ROYAL SOCIETY OPEN SCIENCE

rsos.royalsocietypublishing.org

Research

Cite this article: Chang WH, Lai AG. 2018 Genome-wide analyses of the bHLH superfamily in crustaceans: reappraisal of higher-order groupings and evidence for lineage- specific duplications. *R. Soc. open sci.* **5**: 172433.
http://dx.doi.org/10.1098/rsos.172433

Received: 30 December 2017
Accepted: 8 February 2018

Subject Category:
Biology (whole organism)

Subject Areas:
bioinformatics/genomics

Keywords:
crustacean, *Litopenaeus vannamei*, bHLH, helix-loop-helix, bHLH-PAS, comparative genomics

Author for correspondence:
Alvina G. Lai
e-mail: alvina.lai@ndm.ox.ac.uk

Electronic supplementary material is available online at https://dx.doi.org/10.6084/m9.figshare.c.4015738.

THE ROYAL SOCIETY PUBLISHING

Genome-wide analyses of the bHLH superfamily in crustaceans: reappraisal of higher-order groupings and evidence for lineage-specific duplications

Wai Hoong Chang and Alvina G. Lai

Nuffield Department of Medicine, University of Oxford, Old Road Campus, Roosevelt Drive, Headington Oxford, OX3 7FZ, UK

WHC, 0000-0001-5832-2000; AGL, 0000-0001-8960-8095

The basic helix-loop-helix (bHLH) proteins represent a key group of transcription factors implicated in numerous eukaryotic developmental and signal transduction processes. Characterization of bHLHs from model species such as humans, fruit flies, nematodes and plants have yielded important information on their functions and evolutionary origin. However, relatively little is known about bHLHs in non-model organisms despite the availability of a vast number of high-throughput sequencing datasets, enabling previously intractable genome-wide and cross-species analyses to be now performed. We extensively searched for bHLHs in 126 crustacean species represented across major Crustacea taxa and identified 3777 putative bHLH orthologues. We have also included seven whole-genome datasets representative of major arthropod lineages to obtain a more accurate prediction of the full bHLH gene complement. With focus on important food crop species from Decapoda, we further defined higher-order groupings and have successfully recapitulated previous observations in other animals. Importantly, we also observed evidence for lineage-specific bHLH expansions in two basal crustaceans (branchiopod and copepod), suggesting a mode of evolution through gene duplication as an adaptation to changing environments. In-depth analysis on bHLH-PAS members confirms the phenomenon coined as 'modular evolution' (independently evolved domains) typically seen in multidomain proteins. With the amphipod *Parhyale hawaiensis* as the exception, our analyses have focused on crustacean

transcriptome datasets. Hence, there is a clear requirement for future analyses on whole-genome sequences to overcome potential limitations associated with transcriptome mining. Nonetheless, the present work will serve as a key resource for future mechanistic and biochemical studies on bHLHs in economically important crustacean food crop species.

1. Background

The basic helix-loop-helix (bHLH) transcription factor superfamily is among one of the most ancient gene families shared by eukaryotic organisms [1–3]. Members of this superfamily possess two conserved yet functionally distinct domains made up of a basic DNA binding domain (E box) at the amino-terminal followed by an HLH domain at the carboxy-terminal where the latter confers the possibility for hetero- or homodimerization with other proteins [4,5]. Both domains are required to form active DNA binding complexes whereby the functionality and precision in regulating gene expression networks are determined by the formation of multiple dimer combinations through the HLH domains [6]. This dimerization potential is, arguably, an effective mechanism for gene regulation considering that each dimer pair probably has specific genetic targets. For instance, bHLHs in animals are intimately linked to the regulation of a multitude of physiological and developmental processes. Some of these include cell cycle regulation, neurogenesis, haematopoiesis, myogenesis, differentiation, apoptosis, juvenile hormone signalling in arthropods and sensing of extracellular stimuli [1,2,7–15].

Given their importance as critical transcriptional regulators in major biological processes, it is perhaps not surprising that many studies have focused on the identification and characterization of bHLH orthologues across various plant and animal species. From the first bHLH motif identified, the murine E12 and E47 transcription factors [16], full genetic complements of bHLH proteins have since been reported in various model organisms owing to the availability of complete genome sequences [3,17–25]. Phylogenetic analyses on bHLH orthologues from model bilaterian species (*Homo sapiens*, *Drosophila melanogaster* and *Caenorhabditis elegans*) and early branching metazoans (the demosponge *Amphimedon queenslandica* and cnidarians *Nematostella vectensi*s and *Hydra magnipapillata*) have convincingly demonstrated that diversification of metazoan bHLHs could have arisen in two steps: the first occurring before the divergence of demosponges from other animals and the second before the divergence of cnidarians and bilaterians [3,21]. Moreover, the increasing repertoire of bHLH proteins have enabled the classification of bHLH orthologues into six higher-order groups (A, B, C, D, E and F) based on distinct structural features [1,2,19,26,27].

Major efforts have thus far focused on the classification of bHLH proteins in model organisms. Beyond canonical model species, some headway has been made in elucidating bHLH genes in major metazoan lineages, which include bilaterians (deuterostomes and protostomes) and basal metazoans (hydra and sponge) [3]. Yet, relatively little is known about bHLHs in one of the most important groups of animals that represent a significant portion of aquatic sources of proteins, i.e. crustaceans. To date, studies in crustaceans are not only limited to just a few species but also to specific bHLH groups [13,14]. Systematic and cross-species characterization of crustacean bHLH proteins focusing on major food crop species from the order Decapoda is therefore necessary. In fact, numerous studies have begun to shed light on the importance of group C members, collectively known as bHLH-PAS, as environmental sensors [28]. For example, members of the crustacean bHLH-PAS family have been shown to regulate growth and reproduction through juvenile hormone signalling (Methoprene-tolerant, MET) [15,29], locomotion through circadian timekeeping (aryl hydrocarbon receptor nuclear translocator-like protein 1, BMAL1) [30,31] and response to changes in oxygen tension (hypoxia inducible factor 1, HIF-1) [32,33]. This holds much relevance for shrimp farming industries that are constantly facing the pressures of fluctuating environmental conditions that could potentially compromise aquaculture yield [32,34,35].

Here, we perform a cross-species characterization of the crustacean bHLH superfamily to address this major deficit in the field. The large number of recent crustacean transcriptomic datasets in public repositories has now permitted in-depth studies on key food crop species from the order Decapoda and other species across the broader Crustacea, hence offering important insights into the diversity of crustacean bHLHs (electronic supplementary material, table S1). Using sequence, motif and domain similarity-based approaches, we have conservatively identified 4113 bHLH orthologues, which include 3777 orthologues from 126 crustacean species and 336 orthologues from six other non-crustacean arthropod species (electronic supplementary material, tables S1 and S2 and file S1). Within this key

dataset, we define higher-order bHLH groups in three major decapod species using phylogenetic-based approaches and annotate bHLH-PAS proteins in decapods and basal crustaceans (branchiopods and copepods).

2. Material and methods

2.1. Transcriptome datasets and query sets

We retrieved complete transcriptome datasets for 126 crustacean species available at the time of manuscript preparation from the European Nucleotide Archive (https://www.ebi.ac.uk/ena). Six non-crustacean arthropod proteomes were retrieved from Uniprot (http://www.uniprot.org/). A complete list of accessions used in this study is provided in the electronic supplementary material, table S1. We retrieved a list of query sequences used in subsequent homology searches from Uniprot and GenBank.

2.2. Identification of bHLH orthologues

Based on a previously published workflow [36], we used multiple Basic Local Alignment Search Tool (BLAST)-based approaches such as BLASTp and tBLASTn with varying Blocks Substitution matrices to identify bHLH orthologues. The BLAST results were filtered by e-value of $<10^{-6}$, best reciprocal BLAST hits against the GenBank non-redundant (nr) database and redundant contigs having at least 95% identity were collapsed using CD-HIT (https://github.com/weizhongli/cdhit). We then used HMMER employing hidden Markov models (HMM) profiles [37] to scan for the presence of bHLH Pfam domains [38] on the best reciprocal nr BLAST hits to compile a final non-redundant set of crustacean and arthropod bHLH orthologues. Pfam annotations and associated e-values are provided in electronic supplementary material, table S2.

2.3. Multiple sequence alignment and phylogenetic tree construction

Multiple sequence alignments of bHLH protein sequences were performed using MAFFT [39]. Phylogenetic trees were built from the MAFFT alignment using RAxML WAG + G model to generate best-scoring maximum-likelihood trees [40]. Bayesian inference trees were constructed using MrBayes [41]. Geneious was used to generate graphical representations of Newick trees [42].

3. Results and discussion

3.1. Annotation of putative bHLH genes in crustaceans

Building on our previous analysis of bHLHs in the crustacean *Parhyale hawaiensis* [43], we have identified additional bHLHs from six non-crustacean arthropods: Insecta (three species), Arachnida (two species) and Chilopoda (one species) and 125 additional crustacean species representing three classes: Malacostraca (Amphipoda: 64 species, Decapoda: 19 species, Isopoda: 27 species, Euphausiacea: two species and Mysida: one species), Branchiopoda (three species) and Copepoda (10 species) (figure 1*a*; electronic supplementary material, tables S1 and S2 and file S1). From the complete genome sequences of model arthropod species, we annotated a total of 336 putative bHLHs; *D. melanogaster* (78 genes), *Aedes aegypti* (53 genes), *Anopheles gambiae* (44 genes), *Ixodes scapularis* (43 genes), *Mesobuthus martensii* (63 genes) and *Strigamia maritima* (55 genes) (figure 1). We have also identified 3777 bHLH genes from 126 crustacean species (including *P. hawaiensis*), hence providing a significant coverage of major Crustacea taxa (figure 1; electronic supplementary material, table S2 and file S1).

3.2. Assignment of bHLHs into higher-order groups/families

Considering the large number of putative bHLHs identified, independent analyses on genes from individual species are required. We therefore selected three decapod species for further analyses and assignment of bHLHs into higher-order groups. These species represent three dominant families of economically important food crops: Pacific whiteleg shrimp *Litopenaeus vannamei* (Penaeidae, 87 genes), freshwater crayfish *Cherax quadricarinatus* (Parastacidae, 65 genes) and mud crab *Scylla olivacea* (Portunidae, 66 genes). Maximum-likelihood trees were generated from multiple sequence alignments with bHLHs from *Homo sapiens* to define orthologous groups [21]. Overall, we observed that decapod

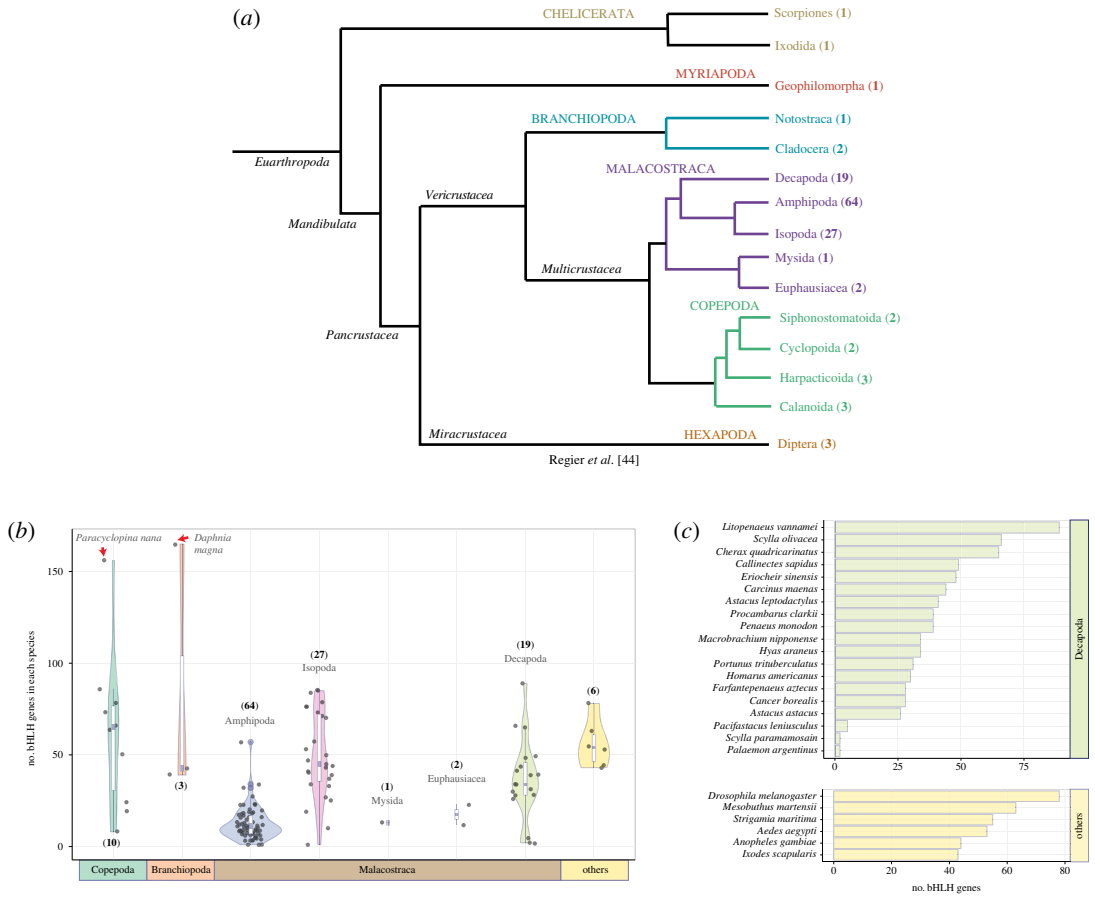

Figure 1. The bHLH superfamily in Crustacea and representative arthropod species. (*a*) Phylogenetic relationship of Arthropoda and Pancrustacea. The number of species within each taxon is denoted in parentheses. (*b*) Number of bHLH orthologues identified in each species is depicted as boxplots, indicating the median and quartiles. Violin plots underlying the boxplots illustrate sample distribution across different crustacean taxa and kernel probability density (width of the shaded areas represent the proportion of data located in these areas). The bHLH orthologues from six non-crustacean species within Arthropoda (others) are also shown. The number of species for each taxon is denoted in parentheses. Number of bHLH genes in *Paracyclopina nana* and *Daphnia magna* are marked with red arrows. (*c*) Bar charts illustrating the number of bHLHs in decapods and six non-crustacean arthropods (others).

bHLHs could be confidently assigned to previously described higher-order groups (A, B, C and E) (figure 2; electronic supplementary material, figure S1). Group A (proteins that bind CACCTG or CAGCTG E boxes) and group E (proteins that bind CACGCG or CACGAG N boxes) bHLHs are monophyletic in all three decapod species (figure 2*a–d*; electronic supplementary material, figure S1); a reappraisal of previous observations on bHLHs from human [21] and sponge [3]. Group B (proteins that bind CACGTG or CATGTTG E boxes) and group C (proteins that contain PAS domains) bHLHs are probably paraphyletic as they constitute members from other groups (figure 2*b–d*) [21]. We observed no evidence of group F bHLHs (figure 2*a*) and only one instance of a group D member from *C. quadricarinatus* (figure 2*d*,*e*). This could suggest that group D and F bHLH members are present but not represented in the transcriptome datasets, that members we have found have evolved convergently and/or a complex phenomenon of loss among certain crustacean taxa.

3.3. Evidence for lineage-specific duplications of bHLHs in Branchiopoda and Copepoda

Our analyses of the full bHLH gene complement from the genomes of seven species have revealed that arthropods/crustacean, on average, possess 56 bHLH genes: *Parhyale hawaiensis* (57 genes), *Ixodes scapularis* (43 genes), *Anopheles gambiae* (44 genes), *Aedes aegypti* (53 genes), *Strigamia maritima* (55 genes), *Mesobuthus martensii* (63 genes) and *Drosophila melanogaster* (78 genes); a finding consistent with another study [3]. Interestingly, we observed that two basal crustacean species, *Daphnia magna* (Branchiopoda) and *Paracyclopina nana* (Copepoda) have a significantly higher number of bHLH genes,

(a)

phylogenetic group	gene families (human orthologues)	description
A	*Achaete-Scute, MyoD, E12/E47, Atonal, Mist, Oligo, Beta3, Neurogenin, NeuroD, Mesp, MyoR, PTF, Hand, Twist, Paraxis, NSCL, SCL*	binding to CACCTG or CAGCTG
B	*AP4, Mnt, Mad, Max, Myc, SRC, MITF, SREBP, USF, MLX, TF4, Figa*	binding to CACGTG or CATGTTG
C	*Bmal, ARNT, Clock, HIF, Sim, Trh, AHR, MET*	binding to ACGTG or GCGTG. Contain PAS domain
D	*Emc*	lack a basic domain. Antagonists of group A
E	*Hey, E(spl), Hairy*	binding to N boxes (CACGCG or CACGAG)
F	*Coe*	contains COE domain

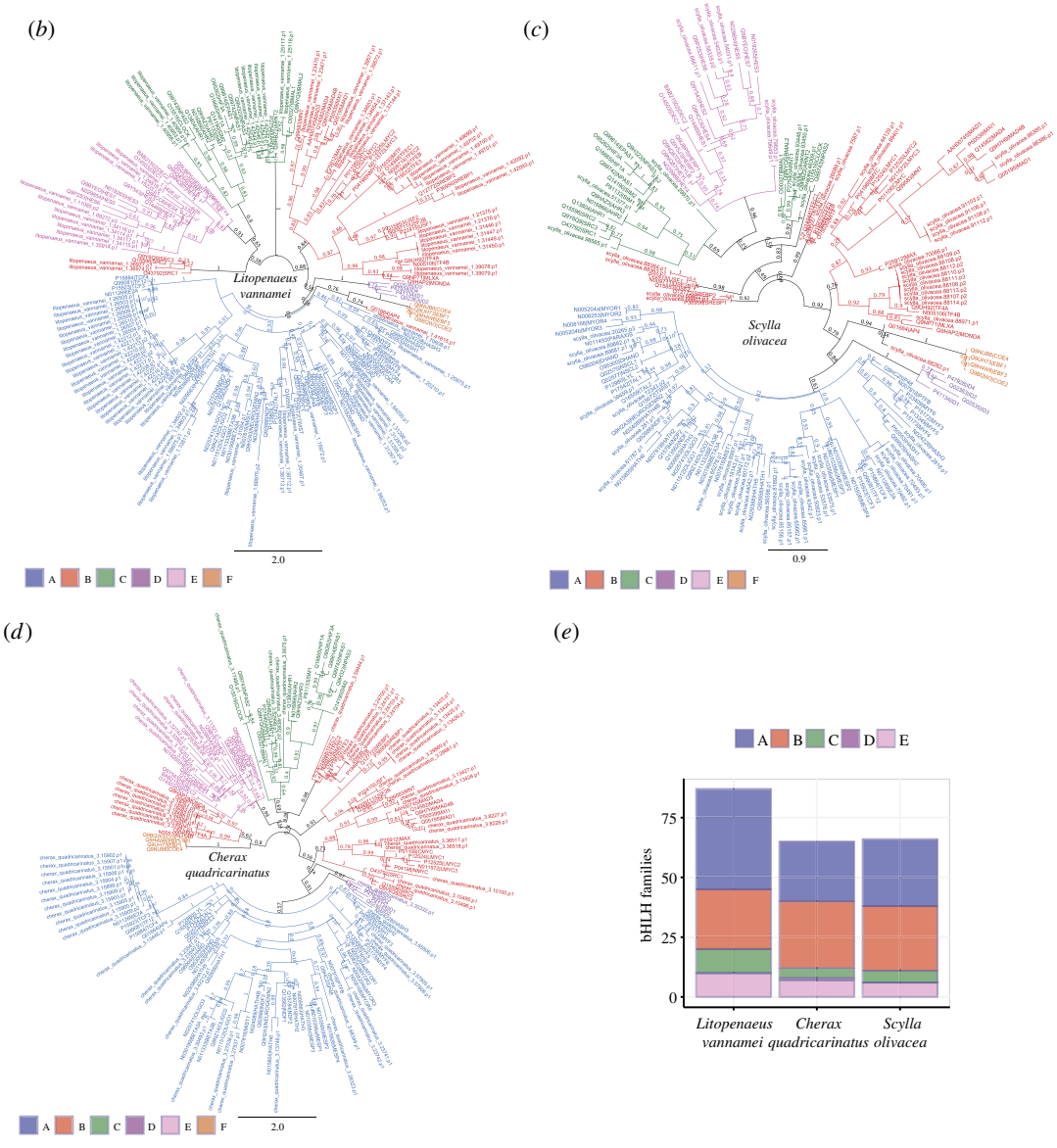

Figure 2. Classification of decapod bHLH proteins into higher-order groups. (*a*) The bHLH proteins can be further classified into six main groups (A–F) based on conservation of residues or the presence of additional domains [2,21,26]. Phylogenetic analyses of bHLHs from decapods (*b*) *Litopenaeus vannamei*, (*c*) *Scylla olivacea* and (*d*) *Cherax quadricarinatus*. Bootstrap support values ($n = 1000$) are denoted as branch labels. (*e*) Number of decapod bHLHs classified into groups A, B, C, D and E.

165 and 156 genes, respectively, than that of other crustacean and arthropod species (figure 1*b*; electronic supplementary material, table S2). This suggests that the expansion of the bHLH superfamily in these two species may have arisen through lineage-specific duplications. Phylogenetic analyses revealed that *P. nana* and *D. magna* bHLH genes could be assigned to four higher-order groups (A, B, C and E) with group E bHLHs exhibiting monophyly in both species (electronic supplementary material, figure S2 and S3). Others have reported that *Daphnia* has an unusually large collection of tandemly duplicated genes

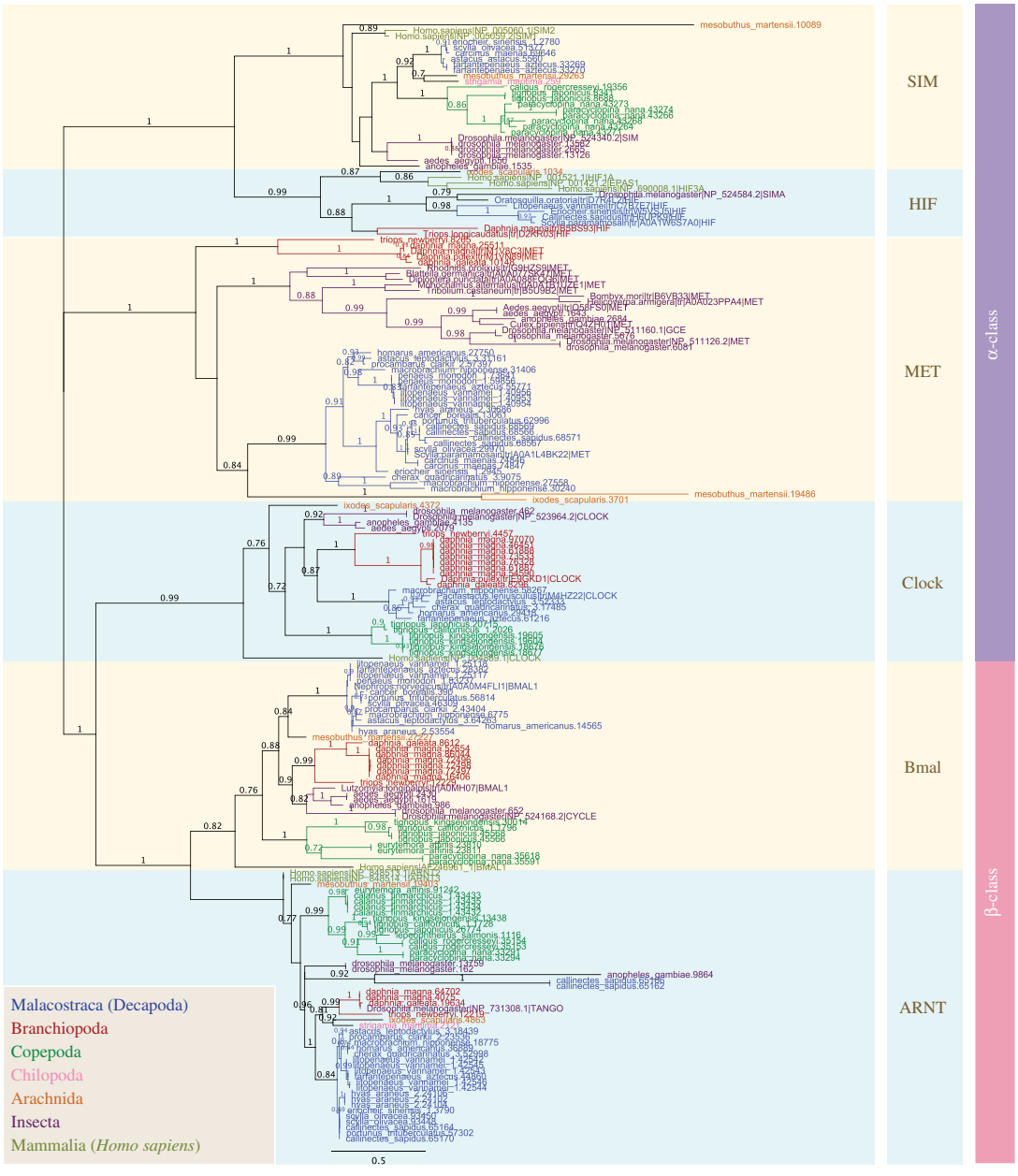

Figure 3. Phylogeny of the bHLH-PAS family in decapod, basal crustaceans and non-crustacean arthropods. The tree was constructed using the maximum-likelihood method from an amino acid multiple sequence alignment. The node labels of each taxon are marked with distinctive colours denoted in the figure inset. Bootstrap support values ($n = 1000$) above 0.7 are denoted as branch labels. The tree illustrates putative α-class and β-class members.

underpinning its intriguing phenotypic plasticity and remarkable ability to adapt to major ecological challenges [45]. The expansion of bHLHs as an adaptation mechanism to adverse environmental conditions can similarly be explained in copepods. These marine planktonic species serve as models for ecotoxicology and environmental genomics due to their high tolerance to a wide range of salinity and temperature [46–49]. As many bHLH members function as sensors of the environment, expansion through gene duplication would serve as an important evolutionary force for functional divergence to genetically adapt and cope with environmental stressors [50].

3.4. Monophyly of β-class bHLH-PAS proteins in decapods and basal crustaceans

The period (PER)-ARNT-SIM (PAS) domain is found in proteins involved in recognizing and transducing environmental signals into appropriate cellular responses pertaining to stress signalling, development

and circadian regulation. Many well-known PAS proteins have been shown to also possess a bHLH domain where they are collectively termed as bHLH-PAS [11,28]. The bHLH-PAS family is further classified into two subgroups: α-class and β-class [11,51]. The α-class proteins act as archetypal sensors of tissue-specific or environmental signals, e.g. single-minded (SIM-development), HIF (oxygen sensing), neuronal PAS domain proteins (NPAS-development) and circadian locomotor output cycles protein kaput (CLOCK-circadian timekeeping) and MET (juvenile hormone signalling) [11,51–53]. Unable to dimerise among themselves, α-class members require β-class proteins, aryl hydrocarbon receptor nuclear translocators (ARNT), as their broad-spectrum binding partners [11,51]. We perform phylogenetic analysis on bHLH-PAS members identified from decapods and basal crustaceans (branchiopods and copepods; figure 3; electronic supplementary material, figures S4 and S5). We have also included bHLH-PAS genes identified from other arthropod species (insects, tick, scorpion and centipede) and previously annotated proteins from Uniprot (figure 3; electronic supplementary material, figures S4 and S5) [21]. Consistent with other reports, we demonstrate that crustacean bHLH-PAS orthologues form two distinct phylogenetic groups, *α* and *β* (figure 3; electronic supplementary material, figures S4 and S5). The β-class ARNT proteins (including the BMAL family) form a well-supported monophyletic group, whereas α-class proteins are polyphyletic: HIF and SIM members form a single cluster while MET and CLOCK form separate monophyletic groups (figure 3; electronic supplementary material, figures S4 and S5). Our observation is consistent with previous reports suggesting that bHLH-PAS paraphyly could be explained through modular evolution [54]. Modular evolution by domain shuffling is commonly seen in bHLH proteins that also possess other highly conserved domains such as PAS. As evidenced by the lack of sequence similarity in flanking regions, it is likely that the association between PAS and bHLH domains occurred multiple times independently through domain duplication or insertion [19,54].

4. Conclusion

Although a majority of our analyses have focused on transcriptome datasets that are dependent on tissue-type and the differential expression of transcripts at the point of sample collection, the power of detection is supported by the fact that (i) the transcriptomes were sequenced to considerable depth (electronic supplementary material, table S1) and (ii) we do find consistent detection of bHLH gene families in whole-genome sequences (electronic supplementary material, table S2). We are unable to entirely rule out the possibility that some genes would remain undetected, hence potential caveats associated with transcriptome mining must be taken into consideration. In summary, we identified 3777 putative bHLH orthologues from 126 crustacean species representing major Crustacea taxa. We also annotated 336 bHLHs from six additional non-crustacean arthropods sampling across broad taxonomic range to include genomes of emerging model species (centipede and scorpion). We observed evidence for lineage-specific gene expansions in branchiopod and copepod suggesting a mechanism for genetic adaptation to adverse environmental factors commonly encountered by these species. Phylogenetic analyses on decapod bHLHs recapitulated the evolutionarily conserved higher-order orthologous groupings seen in other metazoans. Further analysis on group C bHLH-PAS members revealed that although β-class members are monophyletic, this is not true for the remaining α-class members.

Data accessibility. Data supporting the conclusions of this study are provided as the electronic supplementary materials.

Authors' contributions. W.H.C. and A.G.L. conceived and designed the study. W.H.C. performed the data analyses with guidance from A.G.L. W.H.C. and A.G.L. finalized the data, drafted the manuscript and approved the final manuscript for publication.

Competing interests. We declare we have no competing interests.

Funding. This work was supported by the EMBO Fellowship and the Human Frontier Science Program Fellowship to A.G.L. The funding bodies have no roles in design of the study, collection, analysis and interpretation of the data and in writing the manuscript.

References

1. Massari ME, Murre C. 2000 Helix-loop-helix proteins: regulators of transcription in eucaryotic organisms. *Mol. Cell. Biol* **20**, 429–440. (doi:10.1128/MCB.20.2.429-440.2000)
2. Jones S. 2004 An overview of the basic helix-loop-helix proteins. *Genome Biol.* **5**, 226. (doi:10.1186/gb-2004-5-6-226)
3. Simionato E, Ledent V, Richards G, Thomas-Chollier M, Kerner P, Coornaert D, Degnan BM, Vervoort M. 2007 Origin and diversification of the basic helix-loop-helix gene family in metazoans: insights

from comparative genomics. *BMC Evol. Biol.* **7**, 33. (doi:10.1186/1471-2148-7-33)
4. Murre C *et al.* 1989 Interactions between heterologous helix-loop-helix proteins generate complexes that bind specifically to a common DNA sequence. *Cell* **58**, 537–544. (doi:10.1016/0092-8674(89)90434-0)
5. Kadesch T. 1993 Consequences of heteromeric interactions among helix-loop-helix proteins. *Cell Growth Differ.* **4**, 49–55.
6. Kewley RJ, Whitelaw ML, Chapman-Smith A. 2004 The mammalian basic helix–loop–helix/PAS family of transcriptional regulators. *Int. J. Biochem. Cell Biol* **36**, 189–204. (doi:10.1016/S1357-2725(03)00211-5)
7. Amati B, Land H. 1994 Myc—Max—Mad: a transcription factor network controlling cell cycle progression. differentiation and death. *Curr. Opin. Genet. Dev* **4**, 102–108. (doi:10.1016/0959-437X(94)90098-1)
8. Bain G *et al.* 1994 E2A proteins are required for proper B cell development and initiation of immunoglobulin gene rearrangements. *Cell* **79**, 885–892. (doi:10.1016/0092-8674(94)90077-9)
9. Turner DL, Lipnick N, Weintraubt H. 1995 Conversion of Xenopus ectoderm into neurons by NeuroD, a basic helix-loop-helix protein. *Science* **268**, 12.
10. Porcher C, Swat W, Rockwell K, Fujiwara Y, Alt FW, Orkin SH. 1996 The T cell leukemia oncoprotein SCL/tal-1 is essential for development of all hematopoietic lineages. *Cell* **86**, 47–57. (doi:10.1016/S0092-8674(00)80076-8)
11. Gu Y-Z, Hogenesch JB, Bradfield CA. 2000 The PAS superfamily: sensors of environmental and developmental signals. *Annu. Rev. Pharmacol. Toxicol.* **40**, 519–561. (doi:10.1146/annurev.pharmtox.40.1.519)
12. Lüscher B. 2001 Function and regulation of the transcription factors of the Myc/Max/Mad network. *Gene* **277**, 1–14. (doi:10.1016/S0378-1119(01)00697-7)
13. Wheeler SR, Skeath JB. 2005 The identification and expression of achaete-scute genes in the branchiopod crustacean *Triops longicaudatus*. *Gene Expr. Patterns* **5**, 695–700. (doi:10.1016/j.modgep.2005.02.005)
14. Tokishita S, Kimura S, Mandokoro Y, Kato K, Shiga Y, Takahashi Y, Ohta T, Yamagata H. 2006 Tissue-specific expression of a bHLH-PAS protein homologous to ARNT during the development of crustacean *Daphnia magna*. *Gene* **376**, 231–239. (doi:10.1016/j.gene.2006.03.022)
15. Miyakawa H *et al.* 2013 A mutation in the receptor Methoprene-tolerant alters juvenile hormone response in insects and crustaceans. *Nat. Commun.* **4**, 1856. (doi:10.1038/ncomms2868)
16. Murre C, McCaw PS, Baltimore D. 1989 A new DNA binding and dimerization motif in immunoglobulin enhancer binding, daughterless, MyoD, and myc proteins. *Cell* **56**, 777–783. (doi:10.1016/0092-8674(89)90682-X)
17. Moore AW, Barbel S, Jan LY, Jan YN. 2000 A genomewide survey of basic helix–loop–helix factors in *Drosophila*. *Proc. Natl Acad. Sci. USA* **97**, 10 436–10 441. (doi:10.1073/pnas.170301897)
18. Robinson KA, Lopes JM. 2000 Survey and summary: *Saccharomyces cerevisiae* basic helix–loop–helix proteins regulate diverse biological processes. *Nucleic Acids Res.* **28**, 1499–1505. (doi:10.1093/nar/28.7.1499)
19. Ledent V, Vervoort M. 2001 The basic helix-loop-helix protein family: comparative genomics and phylogenetic analysis. *Genome Res.* **11**, 754–770. (doi:10.1101/gr.177001)
20. Peyrefitte S, Kahn D, Haenlin M. 2001 New members of the *Drosophila* Myc transcription factor subfamily revealed by a genome-wide examination for basic helix-loop-helix genes. *Mech. Dev.* **104**, 99–104. (doi:10.1016/S0925-4773(01)00360-4)
21. Ledent V, Paquet O, Vervoort M. 2002 Phylogenetic analysis of the human basic helix-loop-helix proteins. *Genome Biol.* **3**, 1–18. (doi:10.1186/gb-2002-3-6-research0030)
22. Buck MJ, Atchley WR. 2003 Phylogenetic analysis of plant basic helix-loop-helix proteins. *J. Mol. Evol.* **56**, 742–750. (doi:10.1007/s00239-002-2449-3)
23. Heim MA, Jakoby M, Werber M, Martin C, Weisshaar B, Bailey PC. 2003 The basic helix–loop–helix transcription factor family in plants: a genome-wide study of protein structure and functional diversity. *Mol. Biol. Evol.* **20**, 735–747. (doi:10.1093/molbev/msg088)
24. Toledo-Ortiz G, Huq E, Quail PH. 2003 The Arabidopsis basic/helix-loop-helix transcription factor family. *Plant Cell* **15**, 1749–1770. (doi:10.1105/tpc.013839)
25. Li X *et al.* 2006 Genome-wide analysis of basic/helix-loop-helix transcription factor family in rice and Arabidopsis. *Plant Physiol.* **141**, 1167–1184. (doi:10.1104/pp.106.080580)
26. Atchley WR, Fitch WM. 1997 A natural classification of the basic helix—loop—helix class of transcription factors. *Proc. Natl Acad. Sci. USA* **94**, 5172–5176. (doi:10.1073/pnas.94.10.5172)
27. Dubois L, Vincent A. 2001 The COE—Collier/Olf1/EBF—transcription factors: structural conservation and diversity of developmental functions. *Mech. Dev.* **108**, 3–12. (doi:10.1016/S0925-4773(01)00486-5)
28. Crews ST, Fan C-M. 1999 Remembrance of things PAS: regulation of development by bHLH–PAS proteins. *Curr. Opin. Genet. Dev.* **9**, 580–587. (doi:10.1016/S0959-437X(99)00003-9)
29. Miyakawa H, Toyota K, Sumiya E, Iguchi T. 2014 Comparison of JH signaling in insects and crustaceans. *Curr. Opin. Insect. Sci.* **1**, 81–87. (doi:10.1016/j.cois.2014.04.006)
30. Zhang L, Hastings MH, Green EW, Tauber E, Sladek M, Webster SG, Kyriacou CP, Wilcockson DC. 2013 Dissociation of circadian and circatidal timekeeping in the marine crustacean *Eurydice pulchra*. *Curr. Biol.* **23**, 1863–1873. (doi:10.1016/j.cub.2013.08.038)
31. O'Grady JF, Hoelters LS, Swain MT, Wilcockson DC. 2016 Identification and temporal expression of putative circadian clock transcripts in the amphipod crustacean *Talitrus saltator*. *PeerJ* **4**, e2555. (doi:10.7717/peerj.2555)
32. 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. *Comp. Biochem. Physiol. Part C Toxicol. Pharmacol.* **150**, 395–405. (doi:10.1016/j.cbpc.2009.06.005)
33. Soñanez-Organis JG, Racotta IS, Yepiz-Plascencia G. 2010 Silencing of the hypoxia inducible factor 1—HIF-1-obliterates the effects of hypoxia on glucose and lactate concentrations in a tissue-specific manner in the shrimp *Litopenaeus vannamei*. *J. Exp. Mar. Bio. Ecol.* **393**, 51–58. (doi:10.1016/j.jembe.2010.06.031)
34. Seidman ER, Lawrence AL. 1985 Growth, feed digestibility, and proximate body composition of juvenile *Penaeus vannamei* and *Penaeus monodon* grown at different dissolved oxygen levels. J. *World Aquac. Soc.* **16**, 333–346. (doi:10.1111/j.1749-7345.1985.tb00214.x)
35. Rosas C, Martinez E, Gaxiola G, Brito R, Sánchez A, Soto LA. 1999 The effect of dissolved oxygen and salinity on oxygen consumption, ammonia excretion and osmotic pressure of *Penaeus setiferus* (Linnaeus) juveniles. *J. Exp. Mar. Bio. Ecol.* **234**, 41–57. (doi:10.1016/S0022-0981(98)00139-7)
36. Lai AG, Aboobaker AA. 2017 Comparative genomic analysis of innate immunity reveals novel and conserved components in crustacean food crop species. *BMC Genomics* **18**, 389. (doi:10.1186/s12864-017-3769-4)
37. Finn RD, Clements J, Eddy SR. 2011 HMMER web server: interactive sequence similarity searching. *Nucleic Acids Res.* **39**, W29–W37. (doi:10.1093/nar/gkr367)
38. Bateman A *et al.* 2004 The Pfam protein families database. *Nucleic Acids Res.* **32**, D138–D141. (doi:10.1093/nar/gkh121)
39. Katoh K, Asimenos G, Toh H. 2009 Multiple alignment of DNA sequences with MAFFT. *Methods Mol. Biol.* **537**, 39–64. (doi:10.1007/978-1-59745-251-9_3)
40. Stamatakis A. 2014 RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313. (doi:10.1093/bioinformatics/btu033)
41. Ronquist F, Huelsenbeck JP. 2003 MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19**, 1572–1574. (doi:10.1093/bioinformatics/btg180)
42. Kearse M *et al.* 2012 Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28**, 1647–1649. (doi:10.1093/bioinformatics/bts199)
43. Kao D *et al.* 2016 The genome of the crustacean *Parhyale hawaiensis*, a model for animal development, regeneration, immunity and lignocellulose digestion. *Elife* **5**, e20062. (doi:10.7554/eLife.20062)
44. 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. (doi:10.1038/nature08742)
45. Colbourne JK *et al.* 2011 The ecoresponsive genome of *Daphnia pulex*. *Science* **331**, 555–561. (doi:10.1126/science.1197761)
46. Lee KW, Park HG, Lee S-M, Kang H-K. 2006 Effects of diets on the growth of the brackish water cyclopoid copepod *Paracyclopina nana* Smirnov. *Aquaculture* **256**, 346–353. (doi:10.1016/j.aquaculture.2006.01.015)
47. Raisuddin S, Kwok KWH, Leung KMY, Schlenk D, Lee J-S. 2007 The copepod *Tigriopus*: a promising marine model organism for ecotoxicology and environmental genomics. *Aquat. Toxicol.* **83**, 161–173. (doi:10.1016/j.aquatox.2007.04.005)

48. Pascal P-Y, Fleeger JW, Galvez F, Carman KR. 2010 The toxicological interaction between ocean acidity and metals in coastal meiobenthic copepods. *Mar. Pollut. Bull* **60**, 2201–2208. (doi:10.1016/j.marpolbul.2010.08.018)
49. Won E-J, Lee J-S. 2014 Gamma radiation induces growth retardation, impaired egg production, and oxidative stress in the marine copepod *Paracyclopina nana*. *Aquat. Toxicol.* **150**, 17–26. (doi:10.1016/j.aquatox.2014.02.010)
50. Ebert D. 2011 A genome for the environment. *Science* **331**, 539–540. (doi:10.1126/science.1202092)
51. Fribourgh JL, Partch CL. 2017 Assembly and function of bHLH—PAS complexes. *Proc. Natl. Acad. Sci USA* **114**, 5330–5332. (doi:10.1073/pnas.1705408114)
52. Michaud JL, DeRossi C, May NR, Holdener BC, Fan C-M. 2000 ARNT2 acts as the dimerization partner of SIM1 for the development of the hypothalamus. *Mech. Dev.* **90**, 253–261. (doi:10.1016/S0925-4773(99)00328-7)
53. Shin SW, Zou Z, Saha TT, Raikhel AS. 2012 bHLH-PAS heterodimer of methoprene-tolerant and Cycle mediates circadian expression of juvenile hormone-induced mosquito genes. *Proc. Natl Acad. Sci. USA* **109**, 16 576–16 581. (doi:10.1073/pnas.1214209109)
54. Morgenstern B, Atchley WR. 1999 Evolution of bHLH transcription factors: modular evolution by domain shuffling? *Mol. Biol. Evol* **16**, 1654–1663. (doi:10.1093/oxfordjournals.molbev.a026079)