



RESEARCH ARTICLE

A TALE of shrimps: Genome-wide survey of homeobox genes in 120 species from diverse crustacean taxa [version 1; referees: 2 approved, 1 approved with reservations]

Wai Hoong Chang , Alvina G. Lai 

Nuffield Department of Medicine, University of Oxford, Oxford, OX3 7FZ, UK

v1 First published: 17 Jan 2018, 7:71 (doi: [10.12688/f1000research.13636.1](https://doi.org/10.12688/f1000research.13636.1))
 Latest published: 17 Jan 2018, 7:71 (doi: [10.12688/f1000research.13636.1](https://doi.org/10.12688/f1000research.13636.1))

Abstract




The homeodomain-containing proteins are an important group of transcription factors found in most eukaryotes including animals, plants and fungi. Homeobox genes are responsible for a wide range of critical developmental and physiological processes, ranging from embryonic development, innate immune homeostasis to whole-body regeneration. With continued fascination on this key class of proteins by developmental and evolutionary biologists, multiple efforts have thus far focused on the identification and characterization of homeobox orthologs from key model organisms in attempts to infer their evolutionary origin and how this underpins the evolution of complex body plans. Despite their importance, the genetic complement of homeobox genes has yet been described in one of the most valuable groups of animals representing economically important food crops. With crustacean aquaculture being a growing industry worldwide, it is clear that systematic and cross-species identification of crustacean homeobox orthologs is necessary in order to harness this genetic circuitry for the improvement of aquaculture sustainability. Using publicly available transcriptome data sets, we identified a total of 4183 putative homeobox genes from 120 crustacean species that include food crop species, such as lobsters, shrimps, crayfish and crabs. Additionally, we identified 717 homeobox orthologs from 6 other non-crustacean arthropods, which include the scorpion, deer tick, mosquitoes and centipede. This high confidence set of homeobox genes will now serve as a key resource to the broader community for future functional and comparative genomics studies.


Keywords

Crustacean, homeobox, TALE, comparative genomics, arthropod, homeodomain

Open Peer Review

Referee Status:   

	Invited Referees		
	1	2	3
version 1 published 17 Jan 2018	 report	 report	 report

- 1 **Ricardo M. Zayas**, San Diego State University (SDSU), USA
- 2 **Nathan J Kenny**, The Natural History Museum of London, UK
- 3 **Colleen Doherty** , North Carolina State University, USA

Discuss this article

Comments (0)

Corresponding author: Alvina G. Lai (alvina.lai@ndm.ox.ac.uk)

Author roles: **Chang WH:** Conceptualization, Data Curation, Formal Analysis, Investigation, Methodology, Validation, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing; **Lai AG:** Conceptualization, Data Curation, Formal Analysis, Funding Acquisition, Investigation, Methodology, Supervision, Validation, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing

Competing interests: No competing interests were disclosed.

How to cite this article: Chang WH and Lai AG. **A TALE of shrimps: Genome-wide survey of homeobox genes in 120 species from diverse crustacean taxa [version 1; referees: 2 approved, 1 approved with reservations]** *F1000Research* 2018, 7:71 (doi: [10.12688/f1000research.13636.1](https://doi.org/10.12688/f1000research.13636.1))

Copyright: © 2018 Chang WH and Lai AG. This is an open access article distributed under the terms of the [Creative Commons Attribution Licence](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. Data associated with the article are available under the terms of the [Creative Commons Zero "No rights reserved" data waiver](#) (CC0 1.0 Public domain dedication).

Grant information: This work was supported by the EMBO Fellowship and the Human Frontier Science Program Fellowship to AGL. *The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.*

First published: 17 Jan 2018, 7:71 (doi: [10.12688/f1000research.13636.1](https://doi.org/10.12688/f1000research.13636.1))

Introduction

As one of the fastest growing industries, the seafood trade is dominated by fishing and farming of crustaceans, with annual sales exceeding \$40 billion (Stentiford *et al.*, 2012). Crustacean aquaculture is multi-faceted, not only contributing to the ever-increasing demands by international markets, but is also directly linked to the socio-economic aspects of many developing nations through the creation of jobs and infrastructure. Aquaculture practices have intensified in recent years to cope with the demand. Yet, many are not sustainable since the increased densities of farmed shrimps often serve as hotbeds for pathogens if left unabated, causing infectious diseases and the devastation of cultures resulting in massive financial losses. As a result, regulations associated with aquaculture diseases are being enforced with emphasis placed on preventative measures, e.g. enhancement of broodstock and research aiming to further our understanding on crustacean development and ways to utilize the innate ability of crustaceans to combat pathogens (Lai & Aboobaker, 2017; Stentiford *et al.*, 2012).

Several conserved molecular genetic circuitries are well-known for regulating many aspects of development and innate immune homeostasis. One prominent example would be homeobox genes, a family of transcription factors defined by the presence of a homeodomain (Holland, 2013). As one of the most important master controls in development, some headway has already been made in understanding the involvement of homeobox genes in innate immunity; Caudal in *Drosophila melanogaster* is implicated in commensal-gut mutualism (Ryu *et al.*, 2004; Ryu *et al.*, 2008). Given their importance, major efforts have thus far focused on characterization of homeobox genes in well-known model organisms such as humans (Garcia-Fernández, 2005; Holland *et al.*, 2007), *Caenorhabditis elegans* (Bürglin, 1997), *D. melanogaster* (Mukherjee & Bürglin, 2007), planarians (Currie *et al.*, 2016; Felix & Aboobaker, 2010; Garcia-Fernandez *et al.*, 1991), amphioxus (Luke *et al.*, 2003), teleost fish (Mulley *et al.*, 2006) and many more. Although homeobox orthologs have been previously studied in the crustacean *Parhyale hawaiiensis* (Kao *et al.*, 2016), systematic and cross-species characterization of this gene family across the broader Crustacea with focus on food crop species is currently lacking. A better understanding of homeobox genes in crustaceans is therefore required to address this major shortfall, leading us to our present work.

Methods

Transcriptome data sets and query sets

We retrieved complete transcriptome data sets for 120 crustacean species available at the time of manuscript preparation from the European Nucleotide Archive. Six non-crustacean arthropod proteomes were retrieved from Uniprot. A complete list of accessions used in this study is provided in Supplementary Table 1. We retrieved a list of query sequences used in subsequent homology searches from Uniprot and GenBank.

Identification of homeobox orthologs

Based on a previously published workflow (Lai & Aboobaker, 2017), we used multiple Basic Local Alignment Search Tool

(BLAST)-based approaches, such as BLASTp and tBLASTn to identify genes with homeodomain sequences. 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. We then utilized HMMER (version 3.1) employing hidden Markov models (HMM) profiles (Finn *et al.*, 2011) to scan for the presence of Pfam homeodomains (Bateman *et al.*, 2004) on the best reciprocal nr BLAST hits, to compile a final non-redundant set of crustacean and arthropod homeobox gene orthologs (Dataset 1).

Multiple sequence alignment and phylogenetic tree construction

Multiple sequence alignment of homeodomain sequences was performed using MAFFT (version 7) (Katoh *et al.*, 2009). Phylogenetic tree was built from the MAFFT alignment using RAXML WAG + G model to generate a best-scoring maximum likelihood tree (Stamatakis, 2014). Geneious (version 7) was used to generate a graphical representation of Newick tree (Kearse *et al.*, 2012).

Results and discussion

Identification of putative homeobox genes in crustaceans

With the recent availability of a large number of transcriptome data sets, we perform an extensive search for homeobox genes from 120 crustacean species. We focus on species represented across the broader Crustacea sampling from three main crustacean classes, Malacostraca, Branchiopoda and Copepoda, with focus on key food crop species from the order Decapoda (Supplementary Table 1). Using BLAST-based approaches and profile HMM (Bateman *et al.*, 2004; Finn *et al.*, 2011; Finn *et al.*, 2015) for homology searches, we conservatively identified 4183 transcripts with homeodomain sequences from crustaceans (Figure 1; Dataset 1). Additionally, we included six non-crustacean arthropod species in our search and from these species, we identified 717 homeobox orthologs (Figure 1; Dataset 1).

Dataset 1. List of Pfam annotated homeobox genes and associated e-values in crustaceans and other arthropods

<http://dx.doi.org/10.5256/f1000research.13636.d190417>

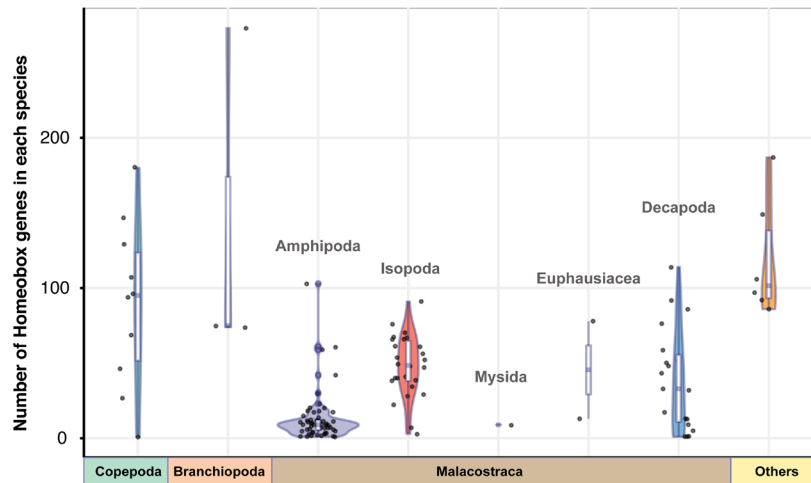
Dataset 2. Fasta file for homeobox gene sequences in crustaceans and other arthropods

<http://dx.doi.org/10.5256/f1000research.13636.d190418>

Classification and phylogenetic analysis of TALE class genes

Concerted efforts to establish evolutionary classification of homeobox genes have resulted in 11 recognised classes (Edvardsen *et al.*, 2005; Holland *et al.*, 2007; Ryan *et al.*, 2006; Zhong *et al.*, 2008; Zhong & Holland, 2011). The Three-Amino acid-Loop Extension (TALE) superclass within the group of homeobox genes is characterized by three additional residues between alpha helices 1 and 2 of the homeodomain (Bertolino *et al.*, 1995). TALE class homeodomain proteins are further divided into

A



B

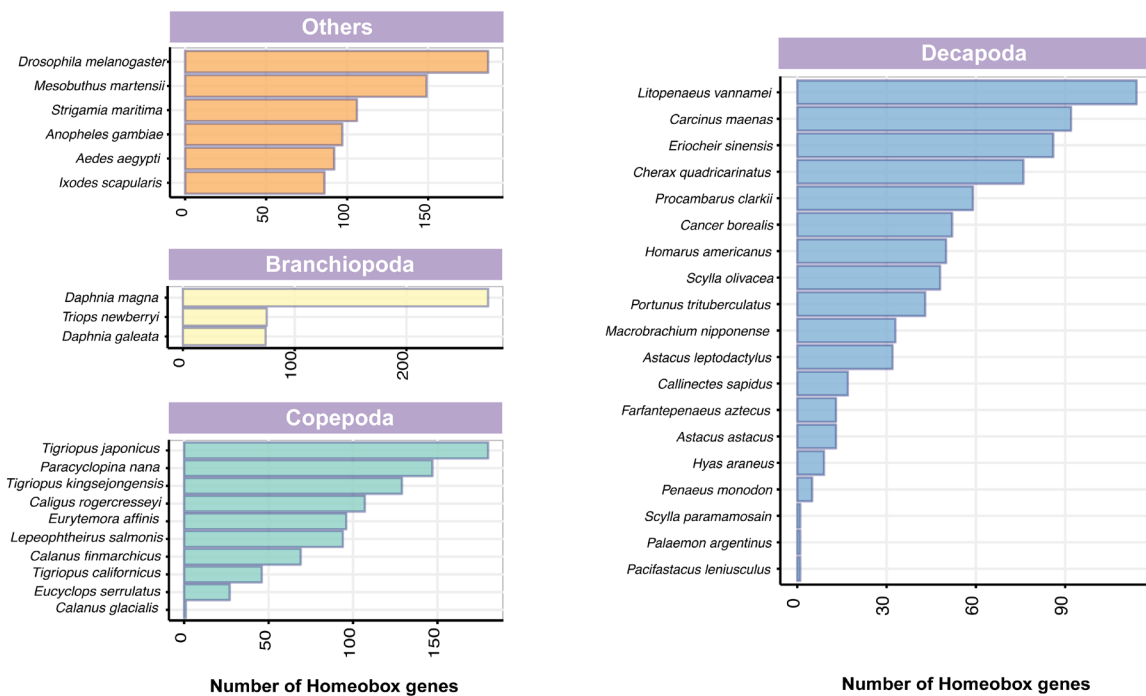


Figure 1. The homeobox superfamily in Crustacea and representative arthropod species. (A) Number of homeobox gene orthologs identified in each species are depicted as boxplots, indicating the median and quartiles. Violin plots underlying the boxplots illustrate sample distribution across different crustacean taxa and kernel probability density (width of the shaded areas represent the proportion of data located in these areas). The homeobox gene orthologs from six non-crustacean species within Arthropoda (others) are also shown. **(B)** Bar charts illustrating the number of homeobox gene orthologs in crustaceans from Decapoda, Branchiopoda and Copepoda along with six non-crustacean arthropods (others).

6 subclasses, Meis, Pknox, Pbc, Irx, Mkx and Tgif characterized by distinct motifs beyond the homeodomain (Bürklin, 1997; Bürklin, 2005; Holland *et al.*, 2007; Mukherjee & Bürklin, 2007). We have classified a total of 165 TALE class orthologs from 15 decapod crustacean species (Figure 2). These genes form

distinct phylogenetic grouping, which allows confident assignment of decapod TALE class orthologs into 6 sub-families (Figure 2). Importantly, the tree topology of crustacean TALE class orthologs recapitulated observations from a previous study (Holland *et al.*, 2007).

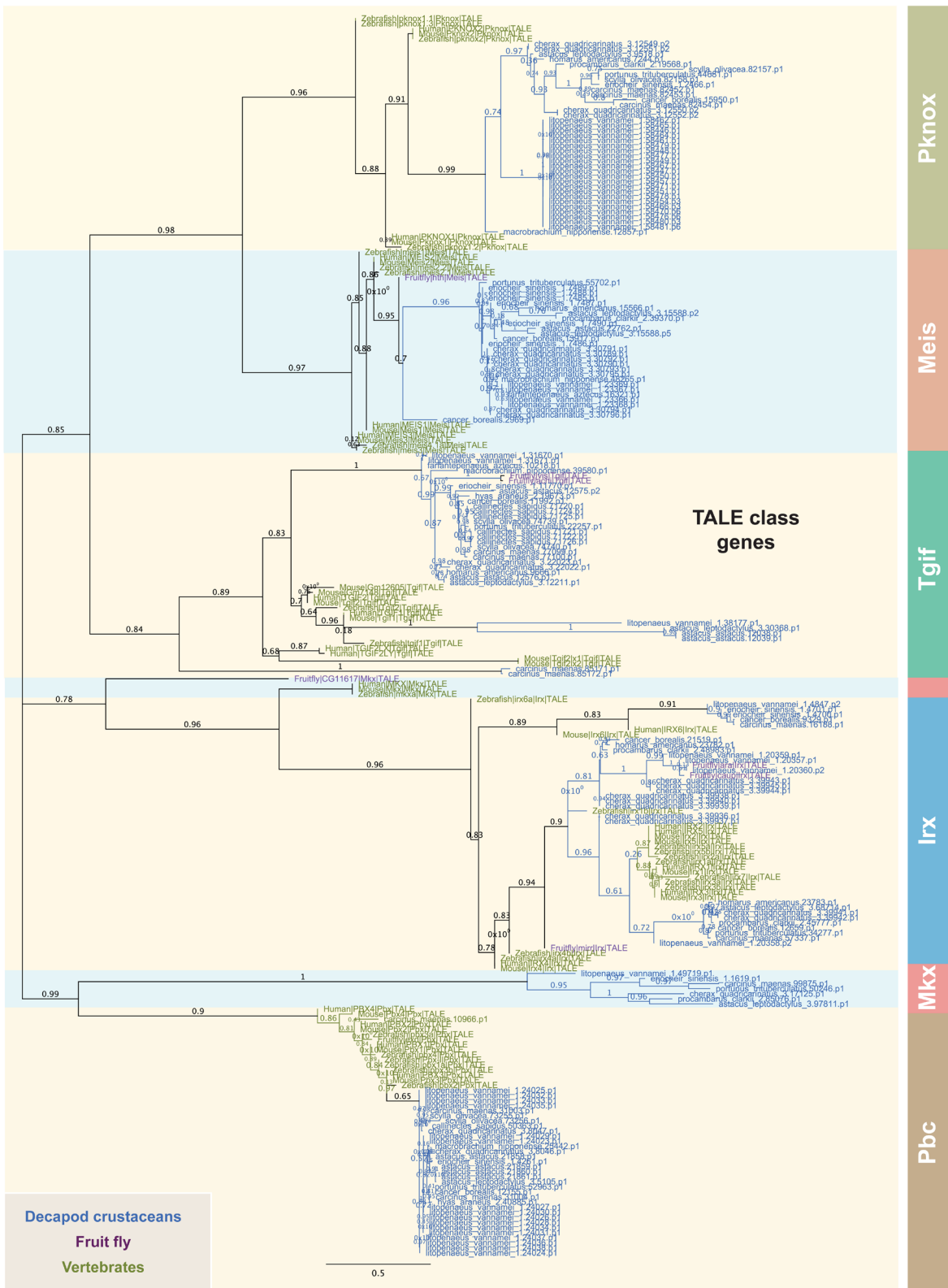


Figure 2. Phylogeny of TALE superclass orthologs in decapod crustaceans. The tree was constructed using the maximum-likelihood method from an amino acid multiple sequence alignment, which include TALE class genes from other species (Zhong *et al.*, 2008 and Zhong & Holland, 2011). TALE orthologs representing 6 subclasses are colour-coded. The node labels of each taxon are marked with distinctive colors denoted in the figure inset. Bootstrap support values (n=1000) are denoted as branch labels.

Conclusion

We identified 4900 homeobox transcripts from 120 crustaceans and 6 non-crustacean arthropod species. Although this data set is non-exhaustive – transcriptomes contain only genes expressed at the point of sample collection – it will now serve as a key resource for future functional studies in the context of crustacean aquaculture. Beyond crustaceans, this work is widely applicable to studies on homeobox genes from other animals and will facilitate evolutionary and comparative genomics investigations.

Data availability

Dataset 1: List of Pfam annotated homeobox genes and associated e-values in crustaceans and other arthropods. DOI, [10.5256/f1000research.13636.d190417](https://doi.org/10.5256/f1000research.13636.d190417) (Chang & Lai, 2018).

Dataset 2: Fasta file for homeobox gene sequences in crustaceans and other arthropods. DOI, [10.5256/f1000research.13636.d190418](https://doi.org/10.5256/f1000research.13636.d190418) (Chang & Lai, 2018).

Competing interests

No competing interests were disclosed.

Grant information

This work was supported by the EMBO Fellowship and the Human Frontier Science Program Fellowship to AGL.

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Supplementary material

Supplementary Table 1: List of accession numbers for species used in this study.

[Click here to access the data.](#)

References

- Bateman A, Coin L, Durbin R, *et al.*: **The Pfam protein families database.** *Nucleic Acids Res.* 2004; **32**(Database issue): D138–41.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Bertolino E, Reimund B, Wildt-Perinic D, *et al.*: **A novel homeobox protein which recognizes a TGT core and functionally interferes with a retinoid-responsive motif.** *J Biol Chem.* 1995; **270**(52): 31178–31188.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Bürglin TR: **Analysis of TALE superclass homeobox genes (MEIS, PBC, KNOX, Iroquois, TGIF) reveals a novel domain conserved between plants and animals.** *Nucleic Acids Res.* 1997; **25**(21): 4173–4180.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Bürglin TR: **Homeodomain proteins.** *Rev Cell Biol Mol Med.* 2005.
- Chang WH, Lai AG: **Dataset 1 in: A TALE of shrimps: Genome-wide survey of homeobox genes in 120 species from diverse crustacean taxa.** *F1000Research.* 2018.
[Data Source](#)
- Chang WH, Lai AG: **Dataset 2 in: A TALE of shrimps: Genome-wide survey of homeobox genes in 120 species from diverse crustacean taxa.** *F1000Research.* 2018.
[Data Source](#)
- Currie KW, Brown DD, Zhu S, *et al.*: **HOX gene complement and expression in the planarian *Schmidtea mediterranea*.** *Evodevo* 2016; **7**: 7.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Edvardsen RB, Seo HC, Jensen MF, *et al.*: **Remodelling of the homeobox gene complement in the tunicate *Oikopleura dioica*.** *Curr Biol.* 2005; **15**(1): R12–R13.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Felix DA, Aboobaker AA: **The TALE class homeobox gene *Smed-prep* defines the anterior compartment for head regeneration.** *PLoS Genet.* 2010; **6**(4): e1000915.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Finn RD, Clements J, Arndt W, *et al.*: **HMMER web server: 2015 update.** *Nucleic Acids Res.* 2015; **43**(W1): W30–W38.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Finn RD, Clements J, Eddy SR: **HMMER web server: interactive sequence similarity searching.** *Nucleic Acids Res.* 2011; **39**(Web Server issue): W29–37.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- García-Fernández J: **The genesis and evolution of homeobox gene clusters.** *Nat Rev Genet.* 2005; **6**(12): 881–892.
[PubMed Abstract](#) | [Publisher Full Text](#)
- García-Fernández J, Baguña J, Saló E: **Planarian homeobox genes: cloning, sequence analysis, and expression.** *Proc Natl Acad Sci U S A.* 1991; **88**(16): 7338–7342.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Holland PW: **Evolution of homeobox genes.** *Wiley Interdiscip Rev Dev Biol.* 2013; **2**(1): 31–45.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Holland PW, Booth HA, Bruford EA: **Classification and nomenclature of all human homeobox genes.** *BMC Biol.* 2007; **5**: 47.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Kao D, Lai AG, Stamatakis E, *et al.*: **The genome of the crustacean *Parhyale hawaiiensis*, a model for animal development, regeneration, immunity and lignocellulose digestion.** *eLife.* 2016; **5**: pii: e20062.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Katoh K, Asimenos G, Toh H: **Multiple alignment of DNA sequences with MAFFT.** *Methods Mol Biol.* 2009; **537**: 39–64.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Kearse M, Moir R, Wilson A, *et al.*: **Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data.** *Bioinformatics.* 2012; **28**(12): 1647–1649.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Lai AG, Aboobaker AA: **Comparative genomic analysis of innate immunity reveals novel and conserved components in crustacean food crop species.** *BMC Genomics.* 2017; **18**(1): 389.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Luke GN, Castro LF, McLay K, *et al.*: **Dispersal of NK homeobox gene clusters in amphioxus and humans.** *Proc Natl Acad Sci U S A.* 2003; **100**(9): 5292–5295.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Mukherjee K, Bürglin TR: **Comprehensive analysis of animal TALE homeobox**

genes: new conserved motifs and cases of accelerated evolution. *J Mol Evol.* 2007; **65**(2): 137–153.

[PubMed Abstract](#) | [Publisher Full Text](#)

Mulley JF, Chiu CH, Holland PW: **Breakup of a homeobox cluster after genome duplication in teleosts.** *Proc Natl Acad Sci U S A.* 2006; **103**(27): 10369–10372.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Ryan JF, Burton PM, Mazza ME, *et al.*: **The cnidarian-bilaterian ancestor possessed at least 56 homeoboxes: evidence from the starlet sea anemone, *Nematostella vectensis*.** *Genome Biol.* 2006; **7**(7): R64.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Ryu JH, Kim SH, Lee HY, *et al.*: **Innate immune homeostasis by the homeobox gene *caudal* and commensal-gut mutualism in *Drosophila*.** *Science.* 2008; **319**(5864): 777–782.

[PubMed Abstract](#) | [Publisher Full Text](#)

Ryu JH, Nam KB, Oh CT, *et al.*: **The homeobox gene *Caudal* regulates constitutive local expression of antimicrobial peptide genes in *Drosophila***

epithelia. *Mol Cell Biol.* 2004; **24**(1): 172–185.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Stamatakis A: **RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies.** *Bioinformatics.* 2014; **30**(9): 1312–1313.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Stentiford GD, Neil DM, Peeler EJ, *et al.*: **Disease will limit future food supply from the global crustacean fishery and aquaculture sectors.** *J Invertebr Pathol.* 2012; **110**(2): 141–157.

[PubMed Abstract](#) | [Publisher Full Text](#)

Zhong YF, Butts T, Holland PW: **HomeoDB: a database of homeobox gene diversity.** *Evol Dev.* 2008; **10**(5): 516–518.

[PubMed Abstract](#) | [Publisher Full Text](#)

Zhong YF, Holland PW: **HomeoDB2: functional expansion of a comparative homeobox gene database for evolutionary developmental biology.** *Evol Dev.* 2011; **13**(6): 567–568.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Open Peer Review

Current Referee Status:



Version 1

Referee Report 23 May 2018

doi:10.5256/f1000research.14814.r33514



Colleen Doherty 

Department of Molecular and Structural Biochemistry, North Carolina State University, Raleigh, NC, USA

In this manuscript, Chang and Lai identify sequences of the homeobox genes in crustaceans from transcriptional data. For TALE family members, they classify these orthologs into the six subfamilies. The introduction provides the justification for establishing this resource in these agriculturally important species. In total, this will be an important reference source for future work in understanding the transcriptional regulation of development in crustacean species. There are a few minor questions and suggestions the authors can address.

The details of the BLAST-based approaches and CD-Hit should be described in more detail. Were default settings applied? Perhaps adding an additional supplemental file or link with the commands would clarify this to facilitate replication in other species.

The authors have the opportunity to address how effective their approach is and potential places for improvement. For example, how do the datasets derived from proteomes compare to that from transcriptomes? Is there an overlapping dataset where the two approaches could be compared? Also, how does the identification of homeobox genes using this approach in *Drosophila* compare to the number annotated in this well-studied species? If all are identified this would suggest that this approach may be sufficient if some are missing it suggests that this conservative starting point could be enhanced in the future by additional refinement.

Minor comments:

Introduction, second paragraph, description of homeobox identification in other species. Either provide references for "and many more" or omit. As is, this phrase is vague and doesn't add constructively to the introduction.

Is the work clearly and accurately presented and does it cite the current literature?

Yes

Is the study design appropriate and is the work technically sound?

Yes

Are sufficient details of methods and analysis provided to allow replication by others?

Partly

If applicable, is the statistical analysis and its interpretation appropriate?

Yes

Are all the source data underlying the results available to ensure full reproducibility?

Yes

Are the conclusions drawn adequately supported by the results?

Yes

Competing Interests: No competing interests were disclosed.

Referee Expertise: Transcriptional networks

I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Referee Report 13 April 2018

doi:[10.5256/f1000research.14814.r32836](https://doi.org/10.5256/f1000research.14814.r32836)



Nathan J Kenny

Department of Life Sciences, The Natural History Museum of London, Cromwell Road, London, SW7 5BD, UK

This work makes a putative assessment of the overall homeodomain complements of the transcriptomes of a number of crustacean species. One class of homeodomain containing genes (TALE) from one order of crustaceans (Decapoda) is assessed in detail, but otherwise no attempt is made to categorise putative hits into gene families. The work is therefore preliminary in scope, and would benefit from the provision of even broadest-level classification of hits into appropriate classes/subclasses of gene, which should be fairly straightforward given the diagnostic residues used to categorise these classes.

The title as it stands is misleading - a genome wide survey is not made, and instead transcriptomic data is used, which will (by necessity) be gappy.

A link is made between aquaculture, innate immunity and homeodomain-containing proteins, but this is very tenuous. Particularly, why the focus is on TALE class genes is unclear, if *Caudal* is the gene used as the exemplar for a link between these fields?

The methods section needs to be more precise. For example: "such as BLASTp and tBLASTn to identify genes with homeodomain sequences". What was done? How were protein sequences derived from nucleotide data for blastp searches? What sequences were used to search your datasets? (Perhaps add these to the sentence: "list of query sequences used in subsequent homology searches from Uniprot and GenBank."). The latter is particularly important as more distant sequences may be missed.

I have several questions about e values.

-Dataset 1 contains several ID'd genes with higher E values than the stated cutoff (< 10⁻⁶). Is this deliberate?

-The e-value of $< 10^{-6}$ will also likely result in larger datasets returning more hits, purely as a consequence of how the E (expect) value is calculated. For example, the *Daphnia magna* transcriptome has 271,000 sequences, *Triops* 12,000. Therefore it is much more likely that sequences will make it through your annotation pathway in *Daphnia magna* rather than *Triops*. This will skew the results shown in Fig 1B. Is it possible to show that homeodomain genes are not artificially excluded, perhaps by giving the E values and best blast hits from "next-best" excluded sequences in your initial searches of small datasets, to prove no homeodomain sequences were artificially excluded? This is crucial, given the short length of the homeodomain, which will be the primary source of signal.

Was HMMR really run on the best reciprocal nr hits, as is suggested by your phrasing? Or was it run on the transcriptome-derived data?

Fig 1A: Violin plots are not appropriate here. Look for instance at the Branchiopod data, where 3 points are used to infer this plot.

The results in the tree in Fig 2 seem to indicate that decapod crustaceans completely lack Mx genes, and the presentation of this is disingenuous in text (note the paraphyly of known Mx homologues with regard to the inferred crustacean Mx). Instead, the crustacean Mx seem to be Pbc?

"Importantly, the tree topology of crustacean TALE class orthologs recapitulated observations from a previous study (Holland et al., 2007)." - this statement does not seem to be correct.

The homeodomain complements, especially of ANTP class genes, of several crustacean species have been described previously, but no attempt is made to place the results observed in the context of the annotated sets of other species. Could this be provided? Particularly, the utility of the re-assessment of non-crustacean datasets is unclear, as these resources have been annotated previously in more detail. Were additional homeodomain-containing genes found by this re-analysis? Or fewer?

In short, this work seems to be partially successful in its aims. With the addition of additional information about the identity of sequences, and the correction of the problems noted above, it will be a coherent addition to extant information on crustacean homeodomain-containing genes.

Minor comments:

\$40 billion - which currency? USD?

There are several areas where the phrasing could be improved, e.g.
-"With continued fascination on this key class of proteins"

Sometimes articles (a/the) are missing from the text, e.g.
- "Phylogenetic tree was built from"

Is the work clearly and accurately presented and does it cite the current literature?

Partly

Is the study design appropriate and is the work technically sound?

Partly

Are sufficient details of methods and analysis provided to allow replication by others?

Partly

If applicable, is the statistical analysis and its interpretation appropriate?

Not applicable

Are all the source data underlying the results available to ensure full reproducibility?

Yes

Are the conclusions drawn adequately supported by the results?

Partly

Competing Interests: No competing interests were disclosed.

I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Referee Report 14 February 2018

doi:[10.5256/f1000research.14814.r29917](https://doi.org/10.5256/f1000research.14814.r29917)



Ricardo M. Zayas

Department of Biology, San Diego State University (SDSU), San Diego, CA, USA

The report by Chang and Lai provide an extensive dataset and phylogenetic analysis crustacean homeobox genes. This data will be useful to individuals interested in studying the evolution and function of homeobox genes in crustacea and other organisms. Overall, this manuscript is well written and appears to be technically sound. I only have some very minor comments the authors could address to improve clarity:

1) In the abstract, the rationale for the work takes an intellectual leap: it is unclear how identification of homeobox genes will be useful for aquaculture sustainability. I think the authors provide some compelling reasons within the text of the manuscript.

2) In the results and discussion there are a few instances where the authors should use the past tense.

3) Page 3, last sentence: even though the authors referred to Peter Holland's work, they should be much more precise about the observations from a "previous study". The authors should expand what they mean.

Is the work clearly and accurately presented and does it cite the current literature?

Yes

Is the study design appropriate and is the work technically sound?

Yes

Are sufficient details of methods and analysis provided to allow replication by others?

Yes

If applicable, is the statistical analysis and its interpretation appropriate?

Yes

Are all the source data underlying the results available to ensure full reproducibility?

Yes

Are the conclusions drawn adequately supported by the results?

Yes

Competing Interests: No competing interests were disclosed.

I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

The benefits of publishing with F1000Research:

- Your article is published within days, with no editorial bias
- You can publish traditional articles, null/negative results, case reports, data notes and more
- The peer review process is transparent and collaborative
- Your article is indexed in PubMed after passing peer review
- Dedicated customer support at every stage

For pre-submission enquiries, contact research@f1000.com

F1000Research