

# Compagen, a comparative genomics platform for early branching metazoan animals, reveals early origins of genes regulating stem-cell differentiation

Georg Hemmrich\* and Thomas C.G. Bosch

## Summary

Large-scale species comparisons at genome and expressed sequence tag (EST) levels have revealed that early branching metazoans such as sponges and cnidarians share many if not most of their genes with the allegedly advanced vertebrates including man. The ancestor of all animals may thus have been much more complex than anticipated. To facilitate and support analysis of genomic and transcriptomic resources in early branching metazoans, we have established a local bio-computational platform, Compagen (<http://www.compagen.org>). The platform contains searchable databases with selected raw genomic and EST sequence datasets from sponges and cnidarians up to the lower vertebrates. In addition to the public datasets, Compagen also provides processed data like CAP3 assembled ESTs or predicted peptides. Evaluating the efficacy of the platform by screening for genes reported to be essential in controlling stem-cell behavior in higher organisms uncovered ancient origins for some but not all components of the vertebrate stem-cell system. *BioEssays* 30:1010–1018, 2008. © 2008 Wiley Periodicals, Inc.

## Introduction

One of the major challenges in developmental biology is to understand how cells become specified and organized in complex tissues. While studies in bilaterian animals such as *Drosophila* and *C. elegans* in the last decade have shown that just a few signaling pathways generate most of the cellular and

morphological diversity during the development of individual organisms,<sup>(1)</sup> only comparative data from more basal animals provide information for reconstructing the early history of bilaterian developmental mechanisms (Fig. 1). Cnidaria are the sister group to the Bilateria<sup>(2,3)</sup> and are the first in evolution to have a defined body plan, a nervous system and a tissue layer construction. One of the surprising discoveries in the last 5 years was the observation that Cnidaria not only have about the same number of genes as human and share most of their genes with human<sup>(4,5)</sup> but that their protein sequences are often more similar to mammalian sequences than to those of fly and worm.<sup>(6,7)</sup> Genes previously thought to have evolved in the deuterostome branch due to their absence in flies and worms, unexpectedly were found to be present and to function in an equivalent manner in cnidarians.<sup>(4,8)</sup> Early-branching metazoans, therefore, not only have preserved much of the genetic complexity of the common metazoan ancestor but also appear to be highly informative with respect to tracing the evolutionary origin of developmental control genes.<sup>(9)</sup>

Until recently, investigations in early-branching metazoans were hindered by the absence of representative molecular datasets like genome sequences or sequenced transcriptomes. Within the last few years, novel computational tools and the development of genomic resources such as large-scale EST Projects as well as genome projects with assembled draft genome sequences (Fig. 1) have brought a new perspective on the developmental capacities of early metazoan animals. To facilitate access and analysis of genomic and transcriptomic data in these organisms, we have established a local bio-computational platform, Compagen. The sequence databases at Compagen contain selected raw genomic and EST sequence datasets from sponges and cnidarians up to the lower vertebrates. In addition to the public datasets, Compagen also provides already processed data like assembled EST datasets or predicted peptides. Compagen allows researchers not only to more quickly refine their data but provides also the “virtual data” for all those interested in

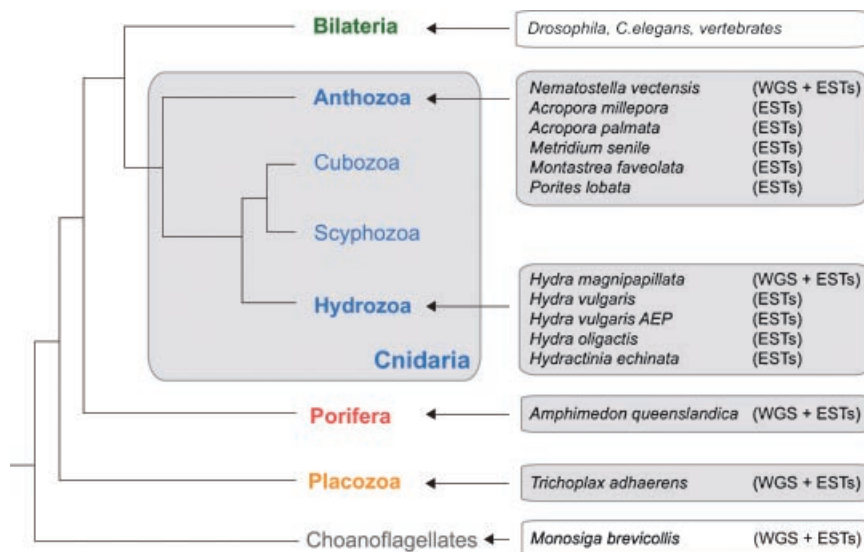
Zoological Institute, Christian-Albrechts-University Kiel, Germany.

Funding agencies: The work was supported in part by grants from the Deutsche Forschungsgemeinschaft (DFG), and grants from the DFG Cluster of Excellence programs “The Future Ocean” and “Inflammation at Interfaces” (to TCGB).

\*Correspondence to: Georg Hemmrich, Zoological Institute, Christian-Albrechts-University Kiel, Olshausenstrasse 40, D-24098 Kiel, Germany. E-mail: ghemmrich@zoologie.uni-kiel.de

DOI 10.1002/bies.20813

Published online in Wiley InterScience ([www.interscience.wiley.com](http://www.interscience.wiley.com)).



**Figure 1.** Phylogenetic relationships of early branching metazoan animals. Placozoans are considered as the earliest branching metazoan animals in the evolutionary tree and represent the sister group to poriferans (sponges), cnidarians (corals, anemones, hydroids, jellyfish) and all bilaterian animals. Cnidarians are amongst the simplest animals at the tissue grade of organization, and are regarded as the closest outgroup to the Bilateria. The sponges (Porifera) are unquestionably animals, but represent a lower level of organization. As non-metazoan outgroup served the unicellular choanoflagellates. For selected taxa/species the availability of genomic (WGS) as well as transcriptomic (EST) sequence data are indicated.

studying the evolutionary origin of developmental control genes.

### Genomic resources in early-branching metazoan animals

The unicellular choanoflagellates represent the closest known relatives of metazoans (Fig. 1).<sup>(10)</sup> Their notable similarity to the feeding cells of sponges, the choanocytes, led taxonomists early on to the idea that there must be a close relationship between both organisms. In the past, several molecular phylogenetic approaches investigating the position of choanoflagellates in the tree of life supported this by comparing their mitochondrial genomes to those of sponges, cnidarians and other metazoan animals.<sup>(10,11)</sup> Finding a proportion of metazoan homologs expressed in choanoflagellates endorses this view.<sup>(12–15)</sup> At the beginning of this year, the draft genome of the choanoflagellate *Monosiga brevicollis* was released<sup>(16)</sup> and promised deeper insights into the transition from unicellular organisms to multicellularity. The assembled genome sequence together with transcriptional data (ESTs) and predicted gene models are accessible and searchable through a common genome browser on the webpage of the Joint Genome Institute.<sup>(16)</sup>

*Trichoplax adhaerens*, a member of the Placozoa (Fig. 1), represents one of the earliest branching metazoan taxa of the animal kingdom.<sup>(17)</sup> Exhibiting an extremely simple body plan generated by only four somatic cell types,<sup>(18)</sup> together with its

taxonomic position, makes it a suitable model to study the transition from unicellular to multicellular animals. The discovery of evolutionary conserved developmental genes<sup>(19,20)</sup> points towards a reduced complexity in gene families compared to higher metazoans. Sequencing of the *Trichoplax* mitochondrial genome<sup>(17)</sup> consolidated the taxonomic position in the tree of life. The recently sequenced genome is accessible through JGI's genome browser, but the community is still awaiting an official publication.

The oldest animals in the sister group of the Placozoa are the Porifera (sponges; Fig. 1). Current research projects mostly investigate the demosponges *Suberites domuncula* and *Amphimedon queenslandica*.<sup>(9)</sup> These multicellular animals consist of at least ten different cell types, including the characteristic choanocytes, but they lack symmetry around a body axis and, thus, have no defined body plan.<sup>(21)</sup> Although established molecular techniques are still limited, preliminary evidence for the presence of conserved genes involved in both developmental processes<sup>(9,22,23)</sup> and the innate immune system<sup>(24,25)</sup> underlines the importance of sponges for comparative studies and the discovery of molecular ancestry and/or secondary gene loss. To unravel the complete ancestral metazoan gene set, the whole genome shotgun (WGS) sequencing project of *Amphimedon queenslandica* and extensive EST sequencing are crucial for further investigations.

Cnidarians as sister-group to all bilaterian animals (Fig. 1) are the first organisms in evolution that have developed a

defined body plan, stem-cell systems, nerve cells and a tissue layer construction. In contrast to the triploblastic Bilateria, cnidarians develop from two germ layers, the ectoderm and the endoderm, and are thus referred to as diploblasts lacking the mesoderm.<sup>(26)</sup> The two body layers are organized around a single (oral–aboral) body axis, forming a gastric cavity that is defined by the mouth opening at one end. The synapomorphic feature of the Cnidaria is the so-called cnidocyte or nematocyte (stinging cell), which is used to catch prey or to defend predators.<sup>(27)</sup> Cnidarians can be grouped (Fig. 1) into the most-basal Anthozoa (corals and anemones) and the Medusozoa, consisting of the Cubozoa (sea wasps), the Scyphozoa (jellyfishes) and the Hydrozoa (hydroids).<sup>(28,29)</sup>

For two cnidarian model systems, the sea anemone *Nematostella vectensis* and the freshwater polyp *Hydra magnipapillata*, extensive molecular resources have been established within the last three years.<sup>(30,31)</sup> Whole genome shotgun (WGS) sequencing approaches generated sequence data for draft genome assemblies with at least six-fold coverage. Two different research groups meanwhile assembled the *Nematostella vectensis* genome and made their results accessible for analysis through classical genome browsers<sup>(30)</sup> or online blast-platforms.<sup>(32)</sup> The recently published draft genome sequence<sup>(30)</sup> revealed that the genomic complexity and gene repertoire in *Nematostella vectensis* are more similar to vertebrates than to model invertebrates like fly and worm.

For *Hydra magnipapillata*, only preliminary genome assemblies are available, which are not yet publicly available. The raw sequencing data are accessible through the NCBI Trace Archive. Both cnidarian genome projects were accompanied by large scale EST sequencing. Whereas the *Nematostella* ESTs are available through the above-described online resources, all *Hydra* sequences were deposited at NCBI dbEST database. Within each particular EST project, several different cDNA libraries derived from several different developmental stages or tissues were generated providing supplementary information.

In addition to the ESTs that have been sequenced accompanying whole genome sequencing projects in selected early branching metazoans, several more anthozoan and hydrozoan species have been subject to smaller EST-based studies<sup>(6,33–35)</sup> so that today we have access to transcript data of more than 10 different cnidarian species. Taken together, the overall situation for early branching metazoan genomics has developed rapidly within the last few years, and new 2<sup>nd</sup> generation sequencing technologies promise more and larger sequence datasets to come in the near future.

To provide a comprehensive working environment for comparative studies using the available sequence data of early branching metazoan animals, we thought that it is necessary to centralize, integrate and pre-analyze these data and make them accessible for the interested researcher. The

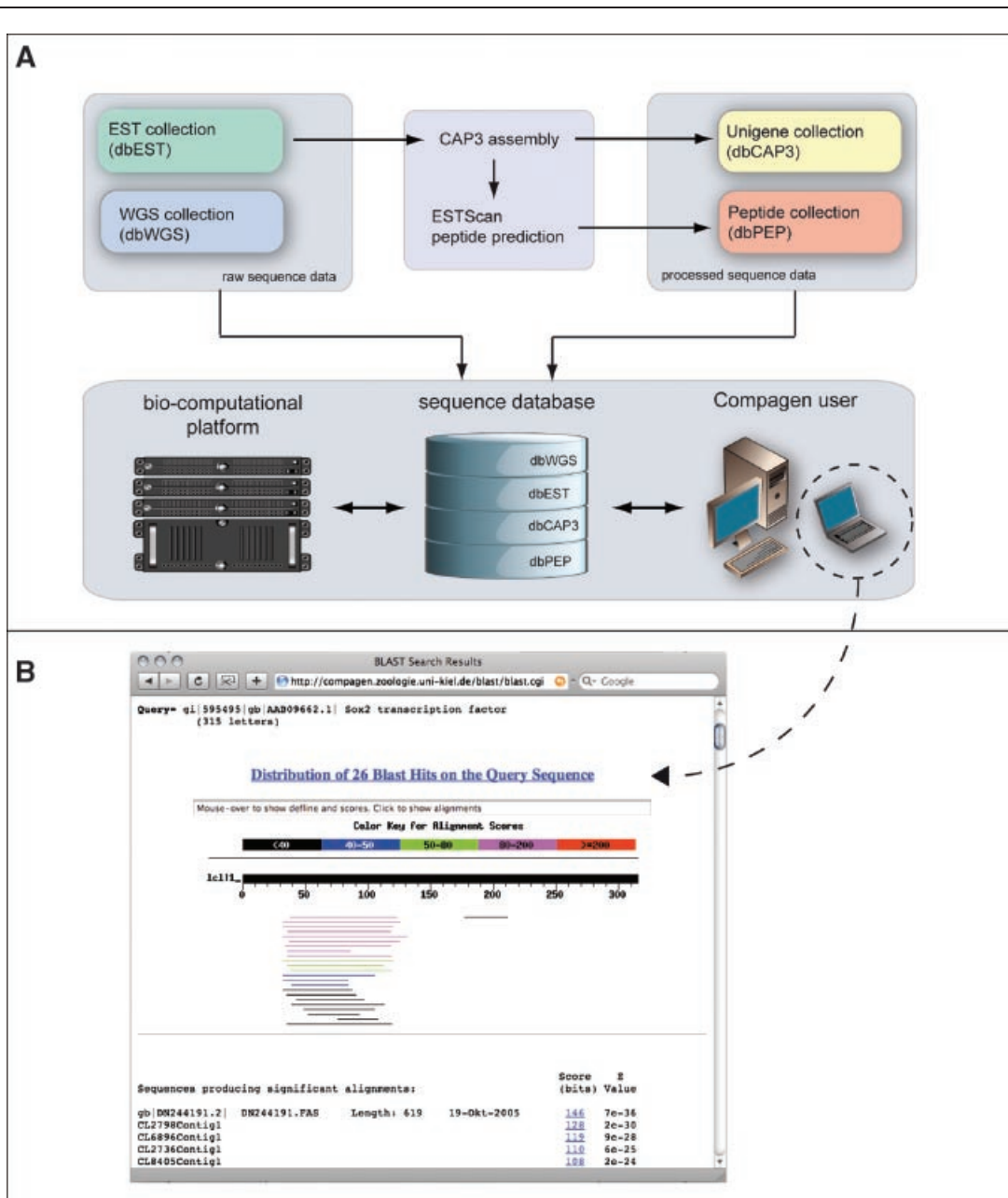
need for such a computational resource for cnidarian model systems and especially for *Hydra* led to the idea to establish a comparative genomics online platform for early branching metazoan animals.

### **Compagen, a bio-computational platform providing easy access to original and processed genomic and transcriptomic data in early branching metazoan animals**

The Compagen bio-computational platform primarily offers an extensive collection of raw and processed genomic as well as transcriptomic sequence datasets (Fig. 2). The idea was to provide centralized and public access to all available sequences derived from various lower metazoan animals and to pre-process redundant sequence data, facilitating analysis for the scientific community. An online Blast-server that can be accessed via the Internet at <http://www.compagen.org> gives the possibility of sequence similarity searches between all available lower metazoan datasets. Allowing a broader comparative perspective and expanding analysis capabilities, additional sequence data from higher metazoan organisms of different taxa as well as from the unicellular choanoflagellates have been included. In addition to the public Blast server, the Compagen-underlying bioinformatics platform enables a variety of local bio-computational methods to be used to further process sequence datasets of different size. Possible computational analyses include different types of sequence assembly for large EST datasets and smaller genomic datasets, EST-based gene and peptide prediction (Fig. 2) as well as general sequence annotation using homology-based gene ontology.

### **Datasets and data organization on Compagen**

At the present time, datasets for 28 different organisms (Table 1) are stored in databases containing about 70 million sequences. As metazoan “outgroup”, genomic and mRNA sequences of two members of the choanoflagellates (*Monosiga brevicollis* and *Monosiga ovata*) have been included into the data collection on Compagen. The raw genomic as well as mRNA sequences of *Trichoplax adhaerens* (Placozoa) serve as most-basal metazoan reference dataset. The Porifera are represented by genomic and mRNA sequences of the marine demosponge *Amphimedon queenslandica*. CAP3 assemblies of both *Trichoplax* and *Amphimedon* EST data are under investigation. Sequences for several anthozoan species, including the raw genomic reads of *Nematostella vectensis*, two *Acropora* species and *Porites lobata*, as well as mRNA sequences for these in addition to *Metridium senile* and *Montastrea faveolata* are available as basal cnidarian datasets. The hydrozoans are represented by EST data for *Hydractinia echinata* and various genomic and mRNA sequence collections for several different *Hydra* species (Table 1). In addition to sequence data derived from early



**Figure 2. A:** Schematic overview of the Compagen bio-computational platform. Raw EST sequence data are processed using the platform-internal assembly pipeline followed by peptide prediction. All raw and processed datasets are stored in the Compagen sequence database and are searchable using the public accessible Blast-server. **B:** Blast-example using human SOX2 transcription factor as query sequence.

branching metazoan animals, a number of reference datasets of higher metazoans, such as molluscs (*Aplysia californica*, *Biomphalaria glabrata*), crustaceans (*Daphnia pulex*, *Daphnia magna*, *Litopenaeus vannamei*, *Penaeus monodon*), planarians (*Dugesia japonica*, *Dugesia ryukyensis*, *Macrostomum lignano*), echinoderms (*Strongylocen-*

*trotus purpuratus*), a cephalochordate (*Branchiostoma floridae*) and a basal chordate (*Petromyzon marinus*) have been included to enable direct comparisons along the evolutionary tree.

All datasets are stored as so-called “flat files” in plain text fasta-format. So far, all datasets have been formatted

**Table 1.** Overview of sequence datasets available on the Compagen-internal Blast server.

Organism	dbWGS # sequences	dbEST # sequences	dbCAP3 # sequences
	Whole Genome Shotgun	Expressed Sequence Tags	CAP3 EST Assemblies
Choanoflagellates:			
<i>Monosiga brevicollis</i>	640.632*	pending	-
<i>Monosiga ovata</i>	N.A.	76.534	12.139
Placozoans:			
<i>Trichoplax adhaerens</i>	1.230.612*	14.571*	pending
Poriferans:			
<i>Amphimedon queensl.</i>	2.823.539*	83.040*	pending
Cnidarians:			
<i>Acropora millepora</i>	14.625	10.247	6.008
<i>Acropora palmata</i>	11.025	4.017	1.421
<i>Hydra magnipapillata</i>	10.272.644	163.221	25.106
<i>Hydra magnipapillata SF-1</i>	N.A.	30.715	8.957
<i>Hydra vulgaris</i>	N.A.	8.993	3.922
<i>Hydra (vulgaris) AEP</i>	N.A.	2.851	267
<i>Hydractinia echinata</i>	N.A.	9.460	—
<i>Metridium senile</i>	N.A.	29.412	8.284
<i>Montastrea faveolata</i>	N.A.	2.156	—
<i>Nematostella vectensis</i>	8.411.866*	166.595*	30.666
<i>Porites lobata</i>	11.450	N.A.	—
Reference datasets:			
<i>Aplysia californica</i>	4.320.600	179.001	
<i>Biomphalaria glabrata</i>	pending	52.624	
<i>Branchiostoma floridae</i>	11.953.628	277.538	
<i>Daphnia pulex</i>	2.724.768	1.548	
<i>Daphnia magna</i>	N.A.	13.134	
<i>Dugesia japonica</i>	N.A.	7.362	
<i>Dugesia ryukyuensis</i>	N.A.	8.988	
<i>Litopenaeus vannamei</i>	N.A.	7.429	
<i>Macrostomum lignano</i>	N.A.	7.617	
<i>Molgula tectiformis</i>	N.A.	106.863	
<i>Penaeus monodon</i>	N.A.	7.330	
<i>Petromyzon marinus</i>	18.808.412	108.847	
<i>Strongylocentrotus purpuratus</i>	7.352.452	141.833	

N.A. = not sequenced; pending = dataset will be intergrated soon; \* these sequence data were produced by and are available through the US Department of Energy (DOE) Joint Genome Institute (<http://www.jgi.doe.gov>). All other sequence data are publicly available in the corresponding database sections at the National Institute of Biotechnology Information (NCBI; <http://www.ncbi.nih.nlm.gov>).

into searchable databases for local and online Blast analysis. The datasets can be divided in (a) raw genomic sequence data (dbWGS), (b) raw EST sequence data (dbEST) and (c) processed EST sequence data (dbCAP3, dbPEP). The “dbWGS” section contains exclusively single whole genome shotgun (WGS) sequencing reads, originating from the corresponding organisms genome-sequencing project. All raw EST sequences are organized in the “dbEST” section. The remaining sections contain CAP3-assembled EST datasets (dbCAP3) and predicted peptides (dbPEP) that enable for conserved domain searches (Fig. 2). To make databases easily distinguishable from each other and easy to work with, a common database-naming convention has been introduced, indicating the type of database, the source of organism and the date of construction (for detailed description see Compagen webpage).

### EST sequence data processing on Compagen

In contrast to the reference datasets that remain in raw format, all early branching metazoan EST datasets have been assembled using the TIGR gene indices clustering tool package.<sup>(36)</sup> Most of the raw sequence data on Compagen originate from public databases. They are obtained using the EBI sequence retrieval system SRS,<sup>(37)</sup> by downloading from NCBI Trace archive or by retrieving directly from the corresponding sequencing center. Using the TGICL *seqclean* program, vector and/or adaptor sequences are clipped and low-quality sequencing reads are removed (Fig. 2). The resulting “cleaned” data are then subjected to the clustering and assembly routine. The purpose of this routine is to cluster and create assemblies (contigs) efficiently from a given set of sequences. During the “clustering phase”, the input dataset is partitioned into smaller groups of sequences (clusters) that

share some similarity in fast MegaBlast<sup>(38)</sup> searches and that potentially come from the same longer original sequence. However, clustering does not produce any multiple alignments but only pairwise alignments. In the “assembly phase”, each cluster is subjected to the CAP3 assembly program,<sup>(39)</sup> which tries to create multiple alignments of the sequences within each cluster. The resulting one or more consensus sequences from the assembly step are then stored as so-called “contig” sequences (or contigs). Sequences that did not fall into clusters or that did not fit in the CAP3-assemblies are afterwards stored as “singletons”. The dbCAP3-datasets used for the Compagen Blast server include the concatenated contig and singleton sequences for each species. The ESTscan program<sup>(40)</sup> is used for the detection of coding regions in assembled EST datasets and for peptide prediction.

### Data availability and future direction

All datasets stored on the Compagen Blast server that contain raw sequence data are available through public domains at NCBI (Trace archive, dbEST) or EBI (ENSEMBL). All processed datasets are available for download directly on the Compagen webpage in the “sequence retrieval” section. New dataset contributions from private investigators are highly welcome and can be integrated into the platform on request. We hope that we can add more data to this resource and extend an invitation to all users to contact us so we can determine if their data can be added.

### Compagen unveils the evolutionary origin of stem-cell-specific genes

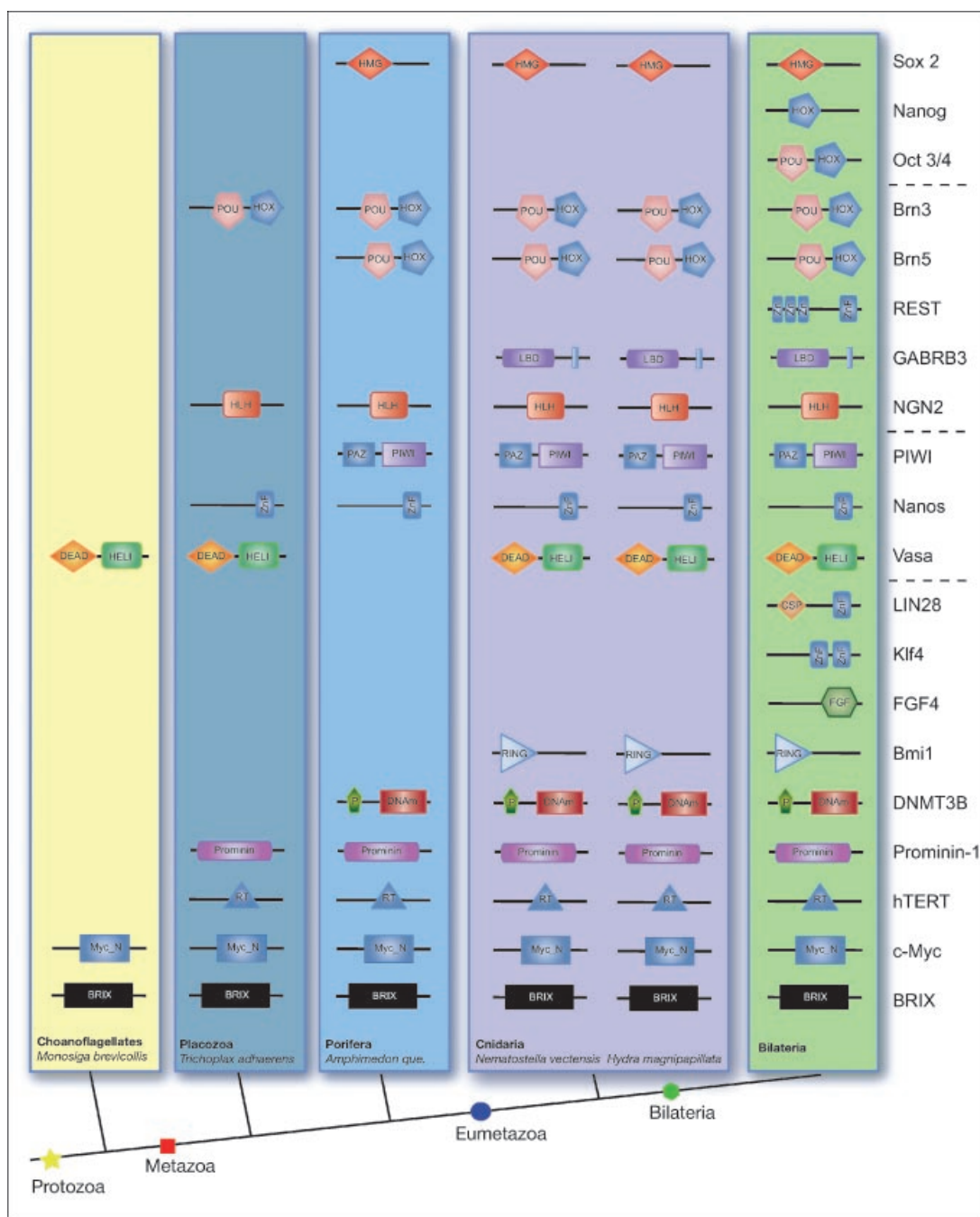
Recent data in a number of vertebrate models suggest that a constellation of intrinsic and extrinsic cellular mechanisms regulates the balance of self-renewal and differentiation in all stem cells. The transcription factors Oct3/4 and Nanog, as well as the LIF-gp130-Stat3, BMP-TGF- $\beta$ -Smad, MAPK-ERK and possibly the Wnt signaling pathways, all have important roles in this process.<sup>(41)</sup> For bone marrow hematopoietic stem cells (HSCs), Bmp signaling is essential; Wnt signaling acts on different developmental stages of the HSC system.<sup>(42)</sup> Thus, an emerging theme is the implementation of the same signaling pathways in distinct stem-cell systems. This is perhaps not surprising given the limited number of such pathways in all of biology. Unfortunately, although all multicellular organisms seem to rely on stem cells and although this seems to be a question of key importance for understanding the evolution of animal life, little is known about stem cells in early-branching taxa.<sup>(43)</sup> Do all stem cells in the animal kingdom follow the same rules? Or are there differences between stem cells in different organisms?

The stem-cell system in the freshwater polyp *Hydra* is arguably the most-tractable cell model to study the mechanisms of regulation of cell differentiation in early branching metazoans, and has certainly provided interesting data-

sets<sup>(44–47)</sup> to evolutionary developmental biologists and cell biologists. Although, however, stem cells in *Hydra* represent one of the most-ancient stem-cell systems in the animal kingdom and, therefore, provide information for reconstructing the early history of stem-cell control mechanisms, the complexity involved in controlling of cell fate choices in stem cells of multiple pathways is poorly understood.

To assess the efficacy of the Compagen platform, we, therefore, conducted a screening in databases of early branching metazoans for genes known to be important in controlling bilaterian stem-cell behavior. We looked for somatic cell reprogramming factors including Sox2, Nanog and Oct3/4.<sup>(48,49)</sup> We also searched for the presence of neuron differentiation factors of the Pou-domain family of proteins such as Brn3 and Brn5<sup>(50,51)</sup> and included other proteins known to be involved in controlling neurogenesis such as RE1 silencing transcription factor (REST),<sup>(52)</sup> gaba receptor (GABRB3)<sup>(53)</sup> and neurogenin 2 (NGN2).<sup>(54)</sup> Furthermore, the Compagen based search included known germ line determinators such as PIWI, Nanos, and vasa.<sup>(55–57)</sup> Other factors reported to be of importance for stem-cell function and, therefore, included in our analysis were microRNA regulator Lin28,<sup>(58)</sup> cell cycle progression control component krueppel-like-factor 4 (Klf4),<sup>(59)</sup> fibroblast growth factor 4 (FGF4),<sup>(60)</sup> epigenetic regulators such as Bmi1 (B lymphoma Mo-MLV insertion region 1 homolog)<sup>(61)</sup> and DNA methyl-transferase DNMT3B.<sup>(62)</sup> Finally, more-general stem-cell markers like Prominin-1,<sup>(63)</sup> telomerase reverse transcriptase hTERT, proto-oncogene c-myc and ribosomal assembly regulator BRIX<sup>(48)</sup> were searched for.

As shown in Fig. 3, screening the Compagen data collection using the platform-internal blast server identified Brn3, nanos, Ngn2, prominin-1, hTERT, c-myc, and BRIX in all early branching metazoan taxa examined including the Placozoa. Interestingly, the dead-box helicase vasa appears to be absent (most likely due to secondary gene loss) in sponges. Fig. 3 also shows that genes for Sox2, Brn5, PIWI and DNMT3B are present in Porifera and Cnidaria but absent in Placozoa and choanoflagellates. GABRB3, Bmi1 can be detected in Cnidaria but not in Porifera and Placozoa. Intriguingly, the classical stem-cell-specific genes encoding Oct3/4 and nanog as well as REST, LIN28, Klf4, FGF4 were not found in any of the basal metazoan taxa examined. Taken together, the preliminary analysis indicates that out of the 20 randomly selected stem-cell marker genes about half can be detected in data bases of early branching metazoans indicating a striking conservation of signaling and transcriptional mechanisms utilized in diverse stem-cell differentiation processes across the animal kingdom. Intriguingly, out of the trinity of nuclear factors (Oct3/4, Nanog and Sox2) governing pluripotency in vivo and in vitro<sup>(46,64)</sup> in vertebrates, only Sox2 can be found in organisms at the base of animal evolution. This implies that not all of the stem-cell regulators



**Figure 3.** Conserved stem-cell marker genes in early branching metazoan animals. Schematic representation of the results obtained by screening early branching metazoan genomes for conserved bilaterian stem-cell marker genes. As non-metazoan outgroup the genome of the choanoflagellate *Monosiga brevicollis* was included into the analysis.

appear to be conserved and that, therefore, unbiased approaches such as stem-cell sorting followed by high-throughput sequencing might discover previously unknown stem-cell regulators.

**Conclusion**

With Compagen, we have established a user-friendly web-based resource that enables researchers to obtain information regarding gene distribution, genomic complexity and gene

expression in early branching metazoans. Here we used this platform to provide the first genome-wide screening for stem-cell differentiation genes in early metazoan animals. The preliminary data indicate that comparative analysis of stem cells in diverse organisms promises new insights into how stem cells act to construct and maintain tissues, and to reveal how the diverse stem-cell systems may have evolved.

## References

- Pires-daSilva A, Sommer R. 2003. The evolution of signalling pathways in animal development. *Nat Rev Genet* 4:39–49.
- Collins AG. 1998. Evaluating multiple alternative hypotheses for the origin of Bilateria: an analysis of 18S rRNA molecular evidence. *Proc Natl Acad Sci USA* 95:15458–15463.
- Philippe H, Lartillot N, Brinkmann H. 2005. Multigene analyses of bilaterian animals corroborate the monophyly of Ecdysozoa, Lophotrochozoa, and Protostomia. *Mol Biol Evol* 22:1246–1253.
- Miller D, Ball E, Technau U. 2005. Cnidarians and ancestral genetic complexity in the animal kingdom. *Trends Genet* 21:536–539.
- Miller DJ, Hemmrich G, Ball EE, Hayward DC, Khalturin K, et al. 2007. The innate immune repertoire in cnidaria—ancestral complexity and stochastic gene loss. *Genome Biol* 8:R59.
- Kortschak R, Samuel G, Saint R, Miller D. 2003. EST analysis of the cnidarian *Acropora millepora* reveals extensive gene loss and rapid sequence divergence in the model invertebrates. *Curr Biol* 13:2190–2195.
- Technau U, Rudd S, Maxwell P, Gordon P, Saina M, et al. 2005. Maintenance of ancestral complexity and non-metazoan genes in two basal cnidarians. *Trends Genet* 21:633–639.
- Kusserow A, Pang K, Sturm C, Hroudá M, Lentfer J, et al. 2005. Unexpected complexity of the Wnt gene family in a sea anemone. *Nature* 433:156–160.
- Larroux C, Luke GN, Koopman P, Rokhsar DS, Shimeld SM, et al. 2008. Genesis and expansion of metazoan transcription factor gene classes. *Mol Biol Evol* 25:980–996.
- Medina M. 2005. Genomes, phylogeny, and evolutionary systems biology. *Proc Natl Acad Sci USA* 102:6630–6635.
- Lang BF, O’Kelly C, Nerad T, Gray MW, Burger G. 2002. The closest unicellular relatives of animals. *Curr Biol* 12:1773–1778.
- King N, Carroll SB. 2001. A receptor tyrosine kinase from choanoflagellates: molecular insights into early animal evolution. *Proc Natl Acad Sci USA* 98:15032–15037.
- King N, Hittinger CT, Carroll SB. 2003. Evolution of key cell signaling and adhesion protein families predates animal origins. *Science* 301:361–363.
- Segawa Y, Suga H, Iwabe N, Oneyama C, Akagi T, et al. 2006. Functional development of Src tyrosine kinases during evolution from a unicellular ancestor to multicellular animals. *Proc Natl Acad Sci USA* 103:12021–12026.
- Snell EA, Brooke NM, Taylor WR, Casane D, Philippe H, et al. 2006. An unusual choanoflagellate protein released by *Hedgehog* autocatalytic processing. *Proc Biol Sci* 273:401–407.
- King N, Westbrook MJ, Young SL, Kuo A, Abedin M, et al. 2008. The genome of the choanoflagellate *Monosiga brevicollis* and the origin of metazoans. *Nature* 451:783–788.
- Dellaporta S, Xu A, Sagasser S, Jakob W, Moreno M, et al. 2006. Mitochondrial genome of *Trichoplax adhaerens* supports placozoa as the basal lower metazoan phylum. *Proc Natl Acad Sci USA* 103:8751–8756.
- Grell KG. 1971. *Trichoplax adhaerens* und die Entstehung der Metazoen. *Naturwiss Rundsch* 160–161.
- Hadrys T, DeSalle R, Sagasser S, Fischer N, Schierwater B. 2005. The *Trichoplax* PaxB gene: a putative Proto-PaxA/B/C gene pre-dating the origin of nerve and sensory cells. *Mol Biol Evol* 22:1569–1578.
- Monteiro A, Schierwater B, Dellaporta S, Holland P. 2006. A low diversity of ANTP class homeobox genes in Placozoa. *Evol Dev* 8:174–182.
- Leys S, Ereskovsky A. 2006. Embryogenesis and larval differentiation in sponges. *Can J Zool* 84:262–287.
- Nichols S, Dirks W, Pearse J, King N. 2006. Early evolution of animal cell signaling and adhesion genes. *Proc Natl Acad Sci USA* 103:12451–12456.
- Simionato E, Ledent V, Richards G, Thomas-Chollier M, Kerner P, et al. 2007. Origin and diversification of the basic helix-loop-helix gene family in metazoans: insights from comparative genomics. *BMC Evol Biol* 7:33.
- Wiens M, Korzhnev M, Perovic-Ottstadt S, Luthringer B, Brandt D, et al. 2007. Toll-like receptors are part of the innate immune defense system of sponges (demospongiae: porifera). *Mol Biol Evol* 24:792–804.
- Hemmrich G, Miller DJ, Bosch TCG. 2007. The evolution of immunity: a low-life perspective. *Trends in Immunol* 28:449–454.
- Ball E, Hayward D, Saint R, Miller D. 2004. A simple plan—cnidarians and the origins of developmental mechanisms. *Nat Rev Genet* 5:567–577.
- Holstein T. 1995. *Cnidaria: Hydrozoa Süßwasserfauna von Mitteleuropa*. 1/2 + 3. Stuttgart, Jena, NY: Gustav Fisher Verlag; 110 p.
- Collins A, Schuchert P, Marques A, Jankowski T, Medina M, et al. 2006. Medusozoan phylogeny and character evolution clarified by new large and small subunit rDNA data and an assessment of the utility of phylogenetic mixture models. *Syst Biol* 55:97–115.
- Dunn CW, Hejnol A, Matus DQ, Pang K, Browne WE, et al. 2008. Broad phylogenomic sampling improves resolution of the animal tree of life. *Nature* 452:745–749.
- Putnam NH, Srivastava M, Hellsten U, Dirks B, Chapman J, et al. 2007. Sea anemone genome reveals ancestral eumetazoan gene repertoire and genomic organization. *Science* 317:86–94.
- Wittlieb J, Khalturin K, Lohmann J, Anton-Erxleben F, Bosch T. 2006. Transgenic *Hydra* allow in vivo tracking of individual stem cells during morphogenesis. *Proc Natl Acad Sci USA* 103:6208–6211.
- Sullivan J, Ryan J, Watson J, Webb J, Mullikin J, et al. 2006. StellaBase: the *Nematostella vectensis* Genomics Database. *Nucleic Acids Res* 34:D495–499.
- Augustin R, Franke A, Khalturin K, Kiko R, Siebert S, et al. 2006. Dickkopf related genes are components of the positional value gradient in *Hydra*. *Dev Biol* 296:62–70.
- Genikhovich G, Kurn U, Hemmrich G, Bosch T. 2006. Discovery of genes expressed in *Hydra* embryogenesis. *Dev Biol* 289:466–481.
- Schwarz JA, Brokstein PB, Voolstra C, Terry AY, Miller DJ, et al. 2008. Coral life history and symbiosis: functional genomic resources for two reef building Caribbean corals, *Acropora palmata* and *Montastraea faveolata*. *BMC Genomics* 9:97.
- Perteza G, Huang X, Liang F, Antonescu V, Sultana R, et al. 2003. TIGR Gene Indices clustering tools (TGICL): a software system for fast clustering of large EST datasets. *Bioinformatics* 19:651–652.
- Kulikova T, Akhtar R, Aldebert P, Althorpe N, Andersson M, et al. 2007. EMBL Nucleotide Sequence Database in 2006. *Nucleic Acids Res* 35: D16–20.
- Zhang Z, Schwartz S, Wagner L, Miller W. 2000. A greedy algorithm for aligning DNA sequences. *J. Comput Biol* 7:203–214.
- Huang X, Madan A. 1999. C. P3:A DNA sequence assembly program. *Genome Res* 9:868–877.
- Iseli C, Jongeneel C, Bucher P. 1999. ESTScan: a program for detecting, evaluating, and reconstructing potential coding regions in EST sequences. *Proc Int Conf Intell Syst Mol Biol* 138–148.
- Moore KA, Lemischka IR. 2006. Stem cells and their niches. *Science* 311:1880–1885.
- Metcalfe D. 2007. Concise review: hematopoietic stem cells and tissue stem cells: current concepts and unanswered questions. *Stem Cells (Dayton, Ohio)* 25:2390–2395.
- Bosch TCG. 2008. *Stem Cells: From Hydra to Man*. Netherlands: Springer.
- Khalturin K, Anton-Erxleben F, Milde S, Plötz C, Wittlieb J, et al. 2007. Transgenic stem cells in *Hydra* reveal an early evolutionary origin for key elements controlling self-renewal and differentiation. *Dev Biol* 309:32–44.
- Siebert S, Anton-Erxleben F, Bosch TCG. 2008. Cell type complexity in the basal metazoan *Hydra* is maintained by both stem cell based mechanisms and transdifferentiation. *Dev Biol* 313:13–24.

46. Bosch TCG. 2007. Symmetry breaking in stem cells of the basal metazoan Hydra. *Prog Mol Subcell Biol* 45:61–78.
47. Bosch TCG. 2007. Why polyps regenerate and we don't: Towards a cellular and molecular framework for Hydra regeneration. *Dev Biol* 303:421–433.
48. Takahashi K, Tanabe K, Ohnuki M, Narito M, Ichisaka T, et al. 2008. Induction of Pluripotent Stem Cells from Adult Human Fibroblasts by Defined Factors. *Cell* 131:861–872.
49. Park I-H, Zhao R, West JA, Yabuuchi A, Huo H, et al. 2008. Reprogramming of human somatic cells to pluripotency with defined factors. *Nature* 451:141–146.
50. Liu W, Khare SL, Liang X, Peters MA, Liu X, et al. 2000. All Brn3 genes can promote retinal ganglion cell differentiation in the chick. *Development* 127:3237–3247.
51. Okamoto K, Wakamiya M, Noji S, Koyama E, Taniguchi S, et al. 1993. A novel class of murine POU gene predominantly expressed in central nervous system. *J. Biol Chem* 268:7449–7457.
52. Ballas N, Grunseich C, Lu DD, Speh JC, Mandel G. 2005. REST and its corepressors mediate plasticity of neuronal gene chromatin throughout neurogenesis. *Cell* 121:645–657.
53. Initiative ISC, Adewumi O, Aflatoonian B, Ahrlund-Richter L, Amit M, et al. 2007. Characterization of human embryonic stem cell lines by the International Stem Cell Initiative. *Nature Biotechnol* 25:803–816.
54. Ma Y-C, Song M-R, Park JP, Henry Ho H-Y, Hu L, et al. 2008. Regulation of motor neuron specification by phosphorylation of neurogenin 2. *Neuron* 58:65–77.
55. Hutvagner G, Simard MJ. 2008. Argonaute proteins: key players in RNA silencing. *Nat Rev Mol Cell Biol* 9:22–32.
56. Suzuki A, Tsuda M, Saga Y. 2007. Functional redundancy among Nanos proteins and a distinct role of Nanos2 during male germ cell development. *Development* 134:77–83.
57. Agata K, Nakajima E, Funayama N, Shibata N, Saito Y, et al. 2006. Two different evolutionary origins of stem cell systems and their molecular basis. *Semin Cell Dev Biol* 17:503–509.
58. Yu J, Vodyanik MA, Smuga-Otto K, Antosiewicz-Bourget J, Frane JL, et al. 2007. Induced pluripotent stem cell lines derived from human somatic cells. *Science* 318:1917–1920.
59. Kim J, Chu J, Shen X, Wang J, Orkin SH. 2008. An extended transcriptional network for pluripotency of embryonic stem cells. *Cell* 132:1049–1061.
60. Mayshar Y, Rom E, Chumakov I, Kronman A, Yayon A, et al. 2008. Fibroblast growth factor 4 and its novel splice isoform have opposing effects on the maintenance of human embryonic stem cell self-renewal. *Stem Cells (Dayton, Ohio)* 26:767–774.
61. Rizo A, Dontje B, Vellenga E, de Haan G, Schuringa JJ. 2008. Long-term maintenance of human hematopoietic stem/progenitor cells by expression of BMI1. *Blood* 111:2621–2630.
62. Okano M, Bell DW, Haber DA, Li E. 1999. DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell* 99:247–257.
63. Kania G, Corbeil D, Fuchs J, Tarasov KV, Blyszczuk P, et al. 2005. Somatic stem cell marker prominin-1/CD133 is expressed in embryonic stem cell-derived progenitors. *Stem Cells (Dayton, Ohio)* 23:791–804.
64. Silva J, Smith A. 2008. Capturing pluripotency. *Cell* 132:532–536.