

CnidBase: The Cnidarian Evolutionary Genomics Database

Joseph F. Ryan* and John R. Finnerty

Boston University

Received August 15, 2002; Revised and Accepted October 27, 2002

ABSTRACT

CnidBase, the Cnidarian Evolutionary Genomics Database, is a tool for investigating the evolutionary, developmental and ecological factors that affect gene expression and gene function in cnidarians. In turn, CnidBase will help to illuminate the role of specific genes in shaping cnidarian biodiversity in the present day and in the distant past. CnidBase highlights evolutionary changes between species within the phylum Cnidaria and structures genomic and expression data to facilitate comparisons to non-cnidarian metazoans. CnidBase aims to further the progress that has already been made in the realm of cnidarian evolutionary genomics by creating a central community resource which will help drive future research and facilitate more accurate classification and comparison of new experimental data with existing data. CnidBase is available at <http://cnidbase.bu.edu/>.

INTRODUCTION

Benefits of studying the phylum Cnidaria

For several reasons, the phylum Cnidaria is an important venue for evolutionary genomics. First, the phylum Cnidaria is an outgroup to the Bilateria (1) which comprises the overwhelming majority of animals. Because of its outgroup status, the Cnidaria is uniquely positioned to help reconstruct the genome of the ancestral bilaterian, a creature that lived some 600–750 million years ago (2), and gave rise to such diverse descendants as humans, fruitflies and soil nematodes. Where bilaterian genomes differ, the Cnidaria can prove decisive in distinguishing the ancestral bilaterian state from one or more derived states. For example, as the genetic makeup of cnidarians is being revealed, many genes that were thought to have recent origins within the vertebrates due to their absence in *Drosophila* and *Caenorhabditis* are being identified in cnidarians, forcing a shift in thinking concerning the origin of complexity (3). Furthermore, the discovery of genetic processes occurring in cnidarians, such as trans-splicing (4), which were thought to have evolved later in the metazoan

phylogeny have a major impact on our understanding of the evolution of complex genetic networks.

Second, modern-day cnidarians (anemones, hydras, corals and jellyfishes) reflect an intermediate stage in the evolution of animal complexity. Cnidarians possess a nervous system, muscle cells, and eyes, making them much more complex than sponges (1). However, cnidarians lack a brain, a through-gut, and a blood-vascular system, making them much simpler than many bilaterian animals, such as fruitflies, nematodes, or vertebrates. Comparing the genomic regulatory systems of cnidarians, sponges, and bilaterians will reveal the underlying genomic architecture of major evolutionary innovations in ontogeny and body plan.

Third, cnidarians display a tremendous degree of developmental plasticity. Cnidarians ‘develop’ in three distinct life history contexts: (1) embryogenesis, (2) asexual reproduction by fission or budding and (3) regeneration following injury. The genomic basis for this kind of flexibility is not understood, but such flexibility can have profound ecological and evolutionary consequences. For example, a single sea anemone, isolated from potential mates by a chance event, could persist indefinitely as an immortal clone of asexually-produced animals. In contrast to cnidarians, the major model systems in developmental genetics are developmentally rigid and stereotyped. Vertebrates, *Drosophila* and *Caenorhabditis* undergo embryogenesis, but they are incapable of reproducing by fission and incapable of complete bi-directional regeneration (5). Elucidating the genomic architecture underlying the developmental plasticity of cnidarians may provide important insights into the developmental rigidity of other model systems.

The need for a Cnidarian-specific database

Cnidarian research is inherently comparative, as no single model system dominates to the extent that, for example, *Drosophila* and *Caenorhabditis* dominate research on arthropods and nematodes. The physiologist August Krogh succinctly stated one of the pillars of comparative biology: ‘For many problems there is an animal in which it can be most conveniently studied’ (6). In the spirit of the ‘Krogh principle’, several species from the phylum Cnidaria have emerged as models for various fundamental biological problems. *Hydra* has long been a model for research into regeneration and asexual budding (7). The sea anemone *Nematostella* (8), the

*To whom correspondence should be addressed. Email: jfryan@bu.edu

coral *Acropora* (3), and the jellyfish *Podocoryne* (9) are emerging models for cnidarian embryogenesis. The colonial hydroids *Hydractinia* (10) and *Eleutheria* (11) are providing insights into colony development. *CnidBase* will foster the advance of cnidarian genomics by synthesizing the data from these distantly related lineages on an explicit phylogenetic framework. This will allow phylogenetic reconstruction of the ancestral cnidarian condition, permitting the discrimination of primitive traits from derived traits and facilitating comparisons to other animal phyla.

An additional motivation for the establishment of *CnidBase* is that cnidarian genomic data do not receive sufficient explanatory annotation in taxonomically generic databases due to several phylum-specific features. For example, cnidarians have complex life histories with three distinctive stages: the planula larva, the polyp and the medusa (12). The polyp, the familiar body plan of a sea anemone, is commonly a sessile benthic creature. The medusa, or jellyfish, is commonly an active pelagic animal. Both the polyp and the medusa can be regarded as adult life stages and a single species may exist as a polyp, or a medusa, or alternate between the polyp and medusa stages (12). A thorough, meaningful description of gene expression in cnidarians should track changes across life stages and allow comparison of equivalent life stages between species.

THE DATABASE

CnidBase stores expression data for cnidarian genes. Each expression assay is assigned a unique accession (i.e. Cnx1). Expression data are cross-referenced according to eight criteria: (1) life history stage, (2) body region, (3) body layer, (4) cell type, (5) expression assay, (6) expression level, (7) cnidarian taxonomy and (8) gene orthology. Life history stage is broken down into six major categories: embryogenesis (spanning gametogenesis through fertilization and larval development), adult polyp, adult medusa, colony, asexual reproduction and regeneration. Body region is broken down according to major regions along the main body axis and subdivisions thereof: in the case of a polyp, the major body regions would be the head, the body column and the foot. Three body layers are recognized: the outer ectoderm or epidermis, the inner endoderm or gastrodermis and the central mesoglea. Twelve cell types are recognized, including neurons, epitheliomuscular cells, intermediate cells and nematocytes. Expression assay includes RT-PCR, Q-PCR, *in situ* hybridization, SAGE, nuclease protection, antibody hybridization and microarray analysis. Expression level is described as 'absent' or 'detected' unless the author notes a more precise description such as 'strong' or 'weak'. Cnidarian taxonomy is arranged according to the internal phylogenetic relationships of the phylum and the corresponding taxonomic hierarchy, including class, order, family, genus and species. Gene orthology is based on explicit phylogenetic analysis where this has been done, or upon sequence similarity according to BLAST searches.

The establishment of gene orthology is a necessary first step to understanding the functional evolution of genes. The best approach to this problem is through rigorous phylogenetic

analysis. At this time, thorough and rigorous phylogenetic analyses have not been performed for most cnidarian genes. The homeobox gene family is an exception, which serves to illustrate the analytical approach of *CnidBase* for establishing gene orthology.

Homeobox genes are an ancient family of transcription factors that are known to regulate important developmental processes in plants, animals and fungi (13). Homeobox genes are easily recognized and easily isolated because they encode a highly conserved DNA-binding domain known as the homeodomain. The functional diversification of homeobox genes is thought to have been critical to the evolution of animals, underlying the origin of key innovations, such as the primary body axis (associated with *Hox* genes), the eyes [associated with *pax6* (14)], or the heart [associated with *tinman* (15)]. Because of the role of homeodomains in patterning, body plan specification and cell differentiation, the homeodomain is the most highly represented cnidarian protein family in GenBank.

Homeobox genes are currently the most broadly studied gene family in Cnidaria, with >90 homeobox genes having been identified in 15 species (Kwong, Burton, Mazza and Finnerty, unpublished results). Approximately 10% of the cnidarian protein sequences in GenBank contain a homeodomain. To address the relationship among cnidarian homeobox genes and between cnidarian and bilaterian homeobox genes, a broadly representative phylogenetic dataset consisting of homeodomain sequences has been compiled. The dataset encompasses a pair of sequences, one deuterostome representative and one protostome representative, for each homeobox gene that is inferred to have been present in the ancestral bilaterian (Kwong, Burton, Mazza and Finnerty, unpublished results). The bilaterian sequences are stored in *CnidBase* and have been cross-referenced (where possible) to NCBI's LocusLink (16). Synonyms for these genes have been imported to make searching for cnidarian genes with bilaterian homologues more robust. This dataset provides a phylogenetic standard of comparison against which any newly isolated cnidarian homeoboxes can be compared.

Data acquisition

The first cnidarian gene expression assays were published about ten years ago, and since then, ~100 studies producing cnidarian gene expression data have been reported. The publication of such data is accelerating rapidly and at the same time, the data are becoming more complicated, involving more species, more kinds of assays, and more precise characterization according to developmental stage, body region, cell-type, etc. The amount of comparative cnidarian data is rapidly outstripping the ability of researchers to deduce evolutionary patterns without the assistance of a relational database. The data presently populating *CnidBase* has been acquired through manual curation of the primary literature. In addition, the database currently allows the submission of gene expression data and we encourage cnidarian researchers to submit their expression results directly to be included in the database. *CnidBase* accession numbers may be provided for use in forthcoming publications and upon request, all pre-published entries may be stored as private records until

Gene Expression Data Query Form

Retrieve Reset Form

Gene Symbol/Name:

Organism:

Expression Level:

Life History Stage: (You can browse stage descriptions)

ANY
Adult Polyp
Asexual Reproduction
Asexual Reproduction/Budding
Asexual Reproduction/Fission

Body Region: (You can browse body region descriptions)

Evagination of Bud or Transplant (Early Stages)
Foot
Foot/Polyp Base
Head
Head/Apical Tip
Head/Developing Head of Bud or Transplant
Head/Hypostome
Head/Hypostome/Around Mouth
Head/Tentacle Zone

Layers:

ANY
Epidermis
Gastrodermis
Mesoglea

Expression Assays

Assay Accession	Assay Type	Pub	Sequence Accession	Sequence Accession Name	Organism
Cnx223	Immunohistochemistry	Cnp26 Cns97	cnox2	Hydractinia symbiolong	
Cnx224	Immunohistochemistry	Cnp26 Cns97	cnox2	Hydractinia symbiolong	
Cnx225	Immunohistochemistry	Cnp26 Cns97	cnox2	Hydractinia symbiolong	
Cnx226	Immunohistochemistry	Cnp26 Cns97	cnox2	Hydractinia symbiolong	
Cnx229	RNA In Situ	Cnp22 Cns67	cnox2	Hydra vulg	
Cnx230	RNA In Situ	Cnp22 Cns67	cnox2	Hydra vulg	
Cnx233	RNA In Situ	Cnp22 Cns67	cnox2	Hydra vulg	
Cnx234	RNA In Situ	Cnp22 Cns67	cnox2	Hydra vulg	

Assay Display: Cnx233

Sequence: [Cns67](#)
Organism: [Hydra vulgaris](#)
Assay Type: RNA In Situ
Life History Stage: Regeneration/lower-head-forming 24 Hour
Citation: Gauchat D et al. **Evolution of Antp-class genes and differential expression of Hydra Hox/paraHox genes in anterior patterning.** Proc Natl Acad Sci U S A. 2000 Apr 25;(9):. PMID: [10781050](#) CnidBase Publication Accession: [Cnp22](#)

Expression Results

Results Accession	Layers	Body Region	Cell Type	Expression Level
Cnr153	Epidermis	Regenerating Tip		detected

Figure 1. CnidBase Gene Expression Query. The left screen depicts the CnidBase query form. Search fields include gene symbol, organism, expression level, life history stage, body region, layers, assay types and cell types. The polyp image changes depending on which body region is chosen from the scroll box. The upper right screen shows the result of a query. Assay accessions are linked to the observations of that particular assay (lower right). Sequence accessions are linked to information about a gene including its known homologues and other expression assays.

publication. Currently, CnidBase stores 105 expression assays and 170 expression results.

Query interface

CnidBase provides a set of web-accessible query tools that initially allow the identification of interesting targets through keyword search sequence homology (BLAST), literature search and detailed expression queries. The keyword search allows case insensitive searches of accession numbers, gene names, gene definition lines and LocusLink synonyms from homologous bilaterian genes. CnidBase BLAST uses NCBI's BLAST program (17) to query a cnidarian database of proteins or nucleic acids that are automatically updated nightly from NCBI. The BLAST results are linked back into CnidBase. The CnidBase literature scan provides another tool to identify potential targets through a query interface to the scientific literature used to populate CnidBase. These queries allow users to identify genes or expression assays associated with particular scientific papers or keywords within abstracts, titles and MESH terms. The literature scan facilitates the search for genes or expression data associated with more fuzzy concepts such as a particular process or line of research.

The CnidBase gene expression search allows for querying expression information by life history stage (i.e. gamete, zygote, polyp, etc.), body region (i.e. hypostome, body column, tentacles, etc.), assay type (i.e. RNA *in situ*, immunohistochemistry, etc.), cell types (i.e. epithelial, neurons, etc.), layers (i.e. epidermis, mesoglea, etc.) and PubMedID. A hierarchical vocabulary has been developed to allow for standardized annotation of expression and to facilitate more flexible querying (Fig. 1). For example, a query for expression 'detected' in the 'body column' will retrieve records for 'moderate' expression in the 'peduncle' (the lower part of the body column), as well as strong expression in the 'gastric region' (the upper part of the body column). Help pages offer descriptions for the vocabulary. In addition, there are graphical diagrams of life-history stages, body layers and body regions. The interface is based on The Gene Expression Database (GXD) (18) which has proved to be a valuable resource in mouse research.

All query results are linked directly to GenBank, PubMed, NCBI Taxonomy Database and LocusLink. Additionally, the front page of CnidBase displays a list of papers involving cnidarian research appearing in PubMed within the last 30 days. Finally, we provide a page of links to other cnidarian-related web sites.

Addressing biological questions

The expression query tool will allow researchers to ask pertinent biological questions and the phylogenetic relationships within the database improve the value of the expression queries. For example, a researcher studying head development in *Hydra* may make the query, 'Display all homeobox genes expressed in the future head region during the planula stage of development'. While the query may not return a gene from *Hydra* expressed in the head region, it may return a gene from a coral that is phylogenetically linked (orthologous) to a *Hydra* gene. While no expression assay has yet been done for the *Hydra* gene itself, the coral expression pattern implies that the gene may have a head-specific pattern of expression in *Hydra*, thereby suggesting an experimental follow-up. Conversely, it is also possible that the orthologous gene has evolved a distinctive role in a different species. The phylogenetic underpinnings of *CnidBase* will emphasize such cases, helping to discriminate functional diversification of genes from functional conservation.

CnidBase will be especially valuable for identifying and resolving incongruous or non-parsimonious situations, where an orthologous gene displays different expression patterns in different taxa, at different life-history stages, or even using different expression assays. The *cnox2* gene presents such a quandary. Expression of *cnox2* (or its synonymous orthologs) has been assayed in six species of Cnidaria (19–25). In every species and every developmental context examined, the gene has been found to exhibit axially-biased expression, characteristic of an axial patterning gene. However, the region of strongest expression along the body axis varies across species. In some cases, *cnox2* is expressed only in the oral region, while it is excluded from the oral region of other species. Has the role of the gene evolved? Differences between studies may be attributable to varying developmental context (adult, regenerating adult, colony, or larva), assay methods (antibody or riboprobe), or taxa (*Hydra* versus sea anemone). Once we have identified the parameters that are most directly correlated with differences in expression, we can test whether these differences are biologically real, or attributable to experimental error or misinterpretation. By instantly summarizing all expression data for a given gene across taxa, developmental contexts and assay methods, *CnidBase* will suggest future experiments that are likely to resolve these sorts of incongruities.

WORK IN PROGRESS—FUTURE DIRECTIONS

The focus of cnidarian research ranges from evolutionary biology, to developmental biology, ecology, toxicology and biomedical research (regeneration and vision). Feedback from these diverse research groups will expand the types of data that are stored and the types of queries that can be made. For instance, we already anticipate the importance of such traits as ecophysiological condition to researchers relating ecology to biochemical characteristics. As such, we plan to add to our expression data the ability to store and query the nutritional state of the animal as well as the effects of external environmental influences such as the presence of predators, the presence of conspecifics and changes in temperature, light and water chemistry.

As certain technologies become more affordable and further perfected, *CnidBase* will respond and expand to accommodate such data. For example, as quantitative cnidarian expression data become available (e.g. Q-PCR), *CnidBase* will provide the tools necessary to take advantage of this data. Moreover, as more mapping and genomic sequence data begins to be generated in cnidarians, *CnidBase* will create the necessary architecture to store and integrate such new information. As cnidarian functional genomics data continues to grow and the functions of a significant number of genes become well-categorized, *CnidBase* will begin to store functional classification data. It is anticipated that such functional classification will be based on existing frameworks established for other model systems by the Gene Ontology Project (23) with revisions to reflect unique characteristics of cnidarians.

As more phylogenetic data are incorporated into *CnidBase*, we will work to establish a stable, phylogenetically-based classification scheme for cnidarian genes. By storing alignments and precomputed pairwise distances from phylogenetic analyses, it is possible to perform fast phylogenetic searches (26). Such an interface to this data will help cnidarian researchers accurately classify and analyze new cnidarian genes.

CONCLUSION

The amount of cnidarian research data generated thus far is modest when compared to model bilaterian systems. Nevertheless, it has reached a point where it is becoming difficult for most researchers to easily keep track of published experimental data. Due to the unique characteristics of cnidarian animals and given the comparative nature of research within this phylum, a specialized database is necessary. *CnidBase* organizes the existing cnidarian experimental data into a central repository and provides the necessary query tools to allow cnidarian researchers to pose biologically relevant questions.

ACKNOWLEDGEMENTS

We would like to thank David Lorenz for a critical reading of this manuscript. We gratefully acknowledge the institutional and financial support of The Department of Biology and The Program in Bioinformatics at Boston University. This research was supported in part by NSF grant #9727244 to M. Q. Martindale and J.R.F. and The NSF Integrative Graduate Education and Research Traineeship program grant to J.F.R. The establishment of *CnidBase* benefited greatly from an unpublished phylogenetic analysis of cnidarian homeobox genes by G. Kwong, P. Burton, M. Mazza and J.R.F.

REFERENCES

1. Nielsen, C. (1995) *Animal Evolution. Interrelationships of the Living Phyla*. Oxford University Press, UK.
2. Ayala, F.J. and Rzhetsky, A. (1998) Origin of the metazoan phyla: molecular clocks confirm paleontological estimates. *Proc. Natl Acad. Sci. USA*, **95**, 606–611.
3. Ball, E.E. et al. (2002) Coral development: from classical embryology to molecular control. *Int. J. Dev. Biol.*, **46**, 671–678.
4. Stover, N.A. and Steele, R.E. (2001) Trans-spliced leader addition to mRNAs in a cnidarian. *Proc. Natl Acad. Sci. USA*, **98**, 5693–5698.

5. Pearson, H. (2001) The regeneration gap. *Nature*, **414**, 388–390.
6. Krogh, A. (1929) Progress of physiology. *Am. J. Physiol.*, **90**, 243–251.
7. Lenhoff, S.G. and Lenhoff, H.M. (1986) *Hydra and the Birth of Experimental Biology, 1744: Abraham Trembley's Memoirs Concerning the Natural History of a Type of Freshwater Polyp with Arms Shaped Like Horns*. Boxwood Press, Pacific Grove, CA.
8. Finnerty, J.R. (1998) Homeoboxes in sea anemones and other nonbilaterian animals: implications for the evolution of the hox cluster and the zootype. *Curr. Top. Dev. Biol.*, **40**, 211–254.
9. Masuda-Nakagawa, L.M., Groer, H., Aerne, B.L. and Schmid, V. (2000) The HOX-like gene *cnox2*-Pc is expressed at the anterior region in all life cycle stages of the jellyfish *Podocoryne carnea*. *Dev. Genes. Evol.*, **210**, 151–156.
10. Cartwright, P. and Buss, L.W. (1999) Colony integration and the expression of the Hox gene, *cnox2*, in *Hydractinia symbiolongicarpus* (Cnidaria: Hydrozoa). *J. Exp. Zool.*, **285**, 57–62.
11. Kuhn, K., Streit, B. and Schierwater, B. (1996) Homeobox genes in the cnidarian *Eleuthera dichotoma*: evolutionary implications for the origin of Antennapedia-class (HOM/Hox) genes. *Mol. Phylogenet. Evol.*, **6**, 30–38.
12. Brusca, R.C. and Brusca, G.J. (1990) *Invertebrates*. Sinauer Associates, Sunderland, MA.
13. Bharathan, G., Janssen, B.J., Kellogg, E.A. and Sinha, N. (1997) Did homeodomain proteins duplicate before the origin of angiosperms, fungi and metazoa? *Proc. Natl Acad. Sci. USA*, **94**, 13749–13753.
14. Callaerts, P., Halder, G. and Gehring, W.J. (1997) PAX-6 in development and evolution. *Annu. Rev. Neurosci.*, **20**, 483–532.
15. Evans, S.M., Yan, W., Murillo, M.P., Ponce, J. and Papalopulu, N. (1995) Tinman, a *Drosophila* homeobox gene required for heart and visceral mesoderm specification, may be represented by a family of genes in vertebrates: XNkx-2.3, a second vertebrate homologue of tinman. *Development*, **121**, 3889–3899.
16. Pruitt, K.D. and Maglott, D.R. (2001) RefSeq and LocusLink: NCBI gene-centered resources. *Nucleic Acids Res.*, **29**, 137–140.
17. Altschul, S.F. et al. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.*, **25**, 3389–3402.
18. Ringwald, M. et al. (2001) The Mouse Gene Expression Database (GXD). *Nucleic Acids Res.*, **29**, 98–101.
19. Schummer, M., Scheurlen, I., Schaller, C. and Galliot, B. (1992) HOM/HOX homeobox genes are present in hydra (*Chlorohydra viridissima*) and are differentially expressed during regeneration. *EMBO. J.*, **11**, 1815–1823.
20. Shenk, M.A., Gee, L., Steele, R.E. and Bode, H.R. (1993) Expression of *cnox2*, a HOM/HOX gene, is suppressed during head formation in hydra. *Dev. Biol.*, **160**, 108–118.
21. Shenk, M.A., Bode, H.R. and Steele, R.E. (1993) Expression of *cnox2*, a HOM/HOX homeobox gene in hydra, is correlated with axial pattern formation. *Development*, **117**, 657–667.
22. Gauchat, D. et al. (2000) Evolution of Antp-class genes and differential expression of Hydra Hox/paraHox genes in anterior patterning. *Proc. Natl Acad. Sci. USA*, **97**, 4493–4498.
23. Yanze, N., Spring, J., Schmidli, C. and Schmid, V. (2001) Conservation of Hox/ParaHox-related genes in the early development of a cnidarian. *Dev. Biol.*, **236**, 89–98.
24. Hayward, D.C. et al. (2001) Gene structure and larval expression of *cnox2Am* from the coral *Acropora millepora*. *Dev. Genes Evol.*, **211**, 10–19.
25. Cartwright, P., Bowsher, J. and Buss, L.W. (1999) Expression of a Hox gene, *cnox2*, and the division of labor in a colonial hydroid. *Proc. Natl Acad. Sci. USA*, **96**, 2183–2186.
26. Zmasek, C.M. and Eddy, S.R. (2002) RIO: Analyzing proteomes by automated phylogenomics using resampled inference of orthologs. *BMC Bioinformatics*, **3**, 14.