Independent losses of the Hypoxia-Inducible Factor (HIF) Pathway within Crustacea

Allie M. Graham*, and Felipe S. Barreto

Department of Integrative Biology, Oregon State University, Corvallis, OR 97330, USA

*Corresponding author: grahaall@oregonstate.edu

Keywords: bHLH-PAS, Cirripedia, Copepoda, gene loss, transcription factor, oxygen sensing

© The Author(s) 2020. Published by Oxford University Press on behalf of the Society for Molecular Biology and Evolution. All rights reserved. For permissions, please e-mail: journals.permissions@oup.com

Downloaded from https://academic.oup.com/mbe/advance-article-abstract/doi/10.1093/molbev/msaa008/5719058 by Oregon State University, grahaall@oregonstate.edu on 04 February 2020

Abstract

Metazoans respond to hypoxic stress via the Hypoxia Inducible Factor (HIF) pathway, a mechanism thought to be extremely conserved due to its importance in monitoring cellular oxygen levels and regulating responses to hypoxia. However, recent work revealed that key members of the HIF pathway have been lost in specific lineages (a tardigrade and a copepod), suggesting this pathway is not as widespread in animals as previously assumed. Using genomic and transcriptomic data from 70 different species across 12 major crustacean groups, we assessed the degree to which the gene *HIF*α, the master regulator of the HIF pathway, was conserved. Mining of protein domains, followed by phylogenetic analyses of gene families, uncovered group-level losses of *HIF*α, including one across three orders within Cirripedia, and in three orders within Copepoda. For these groups, additional assessment showed losses of HIF repression machinery (EGLN, VHL). These results suggest the existence of alternative mechanisms for cellular response to low oxygen, and highlight these taxa as models useful for probing these evolutionary outcomes.

Downloaded from https://academic.oup.com/mbe/advance-article-abstract/doi/10.1093/molbev/msaa008/5719058 by Oregon State University, grahaall@oregonstate.edu on 04 February 2020

Introduction

Multicellular eukaryotes have evolved tight coordination of genes controlling specialized mechanisms that enhance O_2 uptake and distribution. These systems are capable of responding to changes in O_2 availability on local, organismal, and temporal levels. In general, one strategy for surviving in a dynamic environment is to maintain an array of oxygen responders that can regulate function across the spectrum of environmental changes which the organism might experience. These responses are mediated in part through the induction of Hypoxia-Inducible Factors (HIF), composed of regulatory components HIFα and HIFβ/ARNT (Rytkönen, et al. 2011; Graham and Presnell 2017). The mechanism controlling HIFα function is the primary cellular oxygen-sensing and responding pathway in animals. Under normoxia, the constitutively expressed HIFα subunit is hydroxylated by EGLN, a prolyl hydroxylase, which tags HIFα for proteasomal degradation by the von Hippel-Lindau (VHL), thus preventing activation of the pathway. While under low oxygen tension, the O_2-dependent EGLN is disabled, allowing HIFα to form the transcription factor dimer with HIFβ/ARNT. The functional heterodimer binds to DNA regions called HREs (hypoxia response elements) located upstream of certain target genes, effectively manipulating their expression patterns and regulating a network for a system-wide response to low-oxygen (Wenger, et al. 2005).

The HIF pathway appeared early in metazoan evolution (Loenarz, et al. 2011; Mills, et al. 2018), and was previously considered to be functionally and structurally conserved in animals, as it has been identified in every bilaterian as well as placozoan and cnidarian species in which it has been explicitly examined. Recent genomic analyses, however, have documented two losses of this gene, one in a tardigrade (Hashimoto, et al. 2016) and one in a haparcticoid copepod (Graham and Barreto 2019). These findings suggest that the HIF pathway may not be as universally present as previously thought. Therefore, examining the distribution of HIFα, EGLN and VHL loss across animals may identify taxa in which alternative mechanisms have evolved and provide new lines of research for understanding cellular physiological response to hypoxic stress.

The presence of the HIF pathway (via the presence of *HIFα*) is well known in certain crustacean species, but those only represent a few higher-level taxonomic groups, primarily Decapoda and Branchiopoda (Gorr, et al. 2004; Soñanez-Organis, et al. 2009; Hardy, et al. 2012). Crustaceans include many other animal groups that are dominant by biomass, distribution,

Downloaded from https://academic.oup.com/mbe/advance-article-abstract/doi/10.1093/molbev/msaa008/5719058 by Oregon State University, grahaall@oregonstate.edu on 04 February 2020

and species number, especially in aquatic systems (Price, et al. 2011), but genetic resources for this large subphylum as a whole are sparse compared to insects, a group nested within Crustacea. Major crustacean groups such as Amphipoda, Isopoda, Copepoda, and Cirripedia are commonly used for studies of environmental physiology, including organismal responses to hypoxia; the regulatory mechanisms underlying these phenotypes, however, are largely unexplored in these groups, and are likely assumed to be based on the HIF pathway. Based on the recent genomic findings mentioned above (Hashimoto, et al. 2016; Graham and Barreto 2019), we can no longer assume that this pathway is conserved and functional in every group.

In this study, we mined publicly available genomic and transcriptomic resources of 70 species representing most of the major clades within Crustacea, and assayed them for the presence of the HIF pathway. We show that there have been multiple losses of these interacting sets of genes (potentially independently), including specific subgroups of Copepoda (orders Harpacticoida, Cyclopoida and Siphonostomatoida), as well as across three orders of barnacles (Cirripedia: Sessilia, Pedunculata, and Kentrogonida). These results open up exciting lines of research that involve characterizing novel underlying mechanisms associated with oxygen sensing and homeostasis in these groups.

Results

We used a combination of Hidden Markov Model (HMM) searches, BLAST, InterproScan, and phylogenetic analyses to screen for the three main elements of the HIF-pathway (HIFα, EGLN, VHL) across multiple crustacean taxa. We first searched for HIFα across all species. In groups with evidence for loss of *HIFα*, we also searched for its repression machinery (EGLN, VHL), as a way to assess whether the canonical oxygen-sensing and -responding HIF pathway was lost. These proteins were identified based on the presence of specific protein domain elements, including (I) PAS domains, to identify HIFα copies, (II) P4HC domains, in an effort to identify EGLN copies and (III) the VHL-β domain to find VHL.

We used protein sequences from a combination of publicly available assembled genomes and transcriptomes, and several we assembled *de novo* from available raw data. In total, we analyzed data from 70 species across 12 higher-level taxonomic groups (**Table S1**). For some groups, we were able to include multiple members within multiple orders (e.g. Copepoda,

Downloaded from https://academic.oup.com/mbe/advance-article-abstract/doi/10.1093/molbev/msaa008/5719058 by Oregon State University, grahaall@oregonstate.edu on 04 February 2020

Amphipoda, and Decapoda). For others, despite using all known transcriptomic/genomic resources, the coverage was relatively low, with only 1- 4 representatives.

Because of the wide variation in quality and the necessarily incomplete nature of transcriptome and genome data available, our goal was not to assess gene loss at the individual species level. Therefore, we sampled multiple species within a group (when possible) in order to offset any possibility of individual transcriptome missing a HIFα for technical reasons. Ultimately, our determination of HIFα loss (and that of the EGLN or VHL) involved assessment at the group level – if one or more members of group showed evidence of a HIFα copy, the group was conservatively considered to have retained the gene (**Supplemental Methods**).

Overall, BLASTP searches and phylogenetic analyses of bHLH-PAS-containing proteins were highly concordant in identifying HIFα members, but in a few cases where BLAST was unable to detect a HIFα protein, the phylogenetic analysis was more sensitive to identification (**Table S1**). For example, using BLAST, no sequences from either *G. chevreuxi* and *H. gigas* in the Amphipoda showed a significant hit to HIFα, but the phylogenetic groupings identified putative HIFα sequences in these species (**Fig. S1**) that were then confirmed by InterproScan assessment. Representation of the other bHLH-PAS members (NPAS4, NCOA, ARNT, SIM, ARNTL, AhR, NPAS2/CLOCK, NPAS1/3 and Met) were also largely consistent across all groups. However, NPAS4 shows the potential for having been lost in Isopoda, Amphipoda, and Decapoda, but we caution that we did not further scrutinized this pattern since our focus is on HIFα. In addition, sequences which were outliers in the resulting trees were queried further for identity using InterproScan – these were detected to belong to other gene families which contained PAS domains plus domains not associated with the bHLH-PAS-containing families. These included PAS-containing (though not bHLH-PAS) cyclic nucleotide phosphodiesterases, HAMP/histidine kinases, MAP/microtubule affinity regulating kinases, and potassium voltage-gated channel subfamilies, and therefore were not considered potential missing HIF members.

All eight members of four infraorders in Decapoda (Brachyura, Caridae, Dendrobranchiata, Pleocyemata) contained HIFα members (**Fig. S2**). Some, yet not all, of the species in Amphipoda (**Fig. S1**), Isopoda (**Fig. S3**), Mysida and Euphausiacea (**Fig. S4**), Remipedia (**Fig.S5**), and Ostracoda (**Fig. S6**) contained HIFα members – those for which a HIFα was not found include *Grandidierella japonica*, *Eulimnogammarus cruentus Hyalella*

Downloaded from https://academic.oup.com/mbe/advance-article-abstract/doi/10.1093/molbev/msaa008/5719058 by Oregon State University, grahaall@oregonstate.edu on 04 February 2020

azteca (Amphipoda), *Xibalbanus tulumensis* (Remipedia), *Armadillidium vulgare* (Isopoda), *Neomysis awatschensis* (Mysida), and *Conchoecia obtusata* and *Eusarsiella sp.* (Ostracoda). However, those examples are likely due to low-quality sequence/assembly, especially for *N. awatschensis*, *C. obtusata* and *Eusarsiella*, which had a low total number of PAS-containing proteins (**Table S1**). For *Xibalbanus tulumensis,* although a HIFα was identified through BLASTP and InterproScan (Xiba_DN232543_c1_g1_i2; 354 aa), no HIFα grouped with the other remipede species and with anchor HIFα; that sequence instead grouped with NPAS2/CLOCK and suggests a divergent HIF sequence or a misassembled transcript. For *Hyalella azteca*, a HIFα was identified through BLASTP and InterproScan (Hyaztec_GEHV01020431.1; 148 aa), yet the same sequence grouped instead with the Met gene family. In general, these "losses" likely are not biological, but instead represent absences due to issues in sequencing and transcriptome assembly, especially given other members of the same taxonomic group showed clear presence of HIFα. Therefore, our results suggest HIF pathway was retained in Amphipoda (**Fig. S1**), Isopoda (**Fig. S3**), Decapoda (**Fig. S2**), Mysida and Euphausiacea (**Fig. S4**), Remipedia (**Fig. S5**), and Branchiopoda, Cephalocarida, Stomatopoda, and Ostracoda (**Fig. S6**), since we observed that a HIFα family member was unambiguously present in at least one species in each group (**Table S1**).

In the Copepoda, only 5 of the 18 species examined exhibited a copy of HIFα and EGLN. However, these 5 species are all contained within the order Calanoida; therefore, all species available and analyzed from orders Cyclopoida, Harpacticoida and Siphonostomatoida showed no evidence of a canonical HIFα (**Fig. 1; Table S1**) or EGLN (**Fig. S7; Table S2**) protein at any stage of the analyses. Initial screening found evidence for a HIFα and EGLN in the three species of Siphonostomatoida and in one Cyclopoid (*Lernaea cyprinacea*), but further BLAST analyses revealed that these sequences had >97% similarity and > 90% coverage to fish GenBank accessions. Since these four species are obligate fish parasites, this pattern is consistent with contamination from host tissue during RNA isolation. We implemented a step to remove contaminant proteins from these assemblies through a BLAST search against a custom database that included a transcriptome from host fish species (**Supplemental Methods**). After this step, we re-screened the proteomes and found no evidence of a HIFα or EGLN in those groups.

Downloaded from https://academic.oup.com/mbe/advance-article-abstract/doi/10.1093/molbev/msaa008/5719058 by Oregon State University, grahaall@oregonstate.edu on 04 February 2020

With regard to VHL, patterns of loss were spotty – most of Harpacticoida (except *T. californicus*), Calanoida (except *E. affinis*) and Cyclopoida (except *P. nana* and *O. nana*) showed evidence for the presence of a VHL, while none was detected in Siphonostomatoida (**Fig. 3; Fig. 4B; Table S3**). This may suggest either VHL was lost independently in species-specific manner, or that it was simply not captured in the transcriptomic data.

In Cirripedia, all species surveyed across three orders (Sessilia, Pedunculata, Kentrogonida) showed no evidence for the presence of HIFα (**Fig. 2; Table S1**), EGLN (**Fig. S8; Table S2)** or VHL members (**Fig. 3; Fig. 4A; Table S3**).

As a final check, we performed a reciprocal blast between a known decapod HIFα (GenBank accession ACU30154.1), the transcriptome assemblies of the copepod and barnacle taxa, and the Uniprot/Swiss-Prot database. This was done at the transcript stage of the assemblies, instead of the predicted proteins, so that even poorly assembled possible fragments of the gene would not be missed (**Supplemental Methods**). This analysis was entirely consistent with the full screen above (**Table S4**).

Discussion

Hypoxia is a critical physiological constraint, with highly a conserved molecular pathway that monitors and responds to periods of low oxygen. Many crustacean taxa are subjects of studies of hypoxia physiology and tolerance, but the regulatory mechanisms of hypoxia response are largely unexplored in most groups. The presence of *HIFα* is well known in many crustacean species, but those only represent a few higher-level taxonomic groups, primarily Decapoda and Branchiopoda (Gorr, et al. 2004; Soñanez-Organis, et al. 2009; Hardy, et al. 2012). Our analyses in these two groups unsurprisingly showed clear evidence of this gene in all species, and the single target branchiopod examined (*Triops newberryi*) grouped with the model branchiopod *Daphnia pulex* which was included among "anchor" taxa. In the light of our results, presence of a HIF pathway can no longer be assumed, since it is clear that critical members of this pathway, including the main transcription factor (*HIFα*) and part of its regulatory machinery (*EGLN* and *VHL*), has been lost multiple times in Crustacea (**Fig. 3**).

The framework for our understanding of the dynamics and importance of the HIF pathway is largely in the context of vertebrate species, where HIFα members of the pathway have been fully integrated into various elements of embryonic development. However, the

Downloaded from https://academic.oup.com/mbe/advance-article-abstract/doi/10.1093/molbev/msaa008/5719058 by Oregon State University, grahaall@oregonstate.edu on 04 February 2020

physiology, and thus the oxygen requirements of a variety of invertebrates likely differ, with the utility of the HIF pathway differing as well (Harrison 2015; Harrison, et al. 2018). This leaves open the potential for gene loss, or pseudogenization, as a result of unique evolutionary histories and lineage-specific requirements. In general, individual gene loss during speciation and macroevolution is pervasive, although which genes are being lost depends on their individual "dispensability", or their effect on fitness (Albalat and Cañestro 2016). Re-wiring of regulatory networks influenced by transcription factors happens readily (Bhardwaj, et al. 2010; Lynch, et al. 2011; De Smet and Van de Peer 2012; Erkenbrack and Davidson 2015), but a loss of the main regulatory machinery likely means a loss of the pathway. Losses of such a fundamental eukaryotic pathway such as the HIF are largely undiscovered. Prior assessment of this pathway has included a large contingent of invertebrate members, but has remained biased towards terrestrial hexapods, resulting in a substantial gap of knowledge about aquatic crustaceans. Our results are an in-depth assessment of the HIF pathway by examining the presence/absence of its primary transcription factor and its regulatory elements, based on currently available genomic and transcriptomic resources available. Ultimately, we found several instances of independent *HIF*α, *EGLN* and *VHL* loss, within Copepoda and Cirripedia, which were previously unknown.

As mentioned above, our analyses come with the general caveat of the possibility that incompleteness of genome or transcriptome assemblies creates artefactual gene "losses". Our approach to minimize false negative findings (i.e. false losses) was to examine multiple species within each group when possible, and to conservatively claim loss of a gene only when the it was not detected in all species examined within the respective group. Also, while for most groups we examined nearly all species with resources available, these represent only a small fraction of total species in these taxa. Thus, our generalizations regarding loss (or maintenance) of HIF pathway genes in each group are to be interpreted cautiously and should be subject to re-examination when more taxa are sequenced. Finally, the transcriptome data from parasitic copepods species had a large amount of contamination from their fish hosts. We chose to retain these assemblies because in some cases (e.g. order Siphonostomatoida), all species examined are fish parasites, and we did not want to exclude a full taxonomic order. We computationally filtered likely contaminant sequences from these assemblies before performing our protein screening but results from these parasitic species should be interpreted with care.

Downloaded from https://academic.oup.com/mbe/advance-article-abstract/doi/10.1093/molbev/msaa008/5719058 by Oregon State University, grahaall@oregonstate.edu on 04 February 2020

Independent HIF pathway losses within Cirripedia and Copepoda

The phenotypic response to hypoxia in barnacles has been studied to some degree in adults and juveniles. The sessile adults are routinely out of the water during low tides in intertidal species (Davenport and Irwin 2003), and both adults and motile juveniles may encounter pelagic or benthic hypoxic zones (Gilbert, et al. 2010). However, the underlying cellular response to hypoxia has been only sparsely documented (López, et al. 2003; Desai and Prakash 2009; Campanati, et al. 2015), with none describing the transcriptional landscape in any barnacle species. In addition, the current literature in barnacles does not contain work on the HIF pathway explicitly, or its role in mediating the response to low oxygen stress. Our results suggest a loss of *HIF*α, *EGLN* and *VHL* across all species examined in the barnacle orders Sessilia, Pedunculata, and Kentrogonida (**Fig. 2; Fig. S8; Table S1; Table S2; Table S3**). The three orders form a clade along with the order Ibliformes. Therefore, it is likely that only one loss of each gene occurred at the base of this clade. However, given their phylogenetic branching order (Pérez-Losada, et al. 2008), and the lack of transcriptomic data for Ibliformes, the possibility of two or more independent losses cannot be excluded at this point (**Fig. 4A**). In addition, other taxonomic orders within Cirripedia exist, but no transcriptome/genome data were available, so we cannot hypothesize the breadth of this pattern across this group.

As a group, copepods are one of the most successful metazoan taxa, occurring in nearly every aquatic habitat, both in freshwater and saltwater, planktonic and benthic divisions, from polar waters to hot springs, and in bodies of water of wide array of sizes such as swamps, ephemeral ponds, damp moss, and phytoelmata of plants (i.e. water filled recesses), sinkholes and caves, to even being obligate parasites (Boxshall 2000; Hamilton IV, et al. 2000; Boxshall and Defaye 2007; Kiørboe 2010). The most studied forms are pelagic species inhabiting ocean plankton, where they can account for over 80% of total plankton abundance and form an essential trophic level of marine food webs. With such a wide array of environments where low oxygen stress may be encountered, retaining an important pathway would seem crucial; yet, three of the four orders of Copepoda examined (**Fig. 1; Fig. S7**) appear to have lost the use of the HIF pathway, at least in its canonical form. The specific phylogenetic relationship among the three orders examined is not resolved (Eyun 2017; Khodami, et al. 2017), our findings suggest a single loss at the common ancestor of the three orders (**Fig. 4B**). The three orders are diverse in

Downloaded from https://academic.oup.com/mbe/advance-article-abstract/doi/10.1093/molbev/msaa008/5719058 by Oregon State University, grahaall@oregonstate.edu on 04 February 2020

number of species and in the range of environments they inhabit; hence, we cannot speculate on ecological commonalities that might explain their ability to thrive without this critical pathway.

Both barnacles and copepods have converged on the loss of two of the most important regulators of the HIF pathway (*HIFα* and *EGLN*). In addition, there is also potential losses of VHL in both groups, though the extent to which this is the case, seems linage specific; the fact that VHL has also not been fully lost in copepods may be partially driven by the fact it is known to have HIF-independent functions (Calzada, et al. 2006; Berndt, et al. 2009; Gossage, et al. 2015; Nicholson, et al. 2019), thus explaining why VHL seems to have been retained. Some recent work suggests that Copepoda and Thecostraca, which includes Cirripedia, are monophyletic and form the clade Hexanauplia (Oakley, et al. 2012; Lozano-Fernandez, et al. 2019). This does not change our observed pattern that the loss of the HIF pathway occurred independently for barnacles and copepods, although it may suggest that this pathway is more prone to loss within this particular lineage compared to other lineages sampled.

The independent losses of the HIF pathway across two abundant groups of crustaceans raise numerous questions about how these animals are able to sense and regulate oxygen tension on a continuous basis. Alternative regulatory mechanisms of hypoxia response have never been proposed, since our current understanding of this cellular process is based on the regulation by a HIF heterodimer involving a HIFα and a HIFβ subunits, and their repression machinery (EGLN and VHL). The taxa identified in this study will serve as models for new lines of cellular, genetic, and physiological studies aimed at discovering such alternative mechanisms. We hypothesize that a different transcription factor may be involved in regulating a HIF-like pathway, and that this protein is either new (i.e. lineage-specific) or has been co-opted from other existing stress response pathways. Finally, another exciting line of questions include to what degree these lineages have converged upon certain mechanisms, from broad physiological to molecular and genetic scales.

Downloaded from https://academic.oup.com/mbe/advance-article-abstract/doi/10.1093/molbev/msaa008/5719058 by Oregon State University, grahaall@oregonstate.edu on 04 February 2020

Author Contributions

AMG and FB designed the research, analyzed the data, and wrote the manuscript.

Acknowledgements

The material in this article is based upon work supported by the NSF Postdoctoral Research Fellowship in Biology under Grant No. 1812103 to AMG. The authors would also like to acknowledge the core facilities at Center for Genome Research and Biocomputing at Oregon State University, where computing was performed.

Data and Code Availability

The genome and transcriptome sequencing reads are available at either NCBI Short Read Archive (SRA) or NCBI Transcriptome Shotgun Assembly Sequence Database (TSA) (**Table S1**). Protein files from the *de novo* assembled transcriptome assemblies, consensus newick trees and alignments are available on Dryad (DOI: doi:10.5061/dryad.8cz8w9gkq).

Supplemental Figure and Table Legends

Figure S1: Maximum-likelihood phylogenetic tree of bHLH-PAS proteins from Amphipoda containing 12 species (in blue), across 4 infraorders (RAxML; 100 bootstraps). See Table S1 for the full species names and their classification. In orange are the "anchor" sequences of HIFα from insects and *Daphnia pulex* (Dpu) - including *Anophales gambii* (Aga), *Bombyx mori* (Bmo), *Apis mellifera* (Ame), *Acyrthosiphon pisum* (Api), *Dendroctonus ponderosae* (Dpon), *Drosophila melanogaster* (Dme), *Nasonia vitripennis* (Nvi), and *Tribolium castaneum* (Tca).

Figure S2: Maximum-likelihood phylogenetic tree of bHLH-PAS proteins from the Decapoda containing 8 species (in blue), across 4 infraorders (RAxML; 100 bootstraps). See Table S1 for the full species names and their classification. In orange are the "anchor" sequences of HIFα from insects and *Daphnia pulex* (Dpu) - including *Anophales gambii* (Aga), *Bombyx mori* (Bmo), *Apis mellifera* (Ame), *Acyrthosiphon pisum* (Api), *Dendroctonus ponderosae* (Dpon), *Drosophila melanogaster* (Dme), *Nasonia vitripennis* (Nvi), and *Tribolium castaneum* (Tca).

Downloaded from https://academic.oup.com/mbe/advance-article-abstract/doi/10.1093/molbev/msaa008/5719058 by Oregon State University, grahaall@oregonstate.edu on 04 February 2020

Figure S3: Maximum-likelihood phylogenetic tree of bHLH-PAS proteins from the order Isopoda with 6 species (in blue), across 2 suborders (RAxML; 100 bootstraps). See Table S1 for the full species names and their classification. In orange are the "anchor" sequences of HIFα from insects and *Daphnia pulex* (Dpu) - including *Anophales gambii* (Aga), *Bombyx mori* (Bmo), *Apis mellifera* (Ame), *Acyrthosiphon pisum* (Api), *Dendroctonus ponderosae* (Dpon), *Drosophila melanogaster* (Dme), *Nasonia vitripennis* (Nvi), and *Tribolium castaneum* (Tca).

Figure S4: Maximum-likelihood phylogenetic tree of bHLH-PAS proteins from the order Mysida (n = 3 species, across 2 families) and order Euphausiacea (n = 2 species, in 1 family) (RAxML; 100 bootstraps). See Table S1 for the full species names and their classification. In orange are the "anchor" sequences of HIFα from insects and *Daphnia pulex* (Dpu) - including *Anophales gambii* (Aga), *Bombyx mori* (Bmo), *Apis mellifera* (Ame), *Acyrthosiphon pisum* (Api), *Dendroctonus ponderosae* (Dpon), *Drosophila melanogaster* (Dme), *Nasonia vitripennis* (Nvi), and *Tribolium castaneum* (Tca).

Figure S5: Maximum-likelihood phylogenetic tree of bHLH-PAS proteins from the class Remipedia with 4 species, in 1 order (RAxML; 100 bootstraps). See Table S1 for the full species names and their classification. In orange are the "anchor" sequences of HIFα from insects and *Daphnia pulex* (Dpu) - including *Anophales gambii* (Aga), *Bombyx mori* (Bmo), *Apis mellifera* (Ame), *Acyrthosiphon pisum* (Api), *Dendroctonus ponderosae* (Dpon), *Drosophila melanogaster* (Dme), *Nasonia vitripennis* (Nvi), and *Tribolium castaneum* (Tca).

Figure S6: Maximum-likelihood phylogenetic tree of bHLH-PAS proteins from "orphan" groups: class Ostracoda (n = 3 species, across 3 subclasses), class Cephalocarida (n = 1 species), class Branchiopoda (n = 1 species), and order Stomatopoda (n = 1 species) (RAxML; 100 bootstraps). See Table S1 for the full species names and their classification. In orange are the "anchor" sequences of HIFα from insects and *Daphnia pulex* (Dpu) - including *Anophales gambii* (Aga), *Bombyx mori* (Bmo), *Apis mellifera* (Ame), *Acyrthosiphon pisum* (Api), *Dendroctonus ponderosae* (Dpon), *Drosophila melanogaster* (Dme), *Nasonia vitripennis* (Nvi), and *Tribolium castaneum* (Tca).

Downloaded from https://academic.oup.com/mbe/advance-article-abstract/doi/10.1093/molbev/msaa008/5719058 by Oregon State University, grahaall@oregonstate.edu on 04 February 2020

Figure S7: Maximum-likelihood phylogenetic tree of P4HC proteins of Copepoda, containing 18 species (in blue) across 4 orders (RAxML; 200 bootstraps). See Table S1 for the full species names and their classification. In orange are the "anchor" sequences of EGLN from insects and *Daphnia pulex* (Dpu) - including *Anophales gambii* (Aga), *Bombyx mori* (Bmo), *Apis mellifera* (Ame), *Acyrthosiphon pisum* (Api), *Dendroctonus ponderosae* (Dpon), *Drosophila melanogaster* (Dme), *Nasonia vitripennis* (Nvi), and *Tribolium castaneum* (Tca).

Figure S8: Maximum-likelihood phylogenetic tree of P4HC proteins of Cirripedia, containing 10 species (in blue) across 3 orders (RAxML; 200 bootstraps). See Table S1 for the full species names and their classification. In orange are the "anchor" sequences of EGLN from insects and *Daphnia pulex* (Dpu) - including *Anophales gambii* (Aga), *Bombyx mori* (Bmo), *Apis mellifera* (Ame), *Acyrthosiphon pisum* (Api), *Dendroctonus ponderosae* (Dpon), *Drosophila melanogaster* (Dme), *Nasonia vitripennis* (Nvi), and *Tribolium castaneum* (Tca).

Table S1: Assessment of the presence of HIFα sequences. Full list of taxa used. Dataset type refers to the source from which protein sequences were obtained: TSA = Transcriptome Shotgun Assembly from NCBI; SRA = raw Illumina reads from NCBI's Sequence Read Archive, used by the authors for transcriptome assembly de novo; genome = proteins predicted from genome assemblies. General quality of transcriptome assembly and downstream identification of HIFα sequences were assessed (in part) using number of PAS domain containing proteins through the use of HMMER (# PAS). The Identification of HIFα sequences were ultimately assessed through the use of BLAST, InterproScan and phylogenetic tree grouping.

Table S2: Assessment of the presence of EGLN sequences for taxa without a HIFα. Dataset type refers to the source from which protein sequences were obtained: TSA = Transcriptome Shotgun Assembly from NCBI; SRA = raw Illumina reads from NCBI's Sequence Read Archive, used by the authors for transcriptome assembly de novo. General quality of transcriptome assembly and downstream identification of EGLN sequences were assessed (in part) using number of P4HC domain containing proteins through the use of HMMER. The Identification of EGLN sequences were ultimately assessed through the use of BLAST, InterproScan and phylogenetic tree grouping.

Downloaded from https://academic.oup.com/mbe/advance-article-abstract/doi/10.1093/molbev/msaa008/5719058 by Oregon State University, grahaall@oregonstate.edu on 04 February 2020

Table S3: Assessment of the presence of VHL sequences for taxa without a HIFα. Dataset type refers to the source from which protein sequences were obtained: TSA = Transcriptome Shotgun Assembly from NCBI; SRA = raw Illumina reads from NCBI's Sequence Read Archive, used by the authors for transcriptome assembly de novo. General quality of transcriptome assembly and downstream identification of VHL sequences were assessed (in part) using number of VHL domain containing proteins through the use of HMMER. The Identification of VHL sequences were ultimately assessed through the use of BLAST, InterproScan and phylogenetic tree grouping.

Table S4: Results of reciprocal blast used as a final approach to identify HIFα protein in Copepoda and Cirripedia. The validated protein sequence from a decapod shrimp (GenBank accession ACU30154.1) was first used as query in a TBLASTN search against the full transcriptome assembly of each target species. The best hits from the target species were then used as BLASTX query against the full Uniprot/Swiss-prot curated database.

Downloaded from https://academic.oup.com/mbe/advance-article-abstract/doi/10.1093/molbev/msaa008/5719058 by Oregon State University, grahaall@oregonstate.edu on 04 February 2020

References

Albalat R, Cañestro C. 2016. Evolution by gene loss. *Nat. Rev. Genet.* 17:379.

Berndt JD, Moon RT, Major MB. 2009. β-catenin gets jaded and von Hippel-Lindau is to blame. *Trends Biochem. Sci.* 34:101-104.

Bhardwaj N, Kim PM, Gerstein MB. 2010. Rewiring of transcriptional regulatory networks: hierarchy, rather than connectivity, better reflects the importance of regulators. *Sci. Signal.* 3:ra79-ra79.

Boxshall GA. 2000. Parasitic copepods (Copepoda: Siphonostomatoida) from deep-sea and mid-water fishes. *Syst. Parasitol.* 47:173-181.

Boxshall GA, Defaye D. 2007. Global diversity of copepods (Crustacea: Copepoda) in freshwater. In. Freshwater Animal Diversity Assessment: Springer. p. 195-207.

Calzada MJ, Esteban MA, Feijoo-Cuaresma M, Castellanos MC, Naranjo-Suárez S, Temes E, Méndez F, Yánez-Mo M, Ohh M, Landázuri MO. 2006. von Hippel-Lindau tumor suppressor protein regulates the assembly of intercellular junctions in renal cancer cells through hypoxia-inducible factor–independent mechanisms. *Cancer Res.* 66:1553-1560.

Campanati C, Yip S, Lane A, Thiyagarajan V. 2015. Combined effects of low pH and low oxygen on the early-life stages of the barnacle Balanus amphitrite. *ICES J. Mar. Sci.* 73:791-802.

Davenport J, Irwin S. 2003. Hypoxic life of intertidal acorn barnacles. *Mar. Biol.* 143:555-563.

De Smet R, Van de Peer Y. 2012. Redundancy and rewiring of genetic networks following genome-wide duplication events. *Curr. Opin. Plant Biol.* 15:168-176.

Desai DV, Prakash S. 2009. Physiological responses to hypoxia and anoxia in Balanus amphitrite (Cirripedia: Thoracica). *Mar. Ecol. Prog. Ser.* 390:157-166.

Erkenbrack EM, Davidson EH. 2015. Evolutionary rewiring of gene regulatory network linkages at divergence of the echinoid subclasses. *PNAS* 112:E4075-E4084.

Ewers-Saucedo C, Owen CL, Pérez-Losada M, Høeg JT, Glenner H, Chan BK, Crandall KA. 2019. Towards a barnacle tree of life: integrating diverse phylogenetic efforts into a comprehensive hypothesis of thecostracan evolution. *PeerJ* 7:e7387.

Eyun S-i. 2017. Phylogenomic analysis of Copepoda (Arthropoda, Crustacea) reveals unexpected similarities with earlier proposed morphological phylogenies. *BMC Evol. Biol.* 17:23.

Gilbert D, Rabalais N, Diaz R, Zhang J. 2010. Evidence for greater oxygen decline rates in the coastal ocean than in the open ocean. *Biogeosciences* 7:2283.

Downloaded from https://academic.oup.com/mbe/advance-article-abstract/doi/10.1093/molbev/msaa008/5719058 by Oregon State University, grahaall@oregonstate.edu on 04 February 2020

Gorr TA, Cahn JD, Yamagata H, Bunn HF. 2004. Hypoxia-induced synthesis of hemoglobin in the crustacean *Daphnia magna* is hypoxia-inducible factor-dependent. *J. Biol. Chem.* 279:36038-36047.

Gossage L, Eisen T, Maher ER. 2015. VHL, the story of a tumour suppressor gene. *Nat. Rev. Cancer* 15:55-64.

Graham AM, Barreto FS. 2019. Loss of the HIF pathway in a widely distributed intertidal crustacean, the copepod Tigriopus californicus. *PNAS* 116:12913-12918.

Graham AM, Presnell JS. 2017. Hypoxia Inducible Factor (HIF) transcription factor family expansion, diversification, divergence and selection in eukaryotes. *PLoS One* 12:e0179545.

Hamilton IV R, Reid JW, Duffield RM. 2000. Rare copepod, Paracyclops canadensis (Willey), common in leaves of Sarracenia purpurea L. *Northeast. Nat.* 7:17-24.

Hardy KM, Follett CR, Burnett LE, Lema SC. 2012. Gene transcripts encoding hypoxia-inducible factor (HIF) exhibit tissue-and muscle fiber type-dependent responses to hypoxia and hypercapnic hypoxia in the Atlantic blue crab, *Callinectes sapidus*. *Comp. Biochem. Physiol., Part A Mol. Integr. Physiol.* 163:137-146.

Harrison JF. 2015. Handling and use of oxygen by pancrustaceans: conserved patterns and the evolution of respiratory structures. *Integr. Comp. Biol.* 55:802-815.

Harrison JF, Greenlee KJ, Verberk WC. 2018. Functional hypoxia in insects: definition, assessment, and consequences for physiology, ecology, and evolution. *Annu. Rev. Entomol.* 63:303-325.

Hashimoto T, Horikawa DD, Saito Y, Kuwahara H, Kozuka-Hata H, Shin T, Minakuchi Y, Ohishi K, Motoyama A, Aizu T. 2016. Extremotolerant tardigrade genome and improved radiotolerance of human cultured cells by tardigrade-unique protein. *Nat. Commun.* 7:12808.

Khodami S, McArthur JV, Blanco-Bercial L, Arbizu PM. 2017. Molecular phylogeny and revision of copepod orders (Crustacea: Copepoda). *Sci. Rep.* 7:9164.

Kiørboe T. 2010. What makes pelagic copepods so successful? *J. Plankton Res.* 33:677-685.

Loenarz C, Coleman ML, Boleininger A, Schierwater B, Holland PW, Ratcliffe PJ, Schofield CJ. 2011. The hypoxia-inducible transcription factor pathway regulates oxygen sensing in the simplest animal, *Trichoplax adhaerens*. *EMBO Rep.* 12:63-70.

López D, Castro J, González M, Simpfendorfer R. 2003. Physiological responses to hypoxia and anoxia in the giant barnacle, Austromegabalanus psittacus (Molina, 1782). *Crustaceana (Leiden)* 76:533-546.

Lozano-Fernandez J, Giacomelli M, Fleming JF, Chen A, Vinther J, Thomsen PF, Glenner H, Palero F, Legg DA, Iliffe TM. 2019. Pancrustacean evolution illuminated by taxon-rich genomic-scale data sets with an expanded remipede sampling. *Genome Biol Evol* 11:2055-2070.

Downloaded from https://academic.oup.com/mbe/advance-article-abstract/doi/10.1093/molbev/msaa008/5719058 by Oregon State University, grahaall@oregonstate.edu on 04 February 2020

Lynch VJ, Leclerc RD, May G, Wagner GP. 2011. Transposon-mediated rewiring of gene regulatory networks contributed to the evolution of pregnancy in mammals. *Nat. Genet.* 43:1154.

Mills DB, Francis WR, Vargas S, Larsen M, Elemans CP, Canfield DE, Wörheide G. 2018. The last common ancestor of animals lacked the HIF pathway and respired in low-oxygen environments. *Elife* 7:e31176.

Nicholson HE, Tariq Z, Housden BE, Jennings RB, Stransky LA, Perrimon N, Signoretti S, Kaelin WG. 2019. HIF-independent synthetic lethality between CDK4/6 inhibition and VHL loss across species. *Sci. Stke.* 12:eaay0482.

Oakley TH, Wolfe JM, Lindgren AR, Zaharoff AK. 2012. Phylotranscriptomics to bring the understudied into the fold: monophyletic ostracoda, fossil placement, and pancrustacean phylogeny. *Mol. Biol. Evol.* 30:215-233.

Pérez-Losada M, Harp M, Høeg JT, Achituv Y, Jones D, Watanabe H, Crandall KA. 2008. The tempo and mode of barnacle evolution. *Mol Phylogenet Evol* 46:328-346.

Pérez-Losada M, Høeg JT, Crandall KA. 2009. Remarkable convergent evolution in specialized parasitic Thecostraca (Crustacea). *BMC biology* 7:15.

Price PW, Denno RF, Eubanks MD, Finke DL, Kaplan I. 2011. Insect ecology: behavior, populations and communities: Cambridge University Press.

Regier JC, Shultz JW, Zwick A, Hussey A, Ball B, Wetzer R, Martin JW, Cunningham CW. 2010. Arthropod relationships revealed by phylogenomic analysis of nuclear protein-coding sequences. *Nature* 463:1079-1083.

Rytkönen KT, Williams TA, Renshaw GM, Primmer CR, Nikinmaa M. 2011. Molecular evolution of the metazoan PHD–HIF oxygen-sensing system. *Mol. Biol. Evol.* 28:1913-1926.

Schwentner M, Combosch DJ, Nelson JP, Giribet G. 2017. A phylogenomic solution to the origin of insects by resolving crustacean-hexapod relationships. *Curr. Biol.* 27:1818-1824. e1815.

Schwentner M, Richter S, Rogers DC, Giribet G. 2018. Tetraconatan phylogeny with special focus on Malacostraca and Branchiopoda: highlighting the strength of taxon-specific matrices in phylogenomics. *Proceedings of the Royal Society B: Biological Sciences* 285:20181524.

Soñanez-Organis JG, Peregrino-Uriarte AB, Gómez-Jiménez S, López-Zavala A, Forman HJ, Yepiz-Plascencia G. 2009. Molecular characterization of hypoxia inducible factor-1 (HIF-1) from the white shrimp *Litopenaeus vannamei* and tissue-specific expression under hypoxia. *Comparative Biochemistry and Physiology Part C: Toxicology & Pharmacology* 150:395-405.

von Reumont BM, Jenner RA, Wills MA, Dell'Ampio E, Pass G, Ebersberger I, Meyer B, Koenemann S, Iliffe TM, Stamatakis A. 2011. Pancrustacean phylogeny in the light of new phylogenomic data: support for Remipedia as the possible sister group of Hexapoda. *Mol. Biol. Evol.* 29:1031-1045.

Downloaded from https://academic.oup.com/mbe/advance-article-abstract/doi/10.1093/molbev/msaa008/5719058 by Oregon State University, grahaall@oregonstate.edu on 04 February 2020

Wenger RH, Stiehl DP, Camenisch G. 2005. Integration of oxygen signaling at the consensus HRE. *Sci. Stke.* 2005:re12.

Downloaded from https://academic.oup.com/mbe/advance-article-abstract/doi/10.1093/molbev/msaa008/5719058 by Oregon State University, grahaall@oregonstate.edu on 04 February 2020

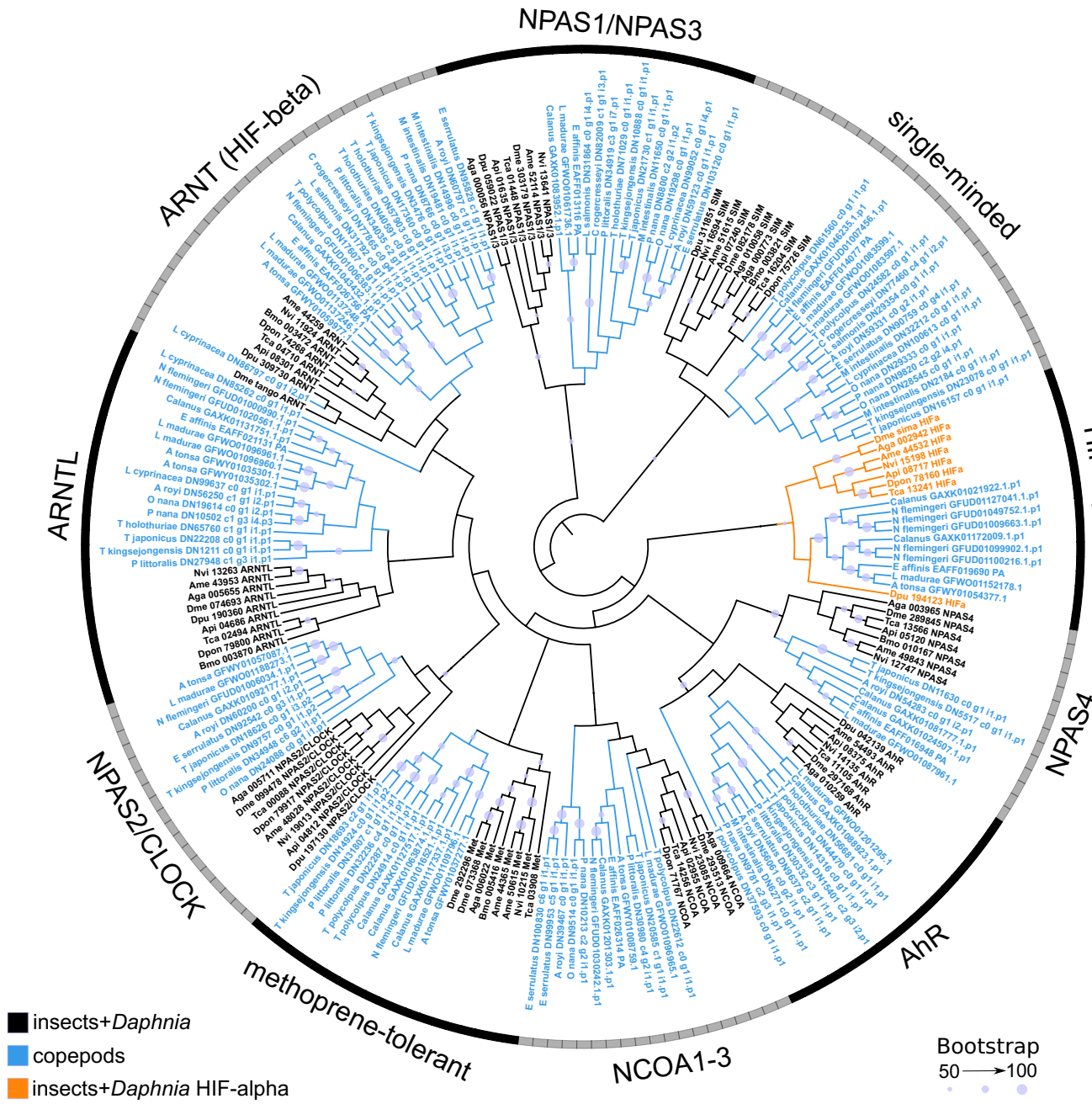

Figure 1. Maximum-likelihood phylogenetic tree of bHLH-PAS proteins of Copepoda, containing 18 species (in blue) across 4 orders (RAxML; 200 bootstraps). Three orders within Copepoda showed a loss of *HIFα*, (Harpacticoida, Cyclopoida, Siphonostomatoida). See Table S1 for the full species names and their classification. In orange are the "anchor" sequences of HIFα from insects and *Daphnia pulex* (Dpu) - including *Anophales gambii* (Aga), *Bombyx mori* (Bmo), *Apis mellifera* (Ame), *Acyrthosiphon pisum* (Api), *Dendroctonus ponderosae* (Dpon), *Drosophila melanogaster* (Dme), *Nasonia vitripennis* (Nvi), and *Tribolium castaneum* (Tca).

Downloaded from https://academic.oup.com/mbe/advance-article-abstract/doi/10.1093/molbev/msaa008/5719058 by Oregon State University, grahaall@oregonstate.edu on 04 February 2020

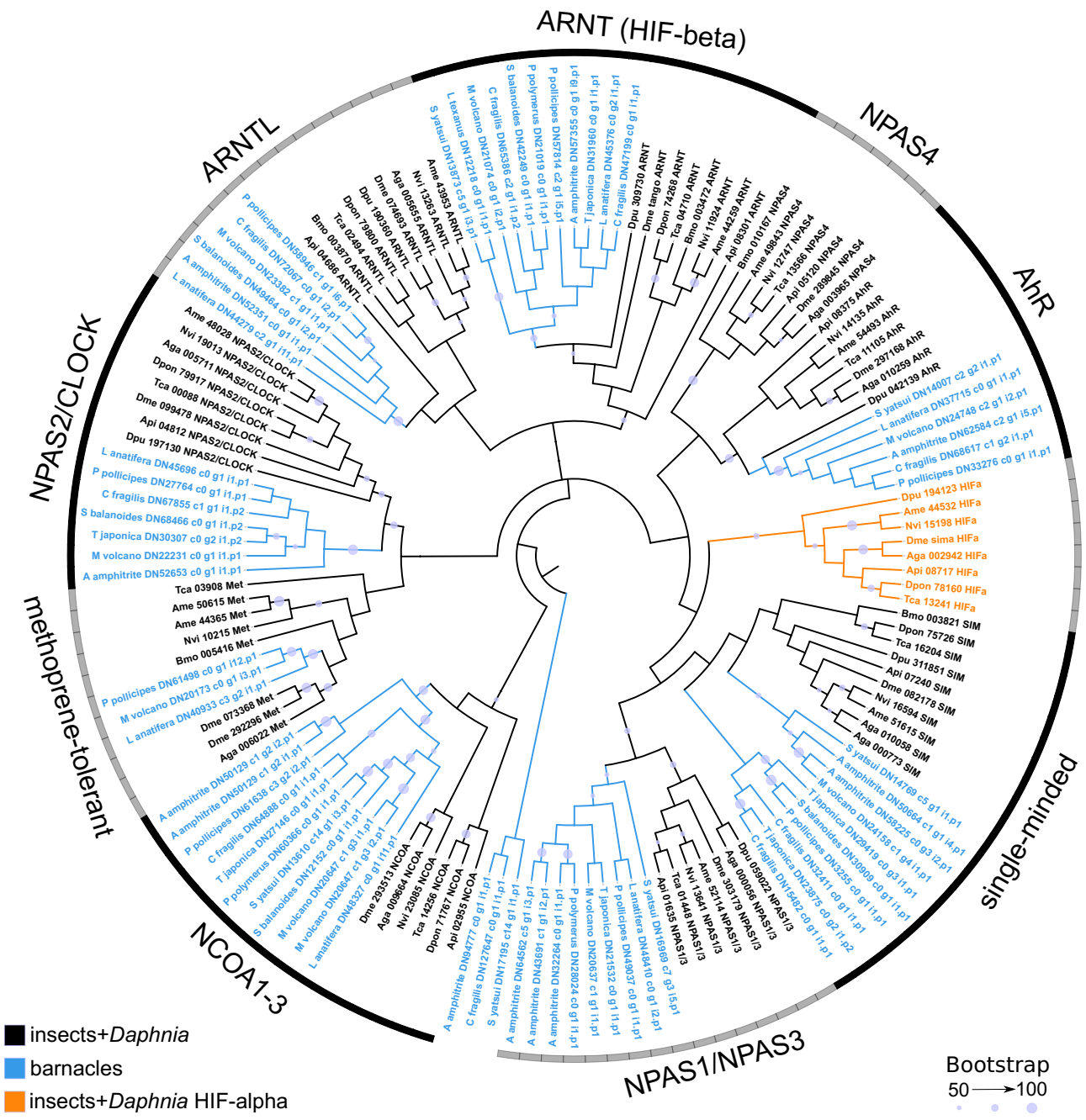

Figure 2. Maximum-likelihood phylogenetic tree of bHLH-PAS proteins of Cirripedia containing 10 species (in blue) across 3 orders (RAxML; 200 bootstraps). See Table S1 for the full species names and their classification. In orange are the "anchor" sequences of HIFα from insects and *Daphnia pulex* (Dpu) - including *Anophales gambii* (Aga), *Bombyx mori* (Bmo), *Apis mellifera* (Ame), *Acyrthosiphon pisum* (Api), *Dendroctonus ponderosae* (Dpon), *Drosophila melanogaster* (Dme), *Nasonia vitripennis* (Nvi), and *Tribolium castaneum* (Tca).

Downloaded from https://academic.oup.com/mbe/advance-article-abstract/doi/10.1093/molbev/msaa008/5719058 by Oregon State University, grahaall@oregonstate.edu on 04 February 2020

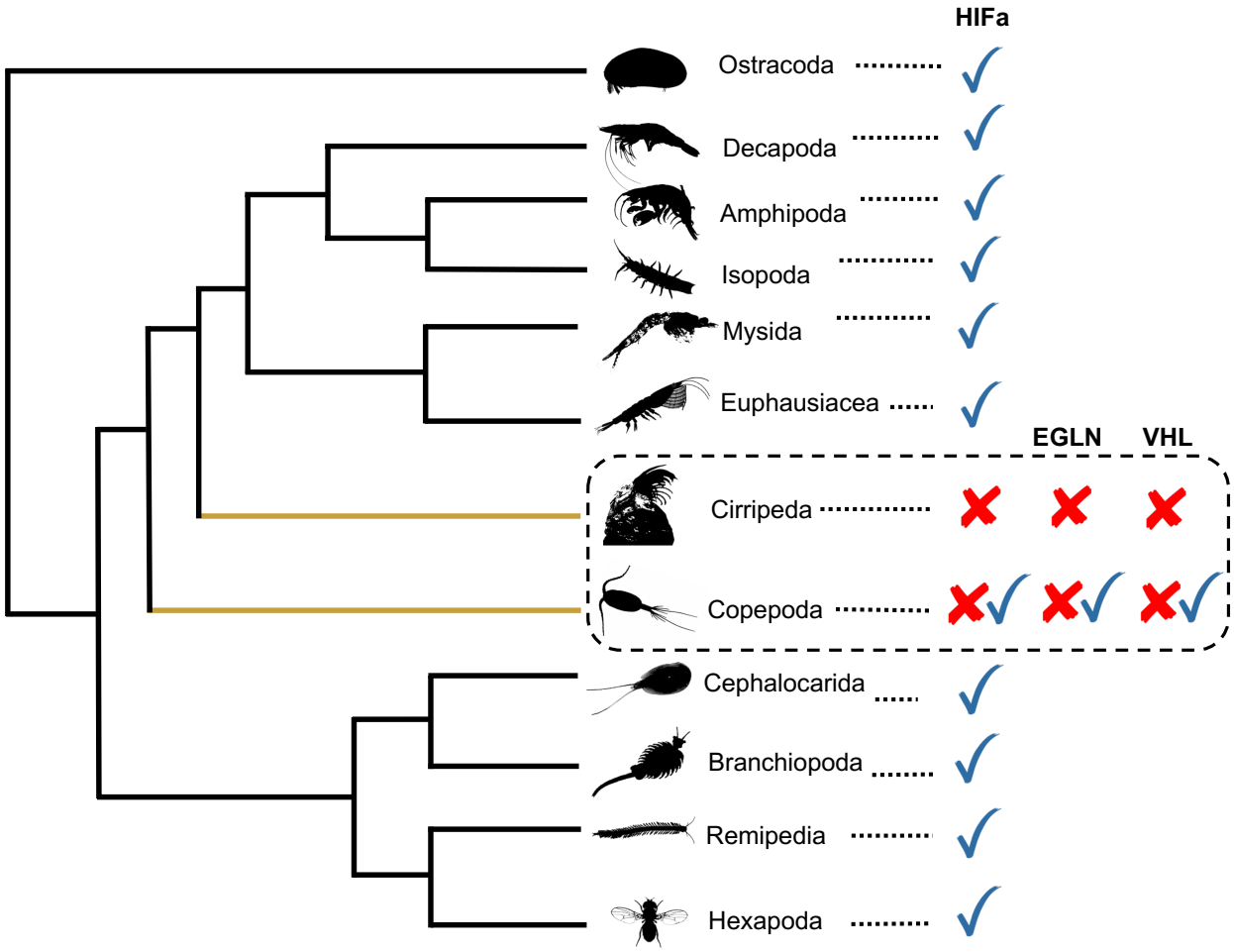

Figure 3. Distribution of HIF-pathway members HIFα, EGLN and VHL across crustacean groups surveyed. Branch lengths in the cladogram are not to scale. General crustacean tree was created based on combination of prior phylogenetic work (Regier, et al. 2010; von Reumont, et al. 2011; Oakley, et al. 2012; Schwentner, et al. 2017; Schwentner, et al. 2018). A check mark represents a confirmed presence, an "X" represents a confirmed absence. Silhouette images are from Phylopic (CC BY-SA 3.0).

Downloaded from https://academic.oup.com/mbe/advance-article-abstract/doi/10.1093/molbev/msaa008/5719058 by Oregon State University, grahaall@oregonstate.edu on 04 February 2020

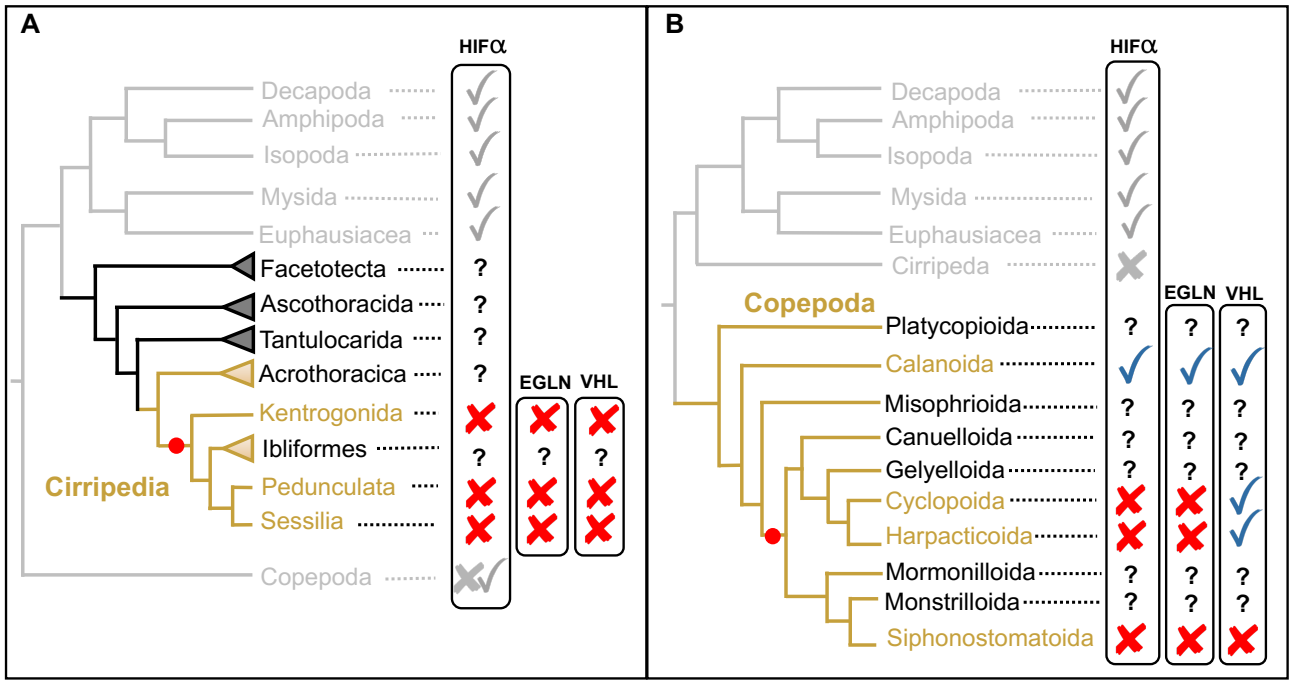

Figure 4. Fine-scale distribution of HIF-pathway members HIFα, EGLN and VHL within the two groups which showed loss. **(A)** Cirripedia (gold lines) and allies (black lines); phylogenetic relationships within cirripedia taken from Ewers-Saucedo, et al. (2019) and Pérez-Losada, et al. (2009). **(B)** Copepoda; phylogenetic relationships within Copepoda taken from Khodami, et al. (2017). A check mark represents a confirmed presence, an "X" represents a confirmed absence, while a question mark represents unknown distribution due to a lack of available sequence data. Red circle in each tree represents our hypothesized point of loss under parsimony.

Downloaded from https://academic.oup.com/mbe/advance-article-abstract/doi/10.1093/molbev/msaa008/5719058 by Oregon State University, grahaall@oregonstate.edu on 04 February 2020