following years, many venomous snakes were shown to feature similar VEGF-like molecules (Yamazaki et al., 2009).

VEGF-E sequences are encoded in the genomes of parapoxviruses and their existence has

been tentatively explained by a single horizontal host-to-virus gene transfer event (Hughes et al., 2010), similar to how the oncogenic v-sis (a homolog of PDGF-B) is thought to have been acquired from its simian host (Doolittle et al., 1983).

Invertebrate VEGFs

In invertebrates, PDGF/VEGF-like molecules have been identified, many of which are referred to as PDGF/VEGF-like growth factors (PVFs), because their exact relationship to the VEGF and PDGF growth factors appeared unclear. Together, the VEGFs and PDGFs form the PDGF/VEGF superfamily. Likely, PDGFs appeared first in the chordate lineage after the divergence from echinoderms (Kipryushina et al., 2015; Singh et al., 1982). Correspondingly, their cognate receptors split before the chordates/tunicates divergence into class III and class V RTKs (Grassot et al., 2006). The fundamental biological change associated with this evolutionary period was the pressurization of the vascular system, and PDGFs are central players in the stabilization of vessels via mural and smooth muscle cells (Hoch and Soriano, 2003). Although many PVFs can be identified from invertebrate genomic sequences, only a few have been subjected to functional analysis, including the *D. melanogaster* PVFs and *C. elegans* PVF-1. In vertebrates, VEGF receptors are expressed by cells of vascular endothelial and hematopoietic lineages, and the molecular integration of the immune and the vascular systems appears to be conserved also in invertebrates (Yan and Hillyer, 2020). A molecular manifestation of this integration is the essential expression of the VEGF receptor-2 (VEGFR-2) by the precursor(s) of both hemopoietic and vascular endothelial lineages (Eichmann et al., 1997). Unsurprisingly, the VEGFR-2 ligand VEGF-C has been shown to be important for various steps in hematopoiesis (Fang et al., 2016, 2020; Thiele et al., 2012).

Opposed to this, no coherent image of the role of PVF signaling in invertebrates has emerged

so far. Three separate studies involve *D. melanogaster* PVFs in immune function (Heino et al., 2001), survival of glia and neural progenitor cells (Read, 2018), and mobilization of storage fat from adipocytes (Zheng et al., 2017), while the *C. elegans* PVF-1 (Tarsitano et al., 2006) functions reportedly as a repressor of Netrin signaling in the patterning of the sensillae of the male tail (Dalpe et al., 2013).

Five studies describe the phylogenetic relationships within the VEGF family of growth factors (Dormer and Beck, 2005; Holmes and Zachary, 2005; Kasap, 2005; He et al., 2014; Kipryushina et al., 2015). However, the studies by Holmes/Zachary and Kasap suffer from a lack of comprehensive data, which was not available in 2005, while the results by Dormer and Beck and He are difficult to parse as they lack sufficient biological context. Thus, we performed a comprehensive analysis of the occurrence of PDGF- and VEGF-like sequences in the animal kingdom and propose, based on our phylogenetic analyses, a likely evolutionary pathway, integrating it with the biological function of the PDGF/VEGF family members.

Results

Coverage

49992 hits were generated using 676 individual blastp searches for homologs of PDGF/VEGF family members. The searches were generated by combining 13 query sequences with 52 animal clades (see Supplementary Figure 1 for the bioinformatics workflow). 8666 of the blast hits were unique. 90.5% of these hits could be programmatically classified as members of the PDGF/VEGF protein family based on explicit manual annotation of the sequence or the PDGF motif (https://www.ncbi.nlm.nih.gov/Structure/cdd/cddsrv.cgi?uid=cd00135). The remaining 9.5% were manually examined and classified. The majority of programmatically unclassified hits

appeared to be homologs of the Balbiani ring 3 protein (BR3P), to which the C-terminal domain of VEGF-C bears a striking homology (Joukov et al., 1996). A very small number of partial sequences were too short to allow classification, in which case they were excluded from further analysis. A summary of the results is shown in Figure 1 and a complete table of all hits is shown as Supplementary Table 1, and an interactive online version of the table is available at https://mjlab.fi/phylo).

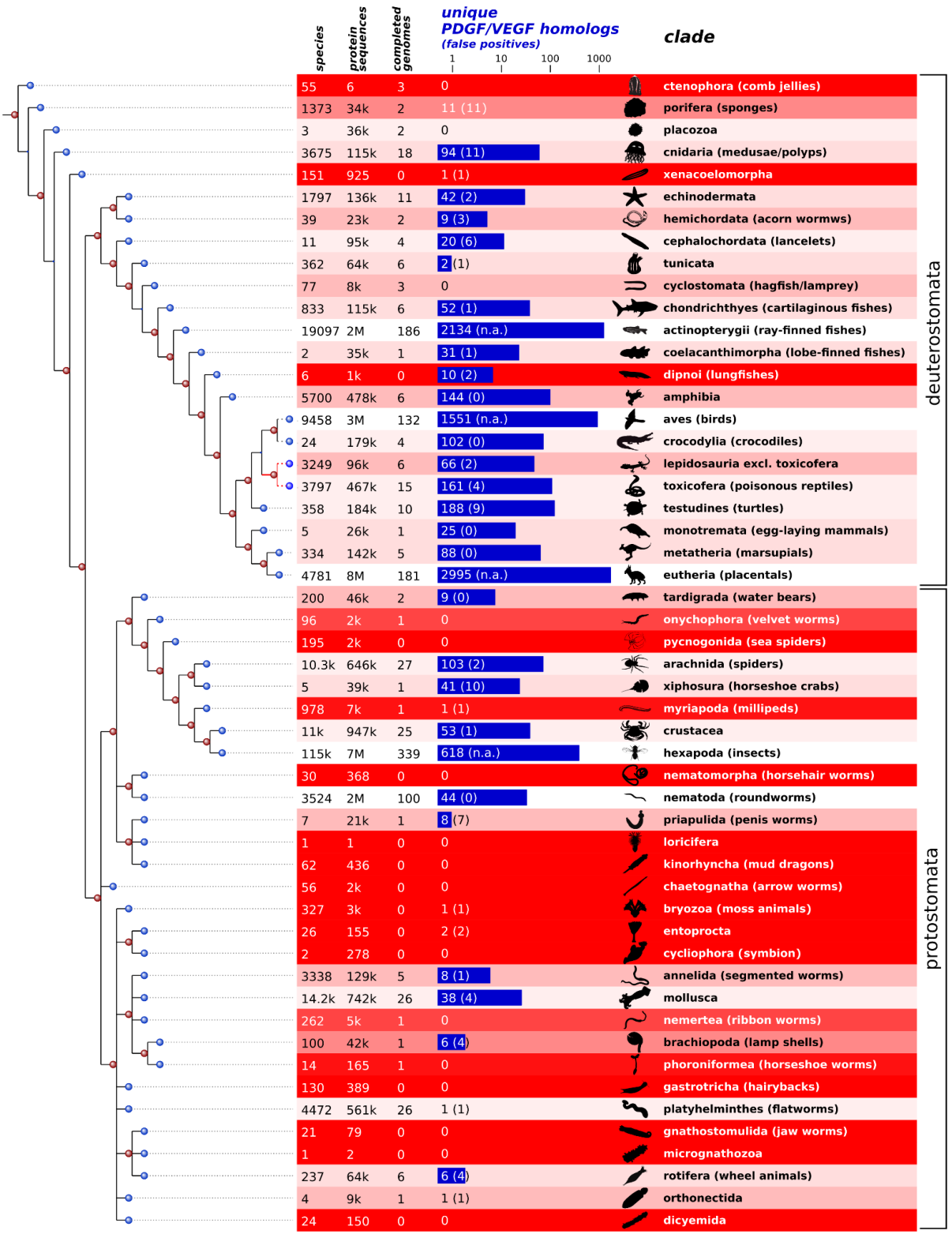


Figure 1. PDGF/VEGF-like blast hits from 52 animal clades and their quantitative representation in the NCBI taxonomy and sequence databases. The number of blast hits from each clade is indicated in blue. The number of false positive hits is indicated in parenthesis. When clades were represented by >500 species, false positives were not manually excluded (n.a.). The darker the red color, the less reliable the analysis results are due to the underrepresentation of the clade in the sequence databases. Most protostome phyla were underrepresented in the NCBI sequence databases.

The current consensus tree of life (Hinchliff et al., 2015) is shown on the left aligned with the clades. Note that the relationship of Lepidosauria to Toxicofera, which might or might not be a clade, is shown with red lines.

Not all of the 52 animal clades were equally well represented in the available sequence databases. In underrepresented clades, the absence of evidence for PDGF/VEGF-like genes was not taken as evidence of absence. To visualize this uncertainty, we used a heuristic formula to indicate the bias in the sampling, which takes into consideration the number of animal species of that clade in the NCBI taxonomy database, the number of sequenced genomes, and the total number of protein sequences for the clade available from the NCBI databases.

PDGF/VEGF-like proteins of the least complex organisms resemble VEGF-C

The least complex animals, where PDGF/VEGF-like proteins were identified are the Cnidaria (which include mostly medusae and corals). While 11 blast hits were from Porifera (sponges), which are less complex compared to Cnidaria, all of these were manually identified as false positives (five of these were genuine BR3P or BR3P-like proteins, see Supplementary Table 2). From the known approximately 115,000 cnidarian protein sequences, 72 were identified as VEGF-like, including the previously described “VEGF” from the marine jellyfish *Podocoryne carnea* and fresh-water polyp *Hydra vulgaris*) (Krishnapati and Ghaskadbi, 2014; Seipel et al., 2004; Turwankar and Ghaskadbi, 2019). When we analyzed their amino acid sequences and compared them to modern-day VEGFs, they appeared more similar to the modern VEGF-C than to VEGF-A, -B, or PlGF. Similar to VEGF-C and VEGF-D, all but one of these contained the characteristic BR3P motif repeats C-terminally to the VEGF homology domain (VHD) (see Figure 2). Cnidarian VEGFs feature typically four BRP3

motif repeats after the VHD. Like *Drosophila melanogaster* PVF-2 and *C. elegans* PVF-1, they frequently lack one or both of the cysteine residues which form the intermolecular disulfide bonds in the mammalian PDGFs/VEGFs. When human VEGF homologs are included in the generation of a phylogenetic tree, they cluster into one branch indicating that PVFs likely originate from a single VEGF precursor gene in the genome of the most recent common ancestor of Vertebrata and Cnidaria (Figure 3). In seven out of the 18 gene-annotated Cnidaria genomes, VEGF-like sequences could be identified (*Acropora digitifera, Exaiptasia pallida, Hydra vulgaris, Nematostella vectensis, Orbicella faveolata, Pocillopora damicornis*, and *Stylophora pistillata*).

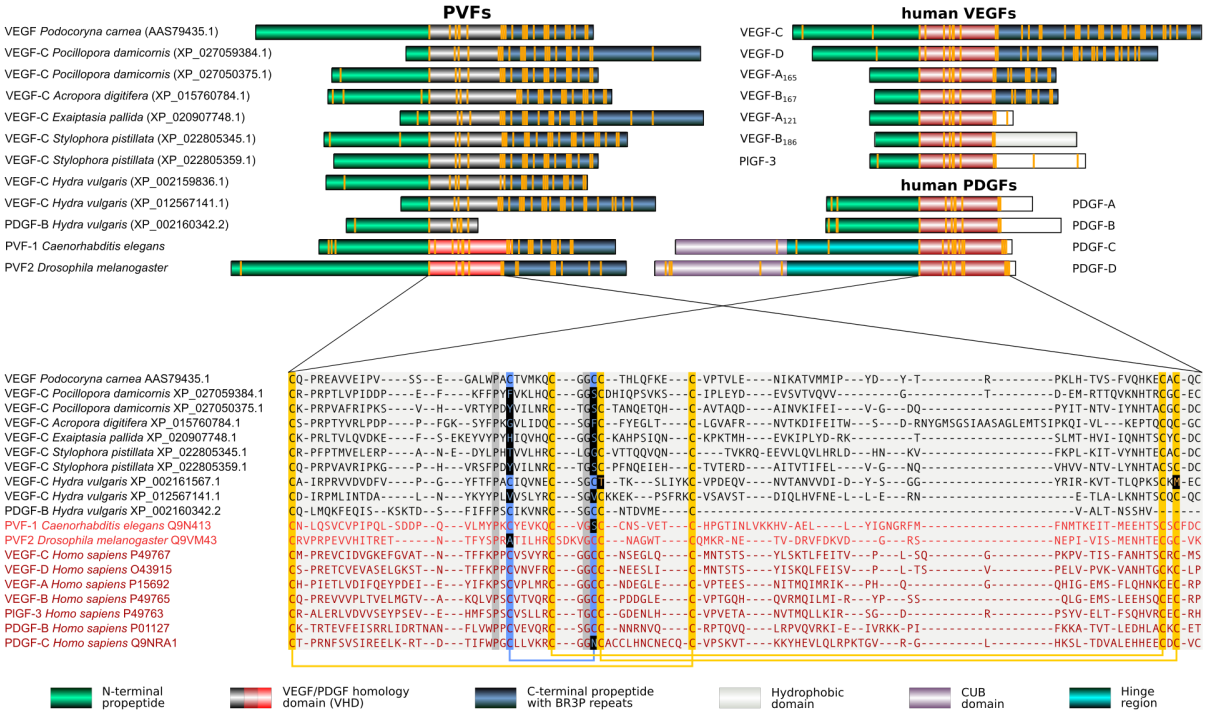


Figure 2. Comparison of the domain structure and alignment of the VEGF homology domain of Cnidaria and human PDGFs/VEGFs. 10 representative cnidarian VEGFs (sequences in black) were aligned with human VEGFs, selected PDGFs, *D. melanogaster* PVF2, and *C. elegans* PVF-1. Most cnidarian VEGFs show a typical cystine knot followed by three to five BR3P motifs similar to the human VEGF-C/D. BR3P motifs are completely absent from mammalian PDGFs and PlGFs, while the longer VEGF-A isoforms and the VEGF-B_{167} isoform contain one complete ($CX_{10}CXCXC$) and one incomplete ($CX_{10}CXC$) BR3P repeat C-terminally to the VEGF homology domain.

In the alignment, the cysteines of the cystine knot are shown on orange background and the cysteines forming the intermolecular disulfide bridges on blue background. The intermolecular disulfide bridges

appear frequently absent in invertebrate VEGFs (marked by inverted coloring), but some of this might be an artifact of the alignment. 100% conserved non-cysteine residues are shown on dark grey background. The bridging pattern of the canonical PDGF/VEGF cysteines is indicated by connecting lines below the alignment.

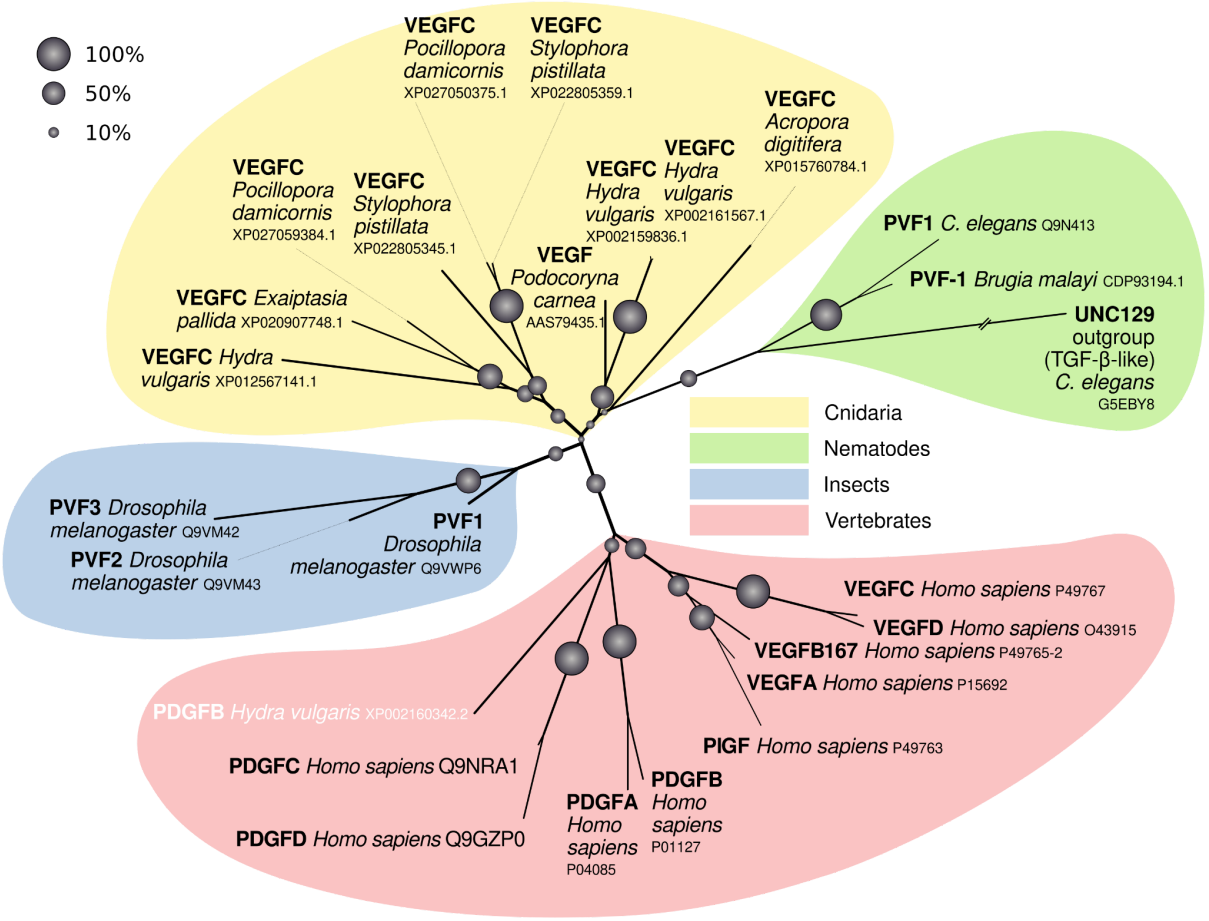


Figure 3. Vertebrate versus invertebrate VEGFs. A phylogenetic tree was calculated from an expanded set of sequences, aligning mostly the PDGF/VEGF homology domains. In this unrooted tree, *Podocoryna carnea* VEGF groups clearly together with all the other Cnidaria VEGF-C-like sequences, the exception being *Hydra vulgaris* PDGF-B, which consistently groups with the mammalian PDGF-C/D group, with which it shares also other characteristics such as the near-complete lack of C-terminal sequences beyond the PDGF/VEGF homology domain. The confidence into branches is given as bootstrap values (% from 1000 repeats).

Distribution of different VEGFs in the deuterostome branch

We did find VEGF-A-like proteins in Echinodermata, Cephalochordata, and Tunicata, all of

which are clades in the deuterostome branch of the animal kingdom. However, all these animals - with the exception of the Tunicata - contained also VEGF-C-like proteins. Among all tunicate sequences, including the six completed tunicate genomes, only one PDGF/VEGF-like gene could be identified. The amino acid sequence of the predicted corresponding tunicate gene product showed a close homology to VEGF-A.

Not being a formal taxonomic group, there is considerable heterogeneity among fish. In bony fish (Osteichthyes), all five mammalian VEGFs (VEGF-A, PlGF, VEGF-B, VEGF-C and VEGF-D) are ubiquitous. Our data shows that, compared to bony fish, the cartilaginous fish (Chondrichthyes: sharks, rays, skates) lack PlGF and VEGF-B, and the jawless fish (Cyclostomata: lamprey and hagfish) lack additionally VEGF-D. When plotted along the branches of the phylogenetic tree of the animal kingdom, there is an overall expansion of VEGF diversity over time, while a few major branches undergo a collapse (Figure 4). The collapse coincides with a reduction in body plan complexity in the case of the Tunicata, which have either reduced the number of VEGF paralogs by eliminating VEGF-C-like sequences (*Ciona intestinalis*) or functional VEGF genes altogether (all other tunicates analyzed). However, a similar reduction of body plan complexity is not seen for the clade Archosauria with its extant members (birds and crocodiles), in which we did not find any signs of functional *VEGFB* genes. The same holds true for Amphibia, in which we did not find any functional genes coding for Placenta growth factor (*pgf*). While we also did not detect any *VEGFB* in Monotremata or any PDGF/VEGF-like proteins in Xenacoelomorpha, we did not consider these findings significant due to the very incomplete sampling of these phyla in the sequence databases.

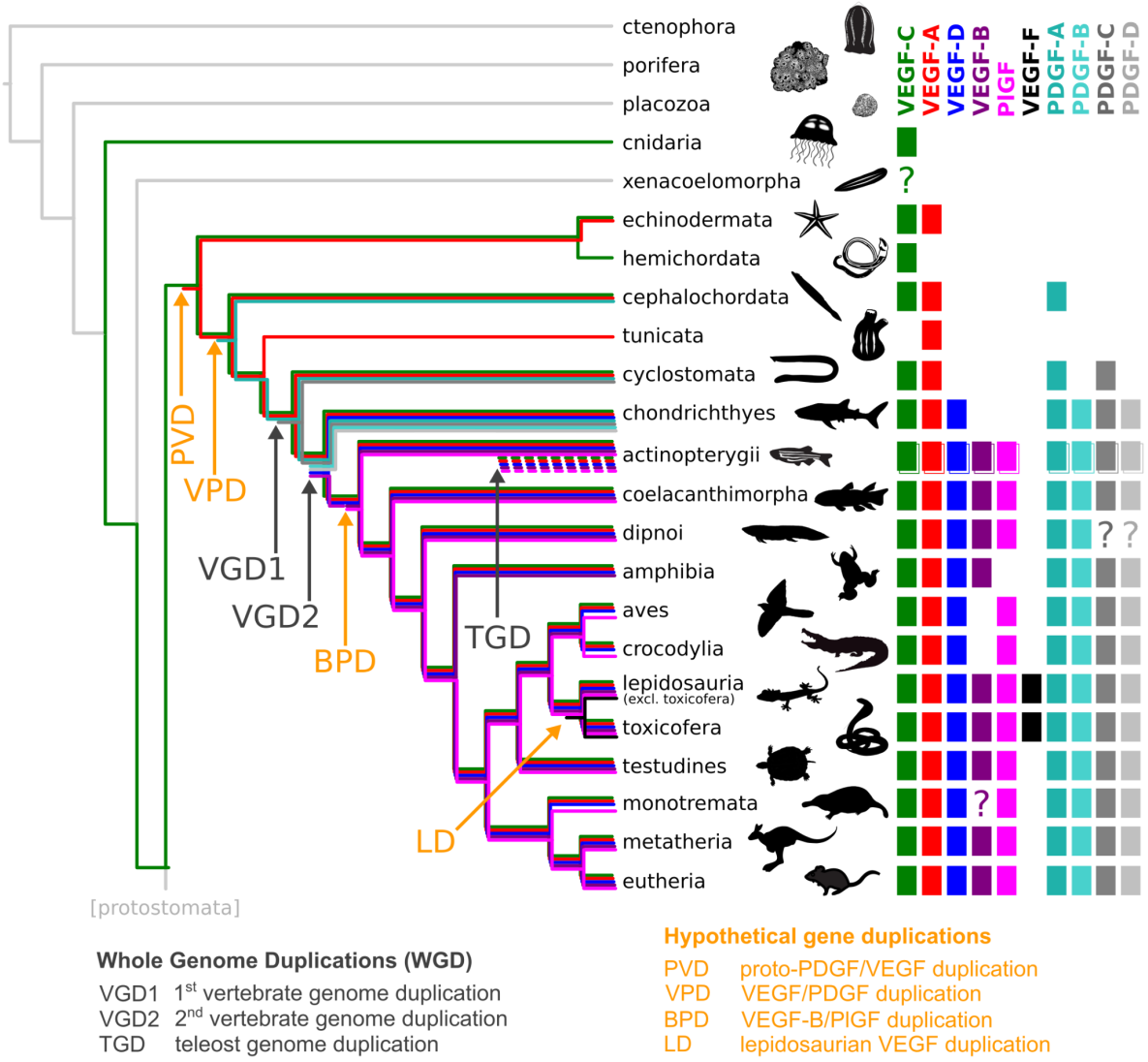


Figure 4. Occurrence of PDGF/VEGF genes in genomes of extant animal clades (excluding protostomes). Inferring from the occurrence of VEGFs in extant animal species, the first VEGF-like protein appeared prior to the deuterostome/protostome split (DPS) and was likely most similar to the modern VEGF-C. The earliest expansion (proto-VEGF-C duplication, CD) appears after the DPS, but prior to the known first vertebrate genome duplication (VGD1). One of the duplicated proto-VEGF-C underwent partial removal of its C-terminal domain, resulting in a VEGF-A-like protein. The subsequent expansion of the VEGF family likely results from VGD2 giving rise to proto-VEGF-D, and the ancestor of VEGF-B/PlGF. The fact that many mammalian genes have two orthologs in zebrafish results from the teleost genome duplication (TGD). Separate, limited gene duplication events (orange arrows) explain the emergence of PlGF and VEGF-B after the VGD2-duplication of proto-VEGF-A (VEGF-B/PlGF duplication, BPD) and, more recently, of VEGF-F (Lepidosauria duplication, LD). Due to gene loss, some clades appear devoid of all or some VEGFs (VEGF-C: Tunicata, PlGF: Amphibia, VEGF-B: Aves, Crocodylia). In tunicates, the gene loss concurs with a massive reduction in morphological complexity, but not in amphibians, birds, and crocodiles. The separation of VEGFs and PDGFs (VEGF/PDGF duplication, VPD) likely happened soon after the first duplication of the VEGFs as already Cephalochordata feature a single PDGF-like gene. Note that the absence of all VEGFs in Xenacoelomorpha, VEGF-B in Monotremata, and PDGF-C/D in Dipnoi could be due to the underrepresentation of these clades in the available sequence data. For reasons of

clarity, the figure does not show a) the PDGFs lines on the branch leading to mammals starting from the VGD2 (since they are consistently present on that branch), b) the Salmonid Genome Duplication (SaGD).

Whole-genome duplications only partially explain the expansion of the PDGF/VEGF family

Whole-genome duplications (WGD) have contributed significantly to the increasing complexity of gene families and vertebrate evolution (Dehal and Boore, 2005; Kasahara, 2007). When we overlayed the established WGDs and the emergence of novel PDGF/VEGF paralogs on the phylogenetic tree of the animal kingdom (Figure 4), the emergence of novel PDGF/VEGF family members coincides only partially with the proposed timing of the three WGDs. The first PDGF/VEGF ("Proto-PDGF/VEGF") appears in the tree before the deuterostome/protostome split (DPS), but diversification is likely to have happened only after the DPS, since the protostome branch lacks the diversification pattern seen in the deuterostome branch. In the deuterostome branch, the first diversification happened likely at the protochordate stage, prior to the first vertebrate whole genome duplication (VGD1). PVFs from species in the protostome branch of the tree (insects, nematodes) do not show a clear diversification into distinct subgroups as can be seen in the vertebrate lineage, where PDGFs, angiogenic VEGFs (VEGF-A, -B, and PlGF), and lymphangiogenic VEGFs (VEGF-C and -D) have formed distinct subgroups. Already species that diverged soon after the DPS (e.g. Echinodermata and Cephalochordata) feature two distinct VEGF homologs, that resemble the shorter, angiogenic VEGFs (VEGF-A-like) and longer, lymphangiogenic VEGFs (VEGF-C-like).

The VGD1 and VGD2 occurred within the direct line to mammals, while the third significant WGD event took place in the common ancestor of the teleost fish lineage (Hughes et al.,

2018), to which also zebrafish (*Danio rerio*) belongs. Although the VGD1 resulted in a duplication of the *proto-VEGFA* and *proto-VEGFC* genes, resulting in four VEGF homologs, we did not find any evidence that any of the duplicated *VEGF* genes permanently became established in the genome. VGD2 likely gave rise to the VEGF-C/VEGF-D subfamily by duplication of the *proto-VEGFC* gene, and to the VEGF-A/PlGF/VEGF-B subfamily by duplication of the *proto-VEGFA* gene. The emergence of three VEGF-A-like VEGFs shortly after VGD2 and the emergence of VEGF-F are most parsimoniously explained by limited duplications in the common ancestor of all Actinopterygii and the common ancestor of all Lepidosauria, respectively.

PDGFs and VEGFs in fishes

Because of recent discoveries in the developmental pathways of the fish vasculature, we wanted to know how successfully PDGF/VEGF ohnologs (WGD-generated homologs) withstood inactivation/pseudogenization, and whether PDGF/VEGF gene duplications can also be found in clades outside the teleost lineage. Thus, we analyzed the RNAseq data from all fish species present in the FishPhylo database (Pasquier et al., 2016) (Figure 5). Despite significant heterogeneity, zebrafish ohnologs of PDGFs/VEGFs could be identified in most other fishes with the expected exception of the Holostei (bowfin and spotted gar). Two notable exceptions in the teleost lineage are *pgfa* and *vegfbb*, for which we did not find a single mRNA contig, and *pdgfba*, which seems to have been lost in five out of six salmonid species. Salmonids, on the other hand, show clear signs of having maintained many of their other PDGF/VEGF ohnologs originating from the Salmonid Genome Duplication (SaGD), most notably VEGF-A ohnologs.

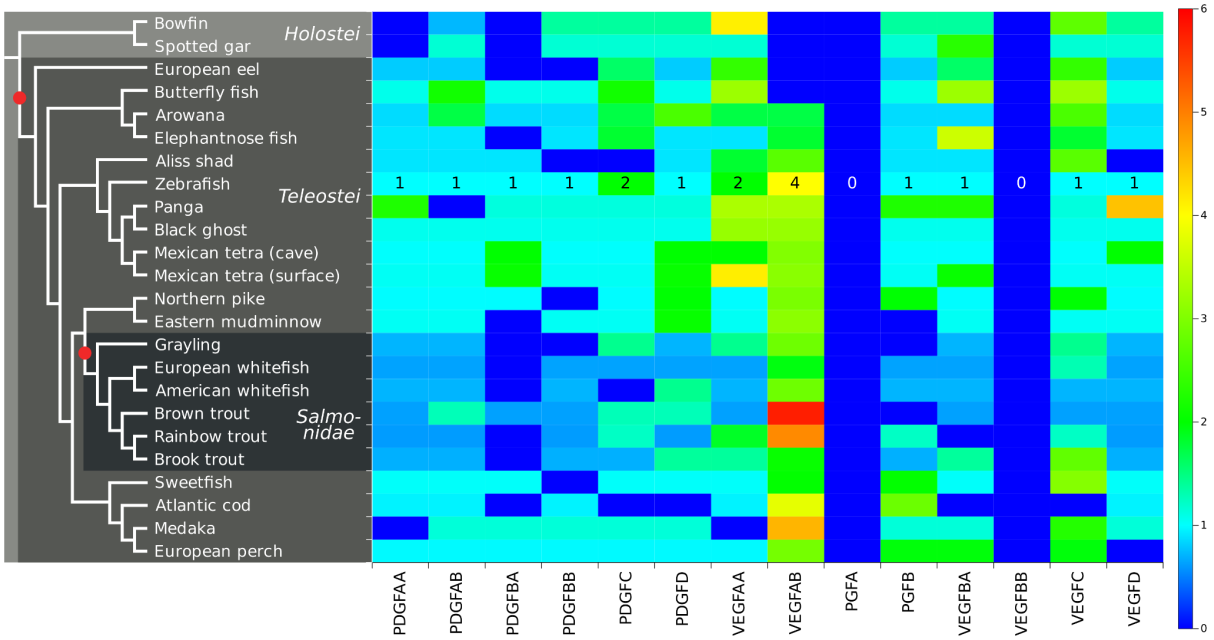


Figure 5. Number of PDGF/VEGF mRNA transcript contigs in 24 fish species as determined by RNA sequencing. While Holostei fishes did not undergo the Teleost Genome Duplication (TGD), they nevertheless feature individually duplicated genes such as *vegfc*. After the TGD, salmonids did undergo one additional round of genome duplication (Salmonid Genome Duplication, SaGD), and consequently, feature - compared to humans - up to 4 times as many genes for each individual PDGF/VEGF, resulting theoretically in up to 36 different *pdgf/vegf* genes. About half of all duplicated genes have been pseudogenized after SaGD (Berthelot et al., 2014; Macqueen and Johnston, 2014), therefore the expected number of functional *pdgf/vegf*-like genes in salmonids is about 20, which is in line with the experimentally determined numbers of 14-22 (Supplementary Table 3). The heatmap has been normalized to zebrafish mRNA transcript numbers in order to compensate for the differences in the total number of mRNA transcript contigs obtained for each species. Whole genome duplications are shown as red dots on the cladogram on the left.

We found VEGF-C duplicated in the fish clade Holostei, which diverged from teleosts before the teleost genome duplication. Only eight extant species comprise the extant Holostei lineage (the bowfin and seven gar species). All three fully sequenced genomes (*Amia calva*, *Atractosteus spatula*, and *Lepisosteus oculatus*) feature a duplicated *vegfc* gene indicating a single gene duplication early in the Holostei lineage (Figure 6A). When testing whether the *vegfc* genes in these fishes are still under purifying selection (i.e. whether gene inactivation or mutation is detrimental), we detected strong pervasive purifying selection throughout the

coding region, with the strongest conservation in the receptor binding domain, followed by the silk homology domain (SHD) (Figure 6B). Individual gene duplications were found to be most common for VEGF-C, but are occasionally also observed for other genes such as VEGF-A (Figure 6C).

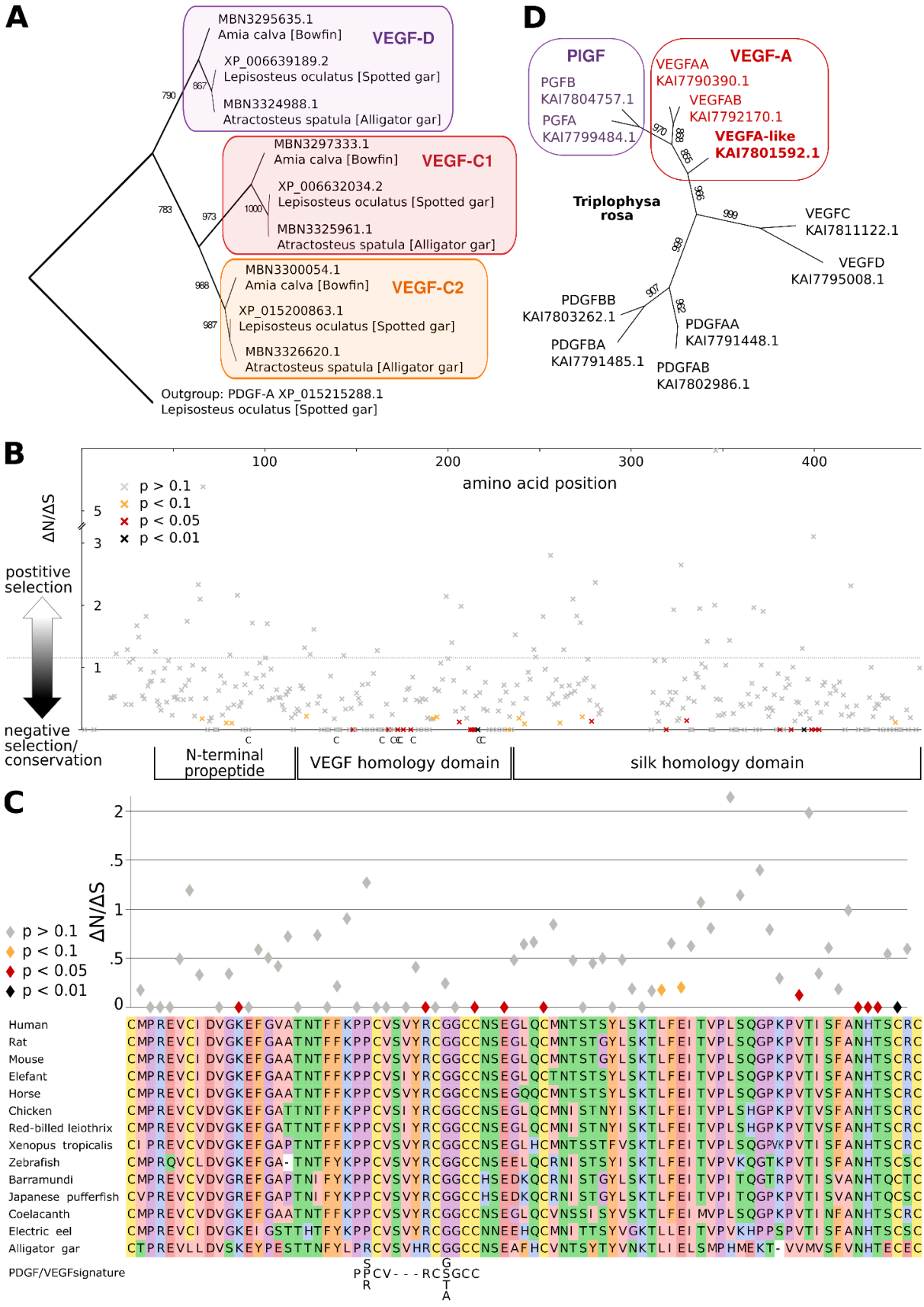


Figure 6. Individual gene duplications contribute to VEGF diversity in fishes. (A) *vegfc* is duplicated in all Holostei species. (B) Holostei *vegfd* and *vegfc* genes are under purifying selection, most notably in the VEGF homology (VHD) and silk homology domain, where there appears to be

strong pressure to maintain the cystine-knot structure. Cysteine residues of the VHD are indicated below the x-axis. (C) Detailed view on the conservation of the VHD, aligned with representative VEGF-C amino acid sequences and the PDGF/VEGF signature. Please note that sites without any synonymous codon changes (ΔS = 0) are not displayed. (D) *vegfa* is also sometimes individually duplicated as shown for *Triplophysa rosa*.

PDGFs/VEGFs are not absent in jawless fishes, but might be absent in parasitic Cnidaria

As shown in Figure 5, individual gene duplications or gene losses were not uncommon in fishes, but a complete apparent absence of PDGFs/VEGFs was seen in the only two extant jawless fishes (Cyclostomata), sea lamprey and hagfish. Based on our previous results, we reasoned that the lack of cyclostomate PDGF/VEGFs in protein databases must be an artifact, perhaps due to a failure of gene prediction or annotation. In fact, manual inspection of the Ensemble (Cunningham et al., 2019) sea lamprey (*Patromyzon marinus*) genome shows - as predicted by our phylogenetic tree - four PDGF/VEGF-like sequences: ENSPMAG00000005730/S4RMA5, ENSPMAG00000006270/S4RNX3 (both VEGF-C-like proteins), and ENSPMAG00000001751/S4R9P2 and ENSPMAG00000007661/S4RT82 (both VEGF-A-like proteins). A very similar setup was found for the hagfish (*Eptatretus burgeri*): one VEGF-A-like gene (ENSEBUT00000003451), a VEGF-C-like gene (ENSEBUT00000000354), and two PDGF-like genes (ENSEBUT00000005231 and ENSEBUT00000023264).

Since we had identified multiple PDGF/VEGF-like proteins in seven Cnidaria species, we reasoned that the lack of these in other Cnidarians could also be an artifact. For in-depth inspection, we chose *Thelohanellus kitauei*, which is a genome-sequenced cnidarian represented with 14792 protein sequences in the Uniprot database. We were not able to identify any genes coding for PDGF/VEGF-like proteins in its genome using very relaxed

degenerate string searches. However, *T. kitauei* belongs to the endoparasitic myxozoa branch of Cnidaria, which is characterized by a reduction in genome size and gene depletion (Chang et al., 2015).

Both VGD1 and VGD2 contribute to PDGF expansion

Kipryushina et al. place the emergence of PDGF after the divergence of Echinoderms and Chordates (Kipryushina et al., 2015), notwithstanding early reports of PDGF/PDGFR signaling in sea urchins (Ramachandran et al., 1995). Concurring with Kipryushina, all PDGF/VEGF-like proteins identified from the known 11 sea urchin genomes are highly homologous to the proto-VEGF-C that we found in Cnidaria, and we found the first bona fide PDGF in Cephalochordata (lancelets). Lancelets feature a single PDGF-like growth factor (XP_002587789.1) in addition to two VEGFs (XP_019625376.1, XP_019621834). While the Cephalochordate proto-PDGF was likely duplicated by VGD1 since Cyclostomata (exemplified by hagfish) already feature orthologs for both PDGF-A (ENSEBUG00000006146) and PDGF-B (ENSEBUG00000014753), another limited gene duplication event is needed to explain the occurrence of PDGF-C (ENSEBUG00000009744) before VGD2, which caused the emergence of PDGF-D by duplicating proto-PDGF-C.

VEGF-F can be found in several Lepidosauria, not only in venomous snakes

In 1999, it was recognized that the hypotensive factor from the venom of *Vipera aspis* (Komori and Sugihara, 1990) is in fact a VEGF-like molecule (Komori et al., 1999). Our analysis shows that VEGF-F is not limited to venomous snakes but is also found in non-venomous snakes (e.g. *Python bivittatus*, XP_025024072.1). Amino acid sequence alignments of VEGF-Fs show a similar high homology to both VEGF-A and PlGF, and based

on phylogenetic trees it appears likely that either VEGF-A or PlGF served as a template for VEGF-F (data not shown). Because we identified VEGF-F orthologs also in lizards, e.g. in the common wall lizard (*Podarcis muralis*, XP_028597744.1) or the gekko (*Gekko japonicus*, XP_015284783.1), the gene duplication likely happened early in the Lepidosauria lineage.

Viral VEGFs

To analyze the relationship between all available viral VEGF sequences (VEGF-E), we constructed a phylogenetic tree of all VEGF-E sequences that we identified in the main analysis (see Figure 7 for the simplified tree and Supplementary Figure 2 for the complete tree). Sequences coding for VEGF-like genes were found in the genomes of at least four different virus clades: orf virus (ORFV), pseudocowpoxvirus (PCPV), bovine pustular stomatitis virus (BPSV), and megalocytivirus (MCV). ORFV, PCPV, and BPSV are known to have been collectively infecting at least 10 mammalian species, while MCVs have been detected in at least eight fish species. The most parsimonious phylogenetic tree suggest that VEGF-Es originate from VEGF-A. We did not find any evidence of recent multiple host-to-virus gene transfers as all vertebrate VEGFs formed tight clusters, which were well separated from the VEGF-E clusters. We did not observe a separation of the host species with the phylogeny. However, only the ORFV cluster contains enough sequences to allow for any separation to become apparent.

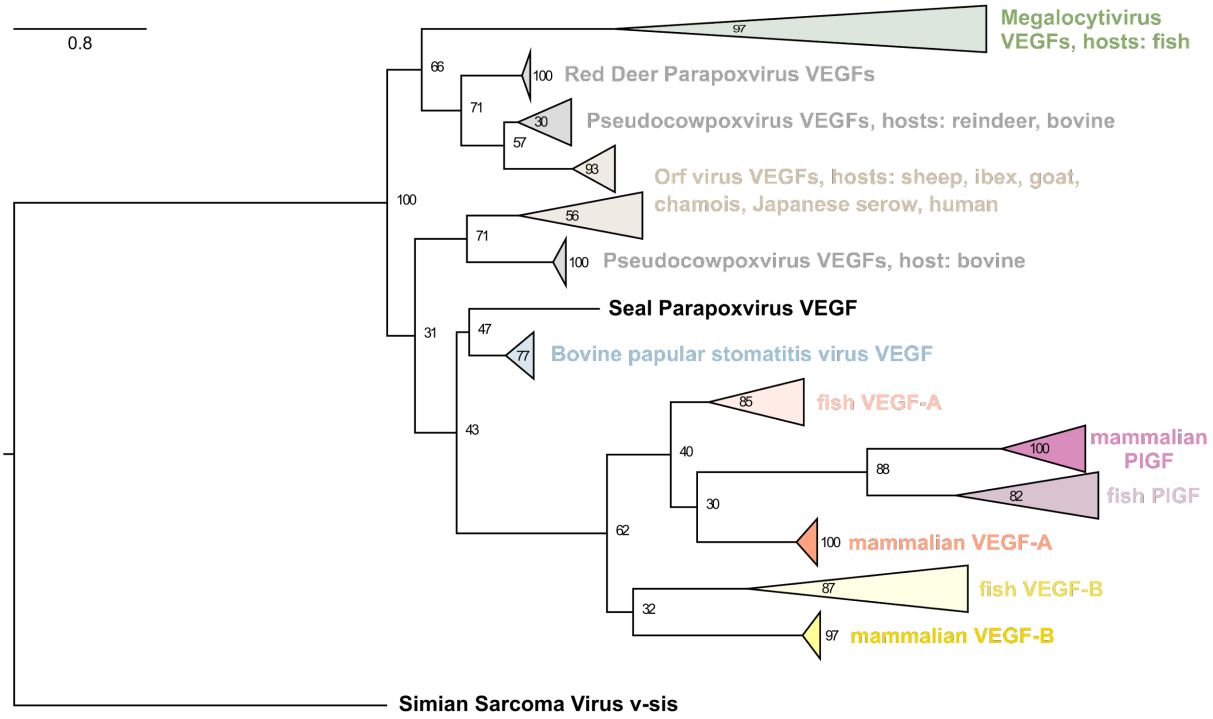


Figure 7. Phylogenetic protein tree of VEGF-E. Without exception, all VEGF-E and all non-viral VEGF sequences cluster together arguing that none of the known VEGF-Es originates from recent host-to-virus gene transfer events. The protein tree is compatible with a single origin of all VEGF-E, but due to the significant distance, convergent evolution cannot be excluded. Based on the VEGF sequence alone, assignment to the Pseudocowpoxvirus or Orf virus group is not possible, since several VEGF sequences derived from different viruses are identical (e.g. reindeer PCPV VEGF is identical to the VEGF sequence from the PCPV reference genome VR634). In this branch of the tree, cross-species transmissions have been reported, including to humans (Huemer et al., 2014). The expanded tree with all leaves is shown in Supplementary Figure 2.

PDGF/VEGF-like molecules of the protostome branch

The sequence coverage of the protostome branch of the animal kingdom was much weaker than that of the deuterostome branch. For 16 out of 29 invertebrate phyla, we could not find a single genomic draft sequence, and almost all available genomic data covered only six phyla: flatworms, roundworms (nematodes), molluscs, crustaceans, insects, and spiders. PDGF/VEGF-like molecules were identified in all these six phyla except for flatworms. Less than half of the 100 completed nematode genomes contained PDGF/VEGF-like genes.

Among these are notably many *Caenorhabditis* PVFs, like *C. elegans* PVF-1 (NP_497461.1), but the majority were found from parasitic nematodes, including intestinal parasites like *Ancylostoma duodenale* (KIH56282.1), lymphatic parasites like *Brugia malayi* (CDP93194.1), and conjunctival parasites like *Loa loa* (XP_003142823.1). While nematodes and flatworms do not possess a blood circulation, insects, spiders, and crustaceans feature a so-called "open circulation", where the hemolymph is circulated inside a body cavity (hemocoel). PDGFs/VEGFs were also found in molluscs and segmented worms (*Annelida*), some of which do feature a closed circulatory system.

Conservation of the VEGF sequences

Except for VEGF-A and VEGF-C, the physiological and pathophysiological role of many VEGF family members is still under debate. Some of them (*Vegfb* and *Vegfd*) can be deleted in mice without any major phenotype (Aase et al., 2001; Baldwin et al., 2005; Bellomo et al., 2000). To compare the evolutionary pressures acting on different VEGFs, we analyzed the conservation of their coding sequences. We compared nonsynonymous and synonymous substitution rates inferred by a maximum-likelihood approach. The VEGF-C sequence was the most strongly conserved, followed by VEGF-D, while PlGF showed the most variability, while still being very conserved. VEGF-A was overall also very conserved but showed a peak of variability at the very end of the VHD corresponding to loop 3, which is a major carrier of the receptor binding epitopes for VEGF.

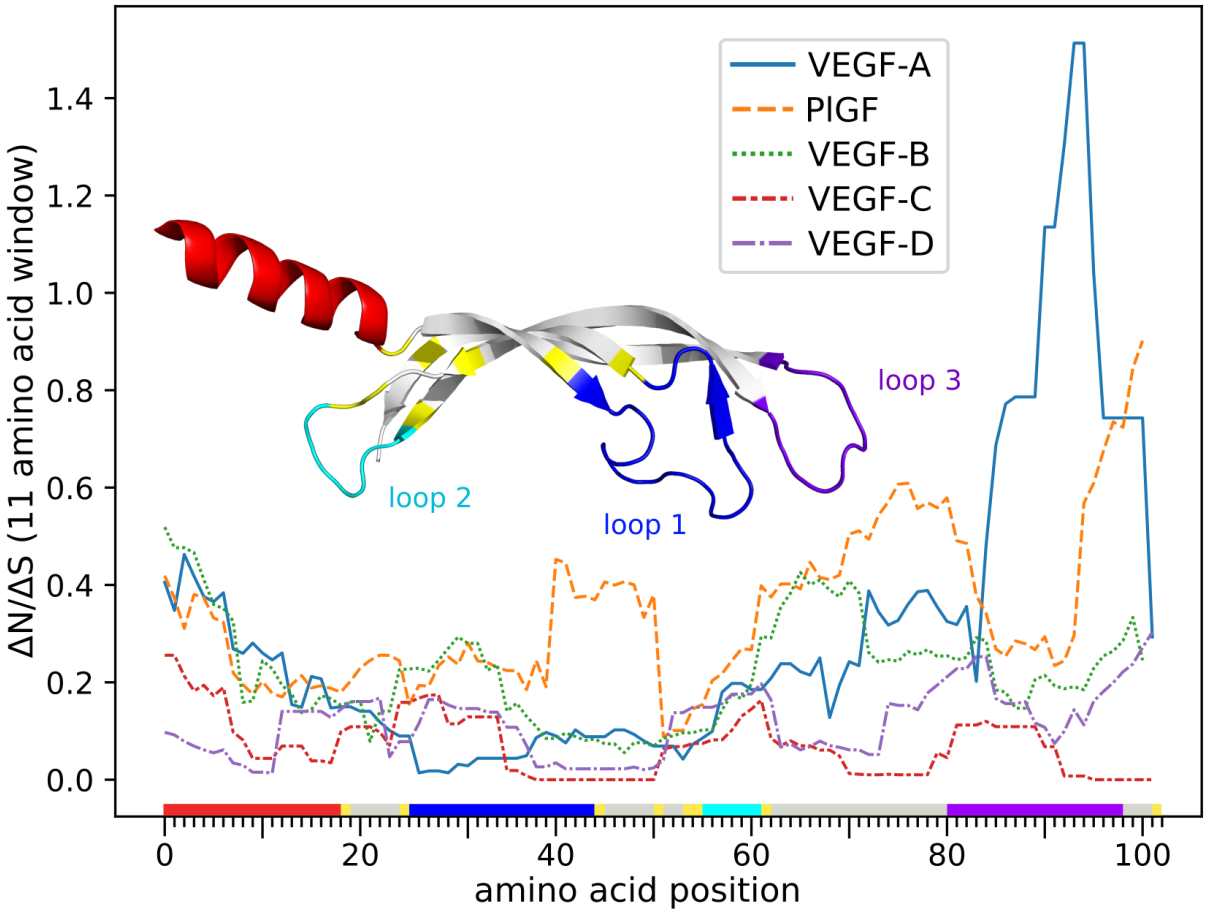


Figure 8. Conservation of the VEGF homology domain. The ratio of non-synonymous to synonymous mutations was inferred for all sites over the whole tree and averaged over an 11-amino acid window. VEGF-C shows the highest conservation while the evolutionary younger family members appear to be more volatile. The amino acid position on the x-axis is color-coded to correspond to the location of the residue in the 3D structure of the monomeric VEGF-C protein (Leppänen et al., 2010). Note, that the peak at the end of the VEGF-A sequence might or might not be an artifact resulting from the difficulty to generate a good alignment for loop 3 (shown in purple). Alternatively, the peak might correspond to adaptations in the receptor interaction, as loop3 carries major determinants for receptor binding (Leppänen et al., 2011; Wiesmann et al., 1997).

Splice isoforms

Splice isoforms are a means of generating diversity at the protein level from a single gene. Within the PDGF/VEGF family, alternative mRNA splicing is very unequally used to generate diversity: the hemangiogenic VEGFs (VEGF-A, PlGF, and VEGF-B) all feature at least two major splice isoforms, while the lymphangiogenic VEGFs (VEGF-C and VEGF-D) seem to generate all of their protein diversity post-translationally by alternative proteolytic processing (Künnapuu et al., 2021). The Uniprot protein entry for human VEGF-A describes no less than 17 different splice isoforms, eight of them with evidence at the protein level. When reimplementing the DIVAA software (Rodi et al., 2004) in Biopython to quantify the diversity of VEGF sequences, we realized that for genes rich in splice isoforms such as VEGF-A, an accurate assignment of isoforms is paramount since indels are not handled well by alignment and treebuilding algorithms. Using the length of the coding region and comparisons with a reference list of known VEGF-A mRNA isoforms, we programmatically sorted all VEGF-A protein sequences into one of four major buckets, corresponding to the 121-, 165-, 189-, and 206-isoforms. When we counted the number of sequences for these four VEGF-A isoforms, we found a ratio of roughly 9:4:3:1 for the 189, 165, 121, and 206 isoforms, indicating that VEGF-A_{189} might be the predominant VEGF-A isoform in most species.

Discussion

In this study, we have identified homologs of PDGFs/VEGFs in most animal phyla that show tissue organization (i.e. excluding sponges) and for which more than a few genome assemblies and gene prediction exist. Despite their pervasive occurrence in many branches of deuterostomes and protostomes, the data clearly support the notion that some animal phyla

are completely or partially devoid of PDGF/VEGF-like molecules and this might above all apply to clades with secondarily reduced body plans like Tunicata, or for the phyla Xenacoelomorpha or Dicyemida. Especially for many of the protostome phyla, not much sequencing data is available. Genome assemblies are often lacking, and the available prediction algorithms might not be very reliable as these animals are rarely the subject of genomic research. For these phyla, the lack of PDGF/VEGF-like proteins is a provisional hypothesis.

After hypothetical proteins are predicted from genomic sequences, programmatic bioinformatics work-flows typically assign them to homology groups (e.g. using PANTHER (Mi et al., 2013)), resulting in automatic annotation like PREDICTED, VEGF-C. Despite this approach, many PDGF/VEGF homologs fail to be programmatically categorized into one of the ortholog groups (VEGF-A, PlGF, VEGF-B, VEGF-C, VEGF-D, VEGF-F, PDGF-A, PDGF-B, PDGF-C, PDGF-D). Our algorithm emulates crowd-sourcing by comparing uncategorized homologs to the most closely related manually and programmatically annotated PDGFs/VEGFs and establishes a majority opinion, allowing for the categorization of the vast majority into one of the ortholog groups. This crowd-sourcing combines human annotation of gene and protein records with the tree-building and clustering methodology used for the protein trees available at Ensemble (https://m.ensembl.org/info/genome/compara/homology_method.html), which are used for the automatic annotation.

Based on the phylogenetic tree of the animal kingdom and our analysis of PDGF/VEGF homologs in different animal clades, the emergence of the earliest PDGF/VEGF-like molecule (“proto-PDGF/VEGF”) predates the establishment of the bilaterian body plan (Lee

et al., 2013) and the split of the animal kingdom into deuterostome and protostome organisms before the start of the Cambrian about 540 MYA (Erwin and Davidson, 2002). Intriguingly, this "proto-PDGF/VEGF" most likely featured a domain structure characteristic for the modern lymphangiogenic VEGF-C/VEGF-D subclass having long N- and C-terminal extensions flanking the VHD and a characteristic repetitive cysteine residue pattern (the BR3P repeat) at its C-terminus. Concurrent with the relatively rapid evolution of different body plans during what has been termed the "Cambrian Explosion", the proto-PDGF/VEGF undergoes diversification establishing a VEGF-A-like and a VEGF-C-like branch, and spinning off the PDGF branch. The deuterostome/protostome split predates both the VEGF diversification and the PDGF spinoff, which explains the difficulty of classifying PDGF/VEGF-like molecules on the protostome branch ("*Drosophila* VEGFs", "*C. elegans* VEGF-C") as either PDGFs or VEGFs.

In the protostome branch, we can detect PDGF/VEGF-like factors (PVFs) in all but one clade for which substantial sequencing data is available. All 26 genome-sequenced flatworms seem to get along without any PVFs. Contrary to this, insects, molluscs, and segmented worms (Annelida) often feature more than one *Pvf* gene, whereas most nematodes feature only one. *C. elegans* PVF-1 is remarkable in that it is still - after more than 500 MYA of evolutionary separation - able to activate the human VEGF receptors-1 and -2 (Tarsitano et al., 2006). Protostome animals do not feature a cardiovascular system with exceptions among molluscs and segmented worms. The latter phylum contains species such as the earthworm *Lumbricus terrestris*, which features a surprisingly sophisticated closed circulatory system displaying vascular specialization and hierarchy including a heart with valves, larger blood vessels, capillaries, blood containing free hemoglobin, and a renal filtration system (Hickman et al.,

2008).

With PlGF, VEGF-B, and VEGF-F, further specializations happen in both the VEGF-A and the VEGF-C lineage after the Cambrian period in the deuterostome branch. Instrumental for this specialization is most likely the VGD2, which increased the number of PDGF/VEGF-like genes from four to seven. When ignoring teleost fishes, whole genome duplications are responsible for about half of the newly emerging PDGFs/VEGFs, the other half requiring at least four duplication events between the gene and chromosome level (Figure 4).

Echinodermata are the most simple animals that display PDGF/VEGF specialization at the gene level. The most parsimonious explanation for the existence of both VEGF-A-like and VEGF-C-like proteins in Echinodermata is that the first gene duplication of the proto-PDGF/VEGF happened prior to the VGD1. For the same reason, the separation of the PDGF lineage also likely predates the VGD1, resulting in PDGFs being present in cephalochordates (lancelets), which are the most simple organisms to have a pressurized vascular system in which the blood is moved around by peristaltic pressure waves created by contractile vessels (Moorman and Christoffels, 2003). In line with this notion is the important role of PDGFs in the supportive layers that stabilize blood vessels (pericytes, smooth muscle cells)(Hellstrom et al., 1999).

Despite the availability of genome assemblies from six different tunicate species, the programmatic approach identified only one VEGF-like molecule in tunicates, *Ciona intestinalis*. This was surprising since there is prior data indicating the role of VEGF/VEGFR signaling in the circulatory system of tunicates (Tiozzo et al., 2008). However, the relationship between the receptor tyrosine kinase that was cloned from the tunicate *Botryllus schlosseri* and PDGF receptors, FGF receptors and c-kit is not clear. The immunological

detection used antibodies directed against human VEGFRs or VEGF-A, and the TK inhibitor PTK787 might have as well inhibited PDGF receptors and c-kit. Although the *C. intestinalis* VEGF-like growth factor appears most similar to VEGF-A, its phylogeny allows also descendence from the PDGF or VEGF-C branch. In that case, the similarity to VEGF-A might have originated from convergent evolution ("long branch attraction"). Since none of the other tunicate species featured any PDGF/VEGF homologs, horizontal gene transfer could alternatively be considered.

Absence of PlGF and VEGF-B from entire animal vertebrate classes, VEGF-F more common than expected

Most striking was the absence of some VEGF family members from entire animal classes. We did not find any PlGF ortholog in Amphibia and also no VEGF-B ortholog in the clade Archosauria, which includes extant birds and crocodiles as well as the extinct dinosaurs. Since bony fish feature both PlGF and VEGF-B, the absence of VEGF-B in extant Archosauria and PlGF in extant Amphibia likely represents an example of lineage-specific gene loss. While five avian protein sequences are annotated as "VEGF-B" or "VEGF-B-like" in the searched database, they did clearly separate on a phylogenetic tree to the same branch as the VEGF-A sequences (data not shown). In addition, their genes did not show the typical exon-intron structure which is characteristic for VEGF-B with overlapping open reading frames leading to two different protein sequences due to a frameshift (Olofsson et al., 1996).

As a counterpoint to the missing PlGF and VEGF-B, we observed to our surprise that VEGF-F is more common than generally thought. Discovered as a snake venom compound (Komori and Sugihara, 1990; Komori et al., 1999), it was also thought to be limited to venomous reptiles. However, we did detect VEGF-F-like sequences in lizards and gekkos,

and thus might be pervasive throughout large parts of the lepidosaurian lineage. At this moment, it is completely unclear, which functions VEGF-F might fulfill beyond its function as a venom compound, where it acts by accelerating venom spread by inducing vascular permeability and by incapacitating the prey by lowering blood pressure.

Complete absence of PDGFs/VEGFs

The absence of individual PDGFs/VEGFs from a species' proteome can be either real or only apparent due to incomplete sampling or an artifact of the bioinformatics analysis pipeline. We generally found very few exceptions to the clade-specific pattern of PDGF/VEGF occurrence in terrestrial vertebrates, which all featured the same set of PDGF/VEGF paralogs, confirming the reliability of the respective genome sequencing and gene prediction pipelines. In our programmatic screen, PDGF/VEGF-like sequences were apparently completely absent in some clades for two different reasons:

1. PDGFs/VEGFs were apparently absent from clades, where there was no comprehensive genomic data or the genomic data had not been analyzed (e.g. sea spiders or velvet worms).
2. PDGFs/VEGFs were absent from clades that are likely truly devoid of VEGF-like molecules (e.g. flatworms, where a substantial number of genomes has been sequenced and analyzed).

Viral VEGFs

While many viruses indirectly induce angiogenesis (Alkharsah, 2018), some viruses encode their own VEGF homologs. These proteins have been collectively termed "VEGF-E". Viral VEGFs have been reported from parapoxviruses, which cause skin lesions in their respective

mammalian hosts (Inder et al., 2007; Lyttle et al., 1994; Ueda et al., 2003). In these viruses, the VEGF-E gene is specifically responsible for the swelling and vascular proliferation (Savory et al., 2000). Based on the sequence homology to VEGF-A, VEGF-E is believed to have been captured from a host during viral evolution (Lyttle et al., 1994) similar to the v-sis oncogene, which is believed to be derived from captured host PDGF-B sequences (Doolittle et al., 1983). Our database search confirms that viral VEGF homologs exist not only in four species of the parapoxvirus genus but also in the very distantly related megalocytiviruses, which infect fish (de Groof et al., 2015). Surprisingly, despite their non-overlapping host range, both the fish and mammalian viral VEGF-Es are likely to originate from one single acquisition from a mammalian host. Unlike megalocytiviruses, which infect fish (and occasionally amphibians), parapoxviruses have a very broad mammalian host range, which includes occasionally humans, but is mostly seen in domesticated and wild ungulates (Essbauer et al., 2010; Hautaniemi et al., 2010). While parapoxvirus infections are typically self-limiting, megalocytiviruses cause considerable economic damage to aquaculture. The pathophysiology of megalocytiviral diseases is not well understood. Infection leads to perivascular cell hypertrophy (Whittington et al., 2010), and VEGF-E might cause easier virus distribution via an increase in vascular permeability.

The “silk homology” domain

The alignment of the accessory domains of VEGFs is non-trivial since these domains contain a variable number of repetitive motifs and thus the evolutionary history (i.e. which repeats got lost or added in which order) is perhaps impossible to deduce with reasonable accuracy. This holds true most notably for the SHD of VEGF-C and VEGF-D, which consists of several complete and incomplete Balbiani ring-3 protein (BR3P) repeats. The C-terminal tails

of the VEGF-A_{165} and VEGF-B_{167} isoforms show a reduced set of these BR3P repeats and in the case of VEGF-A, this domain has been named "heparin binding domain" (HBD). "Heparin binding" (i.e., binding to the extracellular matrix and cell surfaces) is one function of the SHD (Jha et al., 2017), and the HBDs of VEGF-A_{165} and VEGF-B_{167} have developed a stronger heparin affinity compared to VEGF-C or VEGF-D, which perhaps allowed for the reduced size of the HBD compared to the SHD of VEGF-C. In addition, the SHD of VEGF-C is required to keep VEGF-C inactive after secretion, likely by sterical hindrance (Lackner et al., 2019), which is not a requirement for the longer VEGF-A isoforms, where inactivity might be mediated by sequestration mediated by the HBD (Plouët et al., 1997). However, when ECM-bound VEGF-A_{189} or VEGF-A_{206} are in direct contact with endothelial cells, some signaling seems to be possible (Park et al., 1993). Also, the 165-isoform, which ranges in its ECM association between the soluble 121- and the strongly ECM-bound 189- and 206-isoforms, can signal while being ECM-associated, although the signals are distinct from free-floating VEGF-A_{165} (Chen et al., 2010; Lee et al., 2005). Our analysis shows that the SHD was likely an essential part of the proto-PDGF/VEGF. Given that the ECM-binding SHD is larger than the receptor-activating VHD, it is surprising to see that it has been maintained over several hundreds of million years. Since proteins can be kept inactive by much shorter propeptides, it is tempting to speculate that the SHD must have some additional function beyond keeping VEGF-C inactive. One potentially important function could be the establishment of a VEGF-C gradient.

The better known VEGF-A gradients are believed to result from the interaction of the heparin-binding, longer VEGF-A isoforms with extracellular matrix and cell surface heparan sulfate proteoglycans and are believed to be essential for directed vascularization during embryogenesis (Carmeliet et al., 1999; Stalmans et al., 2002; Ruhrberg et al., 2002; Gerhardt

et al., 2003). Although the VEGF-A_{165} isoform is considered the major isoform in humans (Robinson and Stringer, 2001), the slightly stronger ECM-binding VEGF-A_{189} might be the predominant species in many other animals, as we found a strong dominance of this form in sequence databases. However, an increasing number of the sequences in protein databases are hypothetical predictions from genome sequencing projects. Hence it is difficult to say whether the dominance of the 189 isoform in the database reflects true mRNA dominance in most organisms or whether it results from a greedy bias in the splice prediction algorithm favoring the longer or canonical isoforms. Supporting our hypothesis that in many animals the longer VEGF-A isoforms are dominant is the fact, that isoforms corresponding to the human VEGF-A_{165} (and sometimes also VEGF-A_{121}) cannot be computationally mapped from the genomic data. Such animals include cattle, horses, and many birds (data not shown). While the need for VEGF-A diversification is met in mammals exclusively by mRNA splicing, zebrafish accomplish the same by using its two *vegfa* ohnologs (*vegfaa* and *vegfab*), which therefore have become indispensable (Bahary et al., 2007).

Similarly to VEGF-A, VEGF-C might form gradients by the interaction of its SHD domain with the ECM (Jha et al., 2017), and such morphogenetic gradients might be crucial for embryonic lymphangiogenesis, and thus, indirectly by transdifferentiation, also for the cardiovascular system (Das et al., 2022).

Structural differences between protostome and deuterostome PDGFs/VEGFs

PDGF/VEGFs form a family within the superfamily of cystine knot growth factors. Their hallmark is a characteristically spaced pattern of eight cysteine residues, consisting of the 6-cysteine pattern of the cystine knot signature expanded by two cysteines which are responsible for the covalent dimer formation of PDGFs/VEGFs (Iyer and Acharya, 2011).

The 8-cysteine pattern is broken with respect to the intermolecular disulfide bond-forming cysteine by only one vertebrate member of the PDGF/VEGF family, namely PDGF-C. However, in protostomes, missing intermolecular disulfide bonds are rather the rule than the exception (Figure 2). While disulfide bridges increase thermostability, low ambient water temperatures are typical for many marine and freshwater species, for which covalent dimer formation via disulfide bonds might not have any advantage over noncovalent dimer formation. Even at 37°C, the cystine bridge is not strictly necessary for dimer formation: VEGF-C forms also noncovalent dimers (Joukov et al., 1997), and stable VEGF-A can be produced also after the mutation of the intermolecular cystine bridge-forming cysteines (Muller et al., 2002).

Conserved when present, not needed when absent

Differently from VEGF-A and VEGF-C, which are pervasively maintained within the vertebrate lineage, PlGF and VEGF-B are absent from major vertebrate classes. PlGF appears to be absent in cartilaginous fishes and amphibians, and VEGF-B in cartilaginous fishes, birds, and crocodiles. The gene duplication that led to the establishment of the PlGF and VEGF-B genes happened only in the lineage leading to ray-finned fishes (Actinopterygii), after the cartilaginous fishes had already branched off. Different from this, the absence of PlGF and VEGF-B in amphibians, birds, and crocodiles is due secondary gene loss events, which were - very similar to knockout experiments of the same genes in mice (Aase et al., 2001; Bellomo et al., 2000; Tayade et al., 2007) - well tolerated. While VEGF-B plays a role in the regulation of endothelial fatty acid uptake (Hagberg et al., 2013) and vascularization and tissue perfusion via indirect activation of VEGFR-2 (Anisimov et al., 2013), its precise role remains controversial (Moessinger et al., 2020; Zafar et al., 2017). Its evolutionary loss

might have been a net benefit for birds, perhaps even been instrumental to enabling the high metabolic turnover needed for flight (Butler, 2016). In any case, our understanding of PlGF, VEGF-B, and VEGF-D, all having been conserved for 500 MYA despite their apparent present-day redundancy in mice, leaves ample room for future insights.

While very common in plants, polyploidy is tolerated among vertebrates only in fish and amphibians (Woodhouse et al., 2009). This tolerance is also seen at the gene level since individual *pdgf/vegf* gene duplications were frequently found in fish, but essentially absent in higher vertebrates. Holostei fish, a sister clade of the teleost fish, show for example a duplicated *vegfc* gene. Whether the duplicated *vegfc* genes have been maintained in Holostei from one of the prior whole genome duplications, or whether they resulted from a limited gene duplication event early in the Holostei lineage is not known, and perhaps unknowable as the chromosomal context has likely been already lost. It is surprising that both *vegfc* genes continue to be strongly conserved in Holostei. The conservation is strongest in the receptor binding domain, but can also be seen in the SHD (Figure 6B). Interestingly, only two of the conserved residues of the PDGF/VEGF signature were under strong purifying selection, and only three out of the nine residues under strong purifying selection were cysteines, arguing that a better, perhaps more sensitive search pattern for the detection of PDGF/VEGF proteins could be developed by taking conserved non-cysteine residues into consideration. In contrast to this strong conservation is the variability of the immediately N-terminally adjacent region, which is presumably instrumental in the activation of the inactive pro-VEGF-C into the mature VEGF-C by proteolysis (Jha et al., 2019) (see also https://elifesciences.org/articles/44478/figures#fig2s2). Another gene duplication example is the loach *Triplophysa rosa* (Figure 6D), which features three *vegfa* and two *pgf* genes, both of

which could be de-novo duplicated or maintained from one of the previous WGDs.

Evolutionary recent whole genome duplications have been reported for catostomid fishes such as the Chinese Sucker (*Myxocyprinus asiaticus*) (Krabbenhoft et al., 2021; Uyeno and Smith, 1972) and the common carp (*Cyprinus carpio*) (Larhammar and Risinger, 1994; Xu et al., 2019). It would be interesting to analyze the pseudogenization pattern for PDGF/VEGF genes in these two species.

Fish as model organisms

Teleost fish, which comprise most of the extant fish species, have undergone a lineage-specific whole genome duplication 350 MYA (Christoffels et al., 2004; Pasquier et al., 2016), which resulted in presumably 10 active *vegf* and eight active *pdgf* genes immediately after the duplication. In the teleost zebrafish, at least 12 of the duplicated *pdgf/vegf* genes remain functional until today according to our analysis (*pdgfaa/ab, pdgfba/bb, pdgfc, pdgfd, vegfaa/ab, pgfb, vegfba, vegfc, vegfd*). While the Ensemble genome database lists also *vegfbb* as an active zebrafish gene, we did not find any mRNA transcript matching it in 21 fish species, including zebrafish. Similarly, we could not find any *pgfa* transcripts, despite it being programmatically identified as an active gene by the Zebrafish Genome Reference Consortium (*pgfa*). Alternatively, these genes might simply not have been expressed in the tissues that were used for mRNA extraction.

We hypothesize that different teleost lineages underwent different gene elimination patterns; at least this is the most straightforward explanation for the occurrence of e.g. two *vegfc* genes in the European eel (*Anguilliformes*), or the butterfly fish (*Osteoglossiformes*). At least one other lineage, the salmonids, has undergone one additional full genome duplication 88 MYA

(Macqueen and Johnston, 2014). While this might have resulted in theoretically 36 different active *pdgf/vegf* genes immediately after the SaGD, not all of these are active today. For the salmonid *Salmo trutta* (Brown trout), 26 of these genes are identified by the Ensemble analysis pipeline as active (https://www.ensembl.org/Salmo_trutta/Location/Genome?ftype=Domain;id=IPR000072), and for 21 of them we found mRNA transcripts in the PhyloFish database. However, if mRNA and protein data are absent, it is not always possible to reliably distinguish functional from pseudogenes (Rouchka and Cha, 2009).

The partial loss of gene function within the teleost lineage remains a challenge for experimental zebrafish research and generally fish genomic research. Multiple functions that were prior to the fish-specific genome duplications executed by a single protein (e.g. VEGF-A) might be executed by two (zebrafish) or even more (Salmonidae, Catostomidae, *Cyprinus carpio*) ohnologs. There is multiple evidence that the two zebrafish ohnologs *vegfaa* and *vegfab* have diversified in terms of mRNA splicing and, consequently, their angiogenic properties (Bahary et al., 2007; Lange et al., 2022; Weijts et al., 2012). Similarly, the ohnologs *vegfc* and *vegfd* have diversified differently in fishes compared to terrestrial animals in terms of tissue distribution and receptor interaction (Bower et al., 2017; Vogrin et al., 2019). While this makes fish models at times more tedious and difficult to interpret compared to mouse models (Tzahor and Yaniv, 2022), it is at the same time a unique opportunity into a morphological and physiological diversity that cannot be seen in mammals (Das et al., 2022; Jeltsch and Alitalo, 2022).

Materials & Methods

Comprehensive database scan

BLAST searches were executed for a set of 13 reference proteins (human PDGF-A/-B/-C/-D, PlGF-3, VEGF-$A_{121/165/206}$, VEGF-$B_{167/186}$, VEGF-C/-D, and vammin-1) against the non-redundant NCBI protein database (corresponding to RefSeq Release 94). According to the Hit_def of each result, a hit was programmatically categorized based on the Hit_def field in the blast result as a *synonymous hit* (e.g. when the VEGF-D search results in a hit annotated with "VEGF-D" or "FIGF"), a *related hit* (e.g. when a VEGF-D search results in a hit annotated with "VEGF-B" or "PDGF") or an *undefined hit* (e.g. when a search for VEGF-D results in a hit annotated with "hypothetical protein" or similar). To categorize undefined hits, secondary BLASTS were initiated with the sequences for undefined hits, and if more than 50% of the secondary hits agreed in their annotation on a specific PDGF/VEGF, this information was used to categorize the primary hit. The 50% threshold had been empirically determined to be conservative, i.e. never resulting in false negative categorizations with a known set of VEGFs. Undefined hits in secondary BLASTS results are assigned to specific PDGFs/VEGFs in the same fashion, but using computationally more expensive RPS BLAST instead of protein blast. The flow chart of the analysis is show in Supplementary Figure 1.

The full table of BLAST results (Supplementary Table 1) was assembled programmatically from the data generated as described above, and the number of distinct animal species for each clade was obtained from the NCBI taxonomy database. "Fully sequenced" genomes were loosely defined as those that had registered a BioProject with NCBI with the data type "Genome sequencing" or "Genome sequencing and assembly" in the group "animals" and for which results had been published (1049 species at the time of this writing). The number of

protein sequences published for a specific clade was the number of sequences in the corresponding taxon-specific subsection of the NCBI protein database. For all clades with less than 200 unique VEGF homologs, all primary BLAST hits were manually checked for false positives (i.e. when the human-curated protein description specified a named non-PDGF/VEGF protein in the sequence description). The formula for background coloring of Figure 1 and Supplementary Table 1 according to the heuristic reliability due to biased sampling was: $reliability = log10\left(\frac{number\ of\ fully\ sequenced\ genomes^{2.5}}{number\ of\ animal\ species \cdot number\ of\ protein\ sequences+1}\right)$

Alignments, phylogenetic tree building, and conservation analysis

Alignment of cnidarian with human PDGFs/VEGFs (Figure 2)

The mcoffee mode of T-coffee 12.00 was used to align a representative subset of 10 cnidarian PDGF/VEGF-like sequences from all six species in which PDGF/VEGF-like sequences were identified, all seven human PDGF/VEGF orthologs, *C. elegans* PVF-1, and *Drosophila* PVF2. The alignment was trimmed down to 130 amino acid residues of the consensus sequence (corresponding to the VHD) starting from the first of the eight conserved cysteines of the PDGF/VEGF signature (until Proline-132 from VEGF-A). During the alignment, all conserved cysteine residues were anchored according to the alignment by Heino et al. (Heino et al., 2001).

Assessing the relationships between vertebrate and invertebrate PDGFs/VEGFs (Figure 3)

The alignment from Figure 2 above was expanded by including all four PDGFs/VEGFs identified in *Hydra vulgaris*, all three *Drosophila* PVFs, all human PDGF/VEGF paralogs, and the PDGF/VEGF-like molecule identified in the parasite *Brugia malayi*. The TGF-β homolog of *C. elegans* UNC129 was used as an outgroup. To capture information about the ancillary domains/propeptides of the proteins, the sequences included in the alignment were

expanded amino-terminally by 20 amino acids and C-terminally by 30 amino acids beyond the conserved eight cysteine residues of the PDGF/VEGF signature. The mcoffee-provided alignment was trimmed to include 179 amino acid positions corresponding to the first amino acid of the major mature form of human VEGF-C (Joukov et al., 1997) until the last cysteine of the first repeat of the BR3P motif (C-X_{10}-C-X-C-$X_{(1,3)}$-C) (Joukov et al., 1996). Tree building was performed with PhyML 3.0 (Guindon et al., 2010), combining both the original PhyML algorithm and subtree pruning and regrafting for tree topology search. To estimate the reliability of branches, 1000 bootstrapping replicates were performed. The tree was visualized with FigTree version 1.4.4, exported to an SVG file and visually enhanced using Inkscape.

Holostei VEGF-C analysis (Figure 6A-C)

The amino acid alignment and tree building for Holostei VEGF-Cs and VEGF-Ds were performed as described above using T-coffee and PhyML. As an outgroup sequence for tree rooting, we used PDGF-A from one of the Holostei species (*Lepisosteus oculatus*). The corresponding mRNA alignment was generated with PAL2NAL. Treefile and alignment were used by HyPhy version 2.5.1 (Pond et al., 2005) to test for pervasive site-level selection (SLAC). The graphs were generated from the json files using Gnumeric. The alignment of the VHD of representative VEGF-C sequences below the detailed conservation view was generated from an aligned Fasta file with SnapGene Viewer 6.1.1.

Tree building for Triplophysa rosa PDGFs/VEGFs (Figure 6D)

All 11 PDGF/VEGF sequences identified for the sucker *T. rosa* were aligned and treebuilding was performed as described above. For this analysis, we did not include any outgroup resulting in an unrooted tree.

Analysis of viral VEGF homologs

A PSI-BLAST limited to the taxon Viridae (taxid:10239) was run against the starting sequence AAD03735.1 (vascular endothelial growth factor homolog VEGF-E from Orf virus) until no new sequences were found above the 0.005 threshold. The Fasta descriptions were adjusted to include virus names and host species. For each host species, the VEGF-A_{165}, PlGF-1, and VEGF-B_{186} orthologous protein sequences were retrieved (if available) and included in the alignment and tree building, which was performed as described above. Only three out of the eight fish VEGF-B sequences were available. Therefore, we included both zebrafish VEGF-B sequences in the analysis. Similarly, no VEGF sequences were available for *Halichoerus grypus* (grey seal). These were replaced with sequences from the closest species for which VEGF sequences were available (*Zalophus californianus*, California sea lion). The protein sequence alignment was performed with T-coffee, the tree building with PhyML and the visualization of the tree with the ETE Toolkit 3.0. The v-sis sequences from Simian sarcoma virus were used as an outgroup to root the tree. The workflow and all sequences used for the analysis are available from GitHub as a python script (https://github.com/mjeltsch/VEGFE). Tree topology was used to infer the likely origin(s) of viral VEGFs.

Comparison of conservation levels between individual VEGFs

To compare the degree of conservation and to identify possible positive selection, codon analysis (synonymous versus nonsynonymous changes) was deployed. Reference protein and transcript sequences for available ortholog sets were downloaded from NCBI (https://www.ncbi.nlm.nih.gov/gene/XXXX/ortholog, XXXX = 7422, 7423, 7424, 2277, 5228 for VEGF-A, -B, -C, -D and PlGF, respectively). Obviously bogus or truncated transcript predictions (lacking essential exons of the PDGF/VEGF homology domain or

being of low quality) were eliminated manually from the set. Several proteins/transcripts had to be replaced manually to ensure that only transcripts of the same isoform were compared since not all species feature the same set of isoforms (e.g. for VEGF-C: Sus scrofa, for VEGF-A: Mus musculus). Species were only included if they featured the full set of five mammalian VEGFs in the database (VEGFA, PlGF, VEGF-B/-C/-D). The full list (sets of 5 VEGF paralogs for 82 species) was reduced to comprise only the 50 most informative sequences using T-coffee (Notredame et al., 2000); however, all sequences in a set were maintained if only one sequence in a set had been classified as informative. The full set of protein and mRNA sequences is available in Fasta format as Supplementary data. The protein sequences of each VEGF ortholog were aligned using T-coffee's mcoffee mode (Wallace et al., 2006). The final alignments were trimmed manually (keeping the sequence from the first to the last cysteine of the PDGF/VEGF cysteine signature plus 15 amino acid residues N-terminally and 5 amino acid residues C-terminally. To prepare the protein alignment for the analysis of pervasive purifying/adaptive evolution, the corresponding mRNA alignment was obtained with PAL2NAL 14 (Suyama et al., 2006). A maximum-likelihood (ML) approach was used to infer nonsynonymous versus synonymous substitution rates on a per-site basis for the alignment (Kosakovsky Pond and Frost, 2005), using SLAC (Single Likelihood Ancestor Counting) from the HyPhy 2.5.1 software package. The graph was generated with the Python library Seaborn 0.10 and enhanced via Inkscape with a molecular ribbon model of VEGF-C generated by Pymol 2.3.0 based on the PDB structure 2X1X.

Cladograms of evolutionary PDGF/VEGF history

A tree file for the animal kingdom was downloaded from the Open tree of Life (https://tree.opentreeoflife.org), which maintains a consensus tree obtained by

semi-automated synthesis of many individual studies (Hinchliff et al., 2015) (https://tree.opentreeoflife.org/curator). This treefile was used programmatically by the ETE Toolkit to generate Figure 1 and Supplementary Table 1. To generate Figure 4, a cladogram was generated from the same file excluding the protostome branch with FigTree version 1.4.4. The PDF-exported tree was enhanced using Inkscape. When public domain animal silhouettes were available, they were obtained from PhyloPic (http://phylopic.org), otherwise they were drawn by the authors.

Determination of absence of VEGF-B genes in birds

Because we had not come across any avian VEGF-B sequence, we started by blasting the Entrez protein database (ref) with a reptilian VEGF-B protein sequence (Chinese soft shell turtle VEGF-B, Uniprot K7FWR8). The sequence ids of all hits were re-written to include the animal class and common name of the organism to facilitate the identification of potential avian VEGF-B sequences. This set of 4976 sequences was subjected to multiple sequence alignment (MSA) using mcoffee. Gblocks was used with the following in order to manually curate the alignment (Castresana, 2000). After converting the output with Dendroscope 3.8.4 to Newick format, the graphical representation was generated with iTOL (Banerjee and Deshpande, 2016), visually refined using Inkscape and inspected for avian VEGF-B sequences. In order not to miss distantly related sequences, we repeated the above analysis but used three rounds of a PSI-BLAST, based on all available VEGF-B sequences and all avian VEGF-A and PlGF sequences).

Fish mRNA analysis

The PhyloFish mRNA database (http://phylofish.sigenae.org) was queried using tblastn with

all *Danio rerio* PDGF/VEGF protein sequences, resulting in 1547 unique transcripts. mRNA sequences were downloaded for all transcripts. 405 of these transcripts were identified as PDGF/VEGF family members with a relaxed PDGF signature (using the regular expression P.?C.{2,8}C.?G.?C). Individual phylogenetic trees were built for the PDGF/VEGF transcriptome of each species using *Danio rerio* PDGFs/VEGFs as reference sequences. Unannotated mRNA sequences were classified manually based on the nearest reference sequence neighbor on the tree. The total number of unique mRNA transcripts for each species and the number of unique transcript contigs were tabulated using Gnumeric version 1.12.46 (Supplementary Table 3). The numbers were normalized to zebrafish, for which 48158 unique mRNA transcripts had been obtained. The data was visualized using the Gnumeric Matrix plot function. The plot was visually enhanced in Inkscape with a cladogram based on (Pasquier et al., 2016).

Declarations

Acknowledgements

We thank Jeremy Pasquier et al. for public access to the fish mRNA sequencing data.

Funding

This research was funded by the Päivikki and Sakari Sohlberg Foundation, the Novo Nordisk Foundation (#21036), and the Academy of Finland (#337120). M.J. was supported by the Paulo Foundation and the Einar and Karin Stroem Foundation for Medical Research. K.R. was supported by the Otto A. Malm Foundation. H.B. was supported by the Finnish National Agency for Education (EDUFI) and the Finish Pharmaceutical Society.

Conflicts of interest/Competing interests

Not applicable.

Availability of data and material

All of the data used for this study are available from the corresponding GitHub repositories.

Code availability (software application or custom code)

Scripts and algorithms used in this study are available from the following GitHub repositories:

- https://github.com/mjeltsch/VEGFE
- https://github.com/mjeltsch/Holostei
- https://github.com/mjeltsch/VEGFphylo
- https://github.com/mjeltsch/cnidariaVEGFs
- https://github.com/mjeltsch/VEGFselect
- https://github.com/mjeltsch/Fish_mRNA
- https://github.com/mjeltsch/divaa

Authors' contributions

KR and MJ analyzed and interpreted the data, drafted, wrote, and edited the manuscript. MJ developed and performed the bioinformatics analyses, acquired funding, and supervised the study. HB performed the phylogenetic analyses. KR manually curated programmatically uncategorized data.

References

Aase, K., von Euler, G., Li, X., Pontén, A., Thorén, P., Cao, R., Cao, Y., Olofsson, B., Gebre-Medhin, S., Pekny, M., et al. (2001). Vascular Endothelial Growth Factor-B–Deficient Mice Display an Atrial Conduction Defect. Circulation *104*, 358–364. https://doi.org/10.1161/01.CIR.104.3.358.

Alkharsah, K.R. (2018). VEGF Upregulation in Viral Infections and Its Possible Therapeutic Implications. Int. J. Mol. Sci. *19*. https://doi.org/10.3390/ijms19061642.

Anisimov, A., Leppanen, V.-M., Tvorogov, D., Zarkada, G., Jeltsch, M., Holopainen, T., Kaijalainen, S., and Alitalo, K. (2013). The Basis for the Distinct Biological Activities of Vascular Endothelial Growth Factor Receptor-1 Ligands. Sci. Signal. *6*. https://doi.org/10.1126/scisignal.2003905.

Bahary, N., Goishi, K., Stuckenholz, C., Weber, G., LeBlanc, J., Schafer, C.A., Berman, S.S., Klagsbrun, M., and Zon, L.I. (2007). Duplicate VegfA genes and orthologues of the KDR receptor tyrosine kinase family mediate vascular development in the zebrafish. Blood *110*, 3627–3636. https://doi.org/10.1182/blood-2006-04-016378.

Baldwin, M.E., Halford, M.M., Roufail, S., Williams, R.A., Hibbs, M.L., Grail, D., Kubo, H., Stacker, S.A., and Achen, M.G. (2005). Vascular Endothelial Growth Factor D Is Dispensable for Development of the Lymphatic System. Mol. Cell. Biol. *25*, 2441–2449. https://doi.org/10.1128/MCB.25.6.2441-2449.2005.

Banerjee, S., and Deshpande, P.A. (2016). On origin and evolution of carbonic anhydrase isozymes: A phylogenetic analysis from whole-enzyme to active site. Comput. Biol. Chem. *61*, 121–129. https://doi.org/10.1016/j.compbiolchem.2016.01.003.

Bellomo, D., Headrick, J.P., Silins, G.U., Paterson, C.A., Thomas, P.S., Gartside, M., Mould, A., Cahill, M.M., Tonks, I.D., Grimmond, S.M., et al. (2000). Mice lacking the vascular endothelial growth factor-B gene (Vegfb) have smaller hearts, dysfunctional coronary vasculature, and impaired recovery from cardiac ischemia. Circ. Res. *86*, E29-35. https://doi.org/10.1161/01.res.86.2.e29.

Benton, M.J. (1990). Phylogeny of the major tetrapod groups: Morphological data and divergence dates. J. Mol. Evol. *30*, 409–424. https://doi.org/10.1007/BF02101113.

Berthelot, C., Brunet, F., Chalopin, D., Juanchich, A., Bernard, M., Noël, B., Bento, P., Da Silva, C., Labadie, K., Alberti, A., et al. (2014). The rainbow trout genome provides novel insights into evolution after whole-genome duplication in vertebrates. Nat. Commun. *5*, 3657. https://doi.org/10.1038/ncomms4657.

Blair, J.E., and Hedges, S.B. (2005). Molecular Phylogeny and Divergence Times of Deuterostome Animals. Mol. Biol. Evol. *22*, 2275–2284. https://doi.org/10.1093/molbev/msi225.

Bower, N.I., Vogrin, A.J., Guen, L.L., Chen, H., Stacker, S.A., Achen, M.G., and Hogan, B.M. (2017). Vegfd modulates both angiogenesis and lymphangiogenesis during zebrafish embryonic development. Development *144*, 507–518. https://doi.org/10.1242/dev.146969.

Butler, P.J. (2016). The physiological basis of bird flight. Philos. Trans. R. Soc. B Biol. Sci. *371*, 20150384. https://doi.org/10.1098/rstb.2015.0384.

Carmeliet, P., Ferreira, V., Breier, G., Pollefeyt, S., Kieckens, L., Gertsenstein, M., Fahrig, M., Vandenhoeck, A., Harpal, K., Eberhardt, C., et al. (1996). Abnormal blood vessel development and lethality in embryos lacking a single VEGF allele. Nature *380*, 435–439.

https://doi.org/10.1038/380435a0.

Carmeliet, P., Ng, Y.-S., Nuyens, D., Theilmeier, G., Brusselmans, K., Cornelissen, I., Ehler, E., Kakkar, V.V., Stalmans, I., Mattot, V., et al. (1999). Impaired myocardial angiogenesis and ischemic cardiomyopathy in mice lacking the vascular endothelial growth factor isoforms VEGF 164 and VEGF 188. Nat. Med. *5*, 495–502. https://doi.org/10.1038/8379.

Castresana, J. (2000). Selection of Conserved Blocks from Multiple Alignments for Their Use in Phylogenetic Analysis. Mol. Biol. Evol. *17*, 540–552. https://doi.org/10.1093/oxfordjournals.molbev.a026334.

Chang, E.S., Neuhof, M., Rubinstein, N.D., Diamant, A., Philippe, H., Huchon, D., and Cartwright, P. (2015). Genomic insights into the evolutionary origin of Myxozoa within Cnidaria. Proc. Natl. Acad. Sci. *112*, 14912–14917. https://doi.org/10.1073/pnas.1511468112.

Chen, T.T., Luque, A., Lee, S., Anderson, S.M., Segura, T., and Iruela-Arispe, M.L. (2010). Anchorage of VEGF to the extracellular matrix conveys differential signaling responses to endothelial cells. J. Cell Biol. *188*, 595–609. https://doi.org/10.1083/jcb.200906044.

Christoffels, A., Koh, E.G.L., Chia, J., Brenner, S., Aparicio, S., and Venkatesh, B. (2004). Fugu Genome Analysis Provides Evidence for a Whole-Genome Duplication Early During the Evolution of Ray-Finned Fishes. Mol. Biol. Evol. *21*, 1146–1151. https://doi.org/10.1093/molbev/msh114.

Cunningham, F., Achuthan, P., Akanni, W., Allen, J., Amode, M.R., Armean, I.M., Bennett, R., Bhai, J., Billis, K., Boddu, S., et al. (2019). Ensembl 2019. Nucleic Acids Res. *47*, D745–D751. https://doi.org/10.1093/nar/gky1113.

Dalpe, G., Tarsitano, M., Persico, M.G., Zheng, H., and Culotti, J. (2013). C. elegans PVF-1 inhibits permissive UNC-40 signalling through CED-10 GTPase to position the male ray 1 sensillum. Development *140*, 4020–4030. https://doi.org/10.1242/dev.095190.

Das, R.N., Tevet, Y., Safriel, S., Han, Y., Moshe, N., Lambiase, G., Bassi, I., Nicenboim, J., Brückner, M., Hirsch, D., et al. (2022). Generation of specialized blood vessels via lymphatic transdifferentiation. Nature *606*, 570–575. https://doi.org/10.1038/s41586-022-04766-2.

Dehal, P., and Boore, J.L. (2005). Two Rounds of Whole Genome Duplication in the Ancestral Vertebrate. PLOS Biol. *3*, e314. https://doi.org/10.1371/journal.pbio.0030314.

Doolittle, R.F., Hunkapiller, M.W., Hood, L.E., Devare, S.G., Robbins, K.C., Aaronson, S.A., and Antoniades, H.N. (1983). Simian sarcoma virus onc gene, v-sis, is derived from the gene (or genes) encoding a platelet-derived growth factor. Science *221*, 275–277. https://doi.org/10.1126/science.6304883.

Dormer, A., and Beck, G. (2005). Evolutionary analysis of human vascular endothelial growth factor, angiopoietin, and tyrosine endothelial kinase involved in angiogenesis and immunity. In Silico Biol. *5*, 323–339. https://doi.org/10.1007/978-3-540-33177-3_23.

Eichmann, A., Corbel, C., Nataf, V., Vaigot, P., Bréant, C., and Le Douarin, N.M. (1997). Ligand-dependent development of the endothelial and hemopoietic lineages from embryonic mesodermal cells expressing vascular endothelial growth factor receptor 2. Proc. Natl. Acad. Sci. *94*, 5141–5146. https://doi.org/10.1073/pnas.94.10.5141.

Erwin, D.H., and Davidson, E.H. (2002). The last common bilaterian ancestor. Development *129*, 3021–3032. https://doi.org/10.1242/dev.129.13.3021.

Essbauer, S., Pfeffer, M., and Meyer, H. (2010). Zoonotic poxviruses. Vet. Microbiol. *140*, 229–236. https://doi.org/10.1016/j.vetmic.2009.08.026.

Fang, S., Nurmi, H., Heinolainen, K., Chen, S., Salminen, E., Saharinen, P., Mikkola, H.K.A., and Alitalo, K. (2016). Critical requirement of VEGF-C in transition to fetal erythropoiesis. Blood *128*, 710–720. https://doi.org/10.1182/blood-2015-12-687970.

Fang, S., Chen, S., Nurmi, H., Leppänen, V.-M., Jeltsch, M., Scadden, D., Silberstein, L., Mikkola, H., and Alitalo, K. (2020). VEGF-C protects the integrity of the bone marrow perivascular niche in mice. Blood *136*, 1871–1883. https://doi.org/10.1182/blood.2020005699.

Ferrara, N., Carver-Moore, K., Chen, H., Dowd, M., Lu, L., O'Shea, K.S., Powell-Braxton, L., Hillan, K.J., and Moore, M.W. (1996). Heterozygous embryonic lethality induced by targeted inactivation of the VEGF gene. Nature *380*, 439–442. https://doi.org/10.1038/380439a0.

Gerhardt, H., Golding, M., Fruttiger, M., Ruhrberg, C., Lundkvist, A., Abramsson, A., Jeltsch, M., Mitchell, C., Alitalo, K., Shima, D., et al. (2003). VEGF guides angiogenic sprouting utilizing endothelial tip cell filopodia. J. Cell Biol. *161*, 1163–1177. https://doi.org/10.1083/jcb.200302047.

Grassot, J., Gouy, M., Perrière, G., and Mouchiroud, G. (2006). Origin and Molecular Evolution of Receptor Tyrosine Kinases with Immunoglobulin-Like Domains. Mol. Biol. Evol. *23*, 1232–1241. https://doi.org/10.1093/molbev/msk007.

de Groof, A., Guelen, L., Deijs, M., van der Wal, Y., Miyata, M., Ng, K.S., Grinsven, L. van, Simmelink, B., Biermann, Y., Grisez, L., et al. (2015). A Novel Virus Causes Scale Drop Disease in Lates calcarifer. PLOS Pathog. *11*, e1005074. https://doi.org/10.1371/journal.ppat.1005074.

Guindon, S., Dufayard, J.-F., Lefort, V., Anisimova, M., Hordijk, W., and Gascuel, O. (2010). New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0. Syst. Biol. *59*, 307–321. https://doi.org/10.1093/sysbio/syq010.

Hagberg, C., Mehlem, A., Falkevall, A., Muhl, L., and Eriksson, U. (2013). Endothelial Fatty Acid Transport: Role of Vascular Endothelial Growth Factor B. Physiology *28*, 125–134. https://doi.org/10.1152/physiol.00042.2012.

Hautaniemi, M., Ueda, N., Tuimala, J., Mercer, A.A., Lahdenperä, J., and McInnes, C.J. (2010). The genome of pseudocowpoxvirus: comparison of a reindeer isolate and a reference strain. J. Gen. Virol. *91*, 1560–1576. https://doi.org/10.1099/vir.0.018374-0.

He, W., Tang, Y., Qi, B., Lu, C., Qin, C., Wei, Y., Yi, J., and Chen, M. (2014). Phylogenetic analysis and positive-selection site detecting of vascular endothelial growth factor family in vertebrates. Gene *535*, 345–352. https://doi.org/10.1016/j.gene.2013.10.031.

Heino, T.I., Kärpänen, T., Wahlström, G., Pulkkinen, M., Eriksson, U., Alitalo, K., and Roos, C. (2001). The Drosophila VEGF receptor homolog is expressed in hemocytes. Mech. Dev. *109*, 69–77. https://doi.org/10.1016/S0925-4773(01)00510-X.

Hellstrom, M., Kal n, M., Lindahl, P., Abramsson, A., and Betsholtz, C. (1999). Role of PDGF-B and PDGFR-beta in recruitment of vascular smooth muscle cells and pericytes during embryonic blood vessel formation in the mouse. Development *126*, 3047–3055. https://doi.org/10.1242/dev.126.14.3047.

Hickman, C.P., Roberts, L.S., Keen, S.L., Larson, A., I'Anson, H., and Eisenhour, D.J. (2008). Annelids and Allied Taxa, Clade Clitellata. In Integrated Principles of Zoology, (Boston: McGraw-Hill/Higher Education), pp. 371–375.

Hinchliff, C.E., Smith, S.A., Allman, J.F., Burleigh, J.G., Chaudhary, R., Coghill, L.M.,

Crandall, K.A., Deng, J., Drew, B.T., Gazis, R., et al. (2015). Synthesis of phylogeny and taxonomy into a comprehensive tree of life. Proc. Natl. Acad. Sci. *112*, 12764–12769. https://doi.org/10.1073/pnas.1423041112.

Hoch, R.V., and Soriano, P. (2003). Roles of PDGF in animal development. Development *130*, 4769–4784. https://doi.org/10.1242/dev.00721.

Holmes, D.I., and Zachary, I. (2005). The vascular endothelial growth factor (VEGF) family: angiogenic factors in health and disease. Genome Biol. *6*, 209. https://doi.org/10.1186/gb-2005-6-2-209.

Huemer, H.P., Zobl, A., and Windisch, A. (2014). Serological evidence for Parapoxvirus infection in chamois from the Tyrol regions of Austria and Italy. Vet. Ital. 233–236. https://doi.org/10.12834/VetIt.1304.14.

Hughes, A.L., Irausquin, S., and Friedman, R. (2010). The Evolutionary Biology of Poxviruses. Infect. Genet. Evol. J. Mol. Epidemiol. Evol. Genet. Infect. Dis. *10*, 50. https://doi.org/10.1016/j.meegid.2009.10.001.

Inder, M.K., Ueda, N., Mercer, A.A., Fleming, S.B., and Wise, L.M. (2007). Bovine papular stomatitis virus encodes a functionally distinct VEGF that binds both VEGFR-1 and VEGFR-2. J. Gen. Virol. *88*, 781–791. https://doi.org/10.1099/vir.0.82582-0.

Iyer, S., and Acharya, K.R. (2011). Tying the knot: The cystine signature and molecular-recognition processes of the vascular endothelial growth factor family of angiogenic cytokines. FEBS J. *278*, 4304–4322. https://doi.org/10.1111/j.1742-4658.2011.08350.x.

Jeltsch, M., and Alitalo, K. (2022). Lymphatic-to-blood vessel transdifferentiation in zebrafish. Nat. Cardiovasc. Res. 1–3. https://doi.org/10.1038/s44161-022-00073-1.

Jeltsch, M., Kaipainen, A., Joukov, V., Meng, X., Lakso, M., Rauvala, H., Swartz, M., Fukumura, D., Jain, R.K., and Alitalo, K. (1997). Hyperplasia of Lymphatic Vessels in VEGF-C Transgenic Mice. Science *276*, 1423–1425. https://doi.org/10.1126/science.276.5317.1423.

Jeltsch, M., Jha, S.K., Tvorogov, D., Anisimov, A., Leppänen, V.-M., Holopainen, T., Kivelä, R., Ortega, S., Kärpanen, T., and Alitalo, K. (2014). CCBE1 Enhances Lymphangiogenesis via A Disintegrin and Metalloprotease With Thrombospondin Motifs-3–Mediated Vascular Endothelial Growth Factor-C Activation. Circulation *129*, 1962–1971. https://doi.org/10.1161/CIRCULATIONAHA.113.002779.

Jha, S.K., Rauniyar, K., Karpanen, T., Leppänen, V.-M., Brouillard, P., Vikkula, M., Alitalo, K., and Jeltsch, M. (2017). Efficient activation of the lymphangiogenic growth factor VEGF-C requires the C-terminal domain of VEGF-C and the N-terminal domain of CCBE1. Sci. Rep. *7*, 4916. https://doi.org/10.1038/s41598-017-04982-1.

Jha, S.K., Rauniyar, K., Chronowska, E., Mattonet, K., Maina, E.W., Koistinen, H., Stenman, U.-H., Alitalo, K., and Jeltsch, M. (2019). KLK3/PSA and cathepsin D activate VEGF-C and VEGF-D. ELife *8*, e44478. https://doi.org/10.7554/eLife.44478.

Joukov, V., Pajusola, K., Kaipainen, A., Chilov, D., Lahtinen, I., Kukk, E., Saksela, O., Kalkkinen, N., and Alitalo, K. (1996). A novel vascular endothelial growth factor, VEGF-C, is a ligand for the Flt4 (VEGFR-3) and KDR (VEGFR-2) receptor tyrosine kinases. EMBO J. *15*, 290–298. https://doi.org/10.1002/j.1460-2075.1996.tb00359.x.

Joukov, V., Sorsa, T., Kumar, V., Jeltsch, M., Claesson-Welsh, L., Cao, Y., Saksela, O., Kalkkinen, N., and Alitalo, K. (1997). Proteolytic processing regulates receptor specificity

and activity of VEGF-C. EMBO J. *16*, 3898–3911. https://doi.org/10.1093/emboj/16.13.3898.

Kasahara, M. (2007). The 2R hypothesis: an update. Curr. Opin. Immunol. *19*, 547–552. https://doi.org/10.1016/j.coi.2007.07.009.

Kasap, M. (2005). Phylogenetic Analysis of Vascular Endothelial Growth Factor Diversity. Turk. J. Biol. *29*, 217–227.

Kipryushina, Y.O., Yakovlev, K.V., and Odintsova, N.A. (2015). Vascular endothelial growth factors: A comparison between invertebrates and vertebrates. Cytokine Growth Factor Rev. *26*, 687–695. https://doi.org/10.1016/j.cytogfr.2015.04.001.

Komori, Y., and Sugihara, H. (1990). Purification and physiological study of a hypotensive factor from the venom of Vipera aspis aspis (aspic viper). Toxicon *28*, 359–369. https://doi.org/10.1016/0041-0101(90)90073-G.

Komori, Y., Nikai, T., Taniguchi, K., Masuda, K., and Sugihara, H. (1999). Vascular Endothelial Growth Factor VEGF-like Heparin-Binding Protein from the Venom of Vipera aspis aspis (Aspic Viper). Biochemistry *38*, 11796–11803. https://doi.org/10.1021/bi990562z.

Kosakovsky Pond, S.L., and Frost, S.D.W. (2005). Not So Different After All: A Comparison of Methods for Detecting Amino Acid Sites Under Selection. Mol. Biol. Evol. *22*, 1208–1222. https://doi.org/10.1093/molbev/msi105.

Krabbenhoft, T.J., MacGuigan, D.J., Backenstose, N.J.C., Waterman, H., Lan, T., Pelosi, J.A., Tan, M., and Sandve, S.R. (2021). Chromosome-Level Genome Assembly of Chinese Sucker (Myxocyprinus asiaticus) Reveals Strongly Conserved Synteny Following a Catostomid-Specific Whole-Genome Duplication. Genome Biol. Evol. *13*, evab190. https://doi.org/10.1093/gbe/evab190.

Krishnapati, L.-S., and Ghaskadbi, S. (2014). Identification and characterization of VEGF and FGF from Hydra. Int. J. Dev. Biol. *57*, 897–906. https://doi.org/10.1387/ijdb.130077sg.

Künnapuu, J., Bokharaie, H., and Jeltsch, M. (2021). Proteolytic Cleavages in the VEGF Family: Generating Diversity among Angiogenic VEGFs, Essential for the Activation of Lymphangiogenic VEGFs. Biology *10*, 167. https://doi.org/10.3390/biology10020167.

Lackner, M., Schmotz, C., and Jeltsch, M. (2019). The Proteolytic Activation of Vascular Endothelial Growth Factor-C. Lymphol. Forsch. Prax. *23*, 88–98. https://doi.org/10.5281/zenodo.3629263.

Lange, M., Ohnesorge, N., Hoffmann, D., Rocha, S.F., Benedito, R., and Siekmann, A.F. (2022). Zebrafish mutants in vegfab can affect endothelial cell proliferation without altering ERK phosphorylation and are phenocopied by loss of PI3K signaling. Dev. Biol. *486*, 26–43. https://doi.org/10.1016/j.ydbio.2022.03.006.

Larhammar, D., and Risinger, C. (1994). Molecular Genetic Aspects of Tetraploidy in the Common Carp Cyprinus carpio. Mol. Phylogenet. Evol. *3*, 59–68. https://doi.org/10.1006/mpev.1994.1007.

Lawrence, M.G., Lai, J., and Clements, J.A. (2010). Kallikreins on Steroids: Structure, Function, and Hormonal Regulation of Prostate-Specific Antigen and the Extended Kallikrein Locus. Endocr. Rev. *31*, 407–446. https://doi.org/10.1210/er.2009-0034.

Lee, S., Jilani, S.M., Nikolova, G.V., Carpizo, D., and Iruela-Arispe, M.L. (2005). Processing of VEGF-A by matrix metalloproteinases regulates bioavailability and vascular patterning in tumors. J. Cell Biol. *169*, 681–691. https://doi.org/10.1083/jcb.200409115.

Leppänen, V.-M., Prota, A.E., Jeltsch, M., Anisimov, A., Kalkkinen, N., Strandin, T.,

Lankinen, H., Goldman, A., Ballmer-Hofer, K., and Alitalo, K. (2010). Structural determinants of growth factor binding and specificity by VEGF receptor 2. Proc. Natl. Acad. Sci. *107*, 2425–2430. https://doi.org/10.1073/pnas.0914318107.

Leppänen, V.-M., Jeltsch, M., Anisimov, A., Tvorogov, D., Aho, K., Kalkkinen, N., Toivanen, P., Ylä-Herttuala, S., Ballmer-Hofer, K., and Alitalo, K. (2011). Structural determinants of vascular endothelial growth factor-D receptor binding and specificity. Blood *117*, 1507–1515. https://doi.org/10.1182/blood-2010-08-301549.

Lyttle, D.J., Fraser, K.M., Fleming, S.B., Mercer, A.A., and Robinson, A.J. (1994). Homologs of vascular endothelial growth factor are encoded by the poxvirus orf virus. J. Virol. *68*, 84–92. https://doi.org/10.1128/JVI.68.1.84-92.1994.

Macqueen, D.J., and Johnston, I.A. (2014). A well-constrained estimate for the timing of the salmonid whole genome duplication reveals major decoupling from species diversification. Proc. R. Soc. B Biol. Sci. *281*. https://doi.org/10.1098/rspb.2013.2881.

Mattonet, K., and Jeltsch, M. (2015). Heterogeneity of the origin of the lymphatic system. Lymphol. Forsch. Prax. *19*, 84–88.

Mi, H., Muruganujan, A., and Thomas, P.D. (2013). PANTHER in 2013: modeling the evolution of gene function, and other gene attributes, in the context of phylogenetic trees. Nucleic Acids Res. *41*, D377-386. https://doi.org/10.1093/nar/gks1118.

Moessinger, C., Nilsson, I., Muhl, L., Zeitelhofer, M., Heller Sahlgren, B., Skogsberg, J., and Eriksson, U. (2020). VEGF-B signaling impairs endothelial glucose transcytosis by decreasing membrane cholesterol content. EMBO Rep. *21*, e49343. https://doi.org/10.15252/embr.201949343.

Moorman, A.F.M., and Christoffels, V.M. (2003). Cardiac Chamber Formation: Development, Genes, and Evolution. Physiol. Rev. *83*, 1223–1267. https://doi.org/10.1152/physrev.00006.2003.

Muller, Y.A., Heiring, C., Misselwitz, R., Welfle, K., and Welfle, H. (2002). The Cystine Knot Promotes Folding and Not Thermodynamic Stability in Vascular Endothelial Growth Factor. J. Biol. Chem. *277*, 43410–43416. https://doi.org/10.1074/jbc.M206438200.

Mural, R.J., Adams, M.D., Myers, E.W., Smith, H.O., Miklos, G.L.G., Wides, R., Halpern, A., Li, P.W., Sutton, G.G., Nadeau, J., et al. (2002). A Comparison of Whole-Genome Shotgun-Derived Mouse Chromosome 16 and the Human Genome. Science *296*, 1661–1671. https://doi.org/10.1126/science.1069193.

Notredame, C., Higgins, D.G., and Heringa, J. (2000). T-coffee: a novel method for fast and accurate multiple sequence alignment 1 1Edited by J. Thornton. J. Mol. Biol. *302*, 205–217. https://doi.org/10.1006/jmbi.2000.4042.

Ogawa, S., Oku, A., Sawano, A., Yamaguchi, S., Yazaki, Y., and Shibuya, M. (1998). A Novel Type of Vascular Endothelial Growth Factor, VEGF-E (NZ-7 VEGF), Preferentially Utilizes KDR/Flk-1 Receptor and Carries a Potent Mitotic Activity without Heparin-binding Domain. J. Biol. Chem. *273*, 31273–31282. https://doi.org/10.1074/jbc.273.47.31273.

Olofsson, B., Pajusola, K., Kaipainen, A., Euler, G. von, Joukov, V., Saksela, O., Orpana, A., Pettersson, R.F., Alitalo, K., and Eriksson, U. (1996). Vascular endothelial growth factor B, a novel growth factor for endothelial cells. Proc. Natl. Acad. Sci. *93*, 2576–2581. https://doi.org/10.1073/pnas.93.6.2576.

Orlandini, M., Marconcini, L., Ferruzzi, R., and Oliviero, S. (1996). Identification of a c-fos-induced gene that is related to the platelet-derived growth factor/vascular endothelial

growth factor family. Proc. Natl. Acad. Sci. U. S. A. *93*, 11675–11680. https://doi.org/10.1073/pnas.93.21.11675.

Park, J.E., Keller, G.A., and Ferrara, N. (1993). The vascular endothelial growth factor (VEGF) isoforms: differential deposition into the subepithelial extracellular matrix and bioactivity of extracellular matrix-bound VEGF. Mol. Biol. Cell *4*, 1317–1326. https://doi.org/10.1091/mbc.4.12.1317.

Pasquier, J., Cabau, C., Nguyen, T., Jouanno, E., Severac, D., Braasch, I., Journot, L., Pontarotti, P., Klopp, C., Postlethwait, J.H., et al. (2016). Gene evolution and gene expression after whole genome duplication in fish: the PhyloFish database. BMC Genomics *17*, 368. https://doi.org/10.1186/s12864-016-2709-z.

Plouët, J., Moro, F., Bertagnolli, S., Coldeboeuf, N., Mazarguil, H., Clamens, S., and Bayard, F. (1997). Extracellular Cleavage of the Vascular Endothelial Growth Factor 189-Amino Acid Form by Urokinase Is Required for Its Mitogenic Effect. J. Biol. Chem. *272*, 13390–13396. https://doi.org/10.1074/jbc.272.20.13390.

Pond, S.L.K., Frost, S.D.W., and Muse, S.V. (2005). HyPhy: hypothesis testing using phylogenies. Bioinformatics *21*, 676–679. https://doi.org/10.1093/bioinformatics/bti079.

Read, R.D. (2018). Pvr receptor tyrosine kinase signaling promotes post-embryonic morphogenesis, and survival of glia and neural progenitor cells in Drosophila. Development *145*. https://doi.org/10.1242/dev.164285.

Risau, W. (1997). Mechanisms of angiogenesis. Nature *386*, 671–674. https://doi.org/10.1038/386671a0.

Robinson, C.J., and Stringer, S.E. (2001). The splice variants of vascular endothelial growth factor (VEGF) and their receptors. J. Cell Sci. *114*, 853–865. https://doi.org/10.1242/jcs.114.5.853.

Rodi, D.J., Mandava, S., and Makowski, L. (2004). DIVAA: analysis of amino acid diversity in multiple aligned protein sequences. Bioinformatics *20*, 3481–3489. https://doi.org/10.1093/bioinformatics/bth432.

Rouchka, E., and Cha, I. (2009). Current Trends in Pseudogene Detection and Characterization. Curr. Bioinforma. *4*, 112–119. https://doi.org/10.2174/157489309788184792.

Ruhrberg, C., Gerhardt, H., Golding, M., Watson, R., Ioannidou, S., Fujisawa, H., Betsholtz, C., and Shima, D.T. (2002). Spatially restricted patterning cues provided by heparin-binding VEGF-A control blood vessel branching morphogenesis. Genes Dev. *16*, 2684–2698. https://doi.org/10.1101/gad.242002.

Savory, L.J., Stacker, S.A., Fleming, S.B., Niven, B.E., and Mercer, A.A. (2000). Viral Vascular Endothelial Growth Factor Plays a Critical Role in Orf Virus Infection. J. Virol. *74*, 10699–10706. https://doi.org/10.1128/JVI.74.22.10699-10706.2000.

Seipel, K., Eberhardt, M., Müller, P., Pescia, E., Yanze, N., and Schmid, V. (2004). Homologs of vascular endothelial growth factor and receptor, VEGF and VEGFR, in the jellyfish *Podocoryne carnea*. Dev. Dyn. *231*, 303–312. https://doi.org/10.1002/dvdy.20139.

Sereno, P.C. (1999). The Evolution of Dinosaurs. Science *284*, 2137–2147. https://doi.org/10.1126/science.284.5423.2137.

Simons, M., Gordon, E., and Claesson-Welsh, L. (2016). Mechanisms and regulation of endothelial VEGF receptor signalling. Nat. Rev. Mol. Cell Biol. *17*, 611–625.

https://doi.org/10.1038/nrm.2016.87.

Singh, J.P., Chaikin, M.A., and Stiles, C.D. (1982). Phylogenetic analysis of platelet-derived growth factor by radio-receptor assay. J. Cell Biol. *95*, 667–671. https://doi.org/10.1083/jcb.95.2.667.

Stalmans, I., Ng, Y.-S., Rohan, R., Fruttiger, M., Bouché, A., Ÿuce, A., Fujisawa, H., Hermans, B., Shani, M., Jansen, S., et al. (2002). Arteriolar and venular patterning in retinas of mice selectively expressing VEGF isoforms. J. Clin. Invest. *109*, 327–336. https://doi.org/10.1172/JCI14362.

Suyama, M., Torrents, D., and Bork, P. (2006). PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. Nucleic Acids Res. *34*, W609-612. https://doi.org/10.1093/nar/gkl315.

Tarsitano, M., Falco, S.D., Colonna, V., McGhee, J.D., and Persico, M.G. (2006). The C. elegans pvf-1 gene encodes a PDGF/VEGF-like factor able to bind mammalian VEGF receptors and to induce angiogenesis. FASEB J. *20*, 227–233. https://doi.org/10.1096/fj.05-4147com.

Tayade, C., Hilchie, D., He, H., Fang, Y., Moons, L., Carmeliet, P., Foster, R.A., and Croy, B.A. (2007). Genetic Deletion of Placenta Growth Factor in Mice Alters Uterine NK Cells. J. Immunol. *178*, 4267–4275. https://doi.org/10.4049/jimmunol.178.7.4267.

Thiele, W., Krishnan, J., Rothley, M., Weih, D., Plaumann, D., Kuch, V., Quagliata, L., Weich, H.A., and Sleeman, J.P. (2012). VEGFR-3 is expressed on megakaryocyte precursors in the murine bone marrow and plays a regulatory role in megakaryopoiesis. Blood *120*, 1899–1907. https://doi.org/10.1182/blood-2011-09-376657.

Tiozzo, S., Voskoboynik, A., Brown, F.D., and De Tomaso, A.W. (2008). A conserved role of the VEGF pathway in angiogenesis of an ectodermally-derived vasculature. Dev. Biol. *315*, 243–255. https://doi.org/10.1016/j.ydbio.2007.12.035.

Turwankar, A., and Ghaskadbi, S. (2019). VEGF and FGF signaling during head regeneration in hydra. BioRxiv 596734. https://doi.org/10.1101/596734.

Tzahor, E., and Yaniv, K. (2022). How many fish make a mouse? Nat. Cardiovasc. Res. *1*, 2–3. https://doi.org/10.1038/s44161-021-00010-8.

Ueda, N., Wise, L.M., Stacker, S.A., Fleming, S.B., and Mercer, A.A. (2003). Pseudocowpox virus encodes a homolog of vascular endothelial growth factor. Virology *305*, 298–309. https://doi.org/10.1006/viro.2002.1750.

Uyeno, T., and Smith, G.R. (1972). Tetraploid Origin of the Karyotype of Catostomid Fishes. Science *175*, 644–646. https://doi.org/10.1126/science.175.4022.644.

Vitt, U.A., Hsu, S.Y., and Hsueh, A.J.W. (2001). Evolution and Classification of Cystine Knot-Containing Hormones and Related Extracellular Signaling Molecules. Mol. Endocrinol. *15*, 681–694. https://doi.org/10.1210/mend.15.5.0639.

Vogrin, A.J., Bower, N.I., Gunzburg, M.J., Roufail, S., Okuda, K.S., Paterson, S., Headey, S.J., Stacker, S.A., Hogan, B.M., and Achen, M.G. (2019). Evolutionary Differences in the Vegf/Vegfr Code Reveal Organotypic Roles for the Endothelial Cell Receptor Kdr in Developmental Lymphangiogenesis. Cell Rep. *28*, 2023-2036.e4. https://doi.org/10.1016/j.celrep.2019.07.055.

Wallace, I.M., O'Sullivan, O., Higgins, D.G., and Notredame, C. (2006). M-Coffee: combining multiple sequence alignment methods with T-Coffee. Nucleic Acids Res. *34*,

1692–1699. https://doi.org/10.1093/nar/gkl091.

Weijts, B.G.M.W., Bakker, W.J., Cornelissen, P.W.A., Liang, K.-H., Schaftenaar, F.H., Westendorp, B., de Wolf, C.A.C.M.T., Paciejewska, M., Scheele, C.L.G.J., Kent, L., et al. (2012). E2F7 and E2F8 promote angiogenesis through transcriptional activation of VEGFA in cooperation with HIF1. EMBO J. *31*, 3871–3884. https://doi.org/10.1038/emboj.2012.231.

Whittington, R.J., Becker, J.A., and Dennis, M.M. (2010). Iridovirus infections in finfish – critical review with emphasis on ranaviruses. J. Fish Dis. *33*, 95–122. https://doi.org/10.1111/j.1365-2761.2009.01110.x.

Wiesmann, C., Fuh, G., Christinger, H.W., Eigenbrot, C., Wells, J.A., and de Vos, A.M. (1997). Crystal Structure at 1.7 Å Resolution of VEGF in Complex with Domain 2 of the Flt-1 Receptor. Cell *91*, 695–704. https://doi.org/10.1016/S0092-8674(00)80456-0.

Woodhouse, M., Burkart-Waco, D., and Comai, L. (2009). Polyploidy. Nat. Educ. *2*, 1.

Xu, P., Xu, J., Liu, G., Chen, L., Zhou, Z., Peng, W., Jiang, Y., Zhao, Z., Jia, Z., Sun, Y., et al. (2019). The allotetraploid origin and asymmetrical genome evolution of the common carp Cyprinus carpio. Nat. Commun. *10*, 4625. https://doi.org/10.1038/s41467-019-12644-1.

Yan, Y., and Hillyer, J.F. (2020). The immune and circulatory systems are functionally integrated across insect evolution. Sci. Adv. *6*, eabb3164. https://doi.org/10.1126/sciadv.abb3164.

Zafar, M.I., Zheng, J., Kong, W., Ye, X., Gou, L., Regmi, A., and Chen, L.-L. (2017). The role of vascular endothelial growth factor-B in metabolic homoeostasis: current evidence. Biosci. Rep. *37*, BSR20171089. https://doi.org/10.1042/BSR20171089.

Zheng, H., Wang, X., Guo, P., Ge, W., Yan, Q., Gao, W., Xi, Y., and Yang, X. (2017). Premature remodeling of fat body and fat mobilization triggered by platelet-derived growth factor/VEGF receptor in Drosophila. FASEB J. *31*, 1964–1975. https://doi.org/10.1096/fj.201601127R.

Supplementary Materials

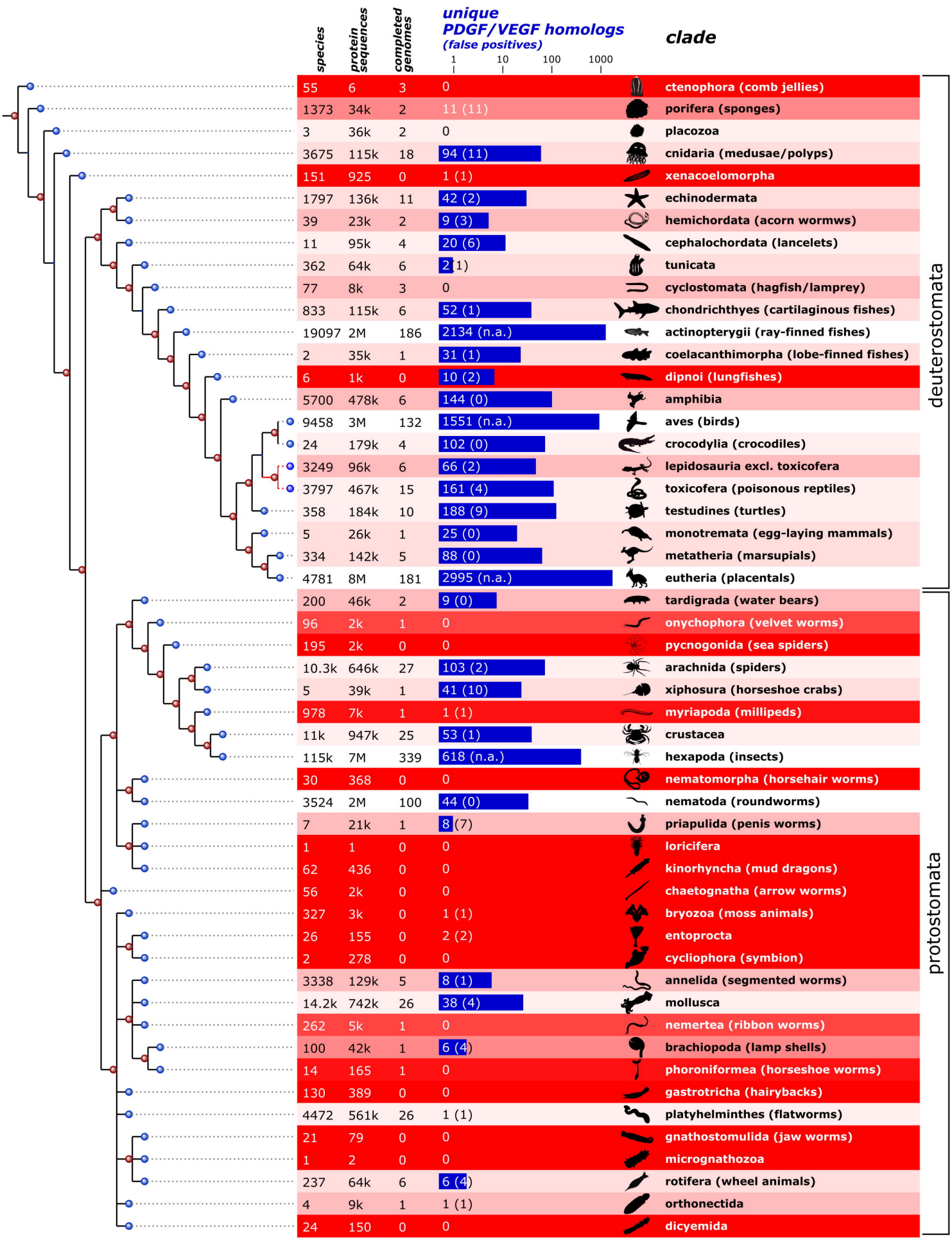
Supplementary Figure 1. Bioinformatics workflow.

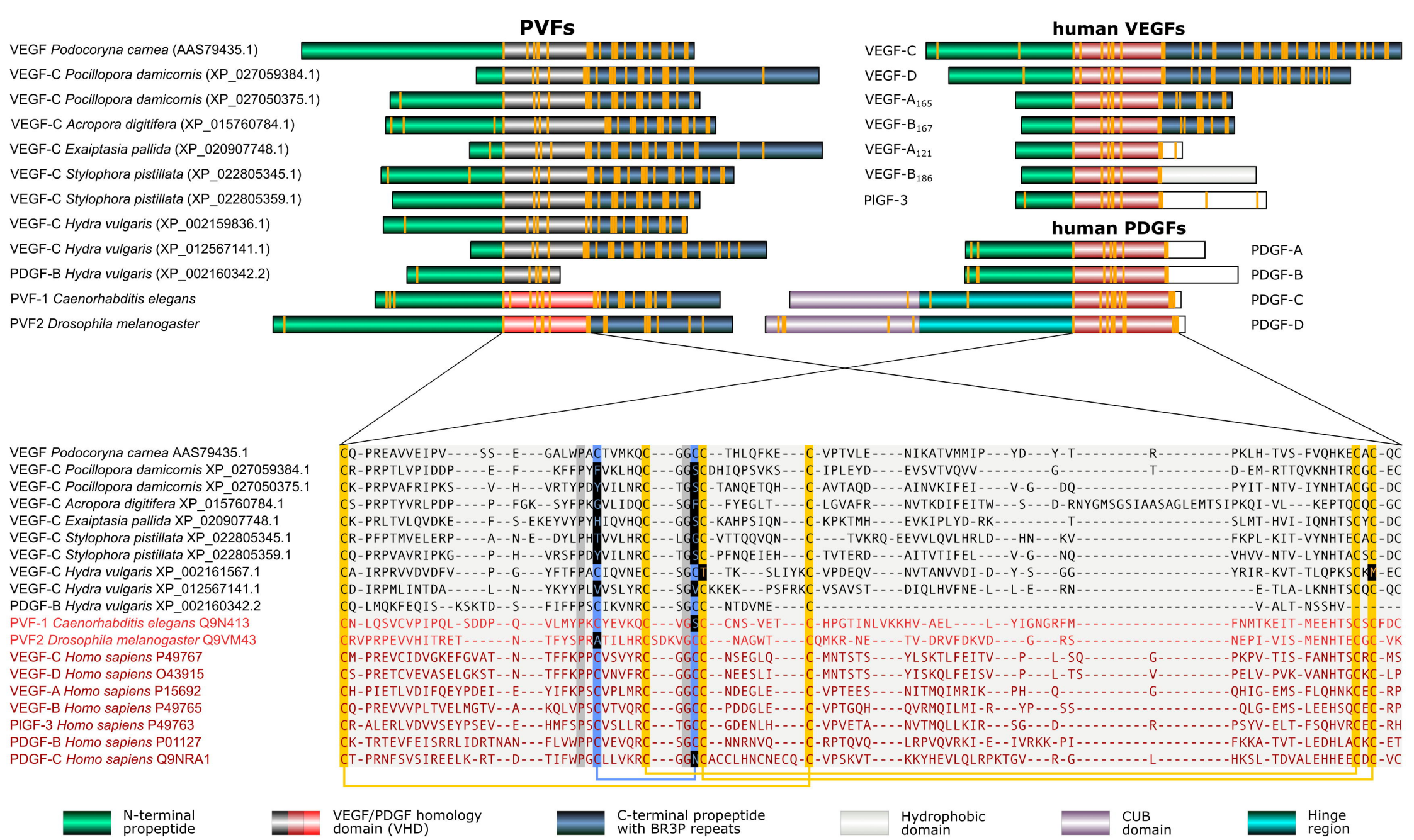
Supplementary Figure 2. Complete phylogenetic tree of VEGF-E.

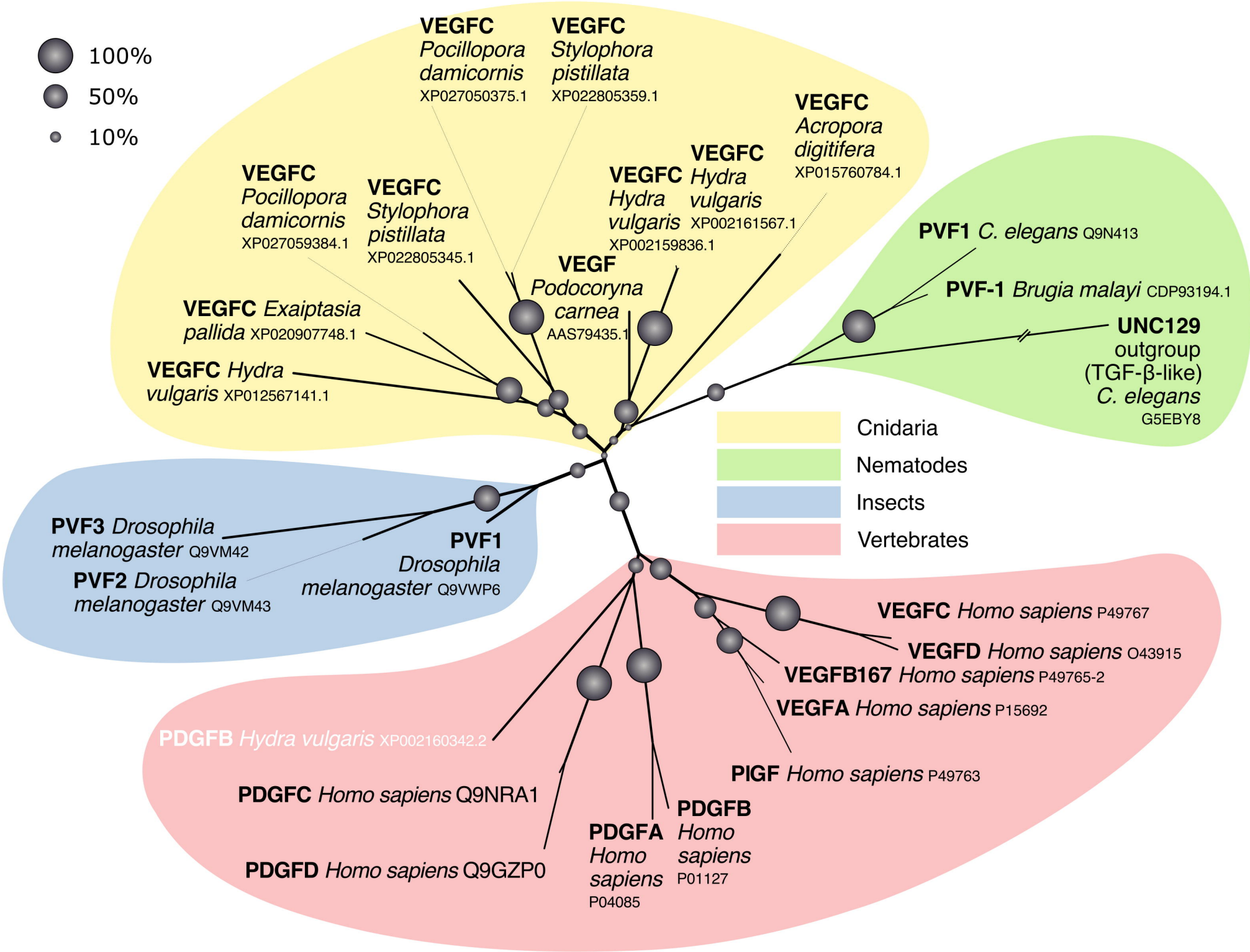
Supplementary Table 1. Complete table of all BLAST hits.

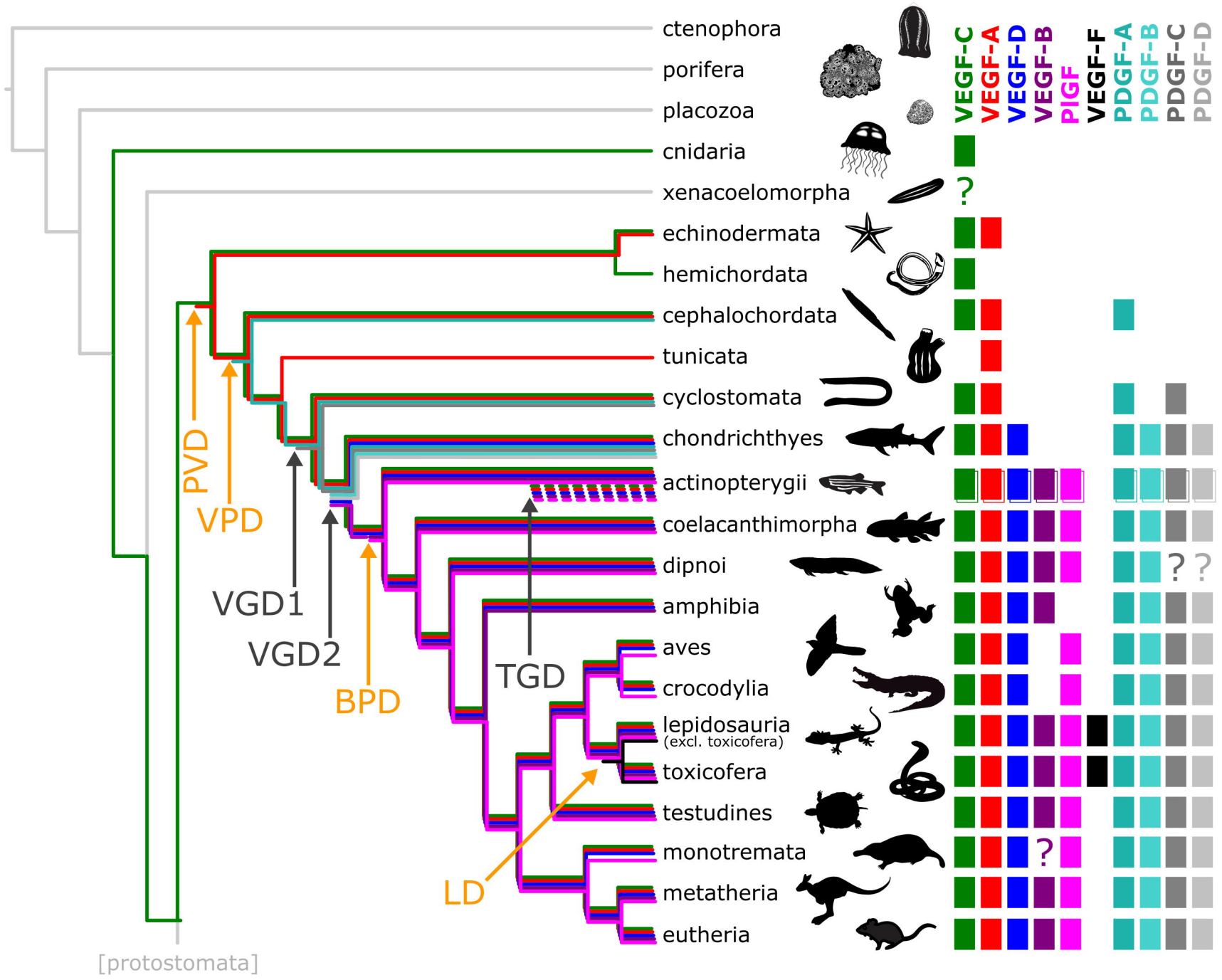
Supplementary Table 2. False-positive BLAST hits of PDGF/VEGF-like sequences from the phylum Porifera.

Supplementary Table 3. Fish PDGF/VEGF mRNA transcript contigs.




PVFs
human VEGFs
human PDGFs
VEGF Podocoryna carnea (AAS79435.1)
VEGF-C Pocillopora damicornis (XP_027059384.1)
VEGF-C Pocillopora damicornis (XP_027050375.1)
VEGF-C Acropora digitifera (XP_015760784.1)
VEGF-C Exaiptasia pallida (XP_020907748.1)
VEGF-C Stylophora pistillata (XP_022805345.1)
VEGF-C Stylophora pistillata (XP_022805359.1)
VEGF-C Hydra vulgaris (XP_002159836.1)
VEGF-C Hydra vulgaris (XP_012567141.1)
PDGF-B Hydra vulgaris (XP_002160342.2)
PVF-1 Caenorhabditis elegans
PVF2 Drosophila melanogaster
VEGF-C
VEGF-D
VEGF-A165
VEGF-B167
VEGF-A121
VEGF-B186
PlGF-3
PDGF-A
PDGF-B
PDGF-C
PDGF-D
VEGF Podocoryna carnea AAS79435.1
VEGF-C Pocillopora damicornis XP_027059384.1
VEGF-C Pocillopora damicornis XP_027050375.1
VEGF-C Acropora digitifera XP_015760784.1
VEGF-C Exaiptasia pallida XP_020907748.1
VEGF-C Stylophora pistillata XP_022805345.1
VEGF-C Stylophora pistillata XP_022805359.1
VEGF-C Hydra vulgaris XP_002161567.1
VEGF-C Hydra vulgaris XP_012567141.1
PDGF-B Hydra vulgaris XP_002160342.2
PVF-1 Caenorhabditis elegans Q9N413
PVF2 Drosophila melanogaster Q9VM43
VEGF-C Homo sapiens P49767
VEGF-D Homo sapiens O43915
VEGF-A Homo sapiens P15692
VEGF-B Homo sapiens P49765
PlGF-3 Homo sapiens P49763
PDGF-B Homo sapiens P01127
PDGF-C Homo sapiens Q9NRA1
N-terminal propeptide
VEGF/PDGF homology domain (VHD)
C-terminal propeptide with BR3P repeats
Hydrophobic domain
CUB domain
Hinge region


100%
50%
10%
VEGFC Pocillopora damicornis XP027050375.1
VEGFC Stylophora pistillata XP022805359.1
VEGFC Acropora digitifera XP015760784.1
VEGFC Pocillopora damicornis XP027059384.1
VEGFC Stylophora pistillata XP022805345.1
VEGFC Hydra vulgaris XP002161567.1
VEGFC Hydra vulgaris XP002159836.1
VEGF Podocoryna carnea AAS79435.1
VEGFC Exaiptasia pallida XP020907748.1
VEGFC Hydra vulgaris XP012567141.1
PVF1 C. elegans Q9N413
PVF-1 Brugia malayi CDP93194.1
UNC129 outgroup (TGF-β-like) C. elegans G5EBY8
Cnidaria
Nematodes
Insects
Vertebrates
PVF3 Drosophila melanogaster Q9VM42
PVF2 Drosophila melanogaster Q9VM43
PVF1 Drosophila melanogaster Q9VWP6
VEGFC Homo sapiens P49767
VEGFD Homo sapiens O43915
VEGFB167 Homo sapiens P49765-2
VEGFA Homo sapiens P15692
PDGFB Hydra vulgaris XP002160342.2
PIGF Homo sapiens P49763
PDGFC Homo sapiens Q9NRA1
PDGFA Homo sapiens P04085
PDGFB Homo sapiens P01127
PDGFD Homo sapiens Q9GZP0

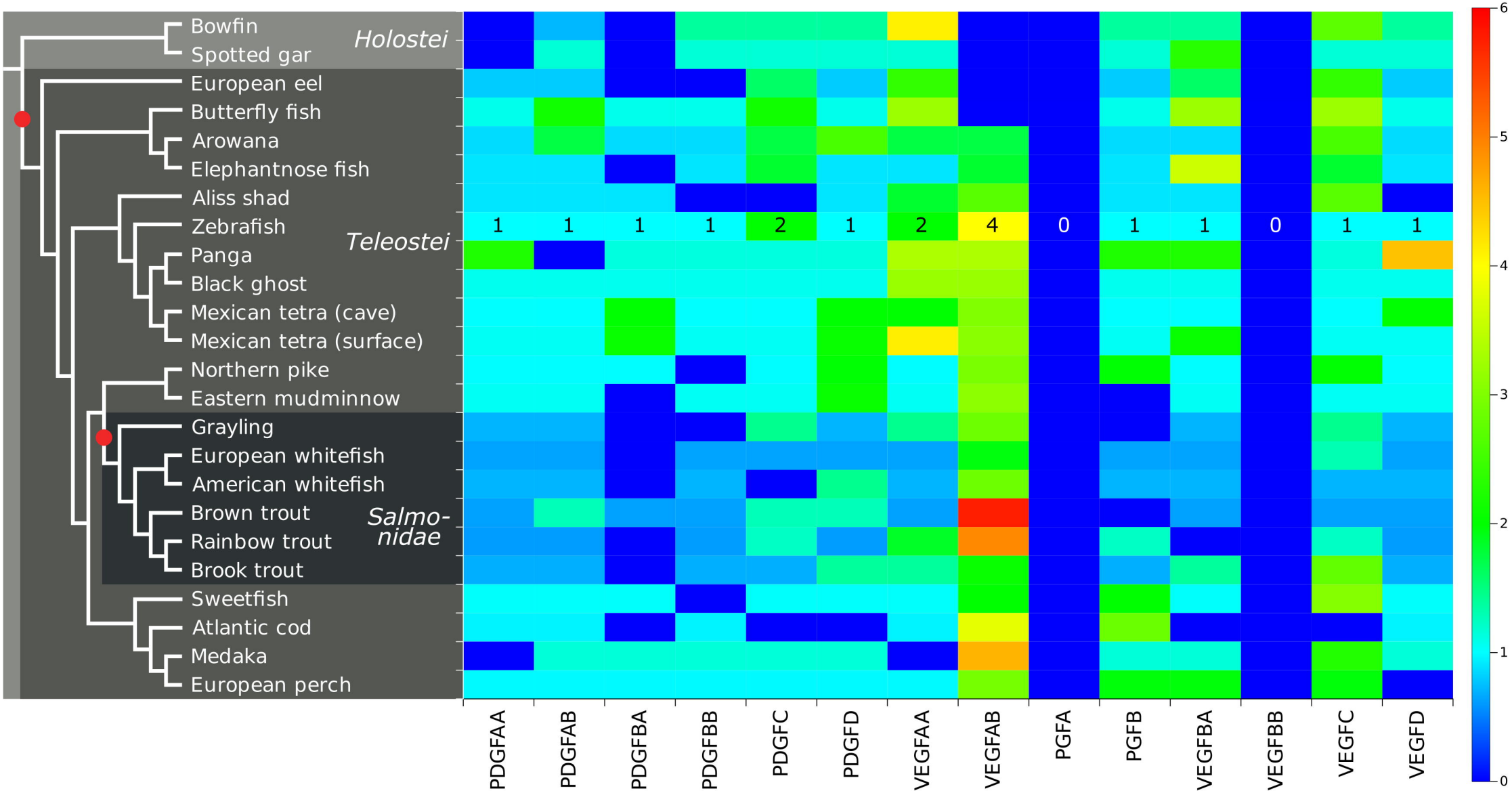


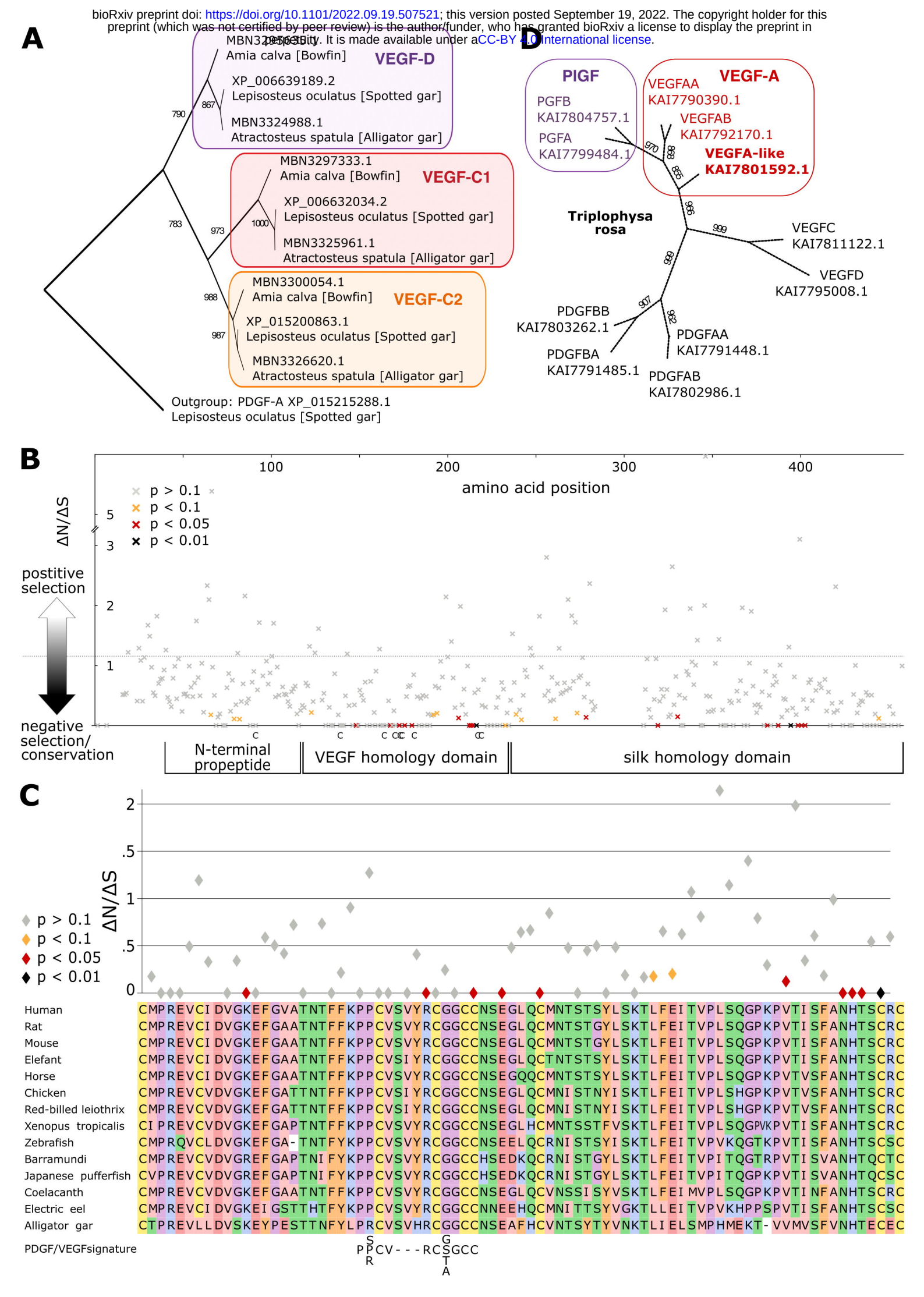
Whole Genome Duplications (WGD)

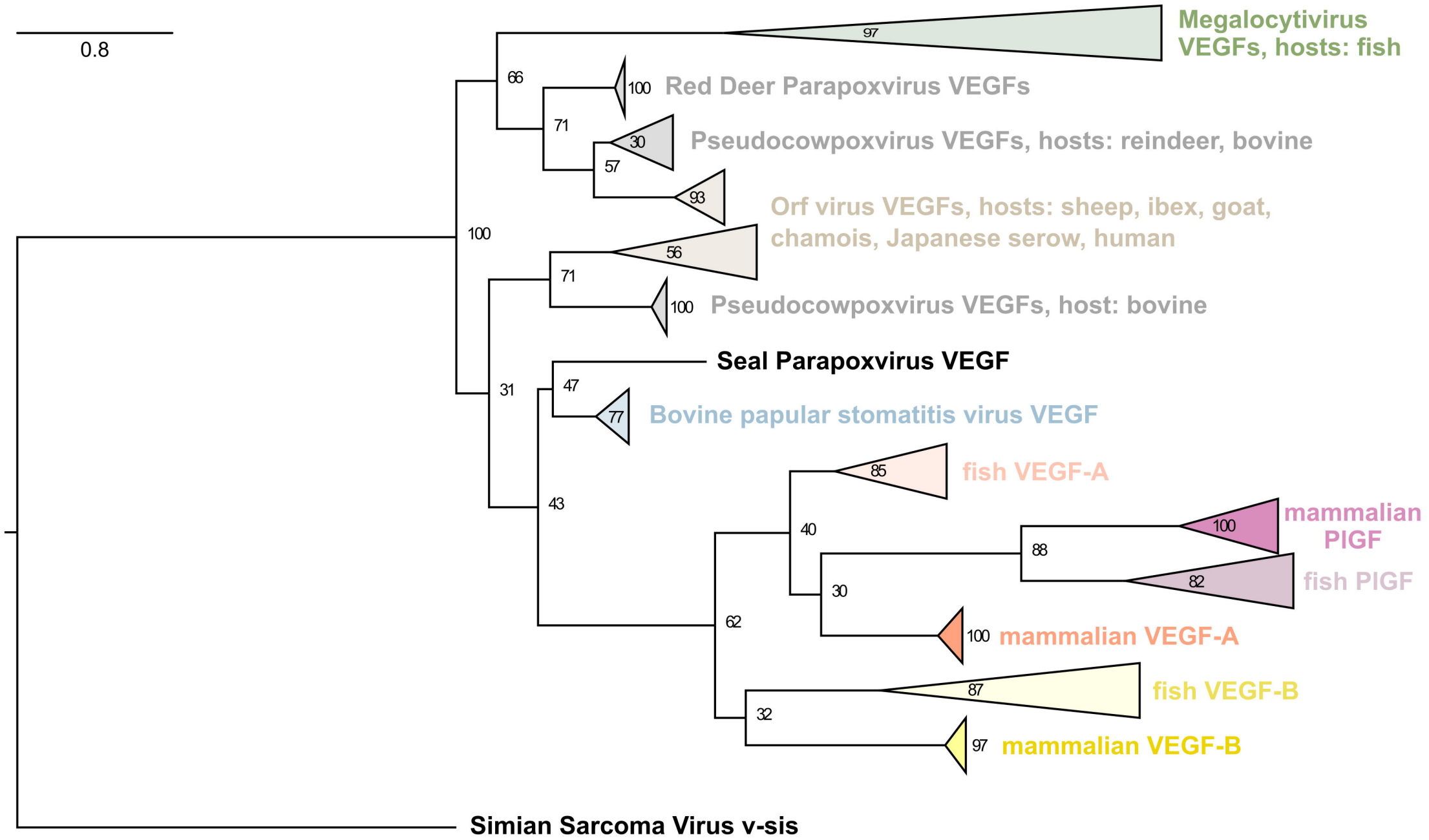
VGD1	1st vertebrate genome duplication
VGD2	2nd vertebrate genome duplication
TGD	teleost genome duplication

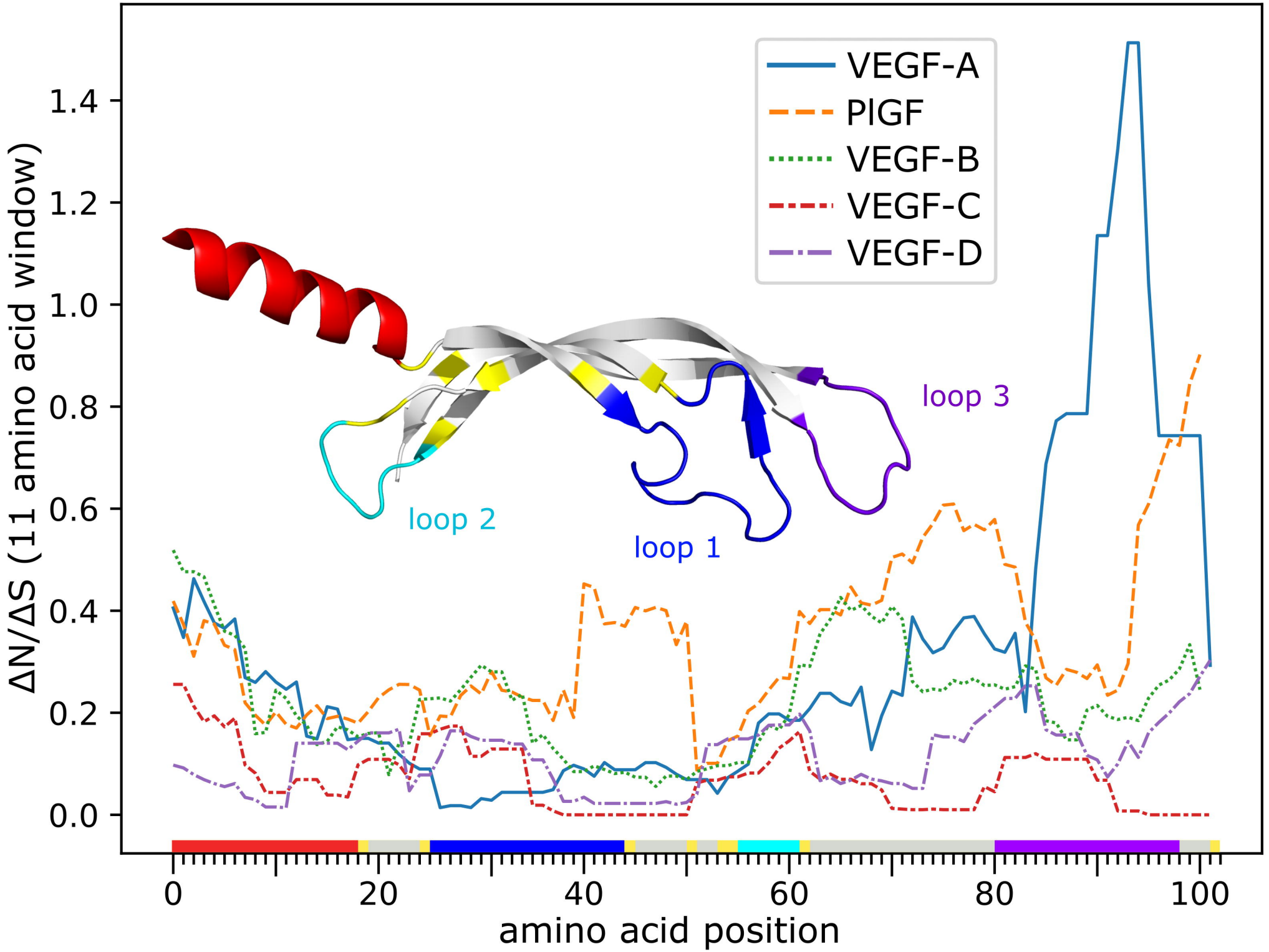
Hypothetical gene duplications

PVD	proto-PDGF/VEGF duplication
VPD	VEGF/PDGF duplication
BPD	VEGF-B/PlGF duplication
LD	lepidosaurian VEGF duplication


Bowfin
Spotted gar
Holostei
European eel
Butterfly fish
Arowana
Elephantnose fish
Aliss shad
Zebrafish
Panga
Black ghost
Mexican tetra (cave)
Mexican tetra (surface)
Northern pike
Eastern mudminnow
Teleostei
Grayling
European whitefish
American whitefish
Brown trout
Rainbow trout
Brook trout
Salmo-nidae
Sweetfish
Atlantic cod
Medaka
European perch
1 1 1 1 2 1 2 4 0 1 1 0 1 1
PDGFAA
PDGFAB
PDGFBA
PDGFBB
PDGFC
PDGFD
VEGFAA
VEGFAB
PGFA
PGFB
VEGFBA
VEGFBB
VEGFC
VEGFD
6
5
4
3
2
1
0




0.8
Megalocytivirus
VEGFs, hosts: fish
Red Deer Parapoxvirus VEGFs
Pseudocowpoxvirus VEGFs, hosts: reindeer, bovine
Orf virus VEGFs, hosts: sheep, ibex, goat,
chamois, Japanese serow, human
Pseudocowpoxvirus VEGFs, host: bovine
Seal Parapoxvirus VEGF
Bovine papular stomatitis virus VEGF
fish VEGF-A
mammalian
PlGF
fish PlGF
mammalian VEGF-A
fish VEGF-B
mammalian VEGF-B
Simian Sarcoma Virus v-sis
97
66
100
71
30
57
93
100
56
71
100
47
77
31
85
43
40
100
88
82
30
100
62
87
32
97


VEGF-A
PlGF
VEGF-B
VEGF-C
VEGF-D
loop 2
loop 1
loop 3
ΔN/ΔS (11 amino acid window)
amino acid position
0.0
0.2
0.4
0.6
0.8
1.0
1.2
1.4
0
20
40
60
80
100