Comparative genomics and computational biology in the basal metazoan *Hydra*

Dissertation zur Erlangung des Doktorgrades
der Mathematisch-Naturwissenschaftlichen Fakultät
der Christian-Albrechts-Universität zu Kiel

vorgelegt von

Dipl.-Biol. Georg Hemmrich

Kiel, im April 2007

Referent: Prof. Dr. Dr. h.c. Thomas C.G. Bosch

Koreferent: _

Tag der mündlichen Prüfung: _

Zum Druck genehmigt: _

_ _

der Dekan

“There is grandeur in this view of life … from so simple a beginning endless forms most beautiful and most wonderful have been, and are being, evolved.”

From: Darwin, C. R. 1872. *The origin of species by means of natural selection, or the preservation of favoured races in the struggle for life*.

TABLE OF CONTENTS

ABBREVIATIONS:

aa	amino acid
amp	ampicillin
ANTP	Antennapedia
APS	ammonium peroxodisulphate
ATP	adenosine triphosphate
BAC	bacterial artificial chromosome
BI	Bayesian inference
BLAST	Basic local alignment search tool
BMP	Bone morphogenic protein
bp	base pairs
BSA	bovine serum albumin
CAP3	sequence assembly program
cDNA	complementary DNA
CO1	Cytochrome oxidase 1
DABCO	1,4-diazabicyclo[2.2.2]octane
db or DB	database
DEPC	diethylpyrocarbonate
DMF	dimethylformamide
Dkk	Dickkopf
DNA	deoxy ribonucleic acid
DPP	Decapentaplegic
DVD	digital video disk
e.g.	for example
EBI	European Bioinformatics Institute
ECM	extracellular matrix
EDTA	ethylendiamine tetraacetic acid
EED	Embryonic Ectoderm Development
EST	expressed sequence tag

et al.	and others
EtOH	ethanol
E-value	expect value
FISH	fluorescence in situ hybridization
GO	gene ontology
GTR	general time reversal
HMM	hidden markov model
HMW	high molecular weight
HTTP	Hypertext transfer protocol
i-cell	interstitial cell
JGI	Joint Genome Institute
kbp	kilo base pairs
LSU	large subunit
MAC/PF	membrane attack / perforin
Mbp	mega base pairs
ML	maximum likelihood
MP	maximum parsimony
mtDNA	mitochondrial DNA
NCBI	National Center of Biotechnology Information
NJ	Neighbor Joining
PBS	phosphate buffered saline solution
PCD	programmed cell death
PcG	Polycomb group
PCR	polymerase chain reaction
PFAM	protein domain database
PMSF	phenylmethanesulphonylfluoride
RNA	ribonucleic acid
RNAi	RNA interference
rRNA	ribosomal RNA
rpm	rounds per minute
RT	room temperature

RTK	receptor tyrosine kinase
SMART	simple modular architecture research tool
SNP	single nucleotide polymorphism
SRS	sequence retrieval system
SSC	sodium chloride / sodium citrate
SSH	supression subtractive hybridization
SSU	small subunit
TAE	tris-acetic EDTA buffer
Taq	*Thermophilus aquaticus*
TBE	tris-boric EDTA buffer
TE	Tris EDTA buffer
TEMED	N,N,N´,N´- tetraethylenediamine
TFBS	transcriptionfactor binding site
TGF	transforming growth factor
TIGR	The Institute of Genomic Research
TIR	Toll / Interleukine receptor domain
TLR	Toll-like receptor
TPA	12-O-tetradecanoylphorbol-13-acetate
UTR	untranslated region
V	Volt
vol.	volume
vs.	versus
WGS	whole genome shotgun
www	world wide web

1 INTRODUCTION

1.1 Evolutionary Developmental Biology – Evo-Devo

"Evo-Devo comprises all that is contained in the black box that lies between genotype and phenotype" (Hall, 2003). Evolutionary developmental biology (Evo-Devo) is a broad scientific field of biology that i) compares developmental processes between organisms to determine their ancestral relationships and ii) tries to understand the evolution of developmental mechanisms, how they account for the generation of novel features and species diversity. Evo-Devo has its origins in evolutionary morphology of the late 19th century when Charles Darwin postulated his concept of evolution to be the result of "descent with modification" through selection in his book "On the Origin of species" (Darwin, 1895). Scientists at that time looked to embryonic and larval stages searching for homologies that would be obscured in the adult. Today, current research investigates the evolution of regulation of the "genetic toolkit" together with gene duplication and gene diversification. And, indeed, finding a very conserved set of regulatory genes playing comparable developmental roles in nearly all organisms (Carroll et al., 2001; Wilkins, 2002) represents a powerful molecular proof for Darwin's concept. However, there remains a big problem: If all developmental genes are the same, how are differences in development and morphology of different organisms accomplished? One possible explanation is that differences between organisms are due to differences in expression of regulatory genes driven by upstream regulators or by changes in the range of downstream target genes (Bosch and Khalturin, 2002; Rudel and Sommer, 2003). But this may not be the only answer. New data from the increasing number of complete genome sequences indicate a substantial number of novel unknown genes (Pires-daSilva and Sommer, 2003) and the obvious question what do these genes code for, is a new and till yet mostly unappreciated issue in current evolutionary developmental research (Bosch and Khalturin, 2002).

1.2 Classical model systems and their limits in current evo-devo research

Our understanding of developmental mechanisms and the evolution of metazoan genomes is mainly based on research in a few complex animals and a small number of complete genome sequences. Recent findings on the evolutionary origin of developmental genes in vertebrates are mostly examples resulting from studies including the insects *Drosophila melanogaster* and *Anopheles gambiae*, the nematode *Caenorhabditis elegans* and the two yeast species *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe* as “basal” invertebrates. Today, we consider these organisms as well established model systems with powerful molecular genetic tools, fully annotated genome sequences, large EST projects and the availability of functional tests. Within the past years extensive research using these model systems led to key findings in evolutionary and developmental biology. Deciphering the genetic control mechanisms of patterning during embryonic development in *Drosophila melanogaster* 1995 (Lewis, 1978; McGinnis and Kuziora, 1994; Nüsslein-Volhard and Wieschaus, 1980), the discovery of key regulators of the cell cycle in *Saccharomyces cerevisiae* 2001 (Hartwell, 2002; Hunt, 2002; Paulovich and Hartwell, 1995; Royer, 2001), as well as unraveling the genetic regulation of organ development and programmed cell death (apoptosis) in *Caenorhabditis elegans* 2002 (Brenner, 2003; Check, 2002; Hoffenberg, 2003) and the discovery of gene silencing by double stranded RNA (RNA interference or RNAi) 2006 (Caplen et al., 2001; Fire et al., 1998; Grishok et al., 2001), were awarded by Nobel prices. Studying the evolution of genes taking part in these processes identified significant examples of conservation of developmental programs, especially between *Drosophila* and vertebrates (Jaruzelska et al., 2003; Sun et al., 2003; Zdobnov et al., 2002).

However, there are also limits of conservation as for example a large number of genes from *D. melanogaster*, *C. elegans* and *S. cerevisiae* appear to be highly derived when compared to vertebrate genomes (1998; Adams et al., 2000; Goffeau et al., 1996). In addition, a significant number of genes have been identified during comparative studies that seem to be present only in one organism and not in others (Gibson, 2001; Hutter et al., 2000; Parkinson et al., 2004; Sommer, 1997; Wood et al., 2002) and thus are referred to as “taxon-specific” genes.

Moreover, finding a significant proportion of mammalian genes being absent within invertebrate genomes initially led to the wrong assumption that these genes are vertebrate innovations instead of gene losses in the invertebrates. With the increasing availability of genome sequences and expressed sequence tags (ESTs) for other basal invertebrates like poriferans or cnidarians, this assumption has repeatedly been shown to be incorrect (Ball et al., 2004; Kortschak et al., 2003; Miller et al., 2005), and the importance of comparative data from model systems other than *Drosophila* and *C. elegans* has been underlined.

1.3 "Non-model" systems in evolutionary and developmental biology

While classical model systems benefit from being completely molecularized, animals representing the so-called "non-models" often suffer from lacking essential functional tests or sequence data from EST or genome projects. But for a number of evolutionary old organisms this situation changed drastically within the last five years. More and more genomes become sequenced mostly accompanied by extensive EST sequencing (see Table 1) and research groups try to apply modern molecular genetic tools and establish functional tests for the animal they are working on.

Organism	**Taxonomy**	**Common name**	**Genome project** (coverage)	**EST project** (# seqs)
Trichoplax adhaerens	Placozoa	-	in process	in process
Reniera sp.	Porifera	sponge	in assembly (n.a.)	83.000
Nematostella vectensis	Cnidaria	sea anemone	completed (7.5 x)	166.000
Hydra magnipapillata	Cnidaria	fresh water polyp	in annotation (6 x)	174.000

Table1: Current status of genome and EST projects for selected "non-model" organisms

Trichoplax adhaerens, a member of the Placozoa (see Figure 1) represents one of the most basal metazoan taxa of the animal kingdom (Dellaporta et al., 2006). Exhibiting an extremely simple body plan generated by only four somatic cell types (Grell, 1971) together with its taxonomic position, it is used as a model to study the transition from unicellular to multicellular animals. Within the last two years the first evolutionary conserved developmental genes were identified (Hadrys et al., 2005; Monteiro et al.,

2006) pointing towards a reduced complexity in gene families compared to higher metazoans. Sequencing of the *Trichoplax* mitochondrial genome (Dellaporta et al., 2006) consolidated the taxonomic position in the tree of life. The ongoing genome and EST project will massively aid in the identification of conserved and novel genes involved in forming such a simplistic animal and will provide new insights in the genome evolution of the metazoa.

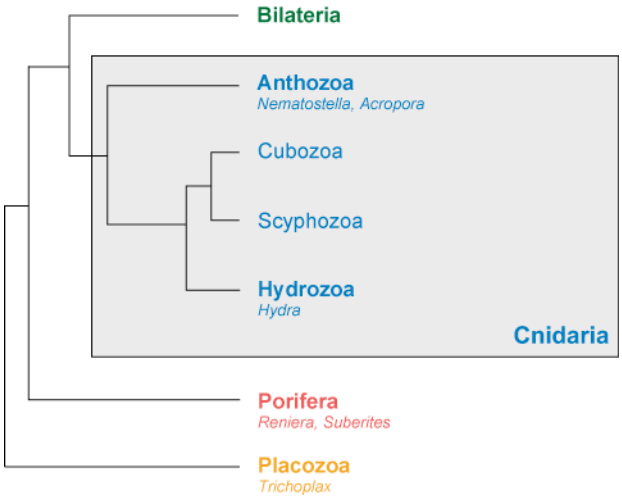


Figure 1: Schematic evolutionary tree of the lower metazoa.
Modified from Miller & Hemmrich et al., 2007.

The oldest animals in the sister group of the Placozoa are the Porifera (sponges; see also Figure 1). Current research projects mostly investigate the demosponges *Suberites domuncula* and *Reniera sp.* 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 (Leys and Ereskovsky, 2006). Although established molecular techniques are still limited, some first evidence about the presence of conserved genes involved in developmental processes (Nichols et al., 2006; Simionato et al., 2007) or components of an ancestral innate immune system (Wiens et al., 2007) already implicates the importance of sponges for comparative studies as they highlight ancestry and/or secondary gene loss. To unravel the complete ancestral metazoan gene set, the active genome-sequencing project and extensive EST data are crucial for further investigations.

1.4 Cnidarians, arising model systems at the base of Bilateria

Cnidarians provide several important features to be subject of active evo-devo research. As sister-group to all bilaterian animals (see Figure 1) they 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 (Ball et al., 2004). 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 co called cnidocyte or nematocyte (stinging cell), which is used to catch prey or to defend predators (Holstein, 1995). Cnidarians can be roughly divided into the most basal Anthozoa (corals & anemones) and the Medusozoa (see Figure 2), consisting of the Cubozoa (sea wasps), the Scyphozoa (jellyfishes) and the Hydrozoa (hydroids) (Collins et al., 2006). Because some of the medusozoans exhibit nearly radially symmetric body plans, cnidarians are often grouped together with the non-cnidarian ctenophores (comb jellyfishes) as the Radiata with two body layers and one axis in contrast to the Bilateria with three body layers and two axes (Martindale et al., 2002).

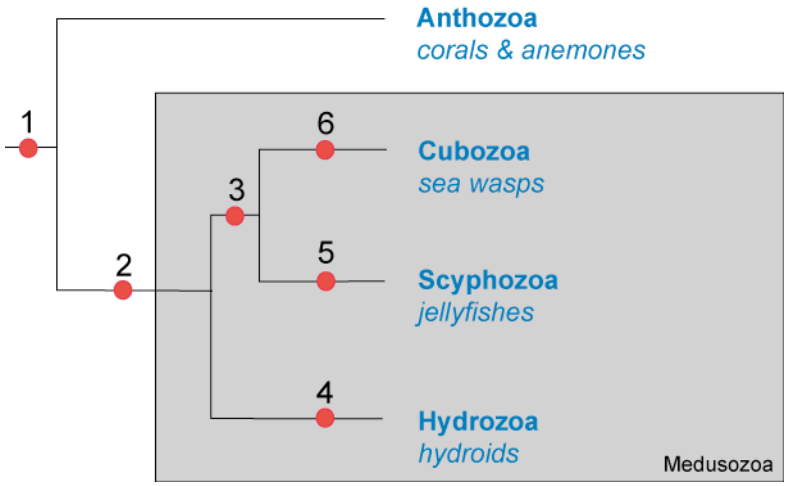


Figure 2: Cnidarian relationships with selected ancestral characters. *1) nematocytes, planula larvae, solitary polyp; 2) pelagic medusa; 3) Rhopalia; 4) medusa produced through lateral budding of the entocodon, epidermal gonads; 5) polydisk strobilation, ephyrae; 6) complex eyes. Modified from Collins et al., Syst. Biol. 2006.*

1.5 State of the art in cnidarian evo-devo research

In current research, cnidarians are often used as model to study various aspects of evolution and development such as patterning, regeneration, embryogenesis, apoptosis or the evolution of metazoan genomes. The most widely used cnidarian organisms are the marine anthozoan *Nematostella vectensis* and the fresh-water hydrozoan *Hydra magnipapillata,* whereas anthozoans are thought to represent the basal and hydrozoans the derived state within this phylum (Figure 2).

1.5.1 Developmental genes in Cnidaria

Since many years Hox genes are studied as key players in patterning processes. The evolutionary origin of cnidarian Hox genes is controversially discussed and remains arguable. Recent publications discuss, whether the bilaterian Hox code was present before the cnidarian/bilaterian split or not (Chourrout et al., 2006; Kamm et al., 2006; Ryan et al., 2007). Kamm et al. in 2006 proposed the term Hox-like genes because it was not clear if the cnidarian Hox genes were paralogs or homologs compared to the bilaterian complement (Kamm et al., 2006). Later studies document the presence of a simple proto-Hox cluster in the anthozoan *Nematostella vectensis* (Chourrout et al., 2006) and in a recent publication Ryan et al. suggest that the bilaterian Hox code was already present before the divergence of Cnidaria (Ryan et al., 2007). Other research projects investigated the possible function of Hox genes in cnidarian development and could show in various cnidarians that indeed there is a correlation between patterning processes along the forming body axis (de Jong et al., 2006; Finnerty et al., 2004; Fröbius et al., 2003). Further studies demonstrated the role of Pax genes in cnidarian ectodermal nerve net development (Matus et al., 2007) and identified the full complement of classical non-Hox ANTP-superclass transcription factors in *Nematostella* (Kamm and Schierwater, 2006).

Another crucial problem in patterning processes is to understand the formation of body axes, positional gradients and the determination of cell fate (Guder et al., 2006a; Lee et al., 2006; Meinhardt, 2006). During the past years extensive research in this field tried to shed some light on these issues. From previous studies it was known that components of the Wnt pathway are involved in axis formation (Hobmayer et al., 2000; Minobe et al., 2000) and the formation of the head organizer in *Hydra magnipapillata*

(Broun et al., 2005). Recent publications showed that an ancient Wnt/Dickkopf antagonism is present in *Hydra* (Guder et al., 2006b) and that Dickkopf genes are also components of the positional value gradient (Augustin et al., 2006). Searching for Wnt genes in *Nematostella* resulted in unexpected Wnt gene complexity and expression data indicate their role in gastrulation and axis formation (Kusserow et al., 2005). Investigating Wnt genes in the marine hydrozoan *Hydractinia echinata* revealed a tight connection of Wnt expression during embryonic development and metamorphosis (Plickert et al., 2006). Besides Wnt, also other genes have been shown to play a role in patterning processes in cnidarians. In a recent study, components of the Notch pathway were found directly linked to nerve cell differentiation in *Hydra* (Käsbauer et al., 2007). Focusing on the TGFbeta signaling cascade, the role of bone morphogenic proteins (BMPs) and their antagonists were investigated in *Hydra* (Hobmayer et al., 2001; Reinhardt et al., 2004), *Nematostella* (Rentzsch et al., 2006; Rentzsch et al., 2007) and the marine hydrozoan *Podocoryne carnea* (Reber-Muller et al., 2006). Signalling via receptor thyrosine kinases (RTKs) is yet mainly studied in *Hydra* and a variety of molecules participating in RTK-related pathways have been identified(Arvizu et al., 2006; Bridge et al., 2000; Cardenas and Salgado, 2003; Reidling et al., 2000; Steele, 2002; Sudhop et al., 2004). Further attempts to identify genes involved in patterning in *Hydra* resulted in the isolation of a secreted peptide governing tentacle formation in *Hydra* (Broun et al., 2005) .

1.5.2 Embryogenesis in Cnidaria

While most scientists study genes and mechanisms tightly linked to patterning processes during embryonic development, some other research projects investigate embryogenesis *per se*. An ultra-structural study of embryogenesis in the sea anemone *Nematostella* identified the cellular mechanisms underlying gastrulation (Kraus and Technau, 2006). In another approach the role of programmed cell death during development of the anemone embryo was investigated (Technau et al., 2003). In *Hydra* not all embryonic stages are easy to access (Alexandrova et al., 2005; Martin et al., 1997) but nevertheless, attempts to isolate genes related to embryogenesis identified interesting candidates that broaden our understanding of the mechanisms during this process (Fröbius et al., 2003; Genikhovich et al., 2006).

1.5.3 Cnidaria as model for regeneration

A remarkable feature of some cnidarians is their capability to regenerate missing body parts. This phenomenon is predominantly present in *Hydra* and subject of research for over 200 years. Scientific findings related to regeneration in *Hydra* have recently been reviewed in Bosch, 2007b. As shown in several independent experiments, regeneration in *Hydra* occurs by morphallaxis (Cummings, 1984; Holstein et al., 1991), a process first described by Thomas Hunt Morgan in 1901 (Morgan, 1901). Various approaches identified genes involved in regenerating the head (Augustin et al., 2006; Kaloulis et al., 2004; Manuel et al., 2006). Recently it has been proposed that also the nervous system in *Hydra* seems to play a role during regeneration (Miljkovic-Licina et al., 2007).

1.5.4 Programmed cell death in Cnidaria

Since the presence of programmed cell death (PCD) or apoptosis in cnidarians has first been reported for the hydrozoan *Hydra* in 1999 (Cikala et al., 1999), this process is under permanent investigation in a variety of cnidarian organisms. Conserved components of caspase signaling could be identified in the anthozoan *Aiptasia pallida* (Dunn et al., 2006) and in the marine hydrozoan *Hydractinia echinata* (Seipp et al., 2001) where metamorphosis was shown to be dependent on caspase signaling (Seipp et al., 2006). Further investigations in apoptosis in *Hydra* revealed a role of PCD during spermatogenesis (Kuznetsov et al., 2002), in regulating cell numbers and during regeneration (Böttger and Alexandrova, 2007).

1.5.5 Genomics and transcriptomics in Cnidaria

As cnidarian research finally entered the age of genomics and transcriptomics a few years ago, scientists also start to investigate subjects such as the evolution of genes and genomes, the appearance of taxon-specific genes, or the evolution of the immune system. In addition large-scale gene expression profiling approaches using ESTs became a valuable tool to isolate genes involved in a certain cell type, a tissue or developmental stage. First implications about ancestral gene structure and cross kingdom conservation came from corals and jellyfish when characterizing the genes encoding for integrins and ion channels as well as components of the DPP/BMP pathway on the genomic level (Hayward et al., 2002; Samuel et al., 2001; Schmitt and Brower, 2001; Spafford et al., 1999). Elaborate EST analysis in the coral *Acropora*

millepora revealed extensive gene loss and a high degree of sequence divergence within the classical models *Drosophila* and *C. elegans* (Kortschak et al., 2003). First reports about ancestral genetic complexity of gene families came from research projects investigating the role of Wnt genes during development (Kusserow et al., 2005; Miller et al., 2005). In another approach a significant number of non-metazoan genes were identified within EST collections for *Acropora* and *Nematostella* (Technau et al., 2005). Recent publications on a variety of subjects focus on the evolution of microRNAs in the bilaterian ancestor (Prochnik et al., 2007) and the presence of clustered developmental genes within cnidarian genomes (Sullivan et al., 2007b). And since genome sequences are available, scientists start to screen for complete gene sets involved in processes like gene regulation (Simionato et al., 2007) or immune response (Miller et al., 2007; Sullivan et al., 2007a).

1.5.6 Molecular resources for model cnidarians

For both *Nematostella* and *Hydra*, extensive molecular resources have been established within the last three years. 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 online platforms (Sullivan et al., 2006). For *Hydra magnipapillata* only preliminary genome assemblies are available which are not yet publicly available. Both genome projects were accompanied by large scale EST sequencing. Whereas a large proportion of the *Nematostella* ESTs are not yet publicly available, all *Hydra* sequences were deposited at NCBI dbEST and are open for analysis. Within each particular EST project, several different cDNA libraries derived from several different developmental stages or tissues were generated providing additional valuable information (see also chapter 2.4). In addition to genomic and transcriptomic data several attempts to construct large insert bacterial artificial chromosome (BAC) libraries resulted in an 8 x coverage *Nematostella* BAC library and a low coverage (3.5 x) library for *Hydra* (Hemmrich and Bosch, unpublished). In addition to these molecular resources several powerful molecular genetic tools have been developed. Gene silencing using RNAi via in vivo electroporation was established for *Hydra* in 1999 and recent publications demonstrate double-stranded RNA feeding experiments resulting in transient gene knock-down

(Chera et al., 2006; Lohmann et al., 1999). In *Nematostella* first gene silencing effects could be shown using RNA morpholinos, but the method is still under investigation (Technau, pers. communication). Finally, the possibility of generating transgenic *Hydra* via microinjection of embryos (Wittlieb et al., 2006) completed the catalogue of methods required for modern functional analysis of genes.

1.6 The cnidarian model system *Hydra*

The fresh-water polyp *Hydra* has a long history as model system in classic developmental biology because of the remarkable plasticity in its differentiation capacity and its ability to regenerate missing body parts (Bode, 2003; Bosch, 2007b; Galliot et al., 2006; Holstein et al., 2003). *Hydra´s* regeneration capacity and the underlying mechanisms, responsible for specification of positional information, present excellent opportunities for understanding how gradients of morphogens could be generated and maintained to control local developmental processes (Meinhardt and Gierer, 2000; Wolpert, 1973; Wolpert et al., 1972).

1.6.1 Systematics of *Hydra*

Within the Cnidaria, *Hydra* belongs to the Hydrozoa (see also Figure 2). Because of the high morphological diversity, the variety of different sensory organs and the complexity in their cnidocytes, hydrozoans are thought to represent the most derived class within the Cnidaria (Collins, 2002; Steele, 2002). The systematics of hydrozoan subtaxa is still far from being complete but the combination of morphological and molecular data help to increasingly clarify the situation (Marques and Collins, 2004). For *Hydra* a vague number of 30 species have been described (Anokhin, 2004) but there is neither clear evidence on the exact number of species nor is it clarified whether *Hydra* is one genus or should be split into several genera. Previous attempts to group different *Hydra* species were all based on general morphological differences in the body plan, different modes of tentacle formation and differences in specific types of cnidocytes (Campbell, 1987; Holstein, 1995) but so far no molecular data was included. Generation of a first comprehensive molecular phylogeny of selected *Hydra* species is part of this thesis (see chapter 2.1) and was published recently in *Molecular Phylogenetics and Evolution* (Hemmrich et al., 2006).

1.6.2 Morphology, histology and life cycle of *Hydra*

Unlike most members of the Hydrozoa, which are typically marine, colonial animals, *Hydra* are solitary living freshwater polyps. Their body plan is organized around a single body axis that can be subdivided in head, body column and foot (see Figure 3A). The head comprises a ring of 4-7 tentacles that are organized around the mouth opening (hypostome) and the foot has a so-called basal disk that is used to attach the polyp to the substrate.

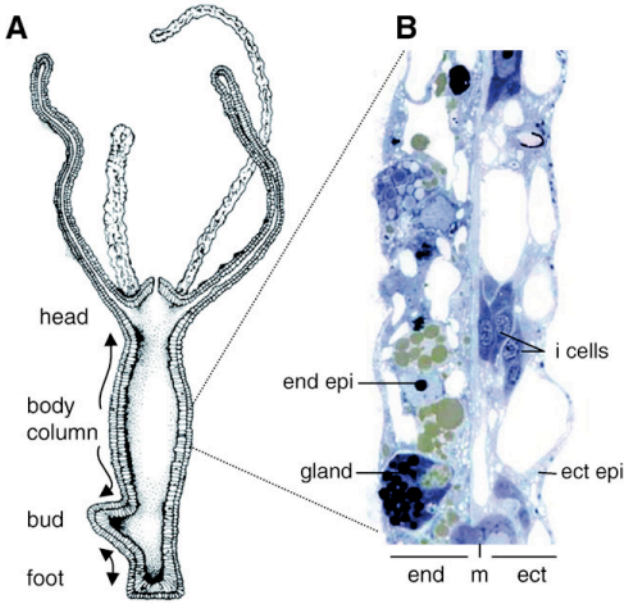


Figure 3: The freshwater polyp Hydra. *(A) Schematic longitudinal cross section indicating the simple epithelial organization. Arrows indicate the direction of tissue displacement. (B) Photograph of a section of part of the epithelial lining of the body column, showing the diploblastic organization. Note how interstitial cells and gland cells are interspersed between ectodermal and endodermal epithelial cells, respectively. End, endoderm; ect, ectoderm; m, mesoglea; Figure taken from Bosch, 2007.*

Hydra is made up of two tissue layers, the ectoderm and the endoderm (Figure 3B). The two layers are separated by a thin extracellular matrix (ECM), the mesoglea. The cellular system of *Hydra* can be divided in three independent cell lineages, the ectormal and the endodermal epithelial cell lineage as well as the intestitial cell lineage. The epithelial cells are epitheliomuscular cells that build the two tissue layers, whereas interstitial stem cells, mainly localized in the interstitial space between ectodermal epithelial cells, give rise to nerve cells, cnidocytes, gland cells and gametes (Bosch, 2007a; Bosch and David, 1986).

In *Hydra,* reproduction is mostly accomplished by clonal propagation in a process called budding during which a new polyp is built from the body column of the adult polyp.

Depending on several different environmental factors like population density, availability of food or water temperature, also gametes are produced from the interstitial cell lineage (Figure 4) (Martin et al., 1997). During oogenesis, interstitial cells proliferate and form a cluster of cells that are connected by cytoplasmic bridges. One of the cells within the cluster gets determined to become the egg cell. All other cells of the cluster are phagocytosed and incorporated into the cytoplasm of the developing oocyte. After external fertilization, the embryo develops directly into an adult polyp without a larval stage in-between. Embryogenesis is finished when a completely developed polyp hatches from the egg. In contrast to most other Hydrozoa, *Hydra,* lacking the medusa stage, has no metagenetic life cycle (Figure 4).

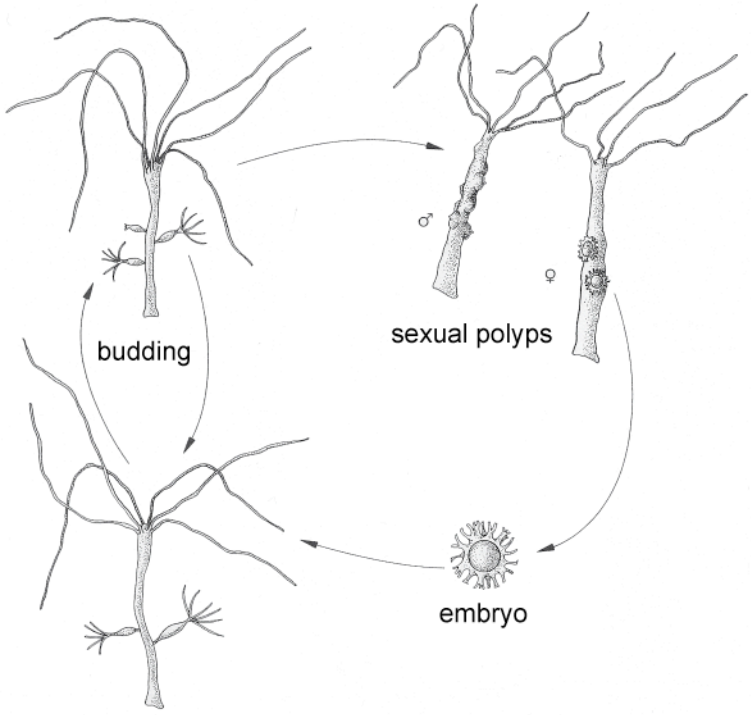


Figure 4: Life cycle of Hydra. *The two ways of propagation in Hydra include budding as asexual mode and the development of sperm and eggs as sexual mode. Embryos develop attached to the female polyp until they are released and completely developed polyps hatch. Figure modified from Westheide & Rieger, 1995.*

1.6.3 Developmental processes in *Hydra*

Whereupon in most other animals patterning processes are restricted to embryonic development, different axial patterning processes are constantly active in the adult *Hydra* and during regeneration or budding even *de novo* patterning processes can be studied (Bode, 2003; Bosch, 2003; Bosch and Fujisawa, 2001; Broun and Bode, 2002; Steele, 2002). The epithelial tissues of *Hydra* are in constant homoeostasis of cell proliferation and cell loss (see Figure 5). Cells are permanently shifted towards the forming buds and the extremities where in the ends of the tentacles, in the hypostome tip and in the basal disk, cells are released into the medium from the ectoderm or released into the gastric cavity from the endoderm.

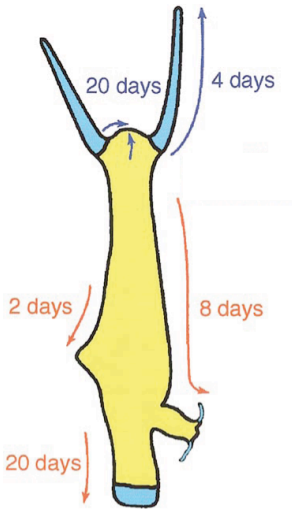


***Figure 5: Tissue dynamics of the adult Hydra**. Arrows indicate directions in which cell are displaced after a certain time. The yellow coloured area indicates the region of epithelial cell proliferation. The blue coloured areas indicate parts of the body where the cells have been transdifferentiated and do not divide any more. Figure modified from Steele, 2002.*

Numerous transplantation and tissue manipulation experiments in the past provided experimental data for the generation of theoretical models, describing patterning processes in *Hydra* (Berking, 2003; Meinhardt, 1993; Meinhardt, 2006; Meinhardt and Gierer, 2000). These models propose, that the morphology of the polyp along the body axis is maintained by a morphogenetic gradient that is maximal in the head and decreases towards the foot (Gierer and Meinhardt, 1972; MacWilliams, 1983; Muller, 1996; Wolpert, 1971; Wolpert et al., 1972). The gradient has been proposed being established by a local autocatalytic activator that produces a long-ranging inhibitor. This inhibitor in turn antagonizes the self-activation (Meinhardt, 2004). But how this gradient is established during development and which molecules account for the gradient is yet to be determined.

1.7 Towards molecularization of *Hydra*: aims of the study

Over the past few years an impressive accumulation of gene sequences, novel tools and genomic resources has brought a new perspective on research in *Hydra* (Bosch, 2007b; Galliot et al., 2006; Holstein et al., 2003). In addition to the already mentioned powerful analytical tools like RNAi and transgenic animals several more valuable methods have been developed. Approaches including suppression subtractive hybridization (SSH) (Augustin et al., 2006; Genikhovich et al., 2006) and phylogenetic footprinting procedures have been established (Siebert et al., 2005) and the genome sizes and corresponding karyotypes of five *Hydra* species have been determined (Zacharias et al., 2004). The huge amount of available genomic and transcriptomic sequence data from various *Hydra* species complement these tools.

But to talk about a completely molecularized model organism, some important features and resources are still missing: i) till yet no molecular phylogeny of genus *Hydra* has been published, ii) most of the EST and genomic sequence data available for *Hydra* are raw data, that require processing and annotation, iii) despite being submitted to NCBI neither for ESTs nor for the genome sequence comprehensive online analytical platforms exist, iv) so far the current organization of molecular data for *Hydra* do not allow application of modern computational biology methods (e.g. conserved domain searches, HMMs, peptide prediction etc.).

To complement the available molecular resources and tools for current research in *Hydra*, a comprehensive molecular phylogeny for selected members of the genus *Hydra* was established. In a separate project, a bioinformatics analytical platform for comparative genetics and genomics in Cnidaria and for high-throughput processing of EST and genomic data was established and used in several approaches. These included an extensive screening for cnidarian genes related to immunity, a large-scale gene expression analysis approach using *Hydra* EST data, and the genomic characterization of a novel, taxon-specific gene family in *Hydra magnipapillata*.

2 RESULTS

2.1 Establishing a molecular phylogeny for selected species of the genus *Hydra*

The phylogeny of the genus *Hydra* for long time has been a controversially discussed and unresolved issue. In all previous attempts, to resolve the phylogenetic relationships of the approximately 30 extant *Hydra* species, only morphological differences were taken into account (Campbell, 1983; Holstein, 1995) whereas molecular data were not included. The lack of such data for *Hydra* led to determine the phylogenetic affinities of the eight most commonly used species and laboratory strains of this genus on the molecular level. Two nuclear (18S rDNA SSU; 28S rDNA LSU) and two mitochondrial (16S rRNA; cytochrome oxidase I, COI) markers were cloned and analyzed by maximum parsimony (MP), maximum likelihood (ML), and Bayesian inference (BI) methods to reconstruct the evolutionary history of these eight species (Hemmrich et al., 2006).

2.1.1 Phylogenetic inference using mitochondrial genes

For the mitochondrial DNA, the data sets included 401 base pairs (bp) of the mitochondrial (mtDNA) 16S ribosomal RNA (rRNA) gene as well as 573 bp of the cytochrome oxidase I (COI) gene. mtDNA sequences of the marine hydrozoans *Obelia geniculata* and the anthozoan *Nematostella vectensis* available on GenBank were included as outgroup. As shown in Figure 6A+B, both single-gene maximum likelihood analyses recovered *Hydra viridissima* as the most basal group. *Hydra circumcincta* and the two members of the “oligactis” group (*Hydra oligactis* and *Hydra robusta*) invariably resolved as the sister groups to the other four *Hydra* species examined. Unexpectedly, all analyses of both mitochondrial genes strongly suggest that *Hydra vulgaris* (strain AEP) is most closely related to *Hydra carnea* and not to *Hydra vulgaris* (srain Basel) or *Hydra magnipapillata*. There were no conflicts between the MP, ML and BI analyses since results from the MP and BI analysis support all of the affinities recovered in the ML analysis (see Appendix, Figures 1+2). Analyses on the combined data sets of both mtDNA genes were also performed. Figure 6C shows that as with the individual gene analyses, *Hydra viridissima* is strongly supported as basal species and *Hydra*

circumcincta and *Hydra oligactis* are the sister taxons to the “vulgaris” group. *Hydra vulgaris* (strain AEP) and *Hydra carnea* form a monophyletic group.

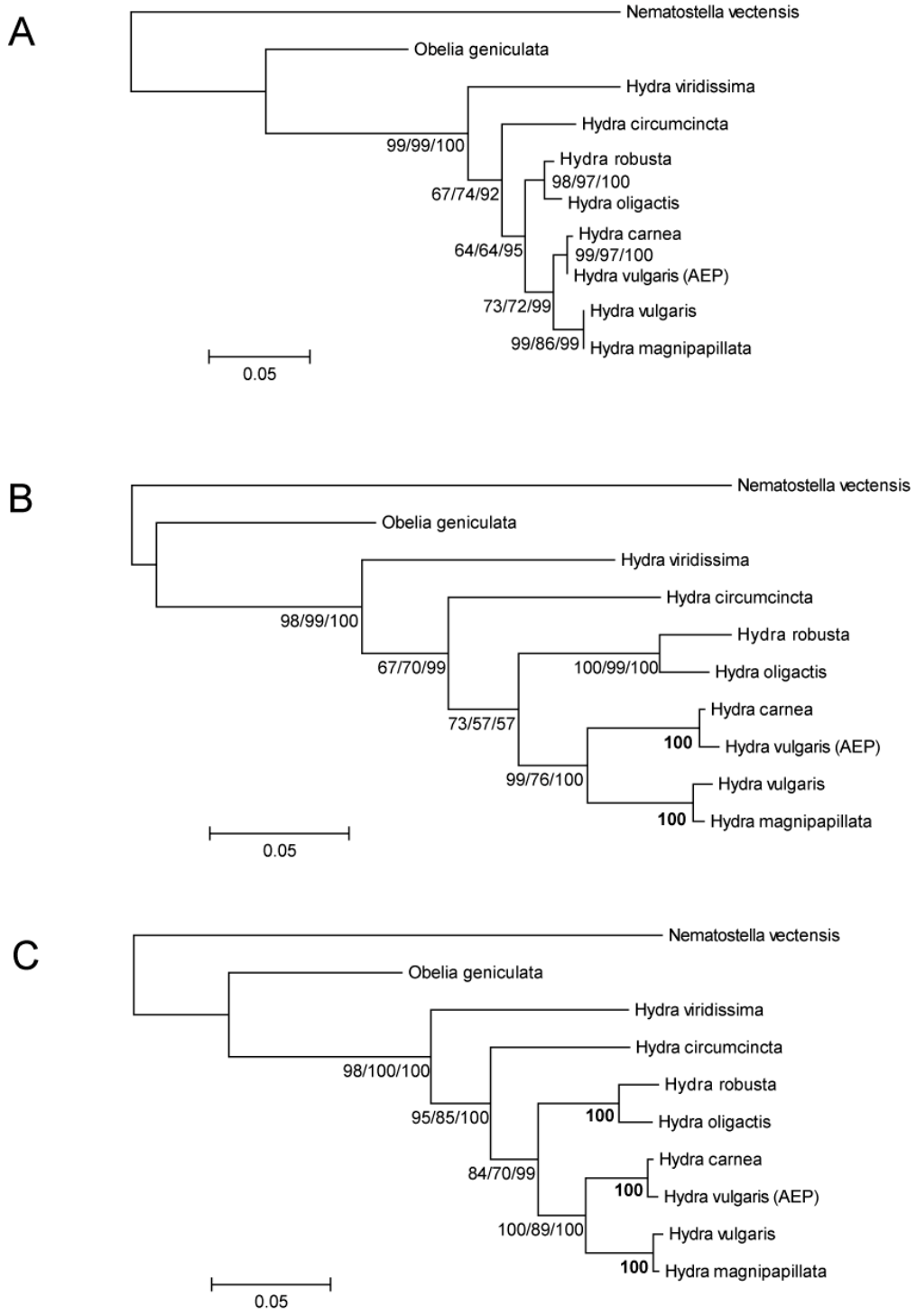


Figure 6: Maximum likelihood phylogenetic trees *inferred of the A) mitochondrial 16s rRNA gene, B) mitochondrial CO1 gene and C) combined mitochondrial dataset. Bootstrap values for ML and MP criteria and Bayesian posterior probabilities (BI) are depicted at the corresponding nodes (order=ML/MP/BI). Single values in bold letters indicate the identical result in all 3 analyses. Branch lengths are scaled to the expected number of substitutions (0.05 substitutions per site). Figure taken from Hemmrich et al., 2006.*

2.1.2 Phylogenetic inference using nuclear genes

Two nuclear genes were used to provide an independent estimate of the evolutionary relationships among the *Hydra* species. The data sets included 1053 bp of the 18S small ribosomal subunit rRNA gene and 1275 bp of the 28S large ribosomal subunit rRNA gene. Corresponding sequences of the marine hydrozoan *Podocoryne carnea* available on GenBank were included as outgroup. As shown in Figure 7A+B, both single-gene maximum likelihood analyses recovered *Hydra viridissima* as the most basal group. *Hydra circumcincta* and the two members of the "oligactis" group (*Hydra oligactis* and *H. robusta*) were recovered as the sister groups to the other four *Hydra* species examined. The only difference between the trees shown in Figure 7 A and B is in the position of *Hydra circumcincta*, as in the 18S rRNA tree it clusters with the "vulgaris" group, while in the the 28S rRNA tree - similar to the trees of mtDNA sequences (see Figure 6) - it is recovered as the sister species to the "oligactis" and "vulgaris" group. Similar to the analyses of mtDNA, phylogenetic trees of both nuclear genes strongly suggest that *Hydra vulgaris* (strain AEP) and *Hydra carnea* form a monophyletic group. Results from the MP and BI analysis support all of the affinities recovered in the ML analysis of the two nuclear genes (see Appendix, Figures 3 and 4). The results of the ML analysis on the combined data sets including the 18S rRNA and the 28S rRNA genes is shown in Figure 7C and indicates that *Hydra circumcincta* should be considered as sister species to *Hydra oligactis* and *Hydra robusta*.

Taken together, in all trees *Hydra viridissima* was significantly differentiated from all the remaining species and recovered as the most basal species. *Hydra circumcincta* and the pair of *Hydra oligactis* and *Hydra robusta* invariably resolved as the sister taxons to *Hydra carnea* and *Hydra vulgaris* (see Figure 7). *Hydra vulgaris* strain AEP clusters with *Hydra carnea* rather than with *Hydra vulgaris* (Basel strain).

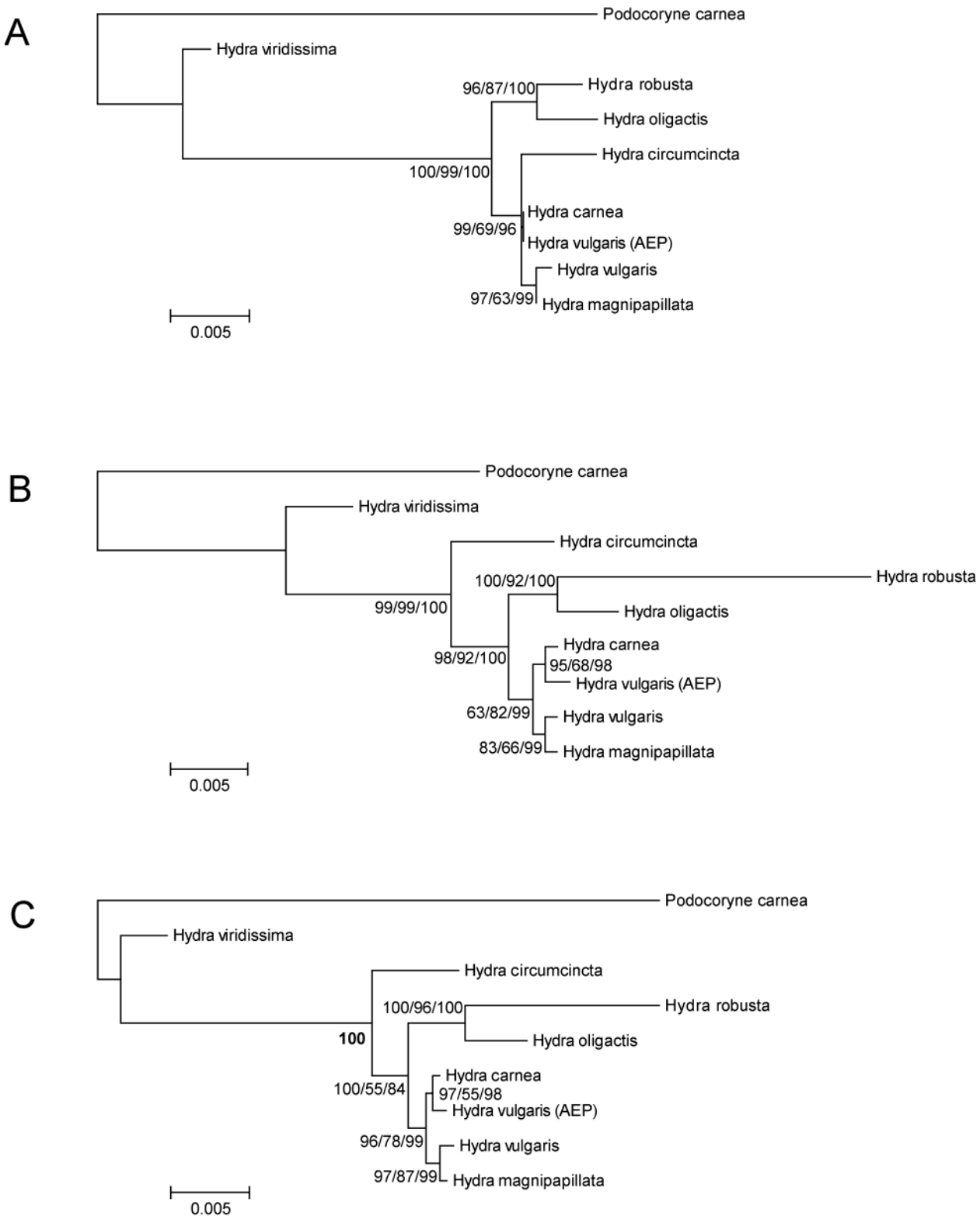


Figure 7: Maximum likelihood phylogenetic trees *inferred of the A) nuclear 18s rRNA, B) nuclear 28s rRNA gene and C) combined nuclear dataset. Bootstrap values for ML and MP criteria and Bayesian posterior probabilities (BI) are depicted at the corresponding nodes (order=ML/MP/BI). Branch lengths are scaled to the expected number of substitutions (0.005 substitutions per site). Figure taken from Hemmrich et al., 2006.*

2.1.3 Phylogenetic inference using morphological characteristics fails to group *Hydra vulgaris* (strain AEP)

The compelling and surprising molecular evidence that *Hydra vulgaris* (strain AEP) is most closely related to *Hydra carnea* and not to *Hydra vulgaris* or *Hydra magnipapillata* prompted us to re-examine morphological characteristics traditionally used for identification purposes within the genus *Hydra*. Beside characters such as general morphology and the order in which tentacles arise on young buds, one of the few diagnostic and reliable features used to classify *Hydra* species is the shape and size of nematocysts (Campbell, 1983). We, therefore, examined the nematocysts in *Hydra vulgaris* (strain AEP) and compared them to the nematocysts in *Hydra carnea* and the other frequently used species. As shown in Figure 8, on the basis of the size and shape of the nematocysts it is impossible to distinguish *Hydra vulgaris* (strain AEP) from the other three species of the “vulgaris” group (*Hydra vulgaris*, *Hydra magnipapillata*, *Hydra carnea*). Other characters such as body form, the order in which tentacles arise, pigments in the epithelium, the mode of sexual reproduction (hermaphroditic versus dioecious), and the genome size also do not allow to assign *Hydra vulgaris* (AEP) to either *Hydra vulgaris* or *Hydra carnea*. Thus, while morphological evidence is not informative to infer the phylogenetic position of *Hydra vulgaris* (strain AEP), molecular evidence strongly suggests, that it is most closely related to *Hydra carnea*. The initial description of this new strain as a strain of the *Hydra vulgaris* species (Martin et al., 1997; Technau and Scholz, 2003) obviously was affected by the lack of molecular data.

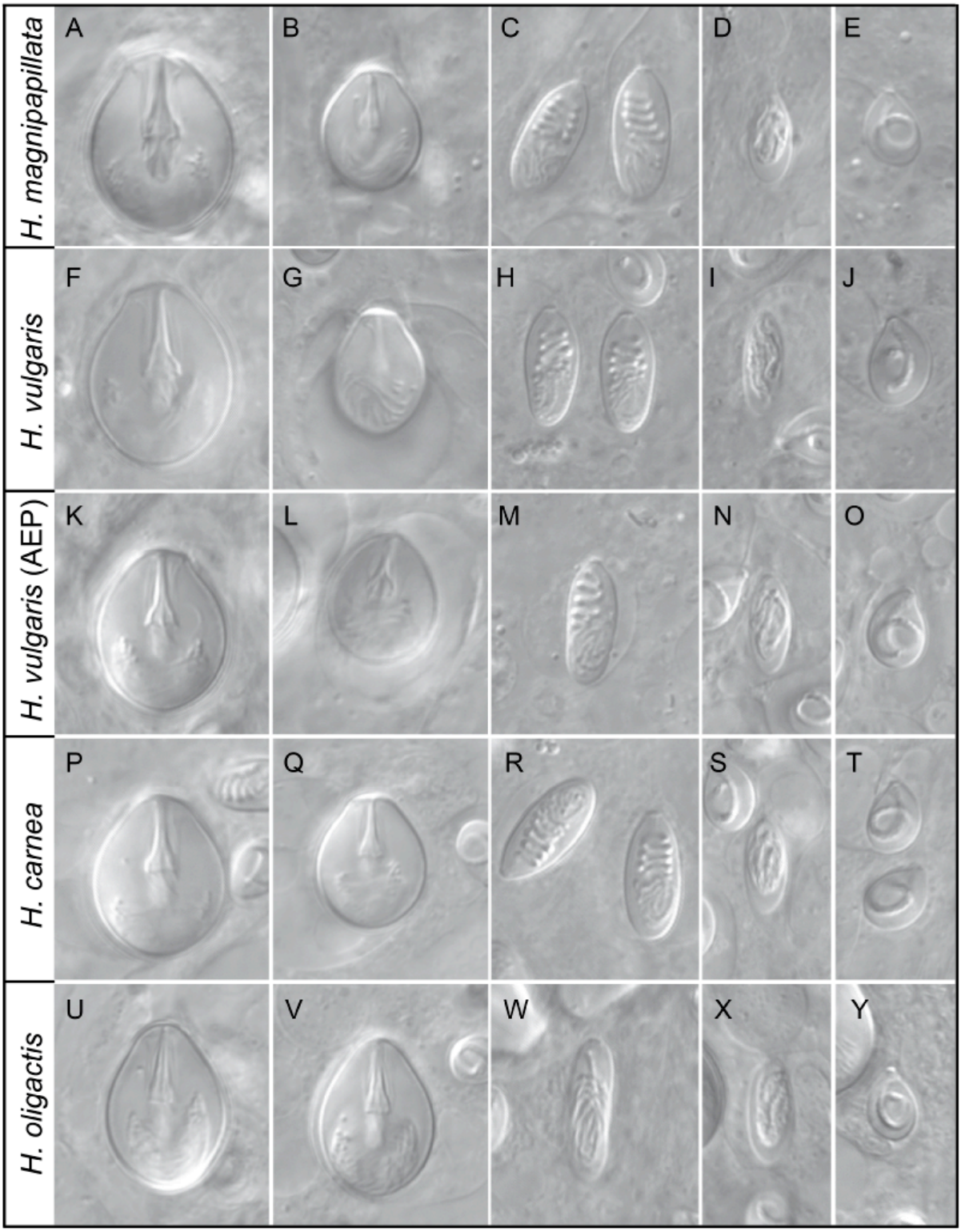


Figure 8: Nematocysts of different species/strains *of the "vulgaris-group" and Hydra oligactis. A-E Hydra magnipapillata; F-J Hydra vulgaris; K-O Hydra vulgaris (AEP); P-T Hydra carnea; U-Y Hydra oligactis; (A,B,F,G,K,L,P,Q,U and V) = stenotels; (C,H,M,R and W) = holotrichous isorhizas; (D,I,N,S and X) = atrichous isorhizas; (E,J,O,T and Y) = desmonemes. Figure taken from Hemmrich et al., 2006.*

presence of symbiotic algae	+	-	-	-	-	-	-	-
presence of stalk	-	-	+	+	-	-	-	-
tentacle length/ body length	similar or shorter	similar or shorter	much longer	much longer	longer	longer	longer	longer
tentacle formation in buds	synchro-nously	synchro-nously	asynchro-nously	asynchro-nously	synchro-nously	synchro-nously	synchro-nously	synchro-nously
mode of sexual reproduction	hermaphro-ditic	hermaphro-ditic	dioecious	dioecious	dioecious	dioecious	dioecious or hermaphro-ditic	dioecious
genome size (Mbp)	380	1150	n.d.	1450	1350	1100	1250	1290

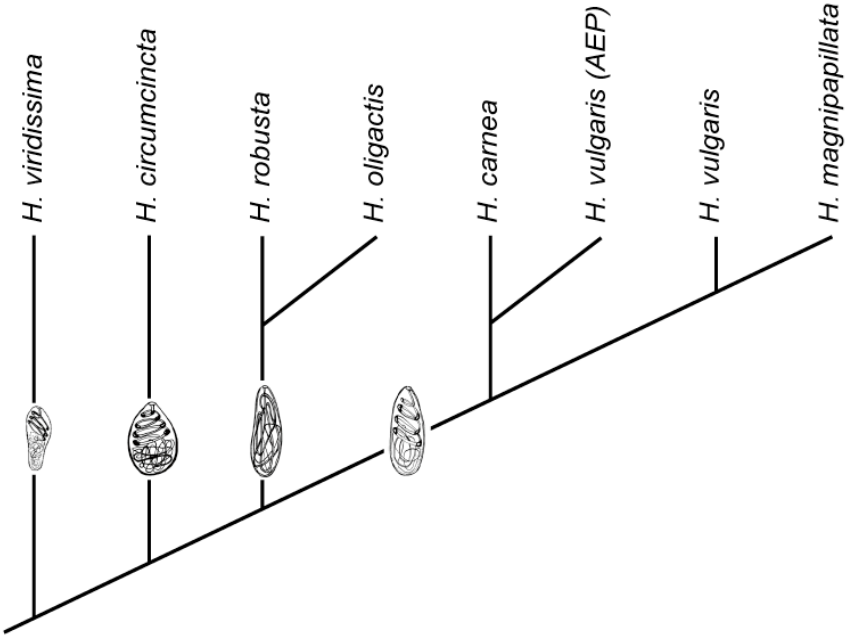


Figure 9: Summary of phylogenetic relations within the genus Hydra *including molecular and morphological data. Schematically depicted in the branches are holotrichous isorhizas of the different groups. Figure taken from Hemmrich et al., 2006.*

2.1.4 Concluding remarks

The results presented in this study represent a preliminary phylogenetic analysis of the *Hydra* species most commonly used in current research. The previously established morphological taxonomy (Campbell, 1983; Holstein, 1995) could be complemented and renewed by the addition of new molecular data, as summarized in Figure 9. Although the work clarifies some of the evolutionary relationships and establishes a solid foundation for future investigations, data from other *Hydra* species are needed to fully understand the evolutionary history and speciation of this group of basal metazoans.

2.2 Establishment of a comparative genomics analysis platform for cnidarian model systems

The availability of online analytical platforms for computational biology today is mainly limited to the well established and widely used model organisms like yeast, fruit fly, earthworm, mouse and man (Bieri et al., 2007; Cherry et al., 1997; Crosby et al., 2007; Eppig et al., 2007; Letovsky et al., 1998). Research groups investigating non-popular models often face the problem of how to get and deal with biological data like genome or EST sequences of their preferred organism. An additional problem lies in the availability and/or accessibility of these data in public domains. Current research projects investigating lower metazoan animals like cnidarians or poriferans (see also Figure 1) are confronted exactly with these problems. Several different sequencing initiatives generated large amounts of (mostly raw) genomic and EST sequence data that are scattered on computers all over the world.

To provide a comprehensive working environment for comparative genomic studies, it is crucial to centralize, integrate and pre-analyze these data and make them publicly available for the interested researcher. The need for such a bioinformatics analytical environment for cnidarian model systems led to the idea of establishing a comparative genomics online platform for basal, evolutionary old metazoan animals.

2.2.1 "Compagen" – a comparative genomics platform for basal metazoa

With "Compagen" I have put together a huge collection of raw and processed genomic and transcriptomic sequence datasets derived from various lower metazoans, generated by public and private sequencing projects. To provide a possible comparative perspective and to enlarge the analysis capability, sequence data from higher metazoan non-model organisms as well as from the unicellular choanoflagellates have been included. Enabling the application of various computational methods, all datasets are organized on a bioinformatics analytical platform on unix based computer systems (Figure 10) situated at the Zoological Institute, University of Kiel.

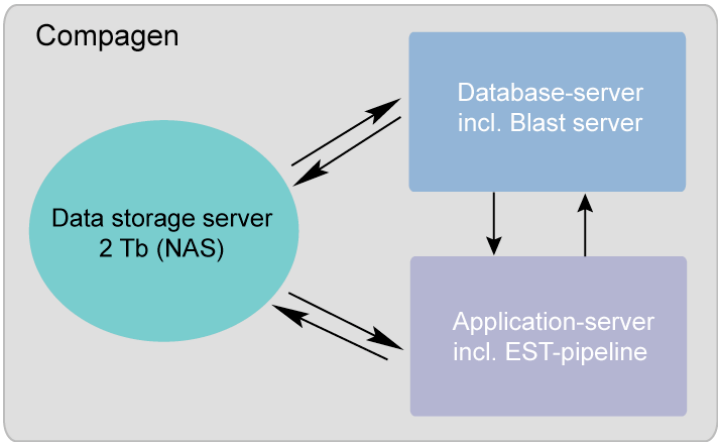


Figure 10: Schematic representation of the "Compagen" platform *at the University of Kiel.*

The platform is split into a database-server, managing sequence data, a storage server for deposition of large amounts of data and an application-server, responsible for data processing. Possible computational analyses include (see also Table 3) sequence assembly of ESTs and small genomic datasets, sequence annotation, gene and peptide prediction, spliced alignments of cDNA to genomic sequence and the prediction of conserved domains via hidden markov models (HMMs) with the possibility to implement additional methods as required. To provide the possibility of sequence similarity searches, all datasets have been made searchable through an online Blast-server (Figure 11) that can be accessed through the Internet at http://www.compagen.org on request.

Figure 11: Homepage of the "Compagen" facility Blast-server*. http://www.compagen.org.*

At the current state, datasets for at least 25 different animal species (as indicated in Figure 12) are stored in databases containing round about 59 million sequences (see also Table 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 organism and the date of construction (Appendix, Table 1).

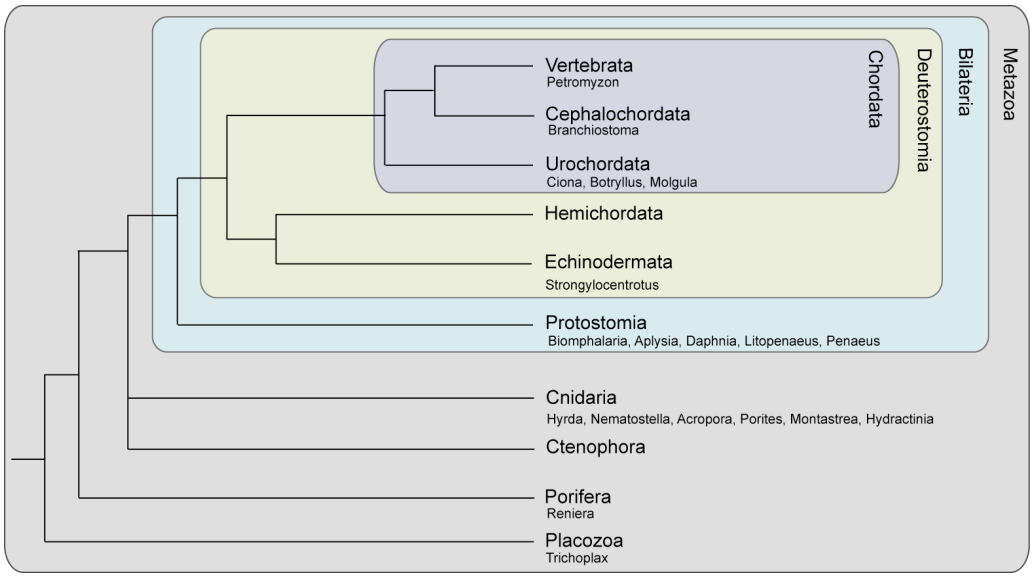


Figure 12: Schematic evolutionary tree of selected metazoan organisms*. For the indicated species, datasets on the "Compagen" database-server are available.*

2.2.2 Datasets and computational tools on the "Compagen" server

Currently the "Compagen" database-server (Figure 10) harbors a collection of different sequence datasets that are stored as so called "flat files" in plain text format. An integration of all sequences into a common relational database scheme is planned but requires more powerful computational resources. So far, all datasets have been formatted into searchable databases for local and online Blast analysis. The datasets can be divided in 1) raw genomic sequence data (dbWGS), 2) raw EST sequence data (dbEST) and 3) processed EST sequence data (dbUNI, dbPEP, dbCAP3). The "dbWGS" section contains exclusively single whole genome shotgun (WGS) sequencing reads, originating from the corresponding organisms genome-sequencing project. In the "dbEST" section all raw EST sequences are organized. The remaining sections contain Unigene collections (dbUNI), predicted peptides (dbPEP) and CAP3-assembled EST

datasets (dbCAP3) that enable for conserved domain searches or gene prediction. Table 2 gives an overview of the different datasets currently available on the “Compagen” platform.

Organism	dbWGS	dbEST	dbUNI	dbPEP	dbCAP3
Hydra magnipapillata	10.272.644	163.221	19.845	19.845	25.106
Hydra mag. SF-1		30.715			
Hydra vulgaris		6.105			
Hydra AEP		2.851			
Hydra viridissima		4.608			
Nematostella vectensis	8.411.866	166.595			30.666
Acropora millepora	14.625	10.247	6.020	5.062	
Acropora palmata	11.025	4.017			
Porites lobata	11.450				
Hydractinia echinata		9.460			
Montastrea faveolata		2.156			
Biomphalaria glabrata	in progress	10.882			
Aplysia californica	4.320.600	179.001			
Daphnia pulex	2.724.768	1.548			
Daphnia magna		11.964			
Litopenaeus vannamei		7.429			
Penaeus monodon		7.330			
Strongylocentrotus purpuratus	7.352.452	17.012			
Petromyzon marinus	18.808.412	108.847			
Monosiga brevicollis	640.632				
Monosiga ovata		7.391			
Trichoplax adhaerens	pending	pending			
Reniera sp.	2.823.539	83.040			
Molgula tectiformis		106.863			
Branchiostoma floridae	11.953.628	277.538			
Total # each:	**58.454.690**	**1.218.820**	**28.865**	**24.907**	**55.772**

Table 2: Databases and corresponding sequence-counts *stored on the “Compagen” facility sever.*

In addition to the sequence data resources on the database-server, the “Compagen” application-server (Figure 10) provides a variety of computational tools that enable for extensive DNA and protein sequence analysis as well as for the inference of phylogeny (see Table 3). For general sequence analysis (pairwise alignment, six-frame translation, restriction site prediction etc.), the two commonly available software suites from NCBI and EMBOSS have been installed.

Software name	Description
ncbi Toolkit	General bioinformatics tools package for sequence analysis
EMBOSS	European Molecular Biology Open Software Suite
blast / wwwBlast	Sequence similarity searches / online application (Blast server)
tgicl	EST clustering and assembly
minimus	Assembly of smaller genomic datasets
MUMmer	Alignment of large sequences and whole genomes
ESTscan	Detection and evaluation of potential coding regions in ESTs
AAT	Generation of spliced alignments (EST vs. genome)
HMMer	Detection of conserved domains using HMMs
Mr. Bayes	Bayesian inference of phylogeny
Phylip	Inference of phylogeny using maximum likelihood methods

Table 3: Summary of important software programs *available on the "Compagen" server.*

The TIGR gene indices clustering tool package (Pertea et al., 2003) serves as backbone for the later described EST analysis pipeline. ESTscan (Iseli et al., 1999) is required for the detection and evaluation of coding regions in assembled ESTs. To generate local genomic assemblies a subprogram of the whole genome shotgun assembler AMOS, called "Minimus" (Sommer et al., 2007), has been implemented. Enabling for the alignment of very long (several 100 kb) sequences, the MUMmer program (Delcher et al., 2002) has been installed and to generate so called spliced alignments between EST and genomic sequences, the AAT suite has been built in. The HMMer software (Eddy, 1998) has been added for the prediction of conserved domains within protein sequence datasets. As phylogenetic analysis applications serve the Tree-Puzzle program (Schmidt et al., 2002) as well as the bayesian inference software Mr. Bayes (Huelsenbeck and Ronquist, 2001).

2.2.3 Establishment of an EST analysis pipeline on the "Compagen" server

For not to analyze thousands of sequences by hand, it is necessary to use *in silico* analytical tools for the analysis of redundant EST sequence data. As large amounts of raw EST data from various cnidarians are available, I established a semi automatic EST analysis pipeline as part of the "Compagen" genomic analysis platform. The pipeline was conceived to handle large and redundant sequence datasets. Major components of the pipeline were previously developed at the bioinformatics section of "The Institute of Genomic Research" (TIGR) in Rockville. Algorithms for preparatory as well as analytical steps were structured into a 5-step procedure shown in Figure 13.

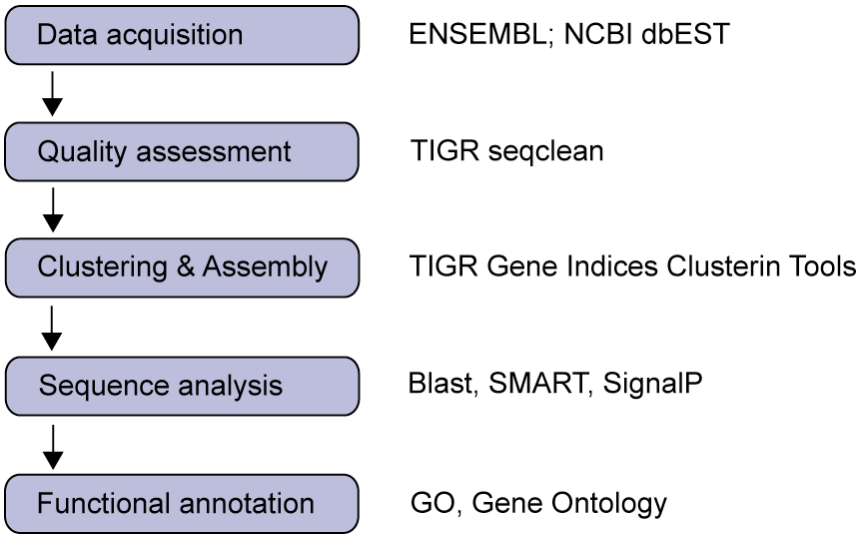


Figure 13: The EST analysis pipeline *for large-scale gene expression profiling*

Raw sequence data mostly were acquired from public databases using the EBI sequence retrieval system SRS (Kulikova et al., 2007) or by downloading directly from the corresponding sequencing center. Using the TGICL *seqclean* program, a careful quality assessment was performed. Vector and/or adaptor sequences were clipped away and low quality sequencing reads were removed. The resulting “cleaned” data then were subjected to the clustering and assembly routine. The purpose of this routine is to efficiently cluster and create assemblies (contigs) 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 (Zhang et al., 2000) 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 (Huang and Madan, 1999) 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”. To check whether the assembly was accurate, the program *clview* was used to visualize the multiple alignments. Eventual misassembly, accidental contig fusions or other mistakes in the previous steps could thus be excluded.

For further analysis of the resulting sequences from the EST assembly, contigs and singletons of each library were subjected to consecutive batch Blast-searches (Altschul et al., 1990) (see Figure 14).

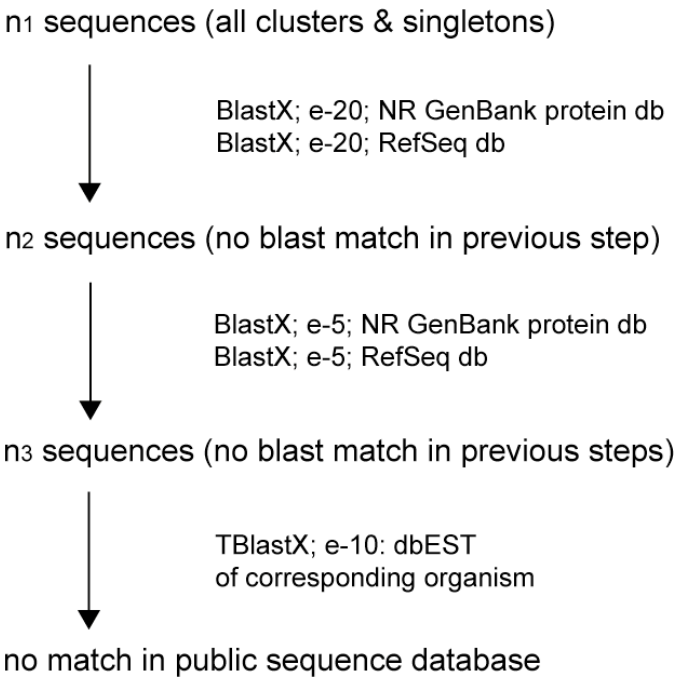


Figure 14: Consecutive Blast pipeline *for the EST analysis pipeline*

In phase n1, BlastX searches using the NCBI non-redundant protein database and the annotated RefSeq (Pruitt et al., 2007) database (threshold e^{-20}) were carried out. All sequences, that gave no homologous match within the first search round were then subjected to phase n2, where the sequences were subjected to BlastX against the same databases as before but with lower similarity threshold (e^{-5}). Sequences that gave no match in both previous searches could then optionally be searched in phase n3 vs. the EST or UniGene database of the corresponding organism, to clarify whether a corresponding similar sequence is already present. Sequences that found homologs in phase n1 were referred to as “strong hits” or real homologous sequences. Blast matches from phase n2 were denoted “weak hits” or highly diverged homologous sequences. Contigs and singletons that gave no Blast match in any databank (n2 and/or n3) were taken as “No Blast match” indicating genes diverged beyond recognition, novel genes or untranslated regions (UTR). Further possible sequence analysis steps

included domain-searches using SMART (Letunic et al., 2004), the prediction of putative signal peptides using SignalP (Bendtsen et al., 2004) or transmembrane domains using TMHMM (Moller et al., 2001). In the optional last step of the pipeline all contig sequences with “strong hits” in the first Blast-search were assigned a specific GO-term indicating a putative functional category of the predicted peptide sequence. This functional annotation was done in most cases by hand, as automatic annotation software such as GoBlet (Groth et al., 2004) or AutoFACT (Koski et al., 2005) only function for highly conserved sequences when compared to the available annotated reference databases like SwissProt (Bairoch and Apweiler, 1996) or RefSeq (Pruitt et al., 2007).

2.2.4 “Compagen”, a growing resource – future perspectives

With regards to the already available variety of different sequence datasets for many different metazoan and non-metazoan organisms and the possibility to subject these data to modern bioinformatics tools, the “Compagen” platform has the potential to develop into a comprehensive, publicly available analysis resource. However, to serve as online platform several important steps have to be taken. As mentioned above, the only usable tool from outside is the Blast-server, which is *per se* helpful. But an internal sequence retrieval system is still missing and sequences have to be copied from long lists. Another problem lies in the availability or retrievability of all other additional analysis results for each single sequence. The most important step to be taken in the future will be the integration of all data into a relational database system, to interconnect all related information, and to provide a graphical interface for analysis and retrieval of all required information.

2.3 Using the "Compagen" platform to unravel the innate immune repertoire in Cnidaria

The availability of whole genome sequences for two cnidarians, the hydrozoan *Hydra magnipapillata* and the basal anthozoan *Nematostella vectensis*, together with elaborate EST datasets for these and for the coral *Acropora millepora,* offered the possibility of getting new insights into the evolution of innate immune systems. In a first research project using the "Compagen" bioinformatics analytical platform, available genomes and transcriptomes of the above-mentioned animals were screened for counterparts of key components of the vertebrate innate immune repertoire (Miller and Hemmrich et al., 2007, in press).

2.3.1 Toll receptors and other TIR-domain containing proteins

Searching the *Hydra* predicted protein collection using PFAM precompiled hidden markov models (HMMs, (Sonnhammer et al., 1998) identified only four TIR domain-containing proteins, two of which are clearly related to MyD88, which functions downstream of TLRs (see Table 4) in the classical Toll signaling pathway. Consistent with their assignment as MyD88 family members, both of these *Hydra* proteins also contain the characteristic DEATH domain.

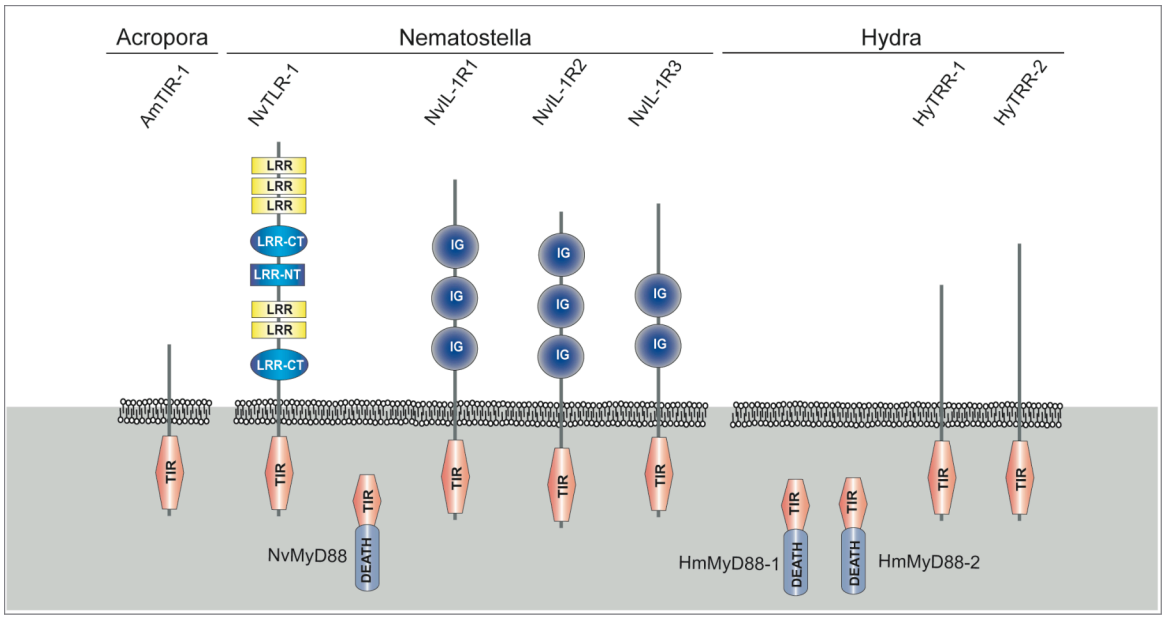


Figure 15: Summary of domain structures of TIR domain containing proteins *identified in selected Cnidaria. Figure taken from Miller and Hemmrich et al., 2007.*

The two other *Hydra* TIR proteins are atypical transmembrane proteins in having relatively short extracellular domains that are devoid of the LRR domains that characterize Toll and the TLRs (see Figure 15). cDNAs encoding these proteins have previously been isolated by the Bosch laboratory (Bosch et al., in prep.) and their functions are presently under investigation; these proteins are annotated as HyTRR-1 and HyTRR-2. Surprisingly, extensive searching of the *Hydra* genome and all available EST/cDNA resources failed to identify any proteins having the canonical Toll/TLR structure, characterized by possession of both LRR and TIR domains.

Whereas only four TIR proteins are present in *Hydra*, substantially more could be identified within the predicted proteins from *Nematostella* using HMM-based search methods. Five of them were sufficiently complete to be included in the analyses presented here. These include a single MyD88 homolog (NvMyD88) and a protein (NvTLR-1) clearly related to members of the Toll/TLR family (Figure 15). Whereas the mammalian TLRs, and some members of the fly Toll/TLR family, have only a C-terminal cysteine-rich motif flanking the LRRs proximal to the membrane, *Nematostella* NvTLR-1 is predicted to contain both C- and N-flanking cysteine-rich motifs in the extracellular part of the protein (Figure 15). This suggests that fly and anemone Toll receptors more closely reflect the ancestral domain structure than do the mammalian TLRs. Moreover, a phylogenetic analysis (see Figure 16) groups the TIR in *Nematostella* NvTLR-1 with its fly and human counterparts, with strong bootstrap support.

Surprisingly, three more of the predicted *Nematostella* TIR proteins also contain multiple immunoglobulin (Ig) domains (Figure 15), and thus reflect the domain structure of mammalian interleukin 1 receptors (IL-1R). NvIL-1R1 and -2 each contain three Ig domains, and NvIL-1R3 contains two predicted Ig domains (Figure 15) but may be incomplete. In the phylogenetic analysis based on TIR domains the *Nematostella* IL-1R like proteins form a clade distinct from both the MyD88 and Toll/TLR types (Figure 16), although these cnidarian TIRs appear to be distinct from those in the vertebrate IL-1 receptors (data not shown). Several other TIR proteins were detected amongst the sequences of *Nematostella* (Appendix, Table 2), but were not subjected to further analysis as the TIR domains were incomplete or the sequences were judged likely to be artefactual. Searching the available coral datasets identified two complete TIRs. The trace archive yielded one TIR from *Acropora palmata* (ApGenomic) and a second was

encoded by an *Acropora millepora* EST (AmTIR-1). These two coral TIRs are most similar to those in the *Nematostella* IL-1R-like proteins (Figure 16), but no linked domains have yet been identified in these cases.

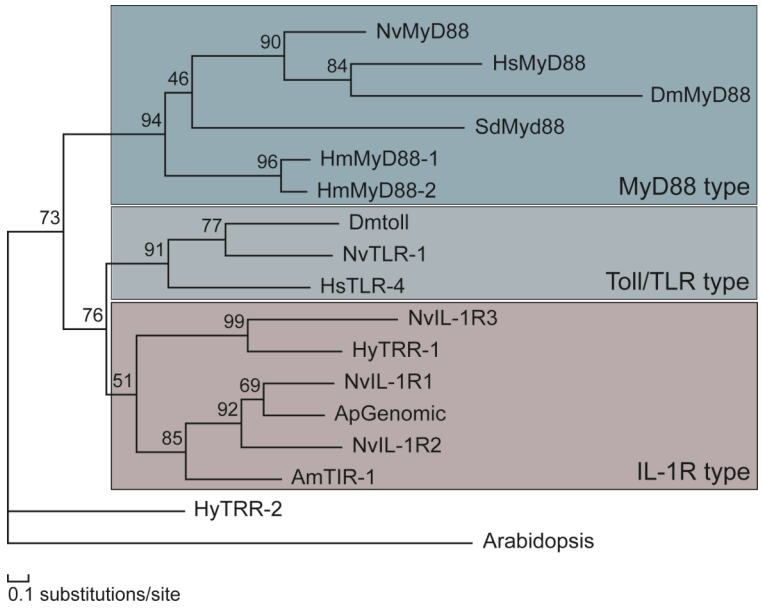


Figure 16: Phylogenetic analysis of cnidarian TIR sequences in comparison to a selection of TIR domains from other species*. The ML tree shown is the result of analysis of an HMM-based alignment of TIR domains. Three clades are resolved by these analyses, corresponding to the TIR domains characteristic of the "MyD88-type", "Toll/TLR-type" and "IL-1R-type". In addition to the TIR domain, the first of these types contains a death domain and the second contains multiple LRRs. Like the mammalian receptors for interleukin 1, the three Nematostella proteins falling into the third clade each also contain multiple immunoglobulin domains. Note that HyTRR1 does not contain such domains and that it is not yet clear whether either of the Acropora proteins does. The Acropora sequences included in the analysis were predicted from A. palmata genomic clones (ApGenomic) and from an A.millepora cDNA clone (AmTIR-1). Hydra lacks a canonical Toll/TLR, having only two MyD88 genes and the two sequences known as TRR-1 and TRR-2; Hydra magnipapillata and Nematostella vectensis sequences are indicated by the prefixes Hy and Nv respectively. Reference sequences: HsMyD88 = human MyD88 (SwissProt:Q99836); DmMyD88 = fly MyD88 (GenBank:AAL56570); SdMyD88 = Suberites MyD88 (EMBL:CAI68016); Dmtoll = fly Toll (SwissProt:P08953); HsTLR4 = human TLR4 (EMBL:CAD99157); Arabidopsis (GenBank:AAN28912). Figure taken from Miller and Hemmrich et al., 2007.*

The Müller group recently reported the identification of MyD88 in a demosponge, *Suberites domuncula* (Wiens et al., 2005). However, while the phylogenetic analysis clearly grouped the TIR in this sponge sequence with those present in unambiguous MyD88 orthologs (Figure 16), domain searching indicates that the predicted sponge protein may not have a functional DEATH domain.

	Anthozoa						**Hydrozoa**		
	Nematostella			*Acropora*			*Hydra*		
		Accn. #	e-value		Accn. #	e-value		Accn. #	e-value
TLR pathway:									
LBP	+	gnl|ti|1139929806	7e-51	n.d.			+	gb|DT619160	2e-13
CD14	-			n.d.			-		
TLR	+	gnl|ti|573160901 gnl|ti|566578628 gnl|ti|558319530 gnl|ti|567085258 gnl|ti|581064934	1e-47	+	gb|EF090256	2e-7	-		
MyD88	+	gnl|ti|1139972660	4e-26	n.d.			+	gb|CV182656	1e-18
IRAK	+	gnl|ti|1146119691	3e-14	n.d.			+	gb|DT608600	2e-10
TRAF6	+	gnl|ti|1135509399	2e-51	+	gb|DY583189	1e-38	+	gb|CV985667	3e-41
TAK1	+	gnl|ti|1135635219	1e-51	+	gb|DY583694	8e-119	+	gb|DN812953	1e-45
$I_\kappa K$	+	gnl|ti|1135636054	5e-68	n.d.			+	gb|CV985420	2e-60
$NF_\kappa B$	+	gnl|ti|1139960940	1e-74	+	gb|DY582971	3e-36	-		
IFN pathway:									
TRAM	+	gnl|ti|1139940977	9e-66	+	gb|DY579224	5e-72	+	gb|DT615400	1e-58
TRIF	+	gnl|ti|1139933368	4e-07	n.d.			?		
TBK-1	?			n.d.			?		
IRF3	+	gnl|ti|1146121907	6e-13	n.d.			+	gb|DT609518	2e-14
p65	-			n.d.			-		
IFN-ß	-			n.d.			-		
ECSIT pathway:									
ECSIT	+	gnl|ti|1139978500	4e-35	n.d.			+	gnl|ti|1223628732	2e-18
MEKK1	+	gnl|ti|1139956887	2e-28	+	gb|DY581138	3e-83	+	gnl|ti|1226566543	3e-25
MKKs	+	gnl|ti|557758729	1e-14	n.d.			+	gnl|ti|1121918104	1e-18
JNK	+	gnl|ti|1135503269	1e-106	n.d.			+	gnl|ti|877334588	2e-33
p38	+	gnl|ti|1139959014	1e-114	+	gb|DY579712	5e-111	+	gnl|ti|686048504	7e-39
AP1	+	gnl|ti|1139792930	3e-10	+	gb|DY581320	3e-09	+	gb|CX771032	7e-10
ATF	+	gnl|ti|1139796564	4e-11	n.d.			+	gb|CN624618	3e-06
Other TLR related proteins:									
HyTRR-1	-			n.d.			+	gb|DQ449929	0
HyTRR-2	-			n.d.			+	gb|DQ449930	0
IL1-R related proteins:									
IL1R-1	+	gnl|ti|573182253	0	n.d.			-		
IL1R-2	+	gnl|ti|557993643	0	n.d.			-		
IL1R-3	+	gnl|ti|567060226	0	n.d.			-		
Complement system related proteins:									
C3/A2M related	+	gnl|ti|557724205 gnl|ti|559738307 gnl|ti|558391450 gnl|ti|573218050 gnl|ti|558266068 gnl|ti|573218146 gnl|ti|586367083 gnl|ti|557912603 gnl|ti|573084165	1e-84	+	gb|EF090257	1e-134	+	gb|DT618439 gb|CN554187 gb|CO376061	
C6/C7/C8	-			n.d.			-		
MAC/PF domain containing proteins:									
Apextrins	-			+	gb|EF091848	6e-15	+	gb|CV185005 gb|DT613346 gb|CF655657 gb|DT620043	4e-04
Tx60-A	+	gnl|ti|1139936806 gb|DY579588	7e-48 3e-35	+	gb|DY579588	9e-48	+	gb|CV464226 gb|CD680300 gb|BP512716 gb|CV464282 gb|DN246811	1e-07
MPEG	+	gnl|ti|613559286	5e-59	n.d.			-		

Table 4: Overview of innate immunity components present or absent in selected Cnidaria*. Plus or minus indicate presence or absence of genes; components marked "n.d." could not be determined within the limited available Acropora dataset; question marks indicate not resolvable Blast results, mostly within kinase domain encoding sequences. All accession numbers originated either from GenBank (gb) or from NCBI trace archive (gnl\|ti). The given e-values were obtained by BlastX searches against the NCBI nr protein database. Taken from Miller and Hemmrich et al., 2007.*

2.3.2 The Toll/TLR pathway is ancestral but some components are missing or highly diverged in *Hydra*

Most of the intracellular mediators of Toll/TLR signalling could be identified in *Nematostella* and *Acropora*, but some key components appear to have either been lost or diverged beyond recognition in *Hydra* (Table 4). The absence of a Toll/TLR protein *sensu stricto* from *Hydra* is discussed above, but in addition only a single highly derived Rel domain could be found in *Hydra* whereas unambiguous NF-kb homologs are present in both *Nematostella* and *Acropora* (Table 4). In addition to the pathway leading to nuclear localisation of NF-kb, Toll/TLR signalling can activate the JNK and p38 MAPK pathways, leading to transcription of a range of target genes via the AP1/ATF factors. Toll/TLR signalling via JNK/MAPK requires the participation of the ECSIT adaptor protein (Kopp et al., 1999), which also provides a link between the Toll/TLR and TGF-b/BMP pathways (Xiao et al., 2003). The presence of ECSIT as well as the key components of the JNK/MAPK pathway in the cnidarian datasets (Table 4, Figure 17) indicates an early origin for this variant of Toll/TLR signalling.

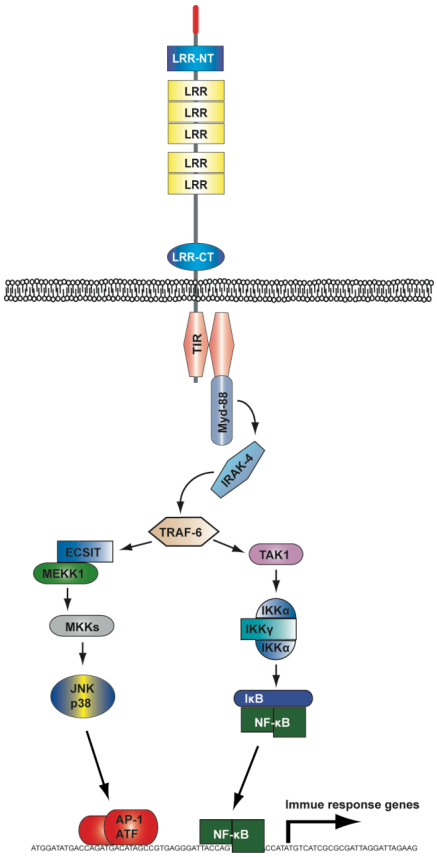


***Figure 17: Signalling pathways downstream of the Toll/TLRs**. Pattern recognition, either indirectly or directly, by Toll/TLRs results in activation of NF-kb (vertebrates) or the Dif/Rel heterodimer (Drosophila) and thus transcription of appropriate immune response genes. At TRAF6, the classical Toll/TIR pathway (shown in the right branch) is linked to the JNK/p38 pathway (shown in the left branch) by the ECSIT protein, which acts as a regulator of MEKK-1 processing (Kopp et al., 1999). Components of both pathways downstream of Toll/TLRs are represented in the cnidarian datasets (Table 1). ECSIT may also act as a link between these and the TGF-b signalling pathway, since it forms complexes with BMP-pathway restricted Smads and is essential for regulation of the BMP-target gene Tlx2 (Xiao et al., 2003). All of the components of the TGF-b signalling pathway are also known from anthozoan cnidarians (Technau et al., 2005). Figure taken from Miller and Hemmrich et al., 2007.*

2.3.3 Cnidarian complement C3 and related proteins

The complement component C3 has recently been reported in another anthozoan cnidarian, the octocoral *Swiftia* (Dishaw et al., 2005), and the corresponding gene has recently been cloned from *Acropora* (Hayward et al., unpublished). The *Acropora* C3 (C3-Am) gene is first expressed strongly in the endoderm of the planula as it elongates following gastrulation (Figure 18A). The endodermal expression is not uniform, being most intense in a subset of dark staining cells that have not yet been characterized. As the planula elongates expression becomes somewhat weaker, with the strongest expression localised to the aboral endoderm (Figure 18B). Post-settlement (Figure 18 C-E) expression is limited to the endoderm and is particularly strong in the endoderm of the polyp as it rises from the calcifying platform at its base (e.g. Figure 18D).

C3 has a complex domain structure. While anthozoan C3s resemble their deuterostome counterparts both in domain structure (Figure 18F) and sequence, not only could no corresponding gene be identified in *Hydra*, but also some of the domains characteristic of C3 (ANATO, C345C; see Figure 18F) could not be detected in any *Hydra* protein. Although lacking a canonical C3, *Hydra* contains a gene encoding A2M related domains. Interestingly, *in situ* hybridisation in *Hydra* using a probe covering these typical A2M-related domains (Figure 18F; A2M-comp/A2M-recep) showed expression restricted to the endodermal epithelium (Figure 18G), as was the case with *Acropora* C3.

2.3.4 MAC/PF domain containing proteins in Cnidaria

Searching for other components of the complement cascade, we identified proteins containing a Membrane Attack Complex/Perforin domain (MAC/PF) similar to that present in complement component C6 and related proteins. HMM searching identified just two MAC/PF domain-containing proteins in *Hydra* (Table 4), whereas four proteins were identified in *Nematostella.* Two MAC/PF proteins were also identified amongst the *Acropora* ESTs. Database searches and analyses of predicted domain structures revealed that most of the cnidarian MAC/PF sequences are likely to fall into three groups corresponding to the known proteins types MPEG, TX-60A and apextrin (Table 4, Figure 18H).

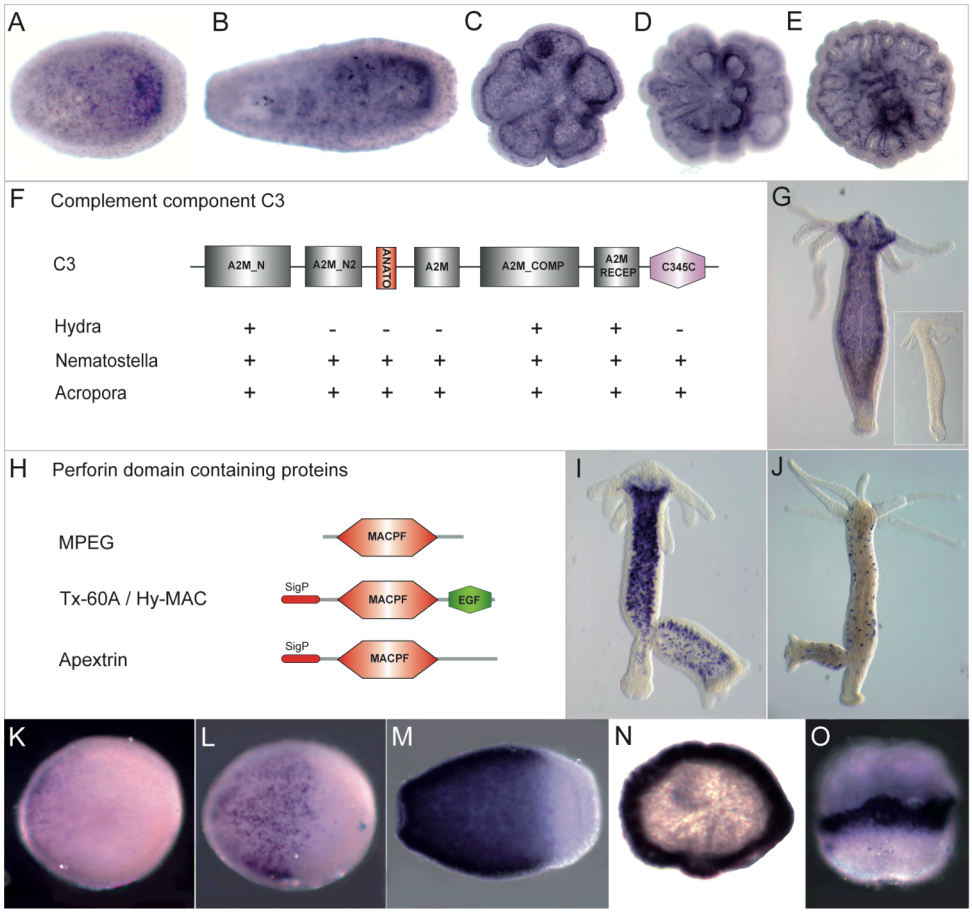


Figure 18: Complement component C3 and MAC/PF domain containing proteins in Cnidaria. *(A-E) In situ hybridisation of C3-Am in Acropora. (F) Domain map and presence/absence data for the various protein domains characteristic of complement C3 components in the Hydra, Nematostella and Acropora datasets. (G) In situ hybridisation of the Hydra magnipapillata A2M-related gene. (H) Domain maps of major cnidarian MAC/PF proteins types. (I) Hydra Tx-60a in situ. The insert shows the sense control. (J) Hydra apextrin in situ. (K-O) Acropora apextrin in situ. Figure taken from Miller and Hemmrich et al., 2007.*

TBlastN-based searches of the *Nematostella* genome identified a gene matching strongly to the human macrophage expressed protein 1 (MPEG1; gbXP_166227) and its abalone homolog abMPEG1 (gbAAR82936) (Mah et al., 2004). A clearly related gene in *Suberites domuncula* has recently been implicated as an effector in a hypothetical sponge innate immune defence pathway (Wiens et al., 2005). Recombinant *Suberites* MPEG has anti-bacterial activity against gram-negative bacteria, and is up-regulated after lipopolysaccharide (LPS) treatment (Wiens et al., 2005). The MPEG1 family clearly has an ancient evolutionary history (the sponge and human sequences have 28% identity and 46% similarity) but only in *Suberites* has any functional characterization been done. Despite the presence of MPEG1 in the sponge and an anthozoan, no corresponding gene could be identified in *Hydra*.

The nematocyst venom of at least some anthozoans contains the protein TX-60A (Oshiro et al., 2004), and two of the *Nematostella* MAC/PF proteins and one of the *Acropora* ESTs clearly correspond to this protein type (Table 1). TX-60A has an EGF domain immediately C-terminal of the MAC/PF domain. In *Hydra*, this domain structure can be found in Hy-MAC, one of the two *Hydra* MAC/PF proteins (Figure 18H and Table 4). However, it is unclear whether the *Hydra* and anthozoan sequences are orthologous, as overall sequence identity is low. *In situ* hybridisation analysis shows that expression of Hy-MAC is restricted to gland-cells that are interspersed throughout the endoderm of *Hydra* (Figure 18I). Since endodermal gland cells and nematocysts are terminally differentiated, this pattern of expression is not easy to reconcile with a common function for the venom TX-60A and Hy-MAC.

2.3.5 *Apextrin*, a gene lost from *Nematostella*

The third class of cnidarian MAC/PF proteins represented in the *Hydra* and *Acropora* ESTs (Figure 18H) contains no other identifiable domains than MAC/PF. These proteins have moderate overall similarity to the echinoderm apextrins (Haag and Raff, 1998; Haag et al., 1999) and to the apicomplexan protein family to which *Plasmodium* MOAP (Kadota et al., 2004) belongs. MOAP is responsible for rupture of epithelial cells in the insect host by the ookinete stage of the parasite. Surprisingly, apextrin seems to be a case of gene loss from *Nematostella* as, despite clearly related genes being present in *Hydra* and *Acropora*, extensive searching of both the predicted protein collection and the anemone genome using a variety of tools failed to identify an apextrin-related gene (Table 4).

2.3.6 Concluding remarks

Taken together, these preliminary analyses of the newly available genomic and transcriptomic datasets indicate that although some immune components have been lost or diverged beyond recognition, a surprising number of key genes of the innate immune system already exist within the Cnidaria, a phylum at the base of metazoan evolution.

2.4 Using "Compagen" for large scale gene expression profiling in *Hydra* and other organisms

Expressed Sequence Tags (ESTs) provide a valuable tool for gene expression studies in the absence of expensive microarray techniques in "non-model" systems like *Hydra*. Sequencing only small (300-600 bp) regions of each transcript of a whole animal, a special tissue or simply a cell, quickly and cost effectively generates large amounts of expression data. For *Hydra* a whole transcriptome EST sequencing project started in 2003 at the Genome Sequencing Center of Washington University in Saint Louis, USA. Studying various aspects of development and evolution on the molecular level, the Bosch lab generated several different cDNA libraries that were sequenced within the *Hydra* EST project or in separate EST sequencing projects (see Table 5). All cDNA libraries were constructed using suppression subtractive hybridization approaches (SSH), a method that allows the qualitative comparison of transcriptomes between different tissues (Diatchenko et al., 1996). Dependent on the focus of each research project, different *Hydra* species with different features were used. *Hydra vulgaris* AEP is known to exhibit an increased level of sexual reproduction (Martin et al., 1997), which predestinates for embryogenetical studies. *Hydra magnipapillata sf-1*, a temperature sensitive mutant strain, loses all i-cells and derivates upon heat shock (Terada et al., 1988). The animals used in other experiments were normal laboratory strains as described by Holstein et al. (Holstein, 1995).

Library:	**Species / Description:**	**# clones:**	**# sequences:**
Kiel 2	*Hydra magnipapillata*, Head regeneration and budding	3.072	3.634
Kiel 3	*Hydra AEP*, Embryogenesis enriched	2.688	2.851
Kiel 4	*Hydra vulgaris*, Pathogen (P. aeroguinosa) induced	2.304	1.715
Kiel 5 +	*Hydra magnipapillata SF-1*, I-cell (+ derivates) enriched	2.304	2.727
Kiel 5 -	*Hydra magnipapillata SF-1*, Epithelial cell enriched	2.304	2.727
Kiel 6	*Hydra oligactis*, species specific	1.152	1.022
Kiel 7	*Hydra magnipapillata*, species specific	1.152	1.104
Kiel 8	*Hydra viridissima*, symbiosis related genes	2.304	4.608
Kiel 9	*Ciona intestinalis*, individual specific variable transcripts	2.304	4.608
Total #		**18.234**	**24.996**

Table 5: Overview of sequenced cDNA libraries *constructed in the Bosch lab.*

In a research project, investigating genes differentially regulated in head regeneration and budding, 3.072 clones were sequenced from a SSH library (Kiel2). Focusing on genes expressed mainly during *Hydra* embryogenesis resulted in sequencing 2.688 clones (Kiel3). For the isolation of genes involved in the innate immune response of *Hydra*, a cDNA library containing genes upregulated upon immune stimulation was constructed and yielded sequences of 2.304 clones (Kiel4). To find genes exclusively expressed in the interstitial cell or epithelial cell lineage, a SSH library enriched in both directions respectively gave 5.454 sequences (Kiel5). The idea of subtracting whole transcriptomes of different *Hydra* species to isolate species-specific genes, led to the construction of *Hydra magnipapillata* and *Hydra oligactis* specific cDNA libraries (Kiel6/7). To gain insights in the regulation of genes involved in the algal symbiosis of the green *Hydra viridissima,* a library of 2.304 clones was sequenced. Taken together over 18.000 clones were sequenced from single or both clone-ends resulting in more than 24.000 EST sequences.

For all above-described cDNA libraries (see Table 5) the “Compagen” EST analysis pipeline (see chapter 2.2.3) was used to generate non-redundant datasets containing the corresponding sequence assemblies (contigs + singletons). In addition, for the first seven libraries, consecutive Blast searches and functional annotations were performed. The complete datasets for each library consisting of raw and processed sequences as well as tables containing Blast-results and annotations are available on the accompanying DVD.

2.4.1 “Compagen” identifies genes differentially expressed during head regeneration and budding

During a research project, investigating genes controlling the processes of regeneration and budding in *Hydra,* a suppression subtractive hybridization based cDNA library was created and sequenced (Augustin et al., 2006). One half of the library contained upregulated (↑) genes during head regeneration and budding, the other half contained the downregulated (↓) ones.

EST sequencing of 3072 clones yielded 3634 sequences. Running the EST analysis pipeline on this dataset generated 448 contigs and 116 singletons in the upregulated part and 504 contigs and 205 singletons in the downregulated part of the library. In total

952 contigs and 321 singletons were obtained. Following the consecutive Blast search procedure resulted in 38% sequences with strong homology to known proteins and 24% sequences exhibiting weak homology, suggesting a high degree of divergence. A surprising fact was the finding of 38% transcripts showing no homology to known protein sequences and thus may represent unknown or novel genes (Figure 19A). Annotation of gene ontology (GO) terms for strong homologous Blast hits could group the sequences in 12 different categories (Figure 19A+B). Apparent quantitative differences in gene up- or downregulation were detectable in the portion of transcripts encoding proteins involved in general cellular metabolism pathways (↑24% / ↓13%), transcripts related to RNA/DNA regulatory pathways (↑14% / ↓8%) and the ribosomal protein category (↑3% / ↓15%). Smaller differences could be detected in the ECM & cytoskeleton portion (↑9% / ↓6%) and in protein metabolism (↑6% / ↓9%).

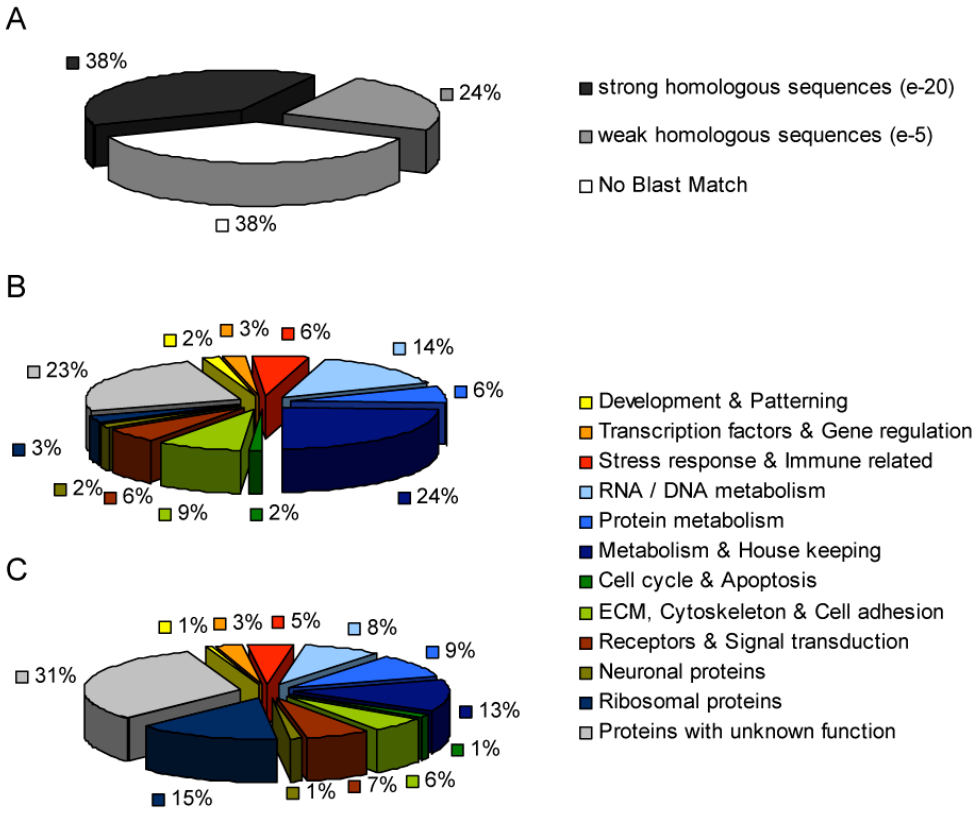


Figure 19: Gene expression profiles of genes differentially regulated during head regeneration and budding*. (A) Results of consecutive Blast analysis; (B) functional annotation of genes upregulated during head regeneration and budding; (C) functional annotation of downregulated genes.*

Two genes *HyDkk1/2/4-A* and *HyDkk1/2/4-C,* identified within this screening approach, belonging to the Dickkopf family (see also Figure 20) were further investigated by Augustin et al. and led to a recent publication (Augustin et al., 2006) underlining their role as components of the positional value gradient in *Hydra*.

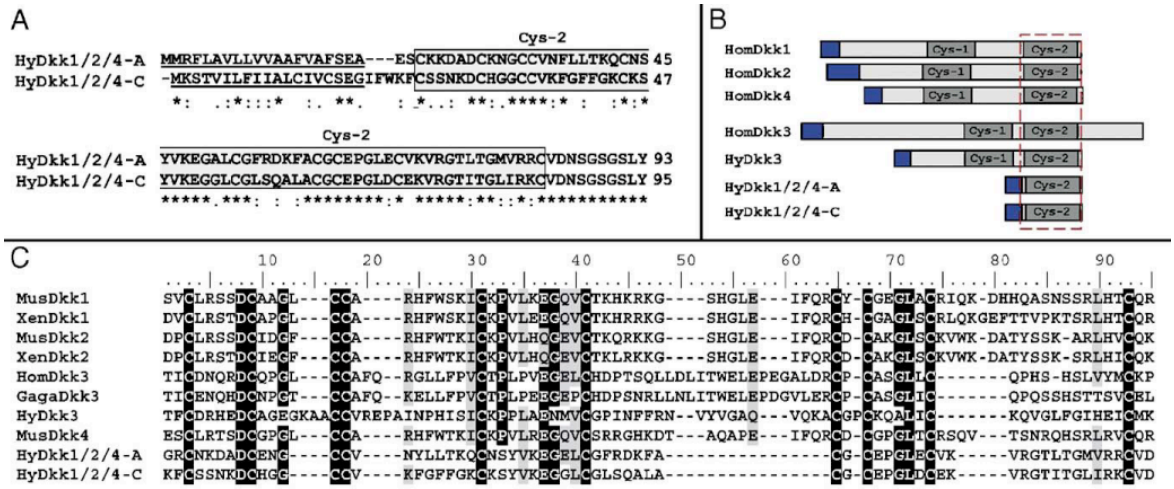


Figure 20: HyDkk1/2/4-C and HyDkk1/2/4-A are similar in structure and related to Dickkopf proteins in vertebrates. *(A) Amino-acid sequence alignment of HyDkk1/2/4-C and HyDkk1/2/4-A. *, Same amino acid residue; :, conserved substitutions; ., semi-conserved substitutions. The signal peptide sequence is underlined and thecysteine-rich domain 2 is shaded. (B) Schematic diagram depicting the structural similarities between HyDkk1/2/4-C and HyDkk1/2/4-A in comparison to HyDkk3and Dkk-1, 2, 3 and 4 in human (HDkk). (C) Multiple amino acid sequence alignment of cystein-rich domain 2 from mouse (Mus), man (Hom), frog (Xen),chicken (Gaga) and Hydra (Hy). black boxed amino acid residues, highly conserved amino acid residues; gray boxed residues, amino acid residues sharing thesame biophysical properties. Figure taken from Augustin et al., 2006.*

2.4.2 "Compagen" identifies genes expressed during embryogenesis in *Hydra*

The primary mode of reproduction in *Hydra* is clonal propagation, called budding. However, *Hydra* also undergoes seasonal sexual phases where eggs and sperm are produced from the interstitial cell lineage (Aizenshtadt and Marshak, 1974; Bosch and David, 1986; Littlefield, 1985; Littlefield and Bode, 1986; Nishimiya-Fujisawa and Sugiyama, 1993). The cellular processes taking place during embryogenesis in *Hydra* are well understood (Aizenshtadt, 1975; Aizenshtadt, 1978; Aizenshtadt and Marshak, 1974; Alexandrova et al., 2005; Honegger, 1989; Martin et al., 1997; Tardent, 1985; Technau et al., 2003) but very little is known about the molecular underpinnings. There is also no clear understanding how different the processes in adult and embryonic patterning are. To isolate genes predominantly or exclusively expressed during

embryogenesis in *Hydra*, an SSH-based cDNA library was generated, subtracting asexually propagating polyps from animals undergoing embryogenesis (Genikhovich et al., 2006). Remaining transcripts should ideally come from genes directly linked to embryogenesis. Construction of the library resulted in 2688 cDNA clones that yielded 2851 EST sequences. Submitting these sequences to the analysis pipeline led to 87 contigs and 47 singletons. The Blast analysis resulted in 42% strong homologous and 24% weak homologous sequences. Again, as well as in the previous analysis, the portion of non-homologous sequences was unexpectedly large containing 34% of the transcripts (Figure 21A). Assignment of GO-terms could group the strong homologous sequences into 11 categories (Figure 21B).

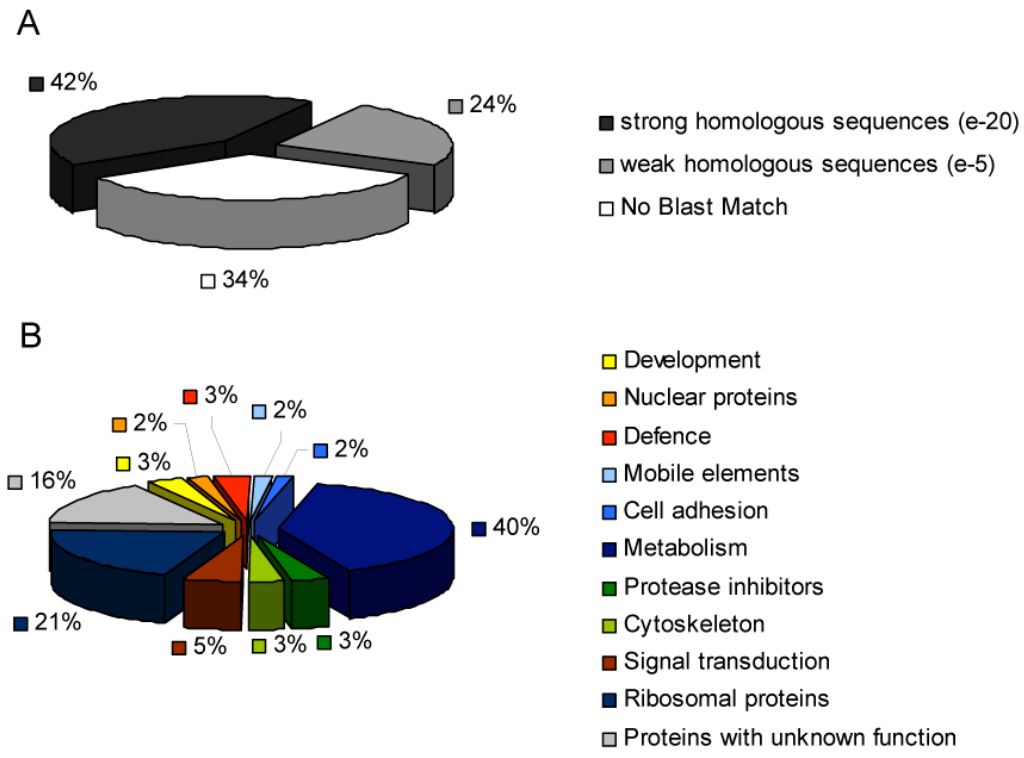


Figure 21: Gene expression profiling for genes differentially regulated during embryogenesis*. A) Result of consecutive Blast analysis; B) functional annotation of genes upregulated during embryogenesis. Figure modified from Genikhovich et al., 2006.*

Further characterization of the different putative embryogenesis specific genes identified several putative candidates. One to be mentioned here is the *Hydra* othologue to *Embryonic Ectoderm Development* (EED), a polycomb group (PcG) gene involved in chromatin modulation and repression of transcription (see also Figure 22). The detailed analysis of *HyEED* and other identified genes were published in Genikhovich et al. (Genikhovich et al., 2006).

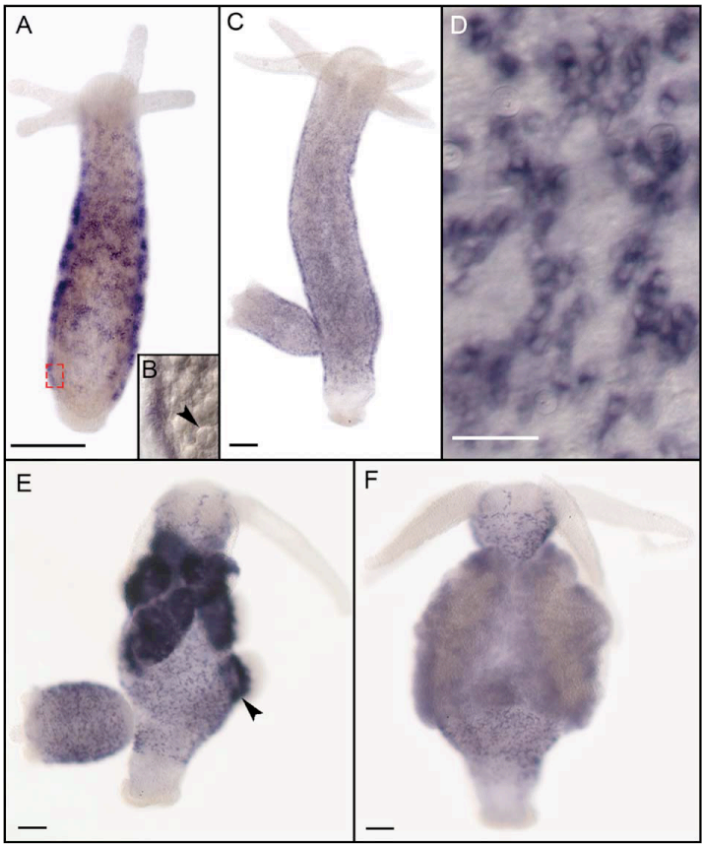


Figure 22: Expression of HyEED in Hydra vulgaris (AEP) polyps. *(A) Newly hatched polyp (scale bar: 240 Am). (B) Close-up of the area boxed in panel A under Nomarsky optics. Embryonic endocytes are still visible in the endodermal cells (arrowhead). (C) Asexual polyp with HyEED-expressing interstitial cells (scale bar: 240 Am). (D) Interstitial cells in a female polyp (scale bar: 30 Am); (E) Male polyp with strong expression of HyEED at the base (arrowhead) of the testis; (F) Female polyp with two egg patches.Figure taken from Genikhovich et al., 2006.*

2.4.3 "Compagen" identifies genes involved in the epithelial defense of *Hydra*

Consisting mainly of two epithelial tissue layers and lacking an impermeable barrier to the outside like a cuticle or an exoskeleton, *Hydra* seems to be highly vulnerable to pathogens in its environment. Living in fresh-water ponds, *Hydra* is constantly exposed to a variety of bacteria, viruses, fungi and protist that may act as potential pathogens. In a large screening approach for potential immune components, the Bosch lab tried to get deeper insights into the immune system of *Hydra*. One approach was to isolate upregulated genes in response to immune stimulation via the suppression subtraction hybridization procedure. To do so, two different SSH libraries were constructed. From previous experiments it was known, that protein extracts of the temperature sensitive mutant *H. magnipapillata sf-1* exhibit higher antimicrobial activity when depleted of all interstitial cells and derivatives than normal animals (Kasahara and Bosch, 2003). Thus, genes could be of interest, which are upregulated during this enhanced antimicrobial activity. In succession, a SSH library was constructed subtracting interstitial cell depleted animals from wild type polyps. 2727 sequences were subjected to the EST analysis pipeline resulting in 297 contigs and 80 singletons. The second library was constructed using animals that were immuno-challenged by bacterial culture supernatant which should induce genes involved in pathogen detection and pathogen defense. Sequencing the library yielded 1715 sequences resulting in 43 contigs and 19 singletons after running through the pipeline. For further analysis both dataset were concatenated (also shown in the statistics Figure 23).

As shown in Figure 23A, nearly half (46%) of the transcripts examined fell into the category of strong homologous sequence. 20% of sequences exhibited only weak homology to know proteins and 34% showed no similarity at all. Comparing the predicted functional categories of the genes identified as strong homologs (also compared to previous findings 2.4.1 and 2.4.2) shows an enlarged proportion of genes involved in antimicrobial- and stress response (Figure 23B). The distribution of other functional gene ontology (GO) categories is comparable to the usual findings.

Interesting candidate genes within this study are *HyPericulin* (from the induced library, see Figure 24) and *HyMacin* (epithelial library) that were later on identified as potent antimicrobial peptides in the *Hydra* immune response. Two other interesting classes of transcripts found in the epithelial library were *HyVinins* (see Figure 24) and *HyDurins*.

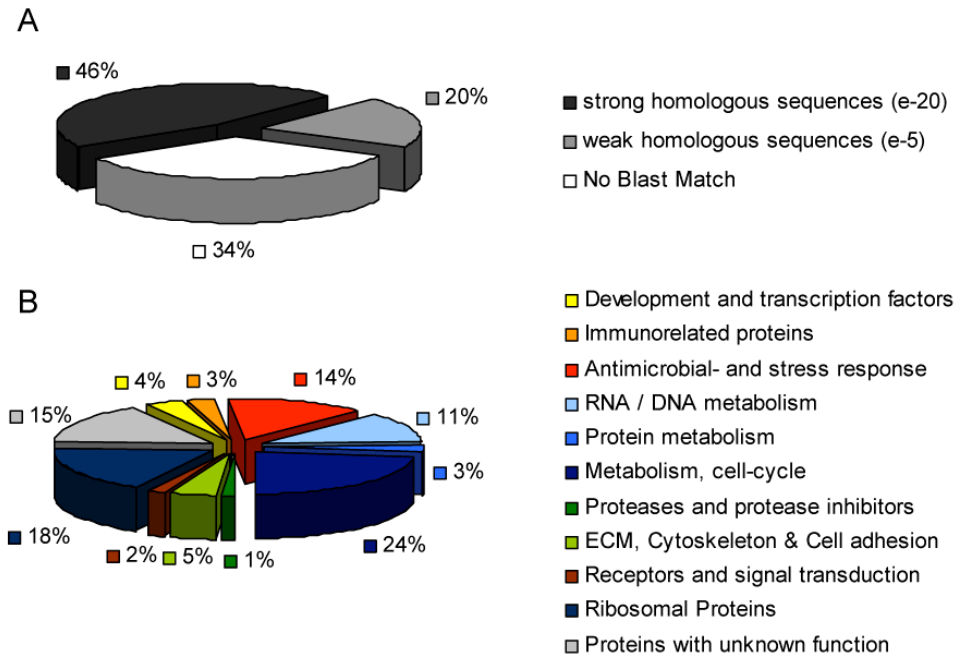


***Figure 23: Gene expression profiles for genes upregulated after immune stimulation in epithelial cells and after pathogen induction**. A) Results of consecutive Blast analysis; B) functional annotation of genes upregulated after immune stimulation.*

These genes were identified in several different “isoforms” and are potentially members of larger gene families. Their domain structure and biophysical features propose a putative role as antimicrobial peptides and their function is currently under investigation. The function of *HyMacin* and *HyPericulin* together with other relevant findings concerning the *Hydra* immune system will be published soon (Bosch et al., in prep.).

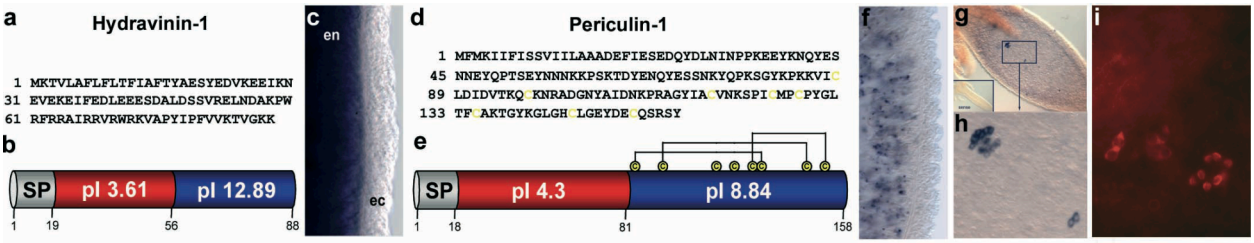


***Figure 24: Hydravinin and periculin, two novel host-defense genes in Hydra.** (a) Hydravinin-1 amino acid sequence; (b) predicted structural features of hydravinin-1. A signal peptide (SP) is followed by an anionic (red,) and a cationic (blue) domain; (c) Hydravinin-1 is expressed exclusively in the endoderm; (d) Periculin-1 amino acid sequence; the 8 cysteines are marked in yellow; (e) structural features of Periculin-1. A signal peptide (SP) is followed by a anionic (red) and a cationic (blue) domain which contains 8 cysteines predicting three disulfide bridges; (f-h) Periculin-1 mRNA is expressed in endodermal cells as well as in interstitial cells in the ectoderm;(i) polyclonal antiserum shows the Periculin-1 peptide localized in the endoderm as well as in some ectodermal interstitial cells. Figure courtesy of T. Bosch.*

2.4.4 "Compagen" identifies genes expressed in cnidocytes

Unraveling the genetic mechanisms that account for animal diversity remains one of the central problems in evolutionary biology. At present there is no comprehensive understanding of how taxon- or species-specific features are encoded. Finding changes in the spatio-temporal use or specificity of regulatory genes being correlated with differences in morphology between different species, it is generally assumed that animal diversity is mainly the cause of differential use of the same conserved components (Duboule and Wilkins, 1998). Another source of creating evolutionary novelty may be differences in the action of downstream or effector genes. A totally neglected fact is the presence of taxon- or species-specific genes.

During an experiment focused on the transcriptomes of interstitial cells and their derivatives (Milde, Hemmrich, Bosch, unpublished), a proportion of new and unknown genes were found to be expressed exclusively in the developing or mature cnidocytes, a cell type restricted to the Cnidaria. The approach included the construction of a SSH library subtracting i-cell containing polyps from i-cell depleted *Hydra magnipapillata* sf-1 and subsequent sequencing of 2304 clones resulting in 2727 cDNA sequences.

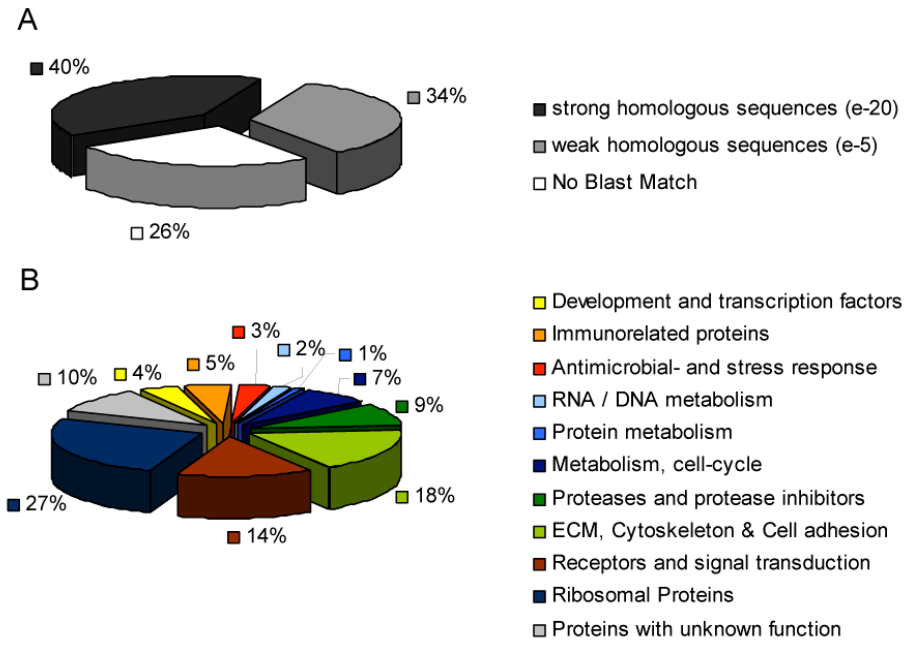


***Figure 25: Gene expression profile for genes differentially regulated in the i-cell lineage**. A) Results of consecutive Blast analysis B) functional annotation of genes expressed in the i-cell lineage.*

These EST sequences were subjected to the analysis pipeline and by that yielded 151 contigs and 163 singletons. The consecutive Blast analysis could identify 40% of the transcripts being homologous to already known proteins, 34% were designated weak homologous sequences and a proportion of 26% of sequences returned no Blast matches (Figure 25A). Among the strong homologous sequences a number of already known nematocyte specific genes was identified. Taking a closer look onto the non-homologous sequences by examining their expression pattern via in situ hybridization (see Figure 26), it was surprising to find nearly all of them expressed in different types and differentiation steps of cnidocytes. The obtained results may provide a good example that novel taxon-specific genes are crucial for the genesis of a taxon-specific structure. To prove their role in this respect, all novel genes identified in this study are currently under in depth investigation (Milde et al., in prep).

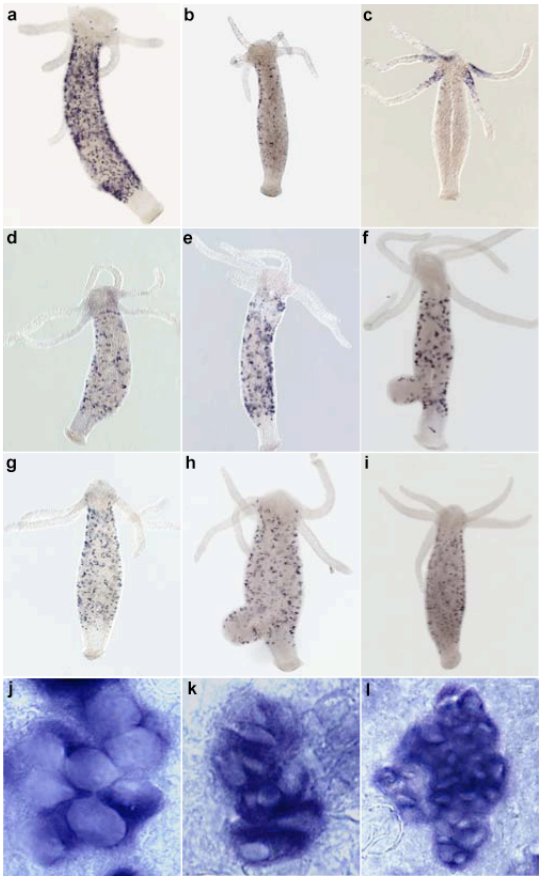


Figure 26: (a-i) Expression of cnidocyte specific genes in adult polyps: *a) CL001, b) CL012, c) CL031, d) CL035, e) CL039, f) CL042, g) CL054, h) CL082 and i) CL092. (j-l) Close-up to different cnidocyte cell types from the in situ hybridisation with CL001: j) stenoteles, k) isorhiza and l) desmonemes. Figure courtesy of S. Milde.*

To complete the EST pipeline, gene ontology annotation of the strong homologous sequences was performed (see Figure 25B). Noteworthy are the relatively high number of transcripts falling into the categories of extra-cellular matrix, cytoskeleton and cell adhesion as well as receptors and signal transduction pointing towards a extensive use of these gene classes in the development of the cnidocyte.

2.4.5 Additional projects

During three other research projects in our laboratory (see also Table 5) the EST analysis pipeline was used to create CAP3 assemblies without subjecting the sequences to subsequent analysis steps.

In an attempt to identify genes involved in encoding taxon-specificity the suppression subtraction hybridization procedure was used to subtract whole transcriptomes of two different *Hydra* species. The resulting putative species-specific cDNA libraries for *Hydra magnipapillata* and *Hydra oligactis* yielded 2126 sequences that were clustered and assembled within the pipeline resulting in 277 / 235 contigs and 209 / 189 singletons, respectively. The genes identified during this approach are currently under investigation and a publication by Khalturin et al. is in preparation.

Another project in the Bosch laboratory studies the genetic basis of symbiosis using *Hydra viridissima* the “green” *Hydra* as model. This species undergoes livelong symbiosis with green algae from the *Chlorella* family (Habetha et al., 2003). The question what genes may be involved in forming and maintaining this symbiotic relationship, led to the construction of a SSH library subtracting symbiotic from non-symbiotic polyps. The obtained 4608 sequences went through the EST pipeline resulting in 1199 contigs and 1041 singletons and are awaiting further analysis

A last study to be mentioned concerning the usage of the EST pipeline so far, is the screening for genes involved in allorecognition in the urochordate *Ciona intestinalis*. Focusing on genes exhibiting high inter- and intra-individual variability SSH was used filtering common transcripts and extracting different/variable ones. The approach resulted in the identification of a highly polymorphic gene vCRL1, which shows structural similarity to vertebrate complement receptors. For further characterization details see Kürn et al., (2007).

2.5 Genomic analysis of a novel, taxon-specific gene-family in *Hydra magnipapillata*

As shown during the above EST analysis, all Blast searches performed so far identified around 35% novel genes. Here I describe one of these novel genes in detail. While many developmental genes have been identified within the cnidarian *Hydra*, little is known about their molecular evolution and their genomic organization. In *Hydra magnipapillata*, one of the genes involved in morphogenesis of the head is *ks1*. The gene was identified in a differential cDNA screening approach focused on genes expressed exclusively in apical/head tissue of *Hydra* (Weinziger et al., 1994). As described by Weinziger et al., (1994), the *ks1* transcript encodes for a 217 amino acid protein (Figure 27) that consists of two highly acidic and three basic domains (Weinziger et al., 1994). *Ks1* is expressed in ectodermal epithelial cells in the upper body column as they enter the tentacle-building zone (Figure 27A). *Ks1* expression continues in the tentacle-base where epithelial cells start to build up a multi-cellular complex, the so-called battery cell, by integrating several different nematocytes. When the battery cell has formed, *ks1*-expression fades off. In previous studies it was shown that the *ks1* gene is regulated by complex interaction of inhibitory factors (Endl et al., 1999) and that loss-of-function polyps generated by dsRNA-mediated interference exhibit defects in head formation indicating that this gene is functionally involved in head development (Lohmann et al., 1999). The concrete function of the *ks1* protein remains unknown. Searching for *ks1* homologous sequences in the available databases from other organisms so far failed to identify a putative counterpart. Even in other cnidarians, such as the anthozoans *Nematostella vectensis* or *Acropora millepora,* a similar gene is not detectable. Thus, *ks1* seems to be a gene restricted to the hydrozoan phylum.

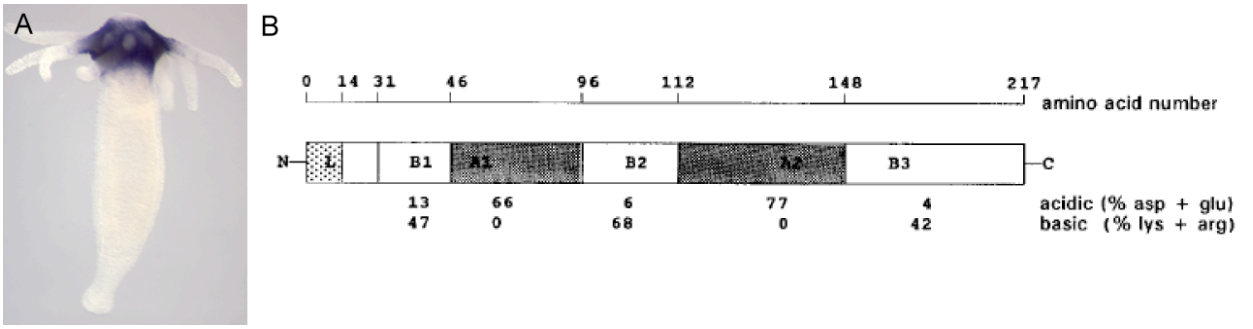


Figure 27: *A) Expression of the ks1 gene in Hydra magnipapillata; B) predicted domain structure of ks1; cross-hatched areas, acidic domains; open boxes, basic domains; stippled, signal peptide; part B taken from Weinziger et al., 1994.*

2.5.1 Different *Hydra* species possess different numbers of *ks1* genes

To get a first impression of how the *ks1* gene is organized on the genomic level, a heterologous Southern blot analysis was conducted on *HindIII* and *XbaI* digested genomic DNA of eight different species of *Hydra*. As heterologous probe a 1,2 kb genomic sequence from the *Hydra magnipapillata Hm_ks1_18A7* gene covering exons 1 to 4 was used for hybridization. As shown in Figure 28A, in *H. magnipapillata*, *H. vulgaris*, *H. vulgaris* AEP and *H. carnea*, several signals were obtained. Since the probe did not contain cutting sites for the used restriction enzymes, these results suggest the presence of multiple *ks1* gene copies in these *Hydra* species.

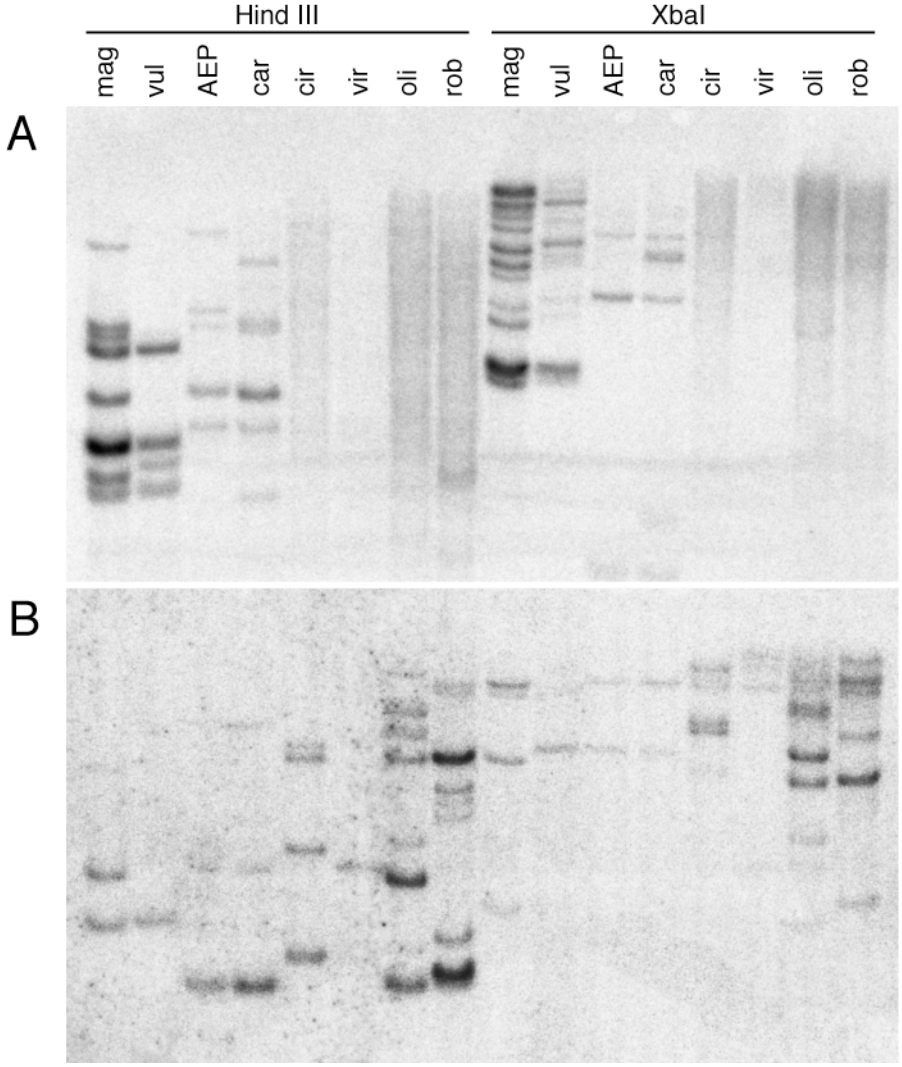


Figure 28: ks1 genes at the genomic level within the genus Hydra. *Southern blots hybridized with (A) the ks1-1 gene probe and (B) the H. magnipapillata ß-actin control probe. Species abbreviations: mag = H. magnipapillata; vul = H. vulgaris; AEP = H. AEP; car = H. carnea; cir = H. circumcincta; vir = H. viridissima; oli = H. oligactis; rob = H. robusta (syn. Pelmatohydra robusta)*

A comparison between the phylogenetically closely related species *H. magnipapillata* and *H. vulgaris*, as well as *H. carnea* and *H. vulgaris* AEP reveals that some of the *ks1* hybridization signals are identical in each species pair respectively, indicating a high degree of sequence similarity. These findings also confirm the results of the phylogenetic analysis described in chapter 2.1 (Hemmrich et al., 2006). Variable numbers of *ks1* hybridization bands in different *Hydra* species may represent differences in the size of the *ks1* gene family of each species. For the remaining four *Hydra* species tested, no *ks1* hybridization signal could be detected neither in *HindIII* nor in *XbaI* restricted DNA. A subsequent hybridization of the same filter with a gene probe for *H. magnipapillata* ß-actin revealed signals in all tested *Hydra* species (see Figure 28B) indicating that DNA was present and restricted properly. Thus, the *ks1* gene, if present at all in *H. circumcincta*, *H. viridissima*, *H. oligactis* and *H. robusta,* must have been significantly diverged in these species.

2.5.2 Characterizing the gene structure of three *ks1* genes in *Hydra magnipapillata* using a BAC library

Discovering extensive complexity of the *ks1* gene family in closely related *Hydra* species was unexpected and surprising and led us to further investigate the genomic background of *ks1* within one selected species, *Hydra magnipapillata*. Lacking a genome project and having only a small set of EST sequences for *Hydra* at that time point, a *Hydra magnipapillata* bacterial artificial chromosome (BAC) library was constructed and screened (see Methods for details) using the 1,2 kb gene probe from *Hm_ks1_18A7* to identify and characterize the different *ks1* gene copies. Among 55.000 checked clones 19 *ks1* positive clones could be identified. All clones were analyzed via restriction digestion fingerprinting and subsequent radioactive hybridization (see Methods for details). Since these analyses revealed that BAC 18A7 and BAC 10A18 comprise small genomic inserts of less than 10 kb enabling for rapid complete sequencing, and that BAC 87C19 contains an insert around 30 kb facilitating isolation of the complete 5'-regulatory region, these three *ks1* positive BAC clones were chosen for a short-term detailed characterization.

To assess the gene position and the exon/intron structure of the three selected *ks1* genes, BAC 18A7 (9049 bp; gb|AJ829761) and BAC 10A18 (4032 bp; gb|AJ841298) were sequenced completely by stepwise primer walking. For BAC 87C19 (3322 bp; gb|AM161050) only the *ks1* gene-containing region of was sequenced. The *ks1* genes on BAC 18A7, 10A18 and 87C19 are referred to as *Hm_ks1-18A7*, *Hm_ks1-10A18* and *Hm_ks1-87C19* respectively. As depicted in Figure 29, all three genes exhibit the same structure with 4 exons interrupted by three introns. A high degree of similarity was observed in exon/intron boundary positions as well as in exon sizes. The only exception was found in exon 2 of gene *Hm_ks1-18A7* where 48 bp (16 aa) compared to the other genes were missing. The splice donor and acceptor sites of all introns within the different genes confirmed to the established (Senapathy et al., 1990) consensus GT at the 5'-end and AG at the 3'-end of the intron. Alignment of the DNA sequence of the three *ks1* genes reveals 86% identity at the nucleotide level in the complete overlapping region and 96% identity in the coding region.

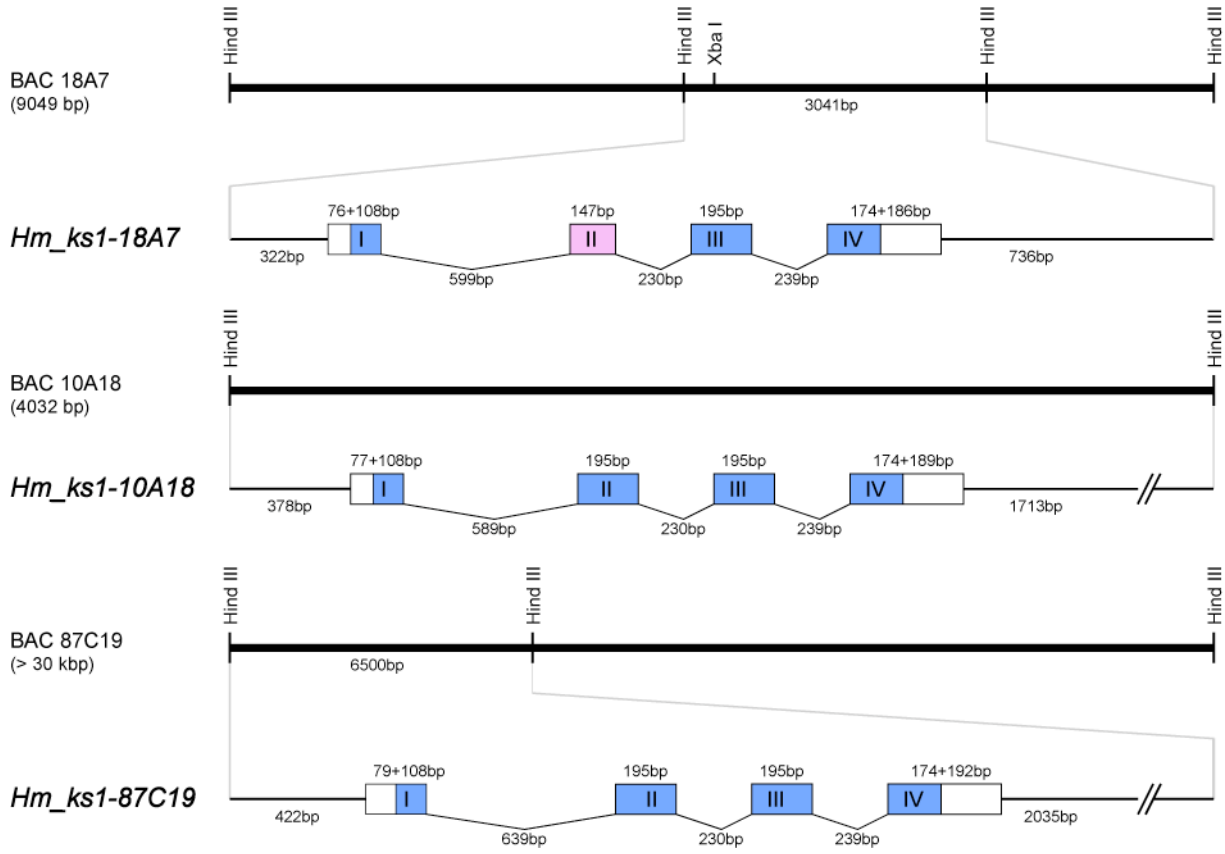


Figure 29: Genomic organization of ks1 genes in H.magnipapillata on selected BAC clones. *Sizes of exons (blue/magenta) and introns or flanking genomic sequences (lines) are indicated in bp. White boxes highlight putative untranslated regions (UTR).*

As shown in Figure 30, comparison of the predicted peptide sequence for the three *ks1* genes revealed three different types of proteins. Two long versions of 224 aa comprising 12 conserved cysteine residues within the acidic domains of the protein and a shortened version of 208 aa missing a stretch with 2 conserved cysteines within the first acidic domain. All other previously described features (Weinziger et al., 1994) of the predicted protein sequences remain unchanged.

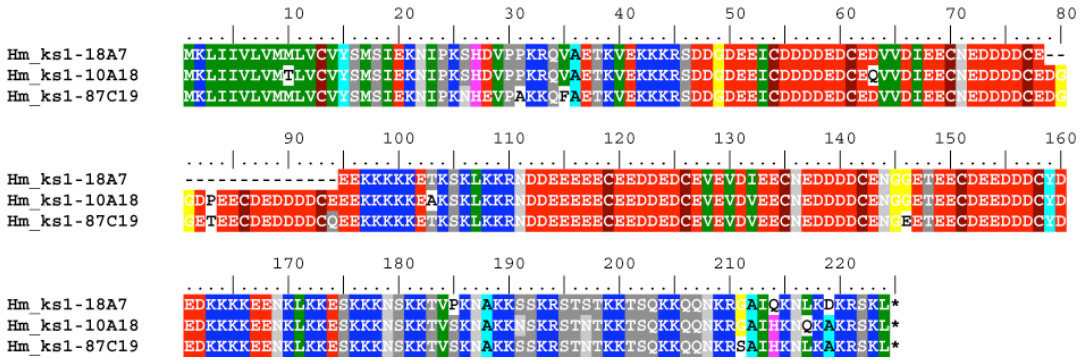


Figure 30: ClustalW multiple alignment of the predicted ks1 peptide sequences deduced from the genes included on the analyzed BAC clones. *Conserved amino acid residues are indicated in colours, white shading of residues highlights variable sites.*

2.5.3 Screening the *Hydra magnipapillata* EST project for *ks1* gene family members

The *Hydra* EST sequencing project at Washington University's Genome sequencing center provided an additional possibility to get information about the expression of different *ks1* genes in *Hydra magnipapillata*. Searching the 174.000 *Hydra* ESTs for *ks1* coding sequences identified 209 single EST sequences covering at least parts of the *ks1* gene. These *ks1* positive ESTs were assembled using the previously described EST analysis pipeline (see also Figure 13, chapter 2.2.3) on the "Compagen" platform with adjusted parameters. To exclusively assemble identical gene sequences and not to mix genes because of their repetitive structure, the identity threshold for the assembly was set to 99%. Of the initial sequence set 192 sequences fell into clusters, resulting in 5 different contig consensus sequences. The remaining 17 sequences could not be grouped into clusters and might thus represent separate transcripts of other *ks1* gene copies. For further analysis only the consensus contigs were used as they encode for the full-length protein sequence. The resulting predicted protein sequences named *Hm_ks1_A* to *Hm_ks1_E* are aligned in Figure 31. Compared to the three predicted

proteins from sequencing the BAC clones (Figure 30), five different predicted *ks1* protein versions could be identified within the ESTs. Similar to the BAC derived sequences, these predicted proteins differ in the amount of conserved cysteine pairs within the acidic domains (red). In addition, *Hm_ks1_D* seems to have lost the small lysine rich basic domain separating the two acidic ones.

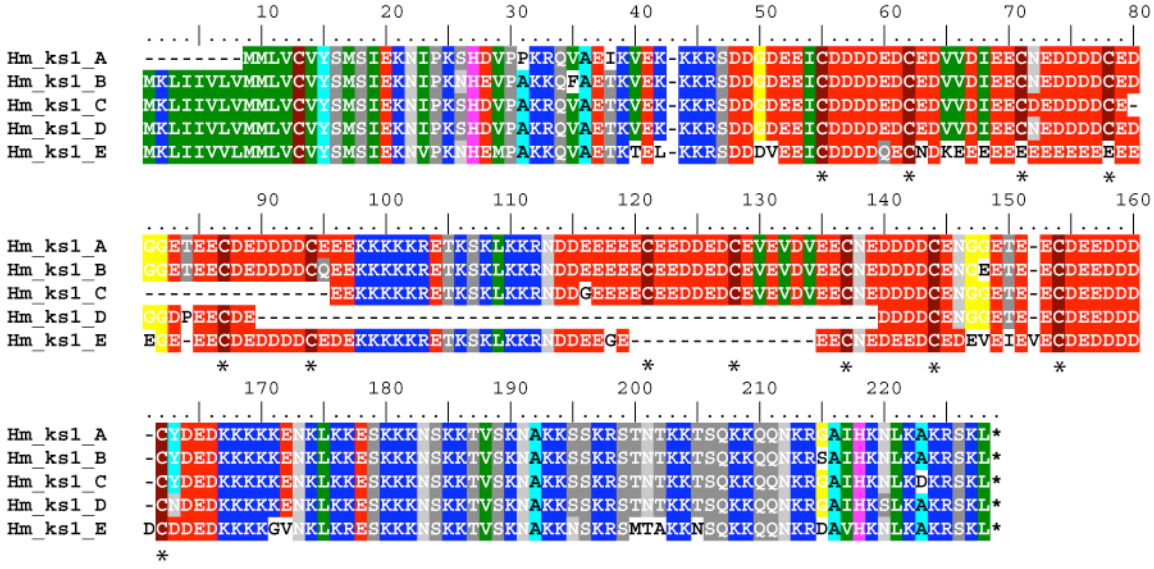


Figure 31: ClustalW multiple alignment of assembled predicted ks1 protein sequences derived from the ESTs. *White shading indicates variable amino acid residues; asterisks mark the position of conserved cysteine residues.*

Two of the previously identified BAC sequences were found to be within this EST derived dataset. Whereas *Hm_ks1_18A7* corresponds to the EST derived variant *Hm_ks1_E*, the corresponding counterpart to *Hm_ks1_87C19* is EST variant *Hm_ks1_B*. Gene *Hm_ks1_10A18* seems to encode for an additional version resulting in a total of 6 different predicted *ks1* protein sequences. The apparent domain composition together with the presence or absence of conserved cysteine residues might point to different protein 3D topologies.

2.5.4 Screening the *Hydra magnipapillata* genome sequence for the *ks1* gene family

During 2006, the genome of *Hydra magnipapillata* was sequenced in a whole genome shotgun approach at the J. Craig Venter Institute (JCVI) in Rockville, USA and the resulting ~12 million single genomic reads recently got assembled at the Institute of genomic research in Rockville (TIGR) and in a second approach by the Joint Genome Institute (JGI, Berkeley). Both the TIGR and the JGI-assembly are not yet publicly available but they were made accessible for trial usage to the *Hydra* community. This offered the possibility to get new and deeper insights into the *ks1* gene-family situation.

Both genome assemblies were screened by TBlastN searches for *ks1* related sequences. Whereas the TIGR assembly exhibited 19 possible loci for *ks1* encoding genes, screening the JGI assembly identified only 5 putative loci. A direct comparison revealed that the 5 predicted loci from JGI were also within the results of the TIGR dataset. Thus, all further analyses were carried out on the basis of the TIGR genome assembly. All predicted loci were carefully hand checked and if possible complete or partial *ks1* genes were annotated on the corresponding genomic scaffold. This procedure identified 14 loci encoding a complete and coding *ks1* gene, two loci coded only for parts of the gene sequence as they were situated at the end of a scaffold and three loci exhibited disrupted or highly diverged *ks1* related sequence. Two loci were found to encode for identical *ks1* genes and further analysis revealed that also the complete scaffolds share 98% identity on nucleotide level (data not shown) and thus most probably represent alleles. To better distinguish the different loci all intact genes were assigned *Hm_ks1_1* to *13*, partial genes were named *ks1_partial_1* and *2,* and diverged loci were assigned *ks1_diverged_1* to *3* (see also Table 6). To check which of the 13 identified genes are expressed, all genomic *ks1* coding sequences were compared to the available *Hydra magnipapillata* EST collection.

Gene name	Scaffold #	gene length ATG–STOP (bp)	length of predicted peptide (aa)	conserved cysteine residues	covered by ESTs	corresponding BAC clone
Hm_ks1_1	1101284898227	1714	224	12	yes / Hm_ks1_A	-
Hm_ks1_2	1101284911599	1730	224	12	yes	*Hm_ks1_10A18*
Hm_ks1_3	1101284871475	1780	224	12	yes / Hm_ks1_B	*Hm_ks1_87C19*
Hm_ks1_4	1101284937399	1723	219	12	no	-
Hm_ks1_5	1101284937399	1663	227	12	no	-
Hm_ks1_6	1101284937399	1780	224	12	yes	-
Hm_ks1_7	1101284939199	1691	208	9	yes	-
	1101284922152	1691	208	9	yes	-
Hm_ks1_8	1101284904964	1687	203	9	yes / Hm_ks1_C	-
Hm_ks1_9	1101284920977	1692	208	9	yes / Hm_ks1_E	*Hm_ks1_18A7*
Hm_ks1_10	1101284882688	1692	203	9	yes	-
Hm_ks1_11	1101284900312	1596	208	8	yes	-
Hm_ks1_12	1101284872160	1587	202	8	no	-
Hm_ks1_13	1101284935645	1359	170	8	yes / Hm_ks1_D	-
ks1_partial_1	1101284920977	(partial) 174	-	-	-	-
ks1_partial_2	1101284935645	(partial) 108	-	-	-	-
ks1_diverged_1	1101284871475	incomplete	2 internal stop codons			
ks1_diverged_2	1101284871475	incomplete	3 internal stop codons			
ks1_diverged_3	1101284936749	incomplete	very diverged fragments originating from 2 clustered genes			

Table 6: Identified putative ks1 gene encoding loci *in the Hydra magnipapillata genome assembly.*

Ten out of thirteen genes were found being covered by cDNA sequence at a minimum of 98% identity on nucleotide level. Also the 5 different *ks1* versions identified during the EST analysis were found among the genes annotated from their corresponding genomic scaffolds. In addition, the previously sequenced BAC clones could be clearly mapped to a corresponding genomic scaffold using the particular BAC-end sequences as query (see Table 6).

For all annotated *ks1*-encoding loci, gene models were generated showing the relative gene structure and the precise exon/intron sizes (see Figure 32). Except of the *Hm_ks1_13* gene that has apparently lost the second intron and fused exons two and three, all other predicted gene models exhibit a four-exon gene architecture with a high degree of conservation in the sizes for the first and the last exon (blue) and variable sizes in the second and third exon (rose). Interestingly *Hm_ks1* genes *1-3* and *6* share identical exon sizes whereas all other genes differ from each other in at least one of the exons. Inferred from the gene structure *Hm_ks1_10*, *11* and *12*, together with the already mentioned *Hm_ks1_13* seem to be the most diverged genes. The identified "partial" *ks1* genes encode only for one exon generally consistent in the corresponding

sizes but from present data it is unclear, whether they belong to a complete gene or remain fragmented. Taking a closer look on the diverged *ks1* encoding loci in *ks1_diverged_1* and *2*, the relative gene structure is still present but due to single nucleotide polymorphisms (SNPs) inserted stop-codons in the second or third exon respectively the continuous open reading frame is interrupted. The third identified diverged *ks1*-encoding locus appeared to be highly derived so that no obvious gene structure could be deduced.

To get insights into the conservation and/or divergence of the 13 complete coding genes, mutiple sequence alignments were generated using the MAFFT program (Katoh et al., 2002) for large sequences (for alignment see Appendix, Figure 7). A surprisingly high degree of conservation could be observed within the coding region (96% on nucleotide level) but also in the smaller intronic regions (92% on nuc. level). The only more variable region was found in the first intron (84% on nuc. level), which is usually around 600 bp in size. Of course some of the more derived gene variants show deletions or insertions like a 325 bp deletion in gene *Hm_ks1_13* but the overall conservation remains unexpectedly high.

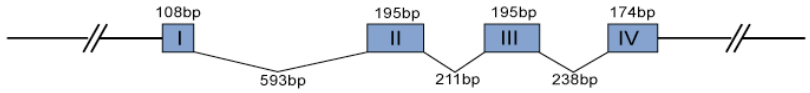
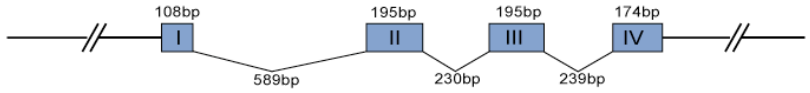
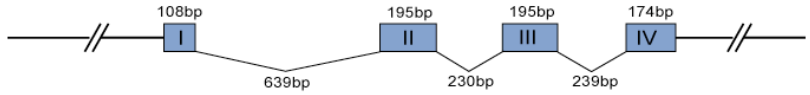
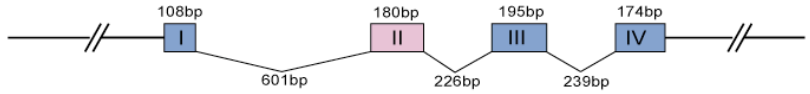
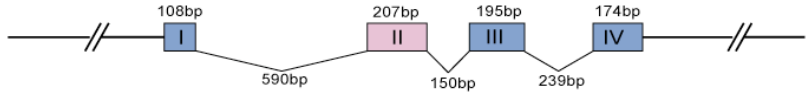
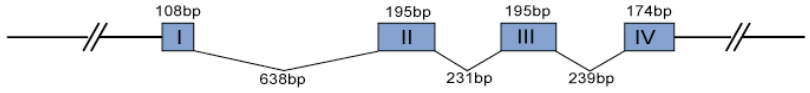
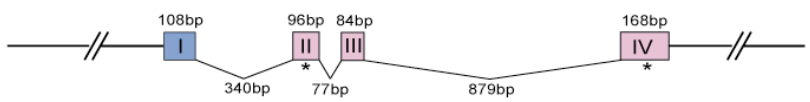
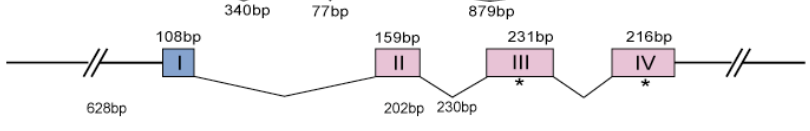
Gene name	EST coverage	BAC coverage	Gene model
Hm_ks1_1	+		
Hm_ks1_2	+	+	
Hm_ks1_3	+	+	
Hm_ks1_4			
Hm_ks1_5			
Hm_ks1_6	+		
Hm_ks1_7	+		
Hm_ks1_8	+		
Hm_ks1_9	+	+	
Hm_ks1_10	+		
Hm_ks1_11	+		
Hm_ks1_12			
Hm_ks1_13	+		
ks1_partial_1			
ks1_partial_2			
ks1_diverged_1			
ks1_diverged_2			

Figure 32: Gene models for the identified ks1 genomic loci *within the Hydra magnipapillata genome.*

To infer the evolutionary relationships of the 13 different coding *ks1* genes, multiple alignments were generated and application of a so-called Neighbor-Joining (NJ) phylogenetic distance method, based on the aligned *ks1* genes, resulted in a clear separation of four distinct groups of sequences (see Figure 33). All nodes are supported by high bootstrap values and a similar tree could be obtained by maximum parsimony (MP) analyses as used for the *Hydra* phylogeny project (see Methods 7.1.4.4). Five different *ks1* gene copies were grouped together in group A with *Hm_ks1_1* being the most basal one. Within group B *Hm_ks1_4* seems to be highly diverged in comparison to the more closely related genes *2* and *5*. Only two gene copies *Hm_ks1_3* and *6* fall into group C. The most distant and diverged sequences from genes *Hm_ks1_11* and *12* resolved in group D. With regard to the previously shown gene models it was surprising that the (from exon/intron structure) apparently derived genes *Hm_ks1_10* and *13* resolved in group A.

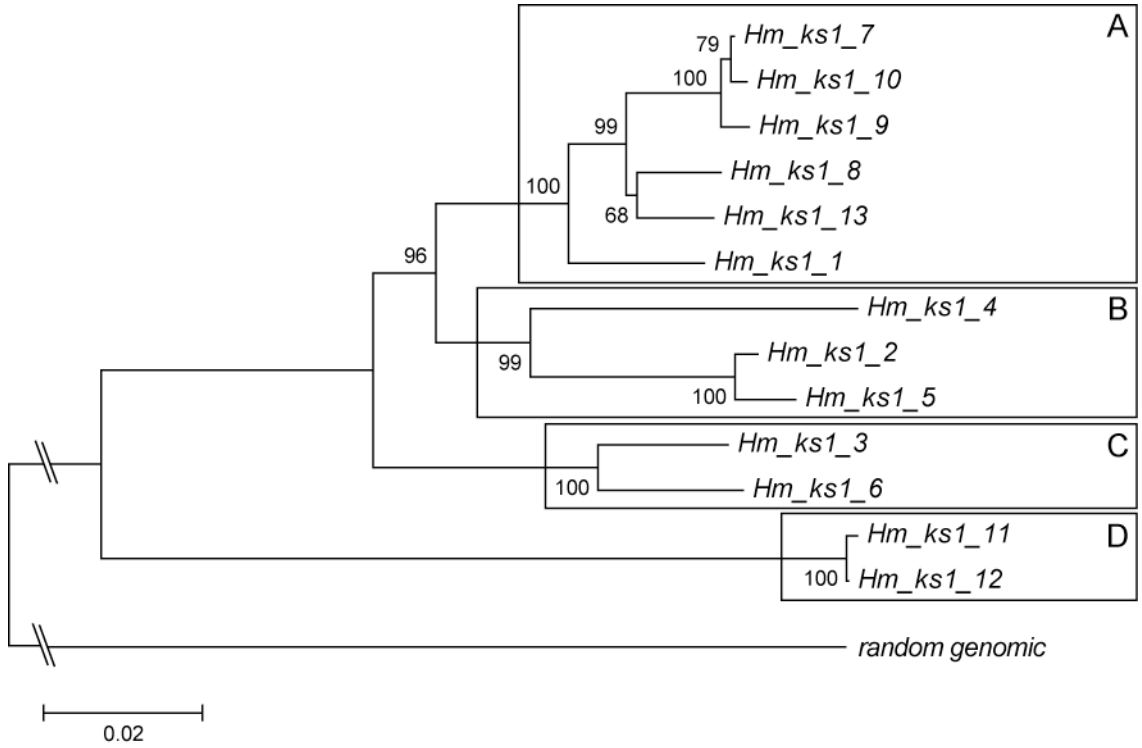


Figure 33: Neighbour-Joining (NJ) tree of complete ks1 gene sequences. *Bootstrap values for NJ criteria are depicted at the corresponding nodes; branch lengths are scaled to the expected number of substitutions (0.02 substitutions per site). As outgroup served a random genomic H. magnipapillata sequence.*

Unfortunately, lacking *ks1* sequence data of more basal *Hydra* species like *Hydra carnea* or *Hydra vulgaris* AEP*,* the phylogenetic tree in Figure 33 does not represent a complete phylogeny of *ks1* gene family. It only displays the affinities of the 13 members

within *Hydra magnipapillata*. But it is yet noteworthy that together with the most basal *Hm_ks1_1* from group A, most genes in groups B and C exhibit similar exon sizes (see Figure 32) and thus code for similar proteins. Taken into consideration that group D genes might not represent the ancestral state because of their obviously high divergence (Figure 32), the conserved architecture of genes *1-3* and *6* instead may represent the evolutionary ancestral state from which the other variants may have evolved.

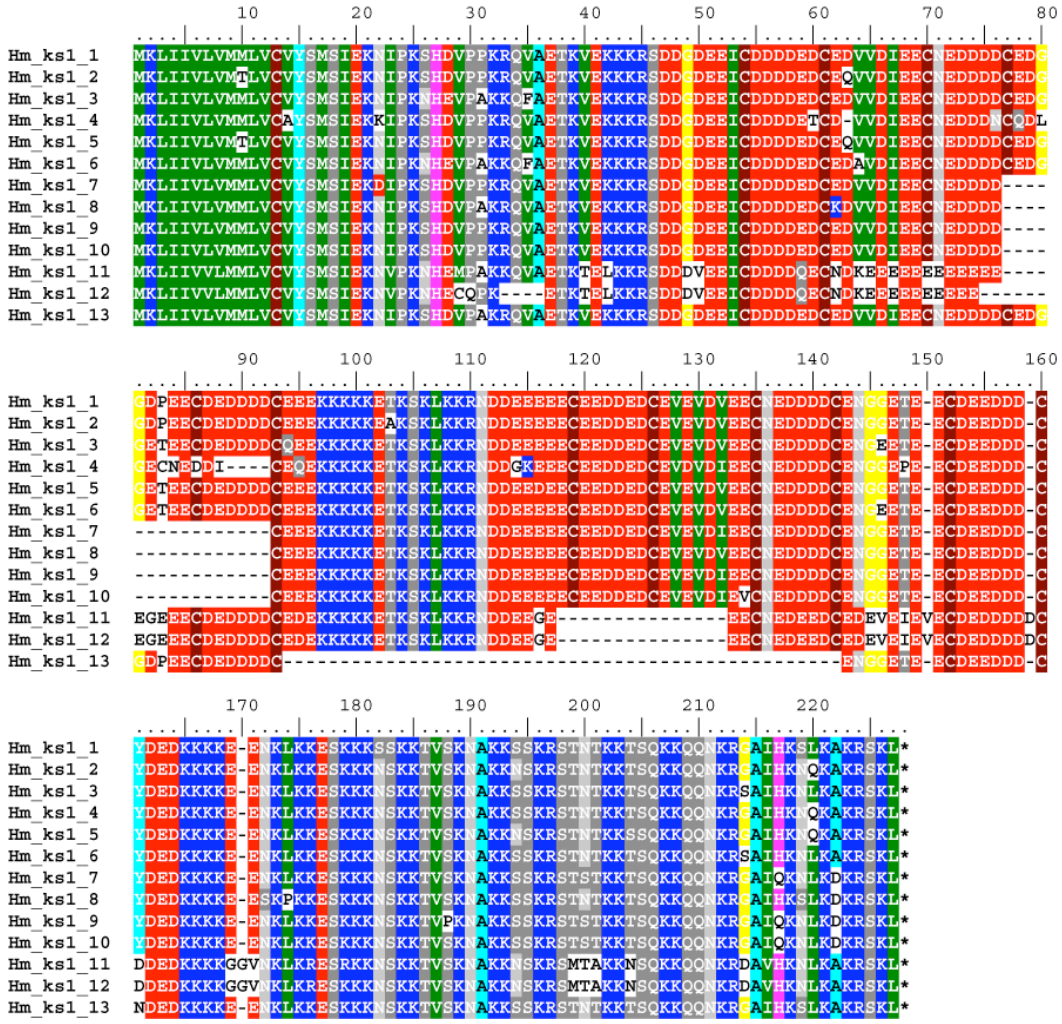


Figure 34: ClustalW multiple alignment of predicted peptide sequences *derived from 13 full-length ks1 gene models.*

To further characterize the 13 identified *ks1* genes, the corresponding predicted peptide sequences were analyzed according to the already known domain features (see Figure 27, chapter 2.5). From the multiple alignment (Figure 34) five different possible types of ks1 proteins could be identified. 1) Genes *Hm_ks1_1- 3*, *5* and *6* encode for a long type (224-227 aa) comprising the classical previously described domain structure. 2) The

predicted proteins of genes 7-10 (203-208 aa) share a deletion within the first acidic domain leading to the loss of 17 amino acid residues. 3) The third type (genes 11+12) exhibits a similar deletion but here the missing residues are in the second acidic domain resulting in a loss of 15 aa. 4) Hm_ks1_4 seems to have been derived within the first acidic domain and thus forms its own derived protein type. 5) Finally, Hm_ks1_13 (170 aa), encoded by only 3 exons, shows the largest deletion, which results in the loss of the complete 2nd basic and large parts of the 2nd acidic domain.

2.5.5 *Ks1* genes are clustered within the *Hydra magnipapillata* genome

Some identified *ks1* encoding loci fell onto the same genomic scaffold and, thus, represent members of multiple *ks1* gene clusters within the *Hydra magnipapillata* genome. As depicted in Figure 35, four different clusters have been identified with two or three involved *ks1* genes. The distances between the single genes range from 7.9 to more than 20 kbp and their orientation differs in each cluster. Interestingly, in both 3-gene-clusters only one gene (*Hm_ks1_3* and *Hm_ks1_6*) is covered by ESTs whereas the other are either diverged or not expressed. Moreover, from the NJ inference it was already shown that *Hm_ks1_3* and *Hm_ks1_6* are most closely related and might represent the result of a gene duplication event whereupon the other members in the cluster seem to be independent duplications. A possible duplication of the whole cluster seems thus to be unlikely.

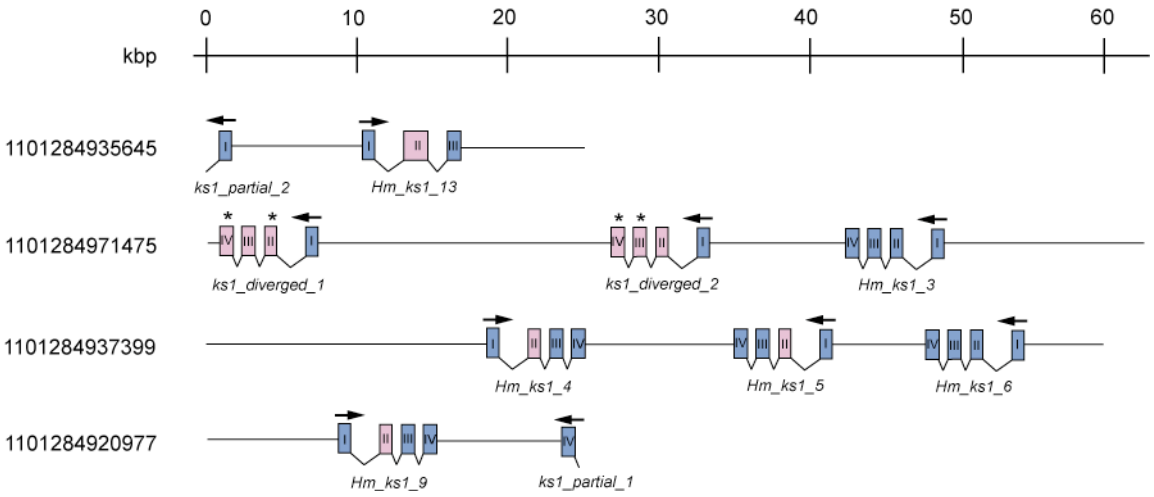


Figure 35: Schematic overview of clustered ks1 genes within the Hydra magnipapillata genome scaffolds*. Black arrows indicate the orientation of the corresponding gene; asterisks mark the positions of additional STOP codons in non-functional duplicates.*

To get an impression of how the *ks1* genes and clusters are distributed within the *Hydra* genome, fluorescent in situ hybridization (FISH) on mitotic chromosome plates of *Hydra magnipapillata* using the 1,2 kb *Hm_ks1_18A7* probe were carried out at different hybridization and washing stringencies. As shown in Figure 36 A-D, a strong hybridization signal was detected on three chromosome pairs using stringent conditions (3 x 5 min. 0,1 x SSC at 61°C). With reduced stringency (3 x 5 min. 0,5 x SSC at 43°C) several additional signals appeared on other chromosome pairs (Figure 36 E-F).

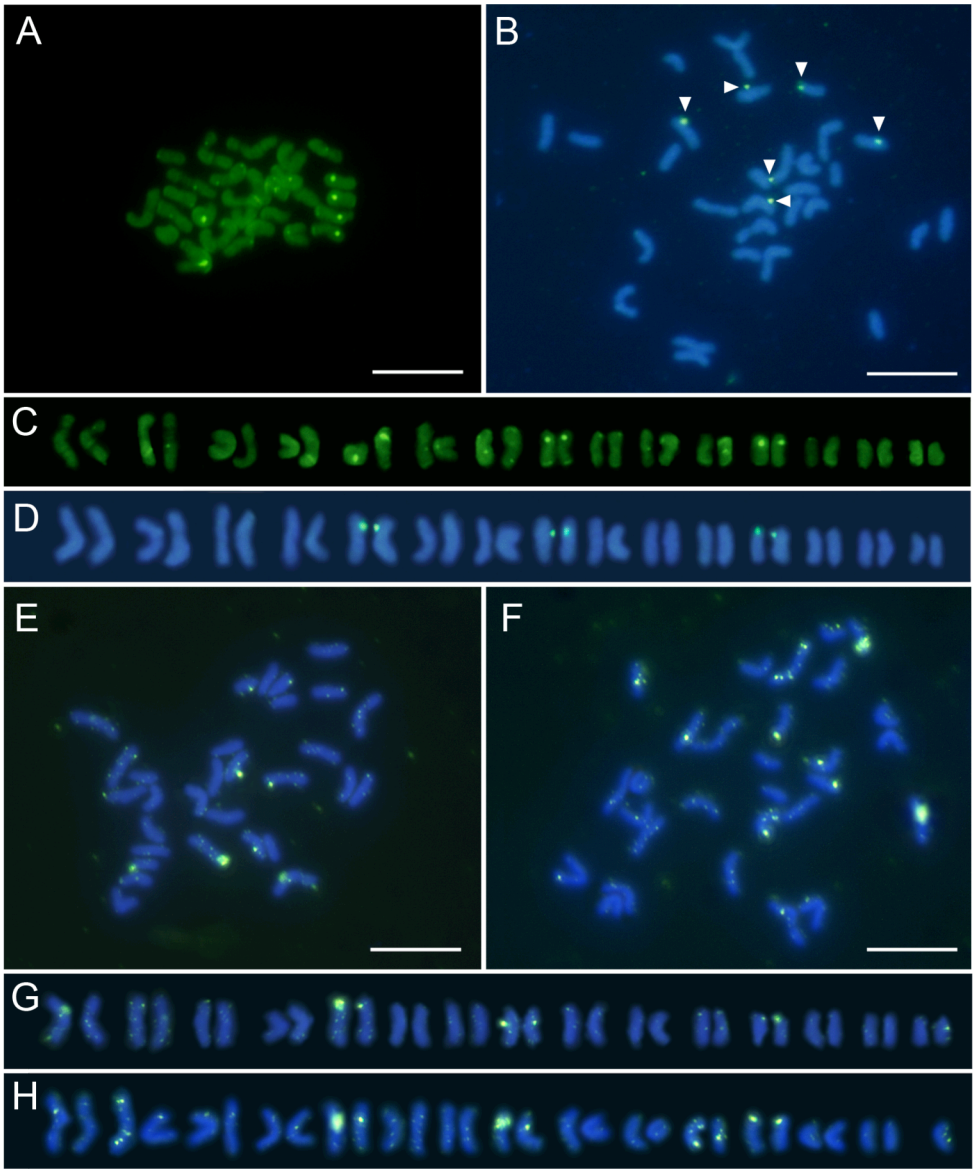


Figure 36: Fluorescence in situ hybridization (FISH) of ks1 on Hydra magnipapillata mitotic metaphase chromosomes. *(A) FISH using a 1,2 kb ks1-1 probe without DAPI counterstaining; (B,E,F) FISH using the ks1-1 probe with DAPI counterstaining. Arrows mark the ks1 positive chromosomes. (C-D) and (G-H) Karyograms of H. magnipapillata metaphase chromosomes. (A-D) Stringent hybridization and washing conditions. (E-H) Low stringency hybridization and washing. Bars indicate 10µm. Pictures courtesey of Dr. B. Anokhin.*

A possible explanation for the strong fluorescent signals present under both stringent and less stringent conditions might be the hybridization of a *ks1* gene cluster in contrast to single gene copies that hybridize only under reduced stringency. Taken together, these findings complement the previously described complexity of the *ks1* gene family.

2.5.6 Insights into the regulation of *ks1* genes by promoter comparison

Since the regulation of *ks1* genes may provide a model for understanding how positional signals control the differentiation of epithelial cells, a further motivation was to get insight into the conservation of regulatory elements in their promoters. For nine of the 13 identified *ks1* genes sufficient genomic 5'-flanking sequence was available. To generate conservation profiles and to infer conserved cis regulatory elements, comparative "phylogenetic footprinting" methods were applied based on the ConSite (Sandelin et al., 2004) and Vista (Frazer et al., 2004) online platforms.

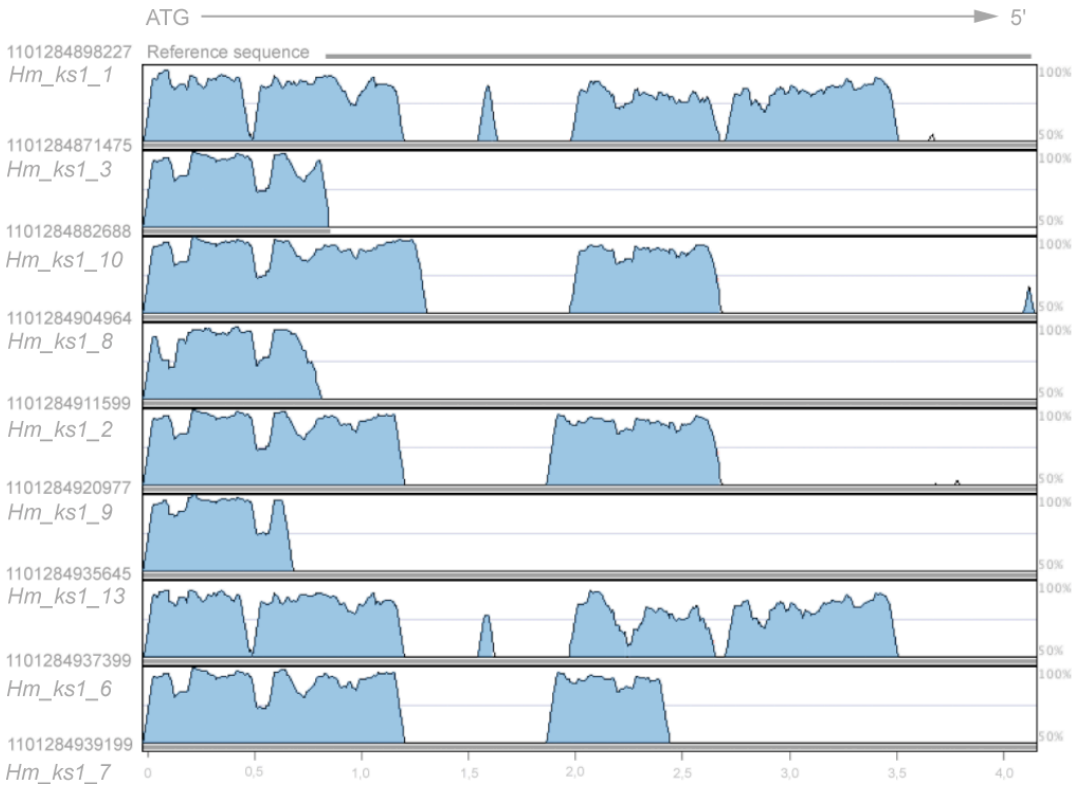


Figure 37: mVista conservation profile for nine ks1 promoters. *The profile covers 4,2 kb of 5'-flanking region; the start codon of the corresponding genes is at position 0 bp; depicted in blue are regions where sequence similarity of each sequence lies above 50% compared to the reference sequence. Sequence numbers correspond to the accession numbers of the genomic scaffolds. Note that for sequence Nr. 1101284882688 only 890 bp of promotor sequence were available for comparison.*

Comparing 4,2 kbp of the selected *ks1* promoters (Figure 37) revealed a high degree of conservation in the proximal 700 bp for all analyzed genes. Furthermore, for six genes this conserved region covers even 1200 bp. In addition to that, some conserved blocks of sequence could be identified in the upstream 4,2 kbp 5'-flanking region of the particular gene. Prediction of putative transcription factor binding sites (TFBS) using ConSite with thresholds of 96% sequence conservation and 90% TF score, resulted in the identification of several conserved binding elements within the most proximal 1 kb (see Figure 38). In all analyzed promoters a TRE-binding motif, a 27bp sequence previously characterized as Hyko-11 site (Endl et al., 1999) and a GATA-binding motif were found to be highly conserved in sequence and position on the first 300 bp upstream of the *ks1* coding sequence. Both TRE- and Hyko-11 motif have been experimentally verified during gel shift experiments (Endl et al., 1999). The short TRE sequence serves as the binding site for transcription factor AP1 and for vertebrate genes numerous experiments have shown its requirement for TPA induction (Angel et al., 1987). For *Hydra* TPA was shown to induce ectopically expression of *ks1* by inhibiting DNA protein interactions at the *ks1* promoter (Endl et al., 1999). The Hyko-11 element has been identified as a binding site for a putative *Hydra* specific transcription factor. Among all other TFBS that appeared to be highly conserved in sequence but variable in position (Figure 38), a FREAC3 binding site, a predicted Ubx binding site, a TCF11 binding element as well as a binding motif for members of the broad-complex (BC) family could be identified. The high degree of conservation for these binding sites implies that their potential for mutations is constricted by their function. Thus, these sequences may contain targets essential for head-specific gene expression.

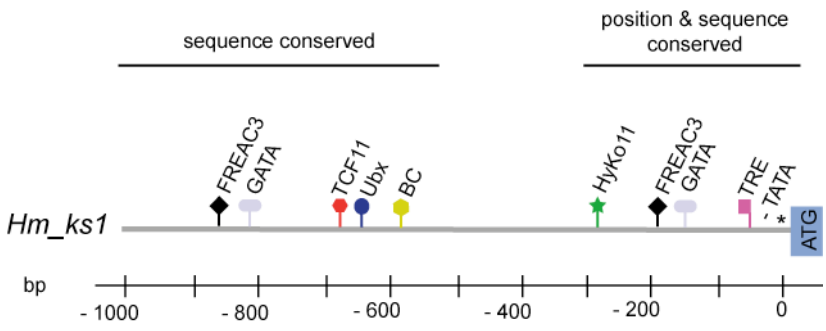


Figure 38: Schematic overview of potential regulatory target sites on the ks1 promoter. *The transcription initiation site (asterisk) is located 42 bp upstream of the translation initiation codon.*

2.5.7 Concluding remarks

The results presented here show that the previously identified head specific gene *ks1* is a member of a large gene family with a complex distribution in different closely related species of the genus *Hydra*. Whereas in basal *Hydra* species no close *ks1* homolog could be identified in heterologous Southern blot experiments, the four most advanced members show different hybridization patterns, indicating different complexity in their *ks1* gene families (Figure 39).

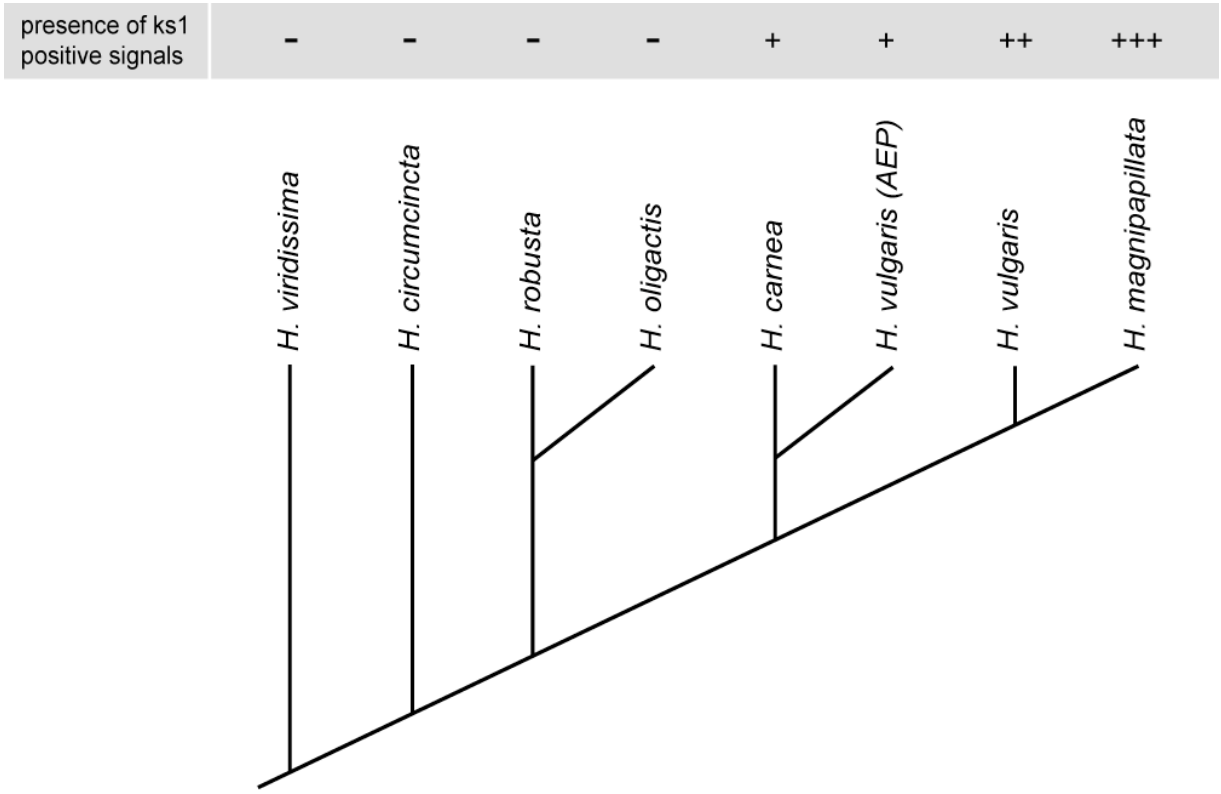


Figure 39: Schematic overview of the distribution of ks1 genes across the genus Hydra. *(+) and (-) indicate the abundance of ks1 positive hybridization signals on the heterologous Southern blot (see Figure 28).*

The most complex situation could be determined in *Hydra magnipapillata*, where the *ks1* gene family consists of 13 members of whom 10 are represented in the *Hydra* EST collection. Comparison of gene structures and sequences revealed high degrees of conservation within the first and the fourth exon, whereas exons two and three exhibit several different modifications. Phylogenetic analyses indicate four groups of related *ks1* genes. Localization of the genes on assembled genome data as well as

fluorescence in situ hybridization suggest that some members of the *ks1* gene family are clustered within the *Hydra magnipapillata* genome. Translated gene sequences propose at least five variants of different *ks1* proteins with alterations in a domain like manner. Comparison of 9 available *ks1* promoters resulted in the identification of a conserved core region including conserved DNA binding sites.

Taken together, the large species-specific complexity of the *ks1* gene family in addition to the complex distribution of *ks1* within the genus *Hydra*, provide a good model to study the evolution of a taxon- specific (in this particular case, genus-specific) gene.

3 DISCUSSION

3.1 Towards a full molecularization of cnidarian models systems

As sister group of the Bilateria, Cnidaria occupy a key place in the evolutionary tree and make it possible to infer the genetic complexity of ancestral metazoans (see Figure 1, chapter 1.3). The lack of genetic background information, however, made it difficult so far to use cnidarians in comparative genetics and genomics. Meanwhile several of these basal metazoan animals are on the way to become fully molecularized, modern model systems.

3.1.1 A molecular phylogeny for *Hydra*

While a substantial number of molecular and genetic resources are already available for *Hydra* and functional tests have been developed and can be applied, several other resources are still missing. In the aim to complement already available resources for the first time a molecular phylogeny for *Hydra* has been generated (Hemmrich et al., 2006) using different marker genes from nuclear (18S, 28S) and mitochondrial (16S, CO1) DNA. Not to unravel the phylogeny of the complete genus *Hydra* that would have required extensive worldwide sampling, the analyses were focused on species and strains that are most commonly used in current research. The obtained grouping of the different species almost completely conformed what was already proposed from morphologically based taxonomy (Campbell, 1983; Holstein, 1995). The only symbiotic species *Hydra vridissima* resolved as most basal, whereas species of the previously postulated “vulgaris group” represent the most derived state. However, the finding that *Hydra vulgaris* AEP, previously described as *Hydra vulgaris* strain (Martin et al., 1997), resolved as close relative to *Hydra carnea* was surprising and unexpected and demonstrated that morphological features failed so far to correctly group that species. This finding is of particular importance as *Hydra vulgaris* AEP is the strain used to generate transgenic animal lines whereas most genomic information is available for *Hydra magnipapillata*. The high degree of similarity of the trees obtained using three different phylogenetic methods – maximum likelihood, maximum parsimony and Bayesian inference – strongly suggests that the resulting phylogenetic estimates are robust and may serve as solid fundament for upcoming comparative approaches.

3.1.2 Computational resources for cnidarian comparative genetics and genomics

The sequencing of several basal metazoan genomes and large EST collections for even a larger number of lower metazoan organisms offers a wide range of possibilities to conduct comparative genetics or genomics approaches. For the sea anemone *Nematostella vectensis* two different genome assemblies became available during the last year. One assembly is completely accessible through a graphical genome browser at the Joint Genome Institute (JGI) in Berkley, whereas the other is at least Blast searchable on the *Nematostella* focused online platform StellaBase (Sullivan et al., 2006). For *Hydra,* the only public resource till yet is a website (www.hydrabase.org) that attempts to govern the sequence data from the *Hydra magnipapillata* EST project. Also temporary access to a preliminary version of a *Hydra* genome browser at JGI proved to be of limited use as the supplied genome assembly was of low quality. A second genome assembly performed at the institute of genomic research (TIGR) provided much better quality but is not publicly available. Thus, most of the available molecular resources for *Hydra* remained collections of raw data deposited in public repository databases that provided limited analytical possibilities.

The development and construction of “Compagen” for the first time provides an up to date comprehensive analytical resource for the application of bioinformatics methods not only to *Hydra* data but also to all other publicly available cnidarian datasets. In addition to that, vast amounts of data provided on “Compagen” are already processed data, such as assembled EST sequences or predicted peptide collections. Moreover, for comparative purposes also data for lower as well as higher metazoan animals are available. The platform can be used for various types of sequence analyses starting from a single DNA or protein sequence up to high-throughput analyses including thousands of sequences. As “Compagen” proved to be a valuable resource in several different projects and collaborations presented here and elsewhere (Augustin et al., 2006; Genikhovich et al., 2006; Kürn et al., 2007; Miller et al., 2007), further development in the future will add not yet implemented but important features like an internal sequence retrieval system and the possibility of online access to all information stored within the “Compagen” database.

Taken together, the newly available molecular and computational resources presented here facilitate research for all cnidarian models and make *Hydra* the most advanced and the farthest molecularized model system within Cnidaria.

3.2 The ancestral genetic toolkit – what we can learn from basal animals

One of the most striking findings in the last few years in comparative genetics and genomics was that the link between morphological complexity and gene number is illusive (Technau et al., 2005). The idea of a simple ancestral genetic toolkit based on the "low" complexity of ancestral animals turned into a contradiction as soon as scientists started to look into their genomes (Kortschak et al., 2003). The common ancestor of cnidarians and higher animals was surprisingly complex on the genetic level and it quickly became clear that the reduced complexity in the genomes of *Drosophila* and *C. elegans* was due to a high level of derivation (Kortschak et al., 2003). To unravel the real ancestral genetic toolkit, it is thus necessary to analyze the genomes of lower metazoan animals and maybe even of their unicellular predecessors.

Today the availability of genome and EST data for two different cnidarians, the sea anemone *Nematostella vectensis* and the freshwater polyp *Hydra,* situated at different phylogenetic positions within the cnidarian phylum, offer new perspectives on the evolution of genes, pathways and developmental mechanisms. In a first screening to unravel the cnidarian repertoire of genes involved in the innate immune systems of higher invertebrate and vertebrate animals, the idea of a genetically complex common ancestor could be strongly supported (Miller et al., 2007). A large variety of genes related to immunity were identified (see chapter 2.3). The Toll/TLR, MyD88 and IL-1R gene families were shown to be distinct before the divergence of the Cnidaria from the Bilateria. Recent findings in sponges suggest that the Toll/TLR pathway even precedes the Porifera/Eumetazoa split (Wiens et al., 2007). Finding genes that code for proteins with the same domain structure as the IL-1R in *Nematostella* indicates that this receptor type even predates chordate origins and that its original ligands may not have been interleukins. In addition, identification of components from the complement system including C3 and multiple MAC/PF proteins suggest the presence of a prototypic effector pathway in these basal metazoans. Another implication on genome evolution

came from the direct comparison of the *Nematostella* and *Hydra* genomes, highlighting the likely extent of gene loss and sequence divergence in the latter. Major components of the Toll signaling cascade such as the Toll-receptor and the nuclear factor NF-kB have been lost or diverged beyond recognition. In addition, *Hydra* appears to have lost a number of MAC/PF proteins and lacks an equivalent of the ancestral complement component C3, which may implicate a degeneration of the prototype complement effector pathway. But gene loss also occurred within the anthozoans, as the gene encoding for the MAC/PF protein apextrin has been lost in *Nematostella* but is still present and expressed in *Acropora millepora*.

Taken together, these preliminary analyses of the newly available genomic and transcriptomic datasets show another example for ancestral genetic complexity but they also highlight the level of divergence present within closely related animal taxa. Thus, simple comparisons between these taxa are unlikely to be informative in terms of understanding the evolution of genes.

An important general implication from these data is, that gene loss may occur stochastically. If genes involved in certain pathways only possess this particular function, then the complete pathway would disappear following the loss of one key component. However, the *Hydra* Toll/TLR data seem to contradict this, as most of the intracellular intermediates are present despite loss of the corresponding receptor, suggesting the invention of an alternative or even novel molecule bypassing this problem. Reconstructing the ancestral genetic toolkit of the common animal ancestor will not be a simple task; it will require the comparison of genome data for a wide range of lower as well as higher animals.

3.3 Novel genes – possible key players in animal diversity

To understand the evolution of novelty is a central problem in evolutionary and developmental biology. Mainly three different evolutionary levels of novelty are currently being discussed: i) the evolution of novel genes, ii) the invention of new regulatory or functional circuits, and iii) the evolution of new morphologies (e.g. animal diversity). For the evolution of new genes and new circuits meanwhile a variety of possible mechanisms have been identified. Certainly gene duplication is the most obvious and

frequently named one, but also others such as exon shuffling, retroposition, gene translocation, mobile elements, lateral gene transfer and “tinkerism” all reviewed in Long, (2001), have been shown to be able to invent new genes, new functions and also new regulatory circuits (Long, 2001). For the evolution of novel morphological features, a “central dogma” has been formulated including the differential spatio-temporal use of the same conserved regulatory genes in all animals as well as changes in the use of regulatory proteins or changes in the action of downstream or effector genes (Carroll et al., 2001).

With the advent of comparative genomics and transcriptomics we learn that this dogma may not be the only possibility to generate morphological novelty. Several recent comparative studies identified a proportion of novel genes that seemed to be restricted to a certain phylum, class, genus or even species. Genes of this category are referred to as taxon- or species-specific genes. Comparative studies investigating vulval patterning in nematodes revealed that the underlying genes and mechanisms are highly species specific (Gibson, 2001; Sommer, 1997). In addition to that, a substantial number of novel genes identified have been suggested being implicated in structures important to all nematodes such as the collagenous cuticle (Hutter et al., 2000). An extensive transcriptomic analysis of the phylum Nematoda including more than 30 different species of nematodes (Parkinson et al., 2004) revealed that more than 50% of the genes were unique to the phylum. In the same way, analysis of two yeast genomes (Goffeau et al., 1996; Wood et al., 2002) showed that over 680 proteins seem to be unique to *Schizosaccharomyces pombe*, while over 1000 proteins were shown to be unique to *Saccharomyces cerevisae*. And when comparing the fungus *Neurospora crassa* to its yeast relatives, more than half of the genes showed no significant similarity (Arnold and Hilton, 2003). A similar trend could also be observed within the large-scale EST analysis approach conducted on several different *Hydra* cDNA libraries shown in this thesis (chapter 2.4). In all seven experiments the proportion of genes with no homologs in other animals was around 30%. A good hint, that these genes might also represent taxon- or species-specific genes came from the analysis of transcripts identified in a SSH cDNA library focused on genes predominantly expressed in cnidocytes (see also chapter 2.4.2). There, 80% of the transcripts that turned out to be non-homologous sequences were shown to be expressed in cnidocyte precursors as well as in mature cnidocytes (Milde et al., in prep.). As cnidocytes are restricted to the

cnidarian phylum and specific cnidocyte types are restricted to single species, these genes represent *bona fide* taxon- or species-specific genes. With regard to the above-mentioned three levels of generating evolutionary novelty, it may be easily possible to explain the mechanisms of how these novel genes have been invented but it is surprising that so far no one tried to interconnect this first level of gene invention with the third level of morphological novelty implicating the role of a taxon specific gene in defining also a taxon specific feature.

Another new example of such a taxon- and even species-specific gene within cnidarians is the *Hydra magnipapillata ks1* gene family. Attempts to find a possible counterpart in all other available cnidarian sequence dataset as well as searches in all other “lower” and “higher” metazoan animals so far failed to identify a possible homologous sequence (see also chapter 2.5). Moreover, it was shown in heterologous southern blot experiments that the *ks1* genes are even absent or at least highly derived in some closely related *Hydra* species. In addition to that, some *Hydra* species seem to have differences in *ks1* gene numbers. Interestingly and consistent with the idea of a novel, species specific gene, functional analysis of the *ks1* promoter in *Hydra vulgaris* revealed that transcriptional regulation of *ks1* also involves novel transcription factor binding sites (Endl et al., 1999). Furthermore, the expression of *ks1* in a very specialized *Hydra* specific cell type, the developing battery cell, also points towards species specificity. Why expression of the *ks1* gene is essential for maintaining head structures in some *Hydra* species (Lohmann et al., 1999) while it is not even present in others remains to be shown.

Taken together, finding numerous novel genes expressed in a taxon specific manner indicate that conserved regulatory genes and signal transduction cascades alone may not be sufficient to explain the advent of a novel, taxon-specific morphological feature. More plausible seems a combination of both conserved and unique components. Obviously many different novel genes were invented in several independent lineages during animal evolution. It might be crucial for the understanding of evolution not only to investigate the similarities between organisms but also the differences.

3.4 Genomic plasticity prior to the divergence of Cnidaria

During evolution, several qualitative and quantitative changes like the expansion of gene families, the appearance of new function and the invention of new regulatory circuits have shaped the metazoan genome. Within *Hydra* we find direct indications for dramatic changes during genome evolution. Measuring and comparing the genome sizes of several different *Hydra* species revealed significant differences (Zacharias et al., 2004). While the most basal *Hydra viridissima* exhibits the smallest genome with 380 mbp, other more derived *Hydra* species dramatically increased the genome size to 1.250 in *Hydra magnipapillata* and even 1.450 Mbp in *Hydra oligactis* (Zacharias et al., 2004).

Further indications come from comparisons of the genetic complexity of several developmental genes. As described in this thesis (chapter 2.5), the extensive genomic characterization of the *ks1* gene family showed high degrees of interspecies variation, gene numbers but also significant levels of complexity within one species. The finding of at least 13 related genes, most of them expressed and some of them even clustered within the genome suggests rapid evolution at surprisingly high levels of conservation in gene structure and also within the promoters. Characterization of *ks1* related genes in more basal *Hydra* species (e.g. *Hydra carnea, Hydra vulgaris* AEP) could even reveal the evolutionary origin of such a taxon- and species-specific gene.

Similar to the *ks1* genes described here, the PPOD gene family was previously shown to be differentially distributed across different *Hydra* species (Thomsen and Bosch, 2006). Whereas only one gene copy was found in *Hydra oligactis* and *Hydra robusta*, several genes were identified in *Hydra vulgaris* and *Hydra magnipapillata*. Moreover, comparing their expression patterns resulted in striking differences. The PPOD gene family, therefore, might represent another example how new evolutionary opportunities are created.

Finally, two recent studies provide first evidence that also in *Hydra* horizontal gene transfer complements the mechanisms that increase the complexity of the genome (Habetha and Bosch, 2005; Steele et al., 2004). Recent publications on research in other cnidarians such as the coral *Acropora* and the sea anemone *Nematostella* report similar mechanisms of genome evolution (Technau et al., 2005). Moreover, analysis of

the available EST collections provides evidence for a significant number of genes only found in non-animal kingdoms. These genes might represent ancient genes that have been lost by all bilaterians rather than genes gained by recent lateral gene transfer (Technau et al., 2005). Taken together, comparative genomics studies presented here and elsewhere point to unexpected genomic plasticity and complexity within groups of closely related species at the base of metazoan evolution.

4 SUMMARY

Among the basal metazoa, cnidarians are used as classical model systems in evolutionary and developmental biology. Entering the age of genomics and transcriptomics, new molecular tools and data resources for model cnidarians, such as the sea anemone *Nematostella vectensis* or the fresh water polyp *Hydra magnipapillata*, become available and offer the possibility of getting new insights in various aspects of evolution and development. To complement the already available molecular resources for *Hydra*, I report in this thesis the establishment of a molecular phylogeny for selected members of the genus *Hydra*, based on the phylogenetic analysis of two nuclear (18S, SSU; 28S, LSU) and two mitochondrial (16S, CO1) markers.

In a second project, complementing resources for cnidarian comparative genomics, “Compagen”, a bioinformatics analysis platform for basal metazoan sequence datasets, was established and used in several different approaches. “Compagen” was used to: i) unravel the cnidarian repertoire of genes related to innate immunity in a comparative genomics study, which resulted in unexpected genetic complexity of the metazoan ancestor and provided evidence for stochastic gene loss in more derived cnidarian species; ii) identify several interesting genes involved in patterning, embryogenesis and immunity during large-scale gene expression profiling approaches using *Hydra* EST data. In addition, a large proportion of non-homologous sequences obtained in each analysis showed to encode for taxon-specific genes; iii) characterize a novel gene, the *Hydra ks1* gene family. *Ks1* was shown to represent a large and partly clustered gene family with 13 conserved members in *Hydra magnipapillata*. During a genus-wide comparison striking differences in the gene distribution in other *Hydra* species could be observed, introducing the *ks1* genes as possible model for gene or even genome evolution within the genus *Hydra*.

The described established resources together with the obtained results contribute to a better understanding of cnidarian evolutionary and developmental biology as well as the evolution of metazoan genomes.

5 ZUSAMMENFASSUNG

Cnidaria stellen als basale Metazoen klassische Modellsysteme in der Entwicklungs- und Evolutionsbiologie dar. Im Zeitalter von Genom- und Transkriptomanalyse sind auch für einzelne Vertreter der Cnidaria, wie z. B. der Anemone *Nematostella vectensis* oder dem Süßwasserpolypen *Hydra magnipapillata*, neue molekulare Werkzeuge und Datenressourcen entwickelt worden, die neue Perspektiven für wissenschaftliche Fragestellungen ergeben. Mit dem Ziel, die bereits vorhandenen molekularen Ressourcen für *Hydra* zu erweitern, wurden die Verwandtschaftsverhältnisse ausgewählter Hydra-Arten durch eine molekularphylogenetische Analyse mittels zweier Kernmarker (18S rDNS, 28S rDNS) und zweier mitochondrieller Marker (16S rDNS, Cytochromoxidase 1) aufgeklärt.

In einem zweiten Projekt wurde eine bioinformatische Analyse-Plattform, genannt „Compagen“, für vergleichende Genomanalyse in basalen Metazoen entwickelt und in den folgenden drei Studien eingesetzt. i) Die vergleichende Suche nach konservierten Komponenten des angeborenen Immunsystems in verschiedenen Vertretern der Cnidaria zeigte eine unerwartete genetische Komplexität des Vorläufers der höheren Metazoen und ergab Hinweise auf zufälligen Verlust von Genen in abgeleiteten Taxa; ii) Der Einsatz von Hochdurchsatz-Transkriptomanalysen ermöglichte die Isolation mehrerer interessanter Gene in *Hydra*, die bei der Musterbildung, in der Embryogenese, und im Immunsystem eine Rolle spielen. Zusätzlich konnte für einige neu identifizierte Gene gezeigt werden, dass diese taxon-spezifisch vorkommen; iii) Die genomische Analyse des *ks1* Gens ergab, dass es sich in *Hydra magnipapillata* um eine große Genfamilie mit wenigstens 13 verschiedenen Genen handelt, von denen einige geclustert im Genom vorliegen. Darüber hinaus wurden signifikante Unterschiede in der Verteilung der *ks1* Gene innerhalb der verschiedenen Arten des Genus *Hydra* gefunden. Die *ks1* Genfamilie kann daher als Modellsystem für Gen- und Genomevolution in *Hydra* dienen.

Die in dieser Arbeit entwickelten molekularen Ressourcen erscheinen hilfreich für die vergleichende Analyse von basalen Metazoen und versprechen, zu einem besseren Verständnis von evolutionären und entwicklungsbiologischen Problemstellungen beizutragen.

6 REFERENCES

Adams, M. D., Celniker, S. E., Holt, R. A., Evans, C. A., Gocayne, J. D., Amanatides, P. G., Scherer, S. E., Li, P. W., Hoskins, R. A., Galle, R. F., George, R. A., Lewis, S. E., Richards, S., Ashburner, M., Henderson, S. N., Sutton, G. G., Wortman, J. R., Yandell, M. D., Zhang, Q., Chen, L. X., Brandon, R. C., Rogers, Y. H., Blazej, R. G., Champe, M., Pfeiffer, B. D., Wan, K. H., Doyle, C., Baxter, E. G., Helt, G., Nelson, C. R., Gabor, G. L., Abril, J. F., Agbayani, A., An, H. J., Andrews-Pfannkoch, C., Baldwin, D., Ballew, R. M., Basu, A., Baxendale, J., Bayraktaroglu, L., Beasley, E. M., Beeson, K. Y., Benos, P. V., Berman, B. P., Bhandari, D., Bolshakov, S., Borkova, D., Botchan, M. R., Bouck, J., Brokstein, P., Brottier, P., Burtis, K. C., Busam, D. A., Butler, H., Cadieu, E., Center, A., Chandra, I., Cherry, J. M., Cawley, S., Dahlke, C., Davenport, L. B., Davies, P., de Pablos, B., Delcher, A., Deng, Z., Mays, A. D., Dew, I., Dietz, S. M., Dodson, K., Doup, L. E., Downes, M., Dugan-Rocha, S., Dunkov, B. C., Dunn, P., Durbin, K. J., Evangelista, C. C., Ferraz, C., Ferriera, S., Fleischmann, W., Fosler, C., Gabrielian, A. E., Garg, N. S., Gelbart, W. M., Glasser, K., Glodek, A., Gong, F., Gorrell, J. H., Gu, Z., Guan, P., Harris, M., Harris, N. L., Harvey, D., Heiman, T. J., Hernandez, J. R., Houck, J., Hostin, D., Houston, K. A., Howland, T. J., Wei, M. H., Ibegwam, C., et al. (2000). The genome sequence of Drosophila melanogaster. *Science* **287,** 2185-95.

Aizenshtadt, T. B. (1975). Investigation of oogenesis in hydra. Communication I. Ultrastructure of interstitial cells at early stages of their transformation into oocytes. *Sov J Dev Biol* **5,** 9-18.

Aizenshtadt, T. B. (1978). [Oogenesis in Hydra. III. The growth and fusion of the oocytes]. *Ontogenez* **9,** 115-23.

Aizenshtadt, T. B., and Marshak, T. L. (1974). [Oogenesis in hydra. II. Cytophotometric determination of DNA concentration in the nuclei of somatic and sex cells]. *Ontogenez* **5,** 446-53.

Alexandrova, O., Schade, M., Bottger, A., and David, C. N. (2005). Oogenesis in Hydra: nurse cells transfer cytoplasm directly to the growing oocyte. *Dev Biol* **281,** 91-101.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. *J Mol Biol* **215,** 403-10.

Angel, P., Imagawa, M., Chiu, R., Stein, B., Imbra, R. J., Rahmsdorf, H. J., Jonat, C., Herrlich, P., and Karin, M. (1987). Phorbol ester-inducible genes contain a common cis element recognized by a TPA-modulated trans-acting factor. *Cell* **49,** 729-39.

Anokhin, B. (2004). Revision of Hydrida (Cnidaria, Hydrozoa): comparative morphological, karyological and taxonomical aspects. *In* "Russian Academy of Science", pp. 190, St. Peterburg.

Arnold, J., and Hilton, N. (2003). Genome sequencing: Revelations from a bread mould. *Nature* **422,** 821-2.

Arvizu, F., Aguilera, A., and Salgado, L. M. (2006). Activities of the protein kinases STK, PI3K, MEK, and ERK are required for the development of the head organizer in Hydra magnipapillata. *Differentiation* **74,** 305-12.

Augustin, R., Franke, A., Khalturin, K., Kiko, R., Siebert, S., Hemmrich, G., and Bosch, T. C. (2006). Dickkopf related genes are components of the positional value gradient in Hydra. *Dev Biol* **296,** 62-70.

Bairoch, A., and Apweiler, R. (1996). The SWISS-PROT protein sequence data bank and its new supplement TREMBL. *Nucleic Acids Res* **24,** 21-5.

Ball, E. E., Hayward, D. C., Saint, R., and Miller, D. J. (2004). A simple plan--cnidarians and the origins of developmental mechanisms. *Nat Rev Genet* **5,** 567-77.

Bendtsen, J. D., Nielsen, H., von Heijne, G., and Brunak, S. (2004). Improved prediction of signal peptides: SignalP 3.0. *J Mol Biol* **340,** 783-95.

Benson, D. A., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J., and Wheeler, D. L. (2007). GenBank. *Nucleic Acids Res* **35,** D21-5.

Berking, S. (2003). A model for budding in hydra: pattern formation in concentric rings. *J Theor Biol* **222,** 37-52.

Bieri, T., Blasiar, D., Ozersky, P., Antoshechkin, I., Bastiani, C., Canaran, P., Chan, J., Chen, N., Chen, W. J., Davis, P., Fiedler, T. J., Girard, L., Han, M., Harris, T. W., Kishore, R., Lee, R., McKay, S., Muller, H. M., Nakamura, C., Petcherski, A., Rangarajan, A., Rogers, A., Schindelman, G., Schwarz, E. M., Spooner, W., Tuli, M. A., Van Auken, K., Wang, D., Wang, X., Williams, G., Durbin, R., Stein, L. D., Sternberg, P. W., and Spieth, J. (2007). WormBase: new content and better access. *Nucleic Acids Res* **35,** D506-10.

Bode, H. R. (2003). Head regeneration in Hydra. *Dev Dyn* **226,** 225-36.

Bosch, T. C. (2003). Ancient signals: peptides and the interpretation of positional information in ancestral metazoans. *Comp Biochem Physiol B Biochem Mol Biol* **136,** 185-96.

Bosch, T. C. (2007a). "Symmetry Breaking in Stem Cells of the Basal Metazoan Hydra." *Progress in Molecular and Subcellular Biology (Alvaro Macieira-Coelho, Ed.)* Springer-Verlag Berlin, Heidelberg.

Bosch, T. C. (2007b). Why polyps regenerate and we don't: Towards a cellular and molecular framework for Hydra regeneration. *Dev Biol*.

Bosch, T. C., and David, C. N. (1986). Male and female stem cells and sex reversal in Hydra polyps. *Proc Natl Acad Sci U S A* **83,** 9478-9482.

Bosch, T. C., and Fujisawa, T. (2001). Polyps, peptides and patterning. *Bioessays* **23,** 420-7.

Bosch, T. C., and Khalturin, K. (2002). Patterning and cell differentiation in Hydra: novel genes and the limits of conservation. *Can. J. Zool.* **80,** 1670-1677.

Brenner, S. (2003). Nobel lecture. Nature's gift to science. *Biosci Rep* **23,** 225-37.

Bridge, D. M., Stover, N. A., and Steele, R. E. (2000). Expression of a novel receptor tyrosine kinase gene and a paired-like homeobox gene provides evidence of differences in patterning at the oral and aboral ends of hydra. *Dev Biol* **220,** 253-62.

Broun, M., and Bode, H. R. (2002). Characterization of the head organizer in hydra. *Development* **129,** 875-84.

Broun, M., Gee, L., Reinhardt, B., and Bode, H. R. (2005). Formation of the head organizer in hydra involves the canonical Wnt pathway. *Development* **132,** 2907-16.

Burge, C., and Karlin, S. (1997). Prediction of complete gene structures in human genomic DNA. *J Mol Biol* **268,** 78-94.

Böttger, A., and Alexandrova, O. (2007). Programmed cell death in Hydra. *Semin Cancer Biol* **17,** 134-146.

Campbell, R. D. (1983). Identifying Hydra species. *In* "Hydra: Research methods" (H. M. Lenhoff, Ed.), pp. 9-28. Chapter 3. Plenum Press, New York and London.

Campbell, R. D. (1987). A new species of Hydra (Cnidaria: Hydrozoa) from North America with comments on species clusters within the genus. *Zool. J. Linn. Soc-Lond.***,** 253-263.

Caplen, N. J., Parrish, S., Imani, F., Fire, A., and Morgan, R. A. (2001). Specific inhibition of gene expression by small double-stranded RNAs in invertebrate and vertebrate systems. *Proc Natl Acad Sci U S A* **98,** 9742-7.

Cardenas, M. M., and Salgado, L. M. (2003). STK, the src homologue, is responsible for the initial commitment to develop head structures in Hydra. *Dev Biol* **264,** 495-505.

Carroll, S. B., Grenier, J. K., and Weatherbee, S. D. (2001). "From DNA to diversity. Molecular genetics and the evolution of animal design." Blackwell Science, Massachusetts.

Check, E. (2002). Worm cast in starring role for Nobel prize. *Nature* **419,** 548-9.

Chera, S., de Rosa, R., Miljkovic-Licina, M., Dobretz, K., Ghila, L., Kaloulis, K., and Galliot, B. (2006). Silencing of the hydra serine protease inhibitor Kazal1 gene mimics the human SPINK1 pancreatic phenotype. *J Cell Sci* **119,** 846-57.

Cherry, J. M., Ball, C., Weng, S., Juvik, G., Schmidt, R., Adler, C., Dunn, B., Dwight, S., Riles, L., Mortimer, R. K., and Botstein, D. (1997). Genetic and physical maps of Saccharomyces cerevisiae. *Nature* **387,** 67-73.

Chourrout, D., Delsuc, F., Chourrout, P., Edvardsen, R. B., Rentzsch, F., Renfer, E., Jensen, M. F., Zhu, B., de Jong, P., Steele, R. E., and Technau, U. (2006). Minimal ProtoHox cluster inferred from bilaterian and cnidarian Hox complements. *Nature* **442,** 684-7.

Cikala, M., Wilm, B., Hobmayer, E., Bottger, A., and David, C. N. (1999). Identification of caspases and apoptosis in the simple metazoan Hydra. *Curr Biol* **9,** 959-62.

Collins, A. G. (2002). Phylogeny of Medusozoa and the evolution of cnidarian life cycles. *J Evol Biol.***,** 418-432.

Collins, A. G., Schuchert, P., Marques, A. C., Jankowski, T., Medina, M., and Schierwater, B. (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.

Crosby, M. A., Goodman, J. L., Strelets, V. B., Zhang, P., and Gelbart, W. M. (2007). FlyBase: genomes by the dozen. *Nucleic Acids Res* **35,** D486-91.

Cummings, S. G., Bode, H.R. (1984). Head regeneration and polarity reversal in Hydra attenuata can occur in the absence of DNA sythesis. *Dev Biol***,** 79-86.

Darwin, C. D. (1895). "On the Origin of Species by Means of Natural Selection." London.

de Jong, D. M., Hislop, N. R., Hayward, D. C., Reece-Hoyes, J. S., Pontynen, P. C., Ball, E. E., and Miller, D. J. (2006). Components of both major axial patterning systems of the Bilateria are differentially expressed along the primary axis of a 'radiate' animal, the anthozoan cnidarian Acropora millepora. *Dev Biol* **298,** 632-43.

Delcher, A. L., Phillippy, A., Carlton, J., and Salzberg, S. L. (2002). Fast algorithms for large-scale genome alignment and comparison. *Nucleic Acids Res* **30,** 2478-83.

Dellaporta, S. L., Xu, A., Sagasser, S., Jakob, W., Moreno, M. A., Buss, L. W., and Schierwater, B. (2006). Mitochondrial genome of Trichoplax adhaerens supports placozoa as the basal lower metazoan phylum. *Proc Natl Acad Sci U S A* **103,** 8751-6.

Diatchenko, L., Lau, Y. F., Campbell, A. P., Chenchik, A., Moqadam, F., Huang, B., Lukyanov, S., Lukyanov, K., Gurskaya, N., Sverdlov, E. D., and Siebert, P. D. (1996). Suppression subtractive hybridization: a method for generating differentially regulated or tissue-specific cDNA probes and libraries. *Proc Natl Acad Sci U S A* **93,** 6025-30.

Dishaw, L. J., Smith, S. L., and Bigger, C. H. (2005). Characterization of a C3-like cDNA in a coral: phylogenetic implications. *Immunogenetics* **57,** 535-48.

Duboule, D., and Wilkins, A. S. (1998). The evolution of 'bricolage'. *Trends Genet* **14,** 54-9.

Dunn, S. R., Phillips, W. S., Spatafora, J. W., Green, D. R., and Weis, V. M. (2006). Highly conserved caspase and Bcl-2 homologues from the sea anemone Aiptasia pallida: lower metazoans as models for the study of apoptosis evolution. *J Mol Evol* **63,** 95-107.

Eddy, S. R. (1998). Profile hidden Markov models. *Bioinformatics* **14,** 755-63.

Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* **32,** 1792-7.

Endl, I., Lohmann, J. U., and Bosch, T. C. (1999). Head-specific gene expression in Hydra: complexity of DNA- protein interactions at the promoter of ks1 is inversely correlated to the head activation potential. *Proc Natl Acad Sci U S A* **96,** 1445-50.

Eppig, J. T., Blake, J. A., Bult, C. J., Kadin, J. A., and Richardson, J. E. (2007). The mouse genome database (MGD): new features facilitating a model system. *Nucleic Acids Res* **35,** D630-7.

Finnerty, J. R., Pang, K., Burton, P., Paulson, D., and Martindale, M. Q. (2004). Origins of bilateral symmetry: Hox and dpp expression in a sea anemone. *Science* **304,** 1335-7.

Fire, A., Xu, S., Montgomery, M. K., Kostas, S. A., Driver, S. E., and Mello, C. C. (1998). Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. *Nature* **391,** 806-11.

Frazer, K. A., Pachter, L., Poliakov, A., Rubin, E. M., and Dubchak, I. (2004). VISTA: computational tools for comparative genomics. *Nucleic Acids Res* **32,** W273-9.

Fröbius, A. C., Genikhovich, G., Kurn, U., Anton-Erxleben, F., and Bosch, T. C. (2003). Expression of developmental genes during early embryogenesis of Hydra. *Dev Genes Evol* **213,** 445-55.

Galliot, B., Miljkovic-Licina, M., de Rosa, R., and Chera, S. (2006). Hydra, a niche for cell and developmental plasticity. *Semin Cell Dev Biol* **17,** 492-502.

Genikhovich, G., Kurn, U., Hemmrich, G., and Bosch, T. C. (2006). Discovery of genes expressed in Hydra embryogenesis. *Dev Biol* **289,** 466-81.

Gibson, G. (2001). Developmental evolution: the unbearable likeness of beings. *Curr Biol* **11,** R345-8.

Gierer, A., Berking, S., Bode, H., David, C. N., Flick, K., Hansmann, G., Schaller, H., and Trenkner, E. (1972). Regeneration of hydra from reaggregated cells. *Nat New Biol* **239,** 98-101.

Gierer, A., and Meinhardt, H. (1972). A theory of biological pattern formation. *Kybernetik* **12,** 30-9.

Gindullis, F., Dechyeva, D., and Schmidt, T. (2001). Construction and characterization of a BAC library for the molecular dissection of a single wild beet centromere and sugar beet (Beta vulgaris) genome analysis. *Genome* **44,** 846-55.

Goffeau, A., Barrell, B. G., Bussey, H., Davis, R. W., Dujon, B., Feldmann, H., Galibert, F., Hoheisel, J. D., Jacq, C., Johnston, M., Louis, E. J., Mewes, H. W., Murakami, Y., Philippsen, P., Tettelin, H., and Oliver, S. G. (1996). Life with 6000 genes. *Science* **274,** 546, 563-7.

Gough, J., Karplus, K., Hughey, R., and Chothia, C. (2001). Assignment of homology to genome sequences using a library of hidden Markov models that represent all proteins of known structure. *J Mol Biol* **313,** 903-19.

Grell, K. G. (1971). Trichoplax adhaerens und die Entstehung der Metazoen. *Naturwiss Rundsch***,** 160-161.

Grishok, A., Pasquinelli, A. E., Conte, D., Li, N., Parrish, S., Ha, I., Baillie, D. L., Fire, A., Ruvkun, G., and Mello, C. C. (2001). Genes and mechanisms related to RNA interference regulate expression of the small temporal RNAs that control C. elegans developmental timing. *Cell* **106,** 23-34.

Groth, D., Lehrach, H., and Hennig, S. (2004). GOblet: a platform for Gene Ontology annotation of anonymous sequence data. *Nucleic Acids Res* **32,** W313-7.

Guder, C., Philipp, I., Lengfeld, T., Watanabe, H., Hobmayer, B., and Holstein, T. W. (2006a). The Wnt code: cnidarians signal the way. *Oncogene* **25,** 7450-60.

Guder, C., Pinho, S., Nacak, T. G., Schmidt, H. A., Hobmayer, B., Niehrs, C., and Holstein, T. W. (2006b). An ancient Wnt-Dickkopf antagonism in Hydra. *Development* **133,** 901-11.

Haag, E. S., and Raff, R. A. (1998). Isolation and characterization of three mRNAs enriched in embryos of the direct-developing sea urchin Heliocidaris erythrogramma: evolution of larval ectoderm. *Dev Genes Evol* **208,** 188-204.

Haag, E. S., Sly, B. J., Andrews, M. E., and Raff, R. A. (1999). Apextrin, a novel extracellular protein associated with larval ectoderm evolution in Heliocidaris erythrogramma. *Dev Biol* **211,** 77-87.

Habetha, M., Anton-Erxleben, F., Neumann, K., and Bosch, T. C. (2003). The Hydra viridis/Chlorella symbiosis. Growth and sexual differentiation in polyps without symbionts. *Zoology (Jena)* **106,** 101-8.

Habetha, M., and Bosch, T. C. (2005). Symbiotic Hydra express a plant-like peroxidase gene during oogenesis. *J Exp Biol* **208,** 2157-65.

Hadrys, T., DeSalle, R., Sagasser, S., Fischer, N., and Schierwater, B. (2005). The Trichoplax PaxB gene: a putative Proto-PaxA/B/C gene predating the origin of nerve and sensory cells. *Mol Biol Evol* **22,** 1569-78.

Hall, B. K. (2003). Evo-Devo: evolutionary developmental mechanisms. *Int J Dev Biol* **47,** 491-5.

Hall, T. A. (1998). BioEdit. Biological sequence alignment editor for Windows.

Harris, M. A., Clark, J., Ireland, A., Lomax, J., Ashburner, M., Foulger, R., Eilbeck, K., Lewis, S., Marshall, B., Mungall, C., Richter, J., Rubin, G. M., Blake, J. A., Bult, C., Dolan, M., Drabkin, H., Eppig, J. T., Hill, D. P., Ni, L., Ringwald, M., Balakrishnan, R., Cherry, J. M., Christie, K. R., Costanzo, M. C., Dwight, S. S., Engel, S., Fisk, D. G., Hirschman, J. E., Hong, E. L., Nash, R. S., Sethuraman, A., Theesfeld, C. L., Botstein, D., Dolinski, K., Feierbach, B., Berardini, T., Mundodi, S., Rhee, S. Y., Apweiler, R., Barrell, D., Camon, E., Dimmer, E., Lee, V., Chisholm, R., Gaudet, P., Kibbe, W., Kishore, R., Schwarz, E. M., Sternberg, P., Gwinn, M., Hannick, L., Wortman, J., Berriman, M., Wood, V., de la Cruz, N., Tonellato, P., Jaiswal, P., Seigfried, T., and White, R. (2004). The Gene Ontology (GO) database and informatics resource. *Nucleic Acids Res* **32,** D258-61.

Hartwell, L. H. (2002). Nobel Lecture. Yeast and cancer. *Biosci Rep* **22,** 373-94.

Hayward, D. C., Samuel, G., Pontynen, P. C., Catmull, J., Saint, R., Miller, D. J., and Ball, E. E. (2002). Localized expression of a dpp/BMP2/4 ortholog in a coral embryo. *Proc Natl Acad Sci U S A* **99,** 8106-11.

Hemmrich, G., Anokhin, B., Zacharias, H., and Bosch, T. C. (2006). Molecular phylogenetics in Hydra, a classical model in evolutionary developmental biology. *Mol Phylogenet Evol*.

Hobmayer, B., Rentzsch, F., and Holstein, T. W. (2001). Identification and expression of HySmad1, a member of the R-Smad family of TGFbeta signal transducers, in the diploblastic metazoan Hydra. *Dev Genes Evol* **211,** 597-602.

Hobmayer, B., Rentzsch, F., Kuhn, K., Happel, C. M., von Laue, C. C., Snyder, P., Rothbacher, U., and Holstein, T. W. (2000). WNT signalling molecules act in axis formation in the diploblastic metazoan Hydra. *Nature* **407,** 186-9.

Hoffenberg, R. (2003). Brenner, the worm and the prize. *Clin Med* **3,** 285-6.

Holstein, T. W. (1995). "Cnidaria: Hydrozoa Süßwasserfauna von Mitteleuropa." Gustav Fischer Verlag, Stuttgart, Jena, NY.

Holstein, T. W., Hobmayer, E., and David, C. N. (1991). Pattern of epithelial cell cycling in hydra. *Dev Biol* **148,** 602-11.

Holstein, T. W., Hobmayer, E., and Technau, U. (2003). Cnidarians: an evolutionarily conserved model system for regeneration? *Dev Dyn* **226,** 257-67.

Honegger, T. G., Zürrer, D., Tardent, P. (1989). Oogenesis in Hydra carnea: a new model based on light and electron microscopic analyses of oocyte and nurse cell differentiation. *Tissue Cell* **21,** 381-393.

Huang, X., Adams, M. D., Zhou, H., and Kerlavage, A. R. (1997). A tool for analyzing and annotating genomic sequences. *Genomics* **46,** 37-45.

Huang, X., and Madan, A. (1999). CAP3: A DNA sequence assembly program. *Genome Res* **9,** 868-77.

Hubbard, T. J., Aken, B. L., Beal, K., Ballester, B., Caccamo, M., Chen, Y., Clarke, L., Coates, G., Cunningham, F., Cutts, T., Down, T., Dyer, S. C., Fitzgerald, S., Fernandez-Banet, J., Graf, S., Haider, S., Hammond, M., Herrero, J., Holland, R., Howe, K., Howe, K., Johnson, N., Kahari, A., Keefe, D., Kokocinski, F., Kulesha, E., Lawson, D., Longden, I., Melsopp, C., Megy, K., Meidl, P., Ouverdin, B.,

Parker, A., Prlic, A., Rice, S., Rios, D., Schuster, M., Sealy, I., Severin, J., Slater, G., Smedley, D., Spudich, G., Trevanion, S., Vilella, A., Vogel, J., White, S., Wood, M., Cox, T., Curwen, V., Durbin, R., Fernandez-Suarez, X. M., Flicek, P., Kasprzyk, A., Proctor, G., Searle, S., Smith, J., Ureta-Vidal, A., and Birney, E. (2007). Ensembl 2007. *Nucleic Acids Res* **35,** D610-7.

Huelsenbeck, J. P., and Ronquist, F. (2001). MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* **17,** 754-5.

Hunt, T. (2002). Nobel Lecture. Protein synthesis, proteolysis, and cell cycle transitions. *Biosci Rep* **22,** 465-86.

Hutter, H., Vogel, B. E., Plenefisch, J. D., Norris, C. R., Proenca, R. B., Spieth, J., Guo, C., Mastwal, S., Zhu, X., Scheel, J., and Hedgecock, E. M. (2000). Conservation and novelty in the evolution of cell adhesion and extracellular matrix genes. *Science* **287,** 989-94.

Iseli, C., Jongeneel, C. V., and Bucher, P. (1999). ESTScan: a program for detecting, evaluating, and reconstructing potential coding regions in EST sequences. *Proc Int Conf Intell Syst Mol Biol***,** 138-48.

Jaruzelska, J., Kotecki, M., Kusz, K., Spik, A., Firpo, M., and Reijo Pera, R. A. (2003). Conservation of a Pumilio-Nanos complex from Drosophila germ plasm to human germ cells. *Dev Genes Evol* **213,** 120-6.

Kadota, K., Ishino, T., Matsuyama, T., Chinzei, Y., and Yuda, M. (2004). Essential role of membrane-attack protein in malarial transmission to mosquito host. *Proc Natl Acad Sci U S A* **101,** 16310-5.

Kaloulis, K., Chera, S., Hassel, M., Gauchat, D., and Galliot, B. (2004). Reactivation of developmental programs: the cAMP-response element-binding protein pathway is involved in hydra head regeneration. *Proc Natl Acad Sci U S A* **101,** 2363-8.

Kamm, K., and Schierwater, B. (2006). Ancient complexity of the non-Hox ANTP gene complement in the anthozoan Nematostella vectensis: implications for the evolution of the ANTP superclass. *J Exp Zoolog B Mol Dev Evol* **306,** 589-96.

Kamm, K., Schierwater, B., Jakob, W., Dellaporta, S. L., and Miller, D. J. (2006). Axial patterning and diversification in the cnidaria predate the Hox system. *Curr Biol* **16,** 920-6.

Kasahara, S., and Bosch, T. C. (2003). Enhanced antibacterial activity in Hydra polyps lacking nerve cells. *Dev Comp Immunol* **27,** 79-85.

Katoh, K., Misawa, K., Kuma, K., and Miyata, T. (2002). MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res* **30,** 3059-66.

Kopp, E., Medzhitov, R., Carothers, J., Xiao, C., Douglas, I., Janeway, C. A., and Ghosh, S. (1999). ECSIT is an evolutionarily conserved intermediate in the Toll/IL-1 signal transduction pathway. *Genes Dev* **13,** 2059-71.

Kortschak, R. D., Samuel, G., Saint, R., and Miller, D. J. (2003). EST analysis of the cnidarian Acropora millepora reveals extensive gene loss and rapid sequence divergence in the model invertebrates. *Curr Biol* **13,** 2190-5.

Koski, L. B., Gray, M. W., Lang, B. F., and Burger, G. (2005). AutoFACT: an automatic functional annotation and classification tool. *BMC Bioinformatics* **6,** 151.

Kraus, Y., and Technau, U. (2006). Gastrulation in the sea anemone Nematostella vectensis occurs by invagination and immigration: an ultrastructural study. *Dev Genes Evol* **216,** 119-32.

Krogh, A. (1997). Two methods for improving performance of an HMM and their application for gene finding. *Proc Int Conf Intell Syst Mol Biol* **5,** 179-86.

Kulikova, T., Akhtar, R., Aldebert, P., Althorpe, N., Andersson, M., Baldwin, A., Bates, K., Bhattacharyya, S., Bower, L., Browne, P., Castro, M., Cochrane, G., Duggan, K., Eberhardt, R., Faruque, N., Hoad, G., Kanz, C., Lee, C., Leinonen, R., Lin, Q., Lombard, V., Lopez, R., Lorenc, D., McWilliam, H., Mukherjee, G., Nardone, F., Pastor, M. P., Plaister, S., Sobhany, S., Stoehr, P., Vaughan, R., Wu, D., Zhu, W., and Apweiler, R. (2007). EMBL Nucleotide Sequence Database in 2006. *Nucleic Acids Res* **35,** D16-20.

Kumar, S., Tamura, K., and Nei, M. (2004). MEGA3: Integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment. *Brief Bioinform* **5,** 150-63.

Kusserow, A., Pang, K., Sturm, C., Hrouda, M., Lentfer, J., Schmidt, H. A., Technau, U., von Haeseler, A., Hobmayer, B., Martindale, M. Q., and Holstein, T. W. (2005). Unexpected complexity of the Wnt gene family in a sea anemone. *Nature* **433,** 156-60.

Kuznetsov, S. G., Anton-Erxleben, F., and Bosch, T. C. (2002). Epithelial interactions in Hydra: apoptosis in interspecies grafts is induced by detachment from the extracellular matrix. *J Exp Biol* **205,** 3809-17.

Käsbauer, T., Towb, P., Alexandrova, O., David, C. N., Dall'armi, E., Staudigl, A., Stiening, B., and Bottger, A. (2007). The Notch signaling pathway in the cnidarian Hydra. *Dev Biol* **303,** 376-90.

Kürn, U., Sommer, F., Hemmrich, G., Bosch, T. C., and Khalturin, K. (2007). Allorecognition in urochordates: Identification of a highly variable complement receptor-like protein expressed in follicle cells of Ciona. *Dev Comp Immunol* **31,** 360-71.

Lee, P. N., Pang, K., Matus, D. Q., and Martindale, M. Q. (2006). A WNT of things to come: evolution of Wnt signaling and polarity in cnidarians. *Semin Cell Dev Biol* **17,** 157-67.

Lenhoff, H. M., and Brown, R. D. (1970). Mass culture of hydra: an improved method and its application to other aquatic invertebrates. *Lab Anim* **4,** 139-54.

Letovsky, S. I., Cottingham, R. W., Porter, C. J., and Li, P. W. (1998). GDB: the Human Genome Database. *Nucleic Acids Res* **26,** 94-9.

Letunic, I., Copley, R. R., Schmidt, S., Ciccarelli, F. D., Doerks, T., Schultz, J., Ponting, C. P., and Bork, P. (2004). SMART 4.0: towards genomic data integration. *Nucleic Acids Res* **32,** D142-4.

Lewis, E. B. (1978). A gene complex controlling segmentation in Drosophila. *Nature* **276,** 565-70.

Leys, S. P., and Ereskovsky, A. V. (2006). Embryogenesis and larval differentiation in sponges. *Can. J. Zool.* **84,** 262-287.

Littlefield, C. L. (1985). Germ cells in Hydra oligactis males. I. Isolation of a subpopulation of interstitial cells that is developmentally restricted to sperm production. *Dev Biol* **112,** 185-93.

Littlefield, C. L., and Bode, H. R. (1986). Germ cells in Hydra oligactis males. II. Evidence for a subpopulation of interstitial stem cells whose differentiation is limited to sperm production. *Dev Biol* **116,** 381-6.

Lohmann, J. U., Endl, I., and Bosch, T. C. (1999). Silencing of developmental genes in Hydra. *Dev Biol* **214,** 211-4.

Long, M. (2001). Evolution of novel genes. *Curr Opin Genet Dev* **11,** 673-80.

MacWilliams, H. K. (1983). Hydra transplantation phenomena and the mechanism of Hydra head regeneration. II. Properties of the head activation. *Dev Biol* **96,** 239-57.

Mah, S. A., Moy, G. W., Swanson, W. J., and Vacquier, V. D. (2004). A perforin-like protein from a marine mollusk. *Biochem Biophys Res Commun* **316,** 468-75.

Manuel, G. C., Reynoso, R., Gee, L., Salgado, L. M., and Bode, H. R. (2006). PI3K and ERK 1-2 regulate early stages during head regeneration in hydra. *Dev Growth Differ* **48,** 129-38.

Marques, A. C., and Collins, A. G. (2004). Cladistic analysis of Medusozoa and cnidarian evolution. *Invertebrate Biology***,** 23-42.

Martin, V. J., Littlefield, C. L., Archer, W. E., and Bode, H. R. (1997). Embryogenesis in hydra. *Biol Bull* **192,** 345-63.

Martindale, M. Q., Finnerty, J. R., and Henry, J. Q. (2002). The Radiata and the evolutionary origins of the bilaterian body plan. *Mol Phylogenet Evol* **24,** 358-65.

Matus, D. Q., Pang, K., Daly, M., and Martindale, M. Q. (2007). Expression of Pax gene family members in the anthozoan cnidarian, Nematostella vectensis. *Evol Dev* **9,** 25-38.

McGinnis, W., and Kuziora, M. (1994). The molecular architects of body design. *Sci Am* **270,** 58-61, 64-6.

Meinhardt, H. (1993). A model for pattern formation of hypostome, tentacles, and foot in hydra: how to form structures close to each other, how to form them at a distance. *Dev Biol* **157,** 321-33.

Meinhardt, H. (2004). Models for the generation of the embryonic body axes: ontogenetic and evolutionary aspects. *Curr Opin Genet Dev* **14,** 446-54.

Meinhardt, H. (2006). Primary body axes of vertebrates: generation of a near-Cartesian coordinate system and the role of Spemann-type organizer. *Dev Dyn* **235,** 2907-19.

Meinhardt, H., and Gierer, A. (2000). Pattern formation by local self-activation and lateral inhibition. *Bioessays* **22,** 753-60.

Miljkovic-Licina, M., Chera, S., Ghila, L., and Galliot, B. (2007). Head regeneration in wild-type hydra requires de novo neurogenesis. *Development* **134,** 1191-201.

Miller, D. J., Ball, E. E., and Technau, U. (2005). Cnidarians and ancestral genetic complexity in the animal kingdom. *Trends Genet* **21,** 536-9.

Miller, D. J., Hemmrich, G., Ball, E. E., Hayward, D. C., Khalturin, K., Funayama, N., Agata, K., and Bosch, T. C. (2007). The innate immune repertoire in Cnidaria - ancestral complexity and stochastic gene loss. *Genome Biol.* **in press**.

Minobe, S., Fei, K., Yan, L., Sarras Jr, M., and Werle, M. (2000). Identification and characterization of the epithelial polarity receptor "Frizzled" in Hydra vulgaris. *Dev Genes Evol* **210,** 258-62.

Moller, S., Croning, M. D., and Apweiler, R. (2001). Evaluation of methods for the prediction of membrane spanning regions. *Bioinformatics* **17,** 646-53.

Monteiro, A. S., Schierwater, B., Dellaporta, S. L., and Holland, P. W. (2006). A low diversity of ANTP class homeobox genes in Placozoa. *Evol Dev* **8,** 174-82.

Morgan, T. H. (1901). Regeneration in the egg, embryo and adult. *Am. Nat.* **35,** 949-973.

Muller, W. A. (1996). Pattern formation in the immortal Hydra. *Trends Genet* **12,** 91-6.

Nichols, S. A., Dirks, W., Pearse, J. S., and King, N. (2006). Early evolution of animal cell signaling and adhesion genes. *Proc Natl Acad Sci U S A* **103,** 12451-6.

Nishimiya-Fujisawa, C., and Sugiyama, T. (1993). Genetic analysis of developmental mechanisms in hydra. XX. Cloning of interstitial stem cells restricted to the sperm differentiation pathway in Hydra magnipapillata. *Dev Biol* **157,** 1-9.

Notredame, C., Higgins, D. G., and Heringa, J. (2000). T-Coffee: A novel method for fast and accurate multiple sequence alignment. *J Mol Biol* **302,** 205-17.

Nüsslein-Volhard, C., and Wieschaus, E. (1980). Mutations affecting segment number and polarity in Drosophila. *Nature* **287,** 795-801.

Oshiro, N., Kobayashi, C., Iwanaga, S., Nozaki, M., Namikoshi, M., Spring, J., and Nagai, H. (2004). A new membrane-attack complex/perforin (MACPF) domain lethal toxin from the nematocyst venom of the Okinawan sea anemone Actineria villosa. *Toxicon* **43,** 225-8.

Page, R. D. (1996). TreeView: an application to display phylogenetic trees on personal computers. *Comput Appl Biosci* **12,** 357-8.

Parkinson, J., Mitreva, M., Whitton, C., Thomson, M., Daub, J., Martin, J., Schmid, R., Hall, N., Barrell, B., Waterston, R. H., McCarter, J. P., and Blaxter, M. L. (2004). A transcriptomic analysis of the phylum Nematoda. *Nat Genet* **36,** 1259-67.

Paulovich, A. G., and Hartwell, L. H. (1995). A checkpoint regulates the rate of progression through S phase in S. cerevisiae in response to DNA damage. *Cell* **82,** 841-7.

Pertea, G., Huang, X., Liang, F., Antonescu, V., Sultana, R., Karamycheva, S., Lee, Y., White, J., Cheung, F., Parvizi, B., Tsai, J., and Quackenbush, J. (2003). TIGR Gene Indices clustering tools (TGICL): a software system for fast clustering of large EST datasets. *Bioinformatics* **19,** 651-2.

Pires-daSilva, A., and Sommer, R. J. (2003). The evolution of signalling pathways in animal development. *Nat Rev Genet* **4,** 39-49.

Plickert, G., Jacoby, V., Frank, U., Muller, W. A., and Mokady, O. (2006). Wnt signaling in hydroid development: formation of the primary body axis in embryogenesis and its subsequent patterning. *Dev Biol* **298,** 368-78.

Prochnik, S. E., Rokhsar, D. S., and Aboobaker, A. A. (2007). Evidence for a microRNA expansion in the bilaterian ancestor. *Dev Genes Evol* **217,** 73-7.

Pruitt, K. D., Tatusova, T., and Maglott, D. R. (2007). NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res* **35,** D61-5.

Reber-Muller, S., Streitwolf-Engel, R., Yanze, N., Schmid, V., Stierwald, M., Erb, M., and Seipel, K. (2006). BMP2/4 and BMP5-8 in jellyfish development and transdifferentiation. *Int J Dev Biol* **50,** 377-84.

Reidling, J. C., Miller, M. A., and Steele, R. E. (2000). Sweet Tooth, a novel receptor protein-tyrosine kinase with C-type lectin-like extracellular domains. *J Biol Chem* **275,** 10323-30.

Reinhardt, B., Broun, M., Blitz, I. L., and Bode, H. R. (2004). HyBMP5-8b, a BMP5-8 orthologue, acts during axial patterning and tentacle formation in hydra. *Dev Biol* **267,** 43-59.

Rentzsch, F., Anton, R., Saina, M., Hammerschmidt, M., Holstein, T. W., and Technau, U. (2006). Asymmetric expression of the BMP antagonists chordin and gremlin in the sea anemone Nematostella vectensis: implications for the evolution of axial patterning. *Dev Biol* **296,** 375-87.

Rentzsch, F., Guder, C., Vocke, D., Hobmayer, B., and Holstein, T. W. (2007). An ancient chordin-like gene in organizer formation of Hydra. *Proc Natl Acad Sci U S A* **104,** 3249-54.

Royer, H. D. (2001). Centenary Nobel Prize in Physiology or Medicine for the cell cycle. *J Mol Med* **79,** 683-5.

Rudel, D., and Sommer, R. J. (2003). The evolution of developmental mechanisms. *Dev Biol* **264,** 15-37.

Ryan, J. F., Mazza, M. E., Pang, K., Matus, D. Q., Baxevanis, A. D., Martindale, M. Q., and Finnerty, J. R. (2007). Pre-Bilaterian Origins of the Hox Cluster and the Hox Code: Evidence from the Sea Anemone, Nematostella vectensis. *PLoS ONE* **2,** e153.

Samuel, G., Miller, D., and Saint, R. (2001). Conservation of a DPP/BMP signaling pathway in the nonbilateral cnidarian Acropora millepora. *Evol Dev* **3,** 241-50.

Sandelin, A., Wasserman, W. W., and Lenhard, B. (2004). ConSite: web-based prediction of regulatory elements using cross-species comparison. *Nucleic Acids Res* **32,** W249-52.

Schmidt, H. A., Strimmer, K., Vingron, M., and von Haeseler, A. (2002). TREE-PUZZLE: maximum likelihood phylogenetic analysis using quartets and parallel computing. *Bioinformatics* **18,** 502-4.

Schmitt, D. M., and Brower, D. L. (2001). Intron dynamics and the evolution of integrin beta-subunit genes: maintenance of an ancestral gene structure in the coral, Acropora millepora. *J Mol Evol* **53,** 703-10.

Schwarzacher, T., Heslop-Harrison, P. (2003). "Practical In Situ Hybridization." BIOS Scientific Publishers, Oxford.

Seipp, S., Schmich, J., and Leitz, T. (2001). Apoptosis--a death-inducing mechanism tightly linked with morphogenesis in Hydractina echinata (Cnidaria, Hydrozoa). *Development* **128,** 4891-8.

Seipp, S., Wittig, K., Stiening, B., Bottger, A., and Leitz, T. (2006). Metamorphosis of Hydractinia echinata (Cnidaria) is caspase-dependent. *Int J Dev Biol* **50,** 63-70.

Senapathy, P., Shapiro, M. B., and Harris, N. L. (1990). Splice junctions, branch point sites, and exons: sequence statistics, identification, and applications to genome project. *Methods Enzymol* **183,** 252-78.

Siebert, S., Thomsen, S., Reimer, M. M., and Bosch, T. C. (2005). Control of foot differentiation in Hydra: phylogenetic footprinting indicates interaction of head, bud and foot patterning systems. *Mech Dev* **122,** 998-1007.

Simionato, E., Ledent, V., Richards, G., Thomas-Chollier, M., Kerner, P., Coornaert, D., Degnan, B. M., and Vervoort, M. (2007). Origin and diversification of the basic helix-loop-helix gene family in metazoans: insights from comparative genomics. *BMC Evol Biol* **7,** 33.

Simossis, V. A., and Heringa, J. (2005). PRALINE: a multiple sequence alignment toolbox that integrates homology-extended and secondary structure information. *Nucleic Acids Res* **33,** W289-94.

Sommer, D. D., Delcher, A. L., Salzberg, S. L., and Pop, M. (2007). Minimus: a fast, lightweight genome assembler. *BMC Bioinformatics* **8,** 64.

Sommer, R. J. (1997). Evolutionary changes of developmental mechanisms in the absence of cell lineage alterations during vulva formation in the Diplogastridae (Nematoda). *Development* **124,** 243-51.

Sonnhammer, E. L., Eddy, S. R., Birney, E., Bateman, A., and Durbin, R. (1998). Pfam: multiple sequence alignments and HMM-profiles of protein domains. *Nucleic Acids Res* **26,** 320-2.

Spafford, J. D., Spencer, A. N., and Gallin, W. J. (1999). Genomic organization of a voltage-gated Na+ channel in a hydrozoan jellyfish: insights into the evolution of voltage-gated Na+ channel genes. *Receptors Channels* **6,** 493-506.

Steele, R. E. (2002). Developmental signaling in Hydra: what does it take to build a "simple" animal? *Dev Biol* **248,** 199-219.

Steele, R. E., Hampson, S. E., Stover, N. A., Kibler, D. F., and Bode, H. R. (2004). Probable horizontal transfer of a gene between a protist and a cnidarian. *Curr Biol* **14,** R298-9.

Sudhop, S., Coulier, F., Bieller, A., Vogt, A., Hotz, T., and Hassel, M. (2004). Signalling by the FGFR-like tyrosine kinase, Kringelchen, is essential for bud detachment in Hydra vulgaris. *Development* **131,** 4001-11.

Sullivan, J. C., Kalaitzidis, D., Gilmore, T. D., and Finnerty, J. R. (2007a). Rel homology domain-containing transcription factors in the cnidarian Nematostella vectensis. *Dev Genes Evol* **217,** 63-72.

Sullivan, J. C., Ryan, J. F., Mullikin, J. C., and Finnerty, J. R. (2007b). Conserved and novel Wnt clusters in the basal eumetazoan Nematostella vectensis. *Dev Genes Evol* **217,** 235-9.

Sullivan, J. C., Ryan, J. F., Watson, J. A., Webb, J., Mullikin, J. C., Rokhsar, D., and Finnerty, J. R. (2006). StellaBase: the Nematostella vectensis Genomics Database. *Nucleic Acids Res* **34,** D495-9.

Sun, Y., Kanekar, S. L., Vetter, M. L., Gorski, S., Jan, Y. N., Glaser, T., and Brown, N. L. (2003). Conserved and divergent functions of Drosophila atonal, amphibian, and mammalian Ath5 genes. *Evol Dev* **5,** 532-41.

Tao, N. (2005). FINDMODEL: a tool to select the best-fit model of nucleotide substitution.

Tardent, P. (1985). "The differentiation of germ cells in Cnidaria." A.R. Liss, New York.

Technau, U., Miller, M. A., Bridge, D., and Steele, R. E. (2003). Arrested apoptosis of nurse cells during Hydra oogenesis and embryogenesis. *Dev Biol* **260,** 191-206.

Technau, U., Rudd, S., Maxwell, P., Gordon, P. M., Saina, M., Grasso, L. C., Hayward, D. C., Sensen, C. W., Saint, R., Holstein, T. W., Ball, E. E., and Miller, D. J.

(2005). Maintenance of ancestral complexity and non-metazoan genes in two basal cnidarians. *Trends Genet* **21,** 633-9.

Technau, U., and Scholz, C. B. (2003). Origin and evolution of endoderm and mesoderm. *Int J Dev Biol* **47,** 531-9.

Terada, H., Sugiyama, T., and Shigenaka, Y. (1988). Genetic analysis of developmental mechanisms in hydra. XVIII. Mechanism for elimination of the interstitial cell lineage in the mutant strain Sf-1. *Dev Biol* **126,** 263-9.

The C. elegans genome sequencing consortium. (1998). Genome sequence of the nematode C. elegans: a platform for investigating biology. *Science* **282,** 2012-8.

Thompson, J. D., Higgins, D. G., and Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* **22,** 4673-80.

Thomsen, S., and Bosch, T. C. (2006). Foot differentiation and genomic plasticity in Hydra: lessons from the PPOD gene family. *Dev Genes Evol* **216,** 57-68.

Weinziger, R., Salgado, L. M., David, C. N., and Bosch, T. C. (1994). Ks1, an epithelial cell-specific gene, responds to early signals of head formation in Hydra. *Development* **120,** 2511-7.

Wiens, M., Korzhev, M., Krasko, A., Thakur, N. L., Perovic-Ottstadt, S., Breter, H. J., Ushijima, H., Diehl-Seifert, B., Muller, I. M., and Muller, W. E. (2005). Innate immune defense of the sponge Suberites domuncula against bacteria involves a MyD88-dependent signaling pathway. Induction of a perforin-like molecule. *J Biol Chem* **280,** 27949-59.

Wiens, M., Korzhev, M., Perovic-Ottstadt, S., Luthringer, B., Brandt, D., Klein, S., and Muller, W. E. (2007). Toll-like receptors are part of the innate immune defense system of sponges (demospongiae: porifera). *Mol Biol Evol* **24,** 792-804.

Wilkins, A. S. (2002). "The Evolution of Developmental Pathways." Sinauer Associates, Massachusetts.

Wittlieb, J., Khalturin, K., Lohmann, J. U., Anton-Erxleben, F., and Bosch, T. C. (2006). Transgenic Hydra allow in vivo tracking of individual stem cells during morphogenesis. *Proc Natl Acad Sci U S A* **103,** 6208-11.

Wolpert, L. (1971). Positional information and pattern formation. *Curr Top Dev Biol* **6,** 183-224.

Wolpert, L. (1973). Role of diffusible gradients in regeneration. *Dev Biol* **30,** concl4-5.

Wolpert, L., Clarke, M. R., and Hornbruch, A. (1972). Positional signalling along hydra. *Nat New Biol* **239,** 101-5.

Wood, V., Gwilliam, R., Rajandream, M. A., Lyne, M., Lyne, R., Stewart, A., Sgouros, J., Peat, N., Hayles, J., Baker, S., Basham, D., Bowman, S., Brooks, K., Brown, D., Brown, S., Chillingworth, T., Churcher, C., Collins, M., Connor, R., Cronin, A., Davis, P., Feltwell, T., Fraser, A., Gentles, S., Goble, A., Hamlin, N., Harris, D., Hidalgo, J., Hodgson, G., Holroyd, S., Hornsby, T., Howarth, S., Huckle, E. J., Hunt, S., Jagels, K., James, K., Jones, L., Jones, M., Leather, S., McDonald, S., McLean, J., Mooney, P., Moule, S., Mungall, K., Murphy, L., Niblett, D., Odell, C., Oliver, K., O'Neil, S., Pearson, D., Quail, M. A., Rabbinowitsch, E., Rutherford, K., Rutter, S., Saunders, D., Seeger, K., Sharp, S., Skelton, J., Simmonds, M., Squares, R., Squares, S., Stevens, K., Taylor, K., Taylor, R. G., Tivey, A., Walsh,

S., Warren, T., Whitehead, S., Woodward, J., Volckaert, G., Aert, R., Robben, J., Grymonprez, B., Weltjens, I., Vanstreels, E., Rieger, M., Schafer, M., Muller-Auer, S., Gabel, C., Fuchs, M., Dusterhoft, A., Fritzc, C., Holzer, E., Moestl, D., Hilbert, H., Borzym, K., Langer, I., Beck, A., Lehrach, H., Reinhardt, R., Pohl, T. M., Eger, P., Zimmermann, W., Wedler, H., Wambutt, R., Purnelle, B., Goffeau, A., Cadieu, E., Dreano, S., Gloux, S., et al. (2002). The genome sequence of Schizosaccharomyces pombe. *Nature* **415,** 871-80.

Xiao, C., Shim, J. H., Kluppel, M., Zhang, S. S., Dong, C., Flavell, R. A., Fu, X. Y., Wrana, J. L., Hogan, B. L., and Ghosh, S. (2003). Ecsit is required for Bmp signaling and mesoderm formation during mouse embryogenesis. *Genes Dev* **17,** 2933-49.

Zacharias, H., Anokhin, B., Khalturin, K., and Bosch, T. C. (2004). Genome sizes and chromosomes in the basal metazoan Hydra. *Zoology (Jena)* **107,** 219-27.

Zdobnov, E. M., von Mering, C., Letunic, I., Torrents, D., Suyama, M., Copley, R. R., Christophides, G. K., Thomasova, D., Holt, R. A., Subramanian, G. M., Mueller, H. M., Dimopoulos, G., Law, J. H., Wells, M. A., Birney, E., Charlab, R., Halpern, A. L., Kokoza, E., Kraft, C. L., Lai, Z., Lewis, S., Louis, C., Barillas-Mury, C., Nusskern, D., Rubin, G. M., Salzberg, S. L., Sutton, G. G., Topalis, P., Wides, R., Wincker, P., Yandell, M., Collins, F. H., Ribeiro, J., Gelbart, W. M., Kafatos, F. C., and Bork, P. (2002). Comparative genome and proteome analysis of Anopheles gambiae and Drosophila melanogaster. *Science* **298,** 149-59.

Zhang, Z., Schwartz, S., Wagner, L., and Miller, W. (2000). A greedy algorithm for aligning DNA sequences. *J Comput Biol* **7,** 203-14.

7 METHODS

7.1 Computational Methods

7.1.1 Comparative genomics analysis tools

7.1.1.1 Sequence alignments

For alignment of nucleotide and amino acid sequences, a variety of algorithms were used. General pair wise or multiple alignments were constructed using ClustalW (Thompson et al., 1994), Muscle (Edgar, 2004), T-Coffe (Notredame et al., 2000) or Praline (Simossis and Heringa, 2005). To generate spliced alignments of cDNA sequences vs. genomic DNA sequence, the MAFFT program (Katoh et al., 2002) or the AAT-algorithm (Huang et al., 1997) were used.

7.1.1.2 Local genomic assemblies

To generate local genomic assemblies based on raw WGS sequencing reads from the NCBI trace archive, the pipeline-based genome-assembler AMOS from TIGR was used (www.amos.sourceforge.net). The included lightweight pipeline Minimus served for generating up to 80 kbp local alignments including important additional sequence information like sequencing-quality or mate-pairs. The Assembly-Viewer was used to visualize resulting sequence assemblies and to check the coverage by mate-pairs.

7.1.1.3 Prediction of ORFs and peptides from EST data and gene prediction from genomic sequence data

To predict the most probable open reading frames and the corresponding predicted peptide sequence from assembled EST sequence data, the ESTscan program (Iseli et al., 1999) was applied. For each investigated organism, a hidden-markov model file (.smat) including among other things the codon usage and Exon/Intron borders was created on the basis of the corresponding NCBI UniGene dataset. Gene predictions from genomic data were performed using either online prediction programs GenScan (Burge and Karlin, 1997) or HMMgene (Krogh, 1997) .

7.1.1.4 Detection of conserved domains in peptide datasets using HMMs

Conserved domains were detected using SMART (Letunic et al., 2004) or a local install of HMMer (Eddy, 1998). Hidden Markov models for local searches were obtained from PFAM (Sonnhammer et al., 1998) and Superfamily (Gough et al., 2001) databases.

7.1.1.5 Prediction of transcription factor binding sites and promoter conservation profiles

For *in silico* analysis of putative promoter sequences, the publicly available online platforms Consite (Sandelin et al., 2004) and Vista tools (Frazer et al., 2004) were used with standard parameters.

7.1.2 Programs and algorithms used to setup the “Compagen” facility

7.1.2.1 Web server and www-Blast-server

To run a private Blast server within the workgroup, the open source HTTP server Apache v.2.2.3 was installed on a RedHat Linux system. The www-Blast package from NCBI was integrated and the server configuration files were adjusted to suite the custom requirements. To open the server to the public, an accompanying small website was implemented at http://www.compagen.org that governs registration of users and provides basic information about the platform.

7.1.2.2 Datasets on the “Compagen” server

Sequence datasets were downloaded in plain fasta-format from the Ensembl trace server (Hubbard et al., 2007) or from Genbank (Benson et al., 2007) and formatted into Blast searchable databases using the *formatdb* script from NCBI´s toolkit. To govern the variety of different databases on the server, a common naming convention has been developed (see Appendix, Table 1). Every database name consists of three qualifiers. The first qualifier indicates the type of sequences within the database (for example dbWGS = whole genome shotgun). The second qualifier provides the name of the animal species the sequences come from (HMAG = *Hydra magnipapillata*). The third qualifier specifies the date of database construction (070825 = YYMMDD). A detailed list of abbreviations can be found in Appendix, Table 1.

7.1.3 Programs and algorithms used in the EST analysis pipeline

Sequence data as well as sequencing quality data was downloaded in flat-file format from the dbEST section of NCBI GenBank (Benson et al., 2007) or from the ENSEMBL database at the European Institute for Bioinformatics (Hubbard et al., 2007).

Raw sequence data was checked for obstructive vector sequences, low quality values (N-stretches) and adaptor sequences, using the TIGR Gene Indices Sequence Cleaning and Validation Script called *seqclean* from the Institute of Genomic Research (Pertea et al., 2003) with the following parameters: psx; –p 1; -n 2000; -i input database; -d cleaning; -v UniVec and/or AdapDB (databases for vector and adapter sequences). Cleaning steps were repeated until no sequences were excluded from the dataset any more.

Cleaned sequence datasets were subjected to a two-step procedure, including sequence clustering and CAP3 assembly (Huang and Madan, 1999) using the TIGR Gene Indices Clustering Tools (TGICL) (Pertea et al., 2003) with the following parameters: -p 95; -l 80; -v 40. Resulting contigs were checked using *clview,* a program for visualization of assembly-files (ace.) (Pertea et al., 2003). Remaining singletons were extracted using *cdbfasta/cdbyank* scripts.

The assembled contigs, as well as the extracted singletons, were subjected to stepwise batch Blast-searches (Altschul et al., 1990). Further possible sequence analysis steps included domain-searches using SMART (Letunic et al., 2004) and the prediction of putative signal peptides using SignalP (Bendtsen et al., 2004).

For sequences with Blast matches in the first round of homology searches (see Figure 14, step n1), a semiautomatic or fully automatic functional annotation according to general GO terms (Harris et al., 2004) was conducted using Goblet (Groth et al., 2004) and/or AutoFACT (Koski et al., 2005).

7.1.4 Molecular phylogenetic analysis

7.1.4.1 Multiple alignments and substitution models

Sequence alignments were generated using ClustalW (Thompson et al., 1994) included into the BioEdit v.7.053 sequence analysis software package (Hall, 1998). Alignments were optimized by hand and converted into required file-formats (.nex, .phy). FindModel (Tao, 2005) was used to estimate the best-fit substitution models for further phylogenetic analyses. To infer phylogenetic relationships among the taxa, three different analytical methods were conducted. Maximum parsimony (MP), maximum likelihood (ML) and Bayesian inference (BI) methods were used for the dataset of each single gene and for the (by concatenation) combined datasets of nuclear and mitochondrial genes, respectively. Trees were drawn using TreeView 1.6.6 (Page, 1996) and MEGA.

7.1.4.2 Maximum parsimony analyses

Maximum parsimony (MP) analyses were performed using the MEGA 3.1 software package (Kumar et al., 2004). A bootstrap test with 100.000 replicates and random seed was conducted to each analyzed dataset. Gaps were set to complete deletion. All three codon positions plus noncoding characters were included. The datasets were tested using the Close-neighbour-interchange (CNI) method with search level 1. Initial trees for CNI searches were build using the Minimal-Mini Heuristic method with a search factor of 100.

7.1.4.3 Maximum likelihood analyses

Maximum likelihood (ML) analyses were performed using the quartett-puzzling method implemented in Tree-Puzzle 5.2 (Schmidt et al., 2002). The analyses included 100.000 puzzling steps. Exact analysis parameters were estimated from each dataset using quartet sampling and NJ trees. Nuclear genes were tested using the Tamura-Nei substitution model. For testing mitochondrial genes the GTR (General Time Reversible) and the HYK (Hasegawa-Kishino-Yano-85) substitution models were used for CO1 and 16s, respectively.

7.1.4.4 Bayesian Inference analyses

Bayesian inference analyses were carried out using Mr. Bayes v.3.0 (Huelsenbeck and Ronquist, 2001). All analyses were run for 100.000 generations and a sample frequency of 100. Trees were inferred at a burn-in of 250. The datasets were tested using the General Time Reversible (GTR) substitution model with 6 substitution types and gamma-shaped rate variation with a proportion of invariable sites. The gamma distribution was approximated using 4 discrete categories.

7.2 Molecular biology methods

7.2.1 Animal culture

Molecular phylogenetic analyses and Southern blots were carried out with *Hydra magnipapillata*, *Hydra vulgaris* (strain Basel), *Hydra carnea*, *Hydra oligactis*, *Hydra robusta*, *Hydra circumcincta*, *Hydra viridissima* and *Hydra vulgaris* (strain AEP) (Martin et al., 1997; Technau et al., 2003). All experiments concerning the BAC library and FISH were carried out with *Hydra magnipapillata* (strain 105). The animals were cultured in mass cultures according to standard conditions (Lenhoff and Brown, 1970) in an air-conditioned room maintained at 18 ± 0.5°C.

7.2.2 Isolation of DNA

7.2.2.1 Isolation of genomic DNA (standard protocol)

Genomic DNA from up to 500 *Hydra* polyps was isolated by tissue homogenization, using a plastic pestle in 500 µl lysis buffer and *proteinase K* treatment (50 µl, 1 mg/ml) for 2 h at 50°C, followed by phenol/chlorophorm and ethanol-precipitation steps. After a washing step in 75 % ethanol, DNA pellets were resolved in 10 mM Tris (pH 8.0).

7.2.2.2 Isolation of genomic DNA from Hydra viridissima (aposymbiotic)

As the standard DNA isolation method for *Hydra viridissima* yields only small amounts of genomic DNA, an alternative protocol was used. Approximately 1000 polyps were cooled on ice and washed in 0.5 x PBS (pH 7.4) by consecutive centrifugation steps at 2000 g. Tissue was homogenized on a horizontal shaker (650 rpm) for 1 h at room temperature in a 0.5 x PBS/ 10 % *trypsin* solution. The resulting cell suspension was

lysed in 500 µl lysis buffer and 100 µg/ml *proteinase K* for 2 h at 60°C. To the resulting lysate 2 µl DEPC was added and incubated for 30 min at 60°C. Adding 50 µl 5 M potassium acetate and incubation for 30 min on ice was followed by 4 consecutive centrifugation steps at 14.000 g for 15 min at 4°C, always carrying over the supernatant. Ethanol precipitation was carried out using 750 µl absolute EtOH at room temperature.

7.2.2.3 Isolation of High-Molecular-Weight (HMW) DNA

High molecular weight DNA (HMW-DNA) was isolated according to Gindullis et al., 2001, with minor modifications. For extraction of HMW-DNA, cells of about 1000 polyps were used. Polyps were first incubated in dissociation medium (Gierer et al., 1972) which contained 1 mM polyamine and pressed gently through a metal screen of 0.5 mm pore size. The resulting cell suspension was centrifuged at 1200 rpm for 5 min at 4° C. The supernatant was discarded and the pellet was washed two times in washing buffer (dissociation medium, containing 1 mM EDTA pH 8.0, 1 mM polyamine), using cell saver tips. The washed pellet was resuspended in 2 ml of 0.75 % low melting point agarose (InCert Agarose, FMC) and poured into 100 µl plug moulds. Solidified plugs were washed overnight in 10 volumes of lysis buffer (0.5 M EDTA, pH 9.0; 1 % Na-laurylsarcosine; 0.1 mg/ml *proteinase K*) at 50°C. After lysis, plugs were washed for 1 h at 50°C in 10-20 vol EDTA (pH 9.0 – 9.3) and once in 0.05 M EDTA (pH 8.0) on ice. Before use, excess EDTA was removed by washing in ice cold TE buffer containing 0.1 mM PMSF and three 1 h washes in ice cold TE without PMSF. Plugs were stored in TE at 4° C.

7.2.3 Isolation of total RNA and mRNA

Total RNA was isolated using TRIzol DNA/RNA extraction reagent from Invitrogen according to the manufacturers protocol. For isolation of mRNA the QuickPrep mRNA Purification Kit (GE Healthcare) was used according to the manufacturers protocol.

7.2.4 Preparation of *Hydra* chromosomes

Chromosome preparations were obtained using the air-drying method: polyps were treated in a hypotonic 0.4 % sodium citrate solution for 25 min and then fixed in 3:1 (v/v) ethanol-glacial acetic acid for 15–30 min. Fixed polyps were homogenized in 0.1–0.3 ml of 70 % acetic acid. The cell suspension was dropped on pre-warmed (40°C) cleaned slides and dried at 37–40°C.

7.2.5 Isolation of plasmid DNA

To isolate plasmid DNA from bacteria, several Kits - depending on the particular application – were used according to the manufacturers protocol. For small insert clones the “NucleoSpin Quick Pure Kit” (Macherey Nagel) and the “Qiafilter Plasmid Midi Kit” (Qiagen) were used. For isolation of large insert clones (BACs) the “BACMAX DNA Purification Kit” (Epicentre) and the “Qiafilter Plasmid Midi Kit” (Qiagen) with a BAC-specialized protocol were used.

7.2.6 Quantification of DNA

Concentration and purity of isolated DNA was determined by measuring the optical density at 260/280/230 nm using an Eppendorf Bio Photometer or a NanoDrop ND-1000 spectrophotometer.

7.2.7 Polymerase Chain Reaction (PCR)

For standard PCRs, *Illustra Taq DNA polymerase* (GE Healthcare) was used. To amplify larger (up to 10 kb) PCR fragments, the Expand Long Template PCR System from Roche was used.

A typical PCR reaction, used for probe generation or insert check, is summarized in the following Table:

Reagent	**volume [µl]**
10 x PCR buffer	2.5
dNTPs [10 mM each]	0.4
forward primer [10 µM]	1
reverse primer [10 µM]	1
Taq polymerase [5 U/µl]	0.2
template	x
water	fill to final volume
Total	25

7.2.8 Cleaning and extraction of PCR products, dialysis

To clean or gel-extract PCR products, the “NucleoSpin Extract 2 Kit” from Macherey Nagel was used. For dialysis of DNA or to change buffers, “YM-50” columns from Millipore were used.

7.2.9 Agarose and pulse-field gel electrophoresis

To separate PCR fragments according to their expected size, 0.7 % – 1.4 % agarose gels were run on BIORAD horizontal gel electrophoresis devices in 1 x TAE buffer at 70-90 V. Ethidium bromide was used to stain and visualize (UV-illumination) the resulting DNA bands. Depending on the expected fragment sizes, different DNA-size standards were added. These included 100 bp – 1 kb ladders from MBI Fermentas, as well as low range markers from New England BioLabs. For fingerprinting BAC clones or for separation of high molecular weight DNA a BIORAD CHEF-2 pulse-field gel electrophoresis system was used according to standard protocols.

7.2.10 Cloning techniques

7.2.10.1 Cloning of PCR products

PCR products were purified as described above and ligated into the pGEM-T vector system (Promega) as described by the manufacturer. Ligation mixtures were dialyzed against water using nitrocellulose filter membranes (0,025 µm) from Millipore to increase the transformation efficiency. Electro competent cells of *E. coli* strains XL-1 blue or DH-10B were transformed (BIORAD electroporation device, 1 mm cuvette) using up to 15 µl ligation mixture and blue/white selected on ampicillin containing LB-plates. Positive (white) colonies were picked and subjected to insert check PCR. Selected clones were cultured over night in 3 ml LB-amp$^+$ medium and plasmids were isolated as described above.

7.2.10.2 Bacterial artificial chromosome (BAC) library construction

Digestion of HMW-DNA was done as described by Gindullis et al., (2001). *HindIII* restriction fragments were resolved by pulse-field electrophoresis (PFGE). Zones with restriction fragments between 80 and 150 kb were cut out of the gel and electro eluted following standard procedures (Gindullis et al., 2001). Size fractionated HMW-DNA (100 ng) was ligated into 25 ng of pCC1H BAC vector using the Copy Control BAC Cloning Kit (Epicentre) according to the manufacturer's manual. The mixture was transformed in competent DH10B cells (Invitrogen) by electroporation. Cells were plated on LB agar plates containing 12.5 µg/ml chloramphenicol, 25 µg/ml isopropylthiogalactoside (IPTG) and 50 µg/ml X-GAL (5-bromo-4-chloro-3-indolyl-b-D-galactoside). Bacterial colonies were picked by the QPix2 automated colony picker (Genetix) in 144 x 384 micro titer well plates, grown overnight and stored at –80°C.

7.2.11 DNA sequencing

Sequencing of cDNA or genomic DNA was performed either on a LI-COR 4300 plate sequencer or using a MegaBace 1000 capillary sequencing system.

Sequencing reactions for the LI-COR system were carried out using the SequiTherm EXCEL II DNA sequencing kit LC (Epicentre) according to the manufacturer's protocol. The 6 % polyacrylamide gel contained 14 ml H_2O, 3.75 ml 40 % Rapid Gel XL solution, 2.5 ml 10 x TBE long run buffer, 10.5 g urea, 38 µl TEMED and 175 µl ammonium persulfate (APS). 5'-IRD700 or IRD800 labeled sequencing primers were used for standard sequencing runs. Basecalling and quality assessment was conducted using e-Seq software package provided with the sequencer.

Sequencing reactions for the MegaBace 1000 capillary system were carried out using the DYEnamic ET Terminator Cycle sequencing kit (GE Healthcare) following the manufacturer's protocol. Amersham Bioscience provided the required sequencing matrix and buffers as well as basecalling and sequence assembly software.

7.2.12 Hybridization techniques

7.2.12.1 Southern Blot Hybridization

20 µg of isolated DNA was digested with 4 Units of *HindIII* and *XbaI* respectively. Nucleic acids were transferred to Hybond N+ nylon membranes (Amersham Biosciences). Hybridizations were carried out over night in Church buffer at 55°C, followed by washes in 0.2 x SSC / 0.1 % SDS at room temperature, 42°C and 60°C for 2 x 30 minutes, depending on the signal/ background ratio. Autoradiographies were performed using imaging-plates for the Phosphoimager FLA-5000 (FUJI). DNA-probes were radioactively labelled with P–[32P]-dCTP using the Megaprime DNA labelling System (Amersham Biosciences).

7.2.12.2 Screening of the Hydra magnipapillata BAC library

For the identification of BAC clones containing specific genes, BAC colonies on Performa high-density membrane (Genetix) were hybridized with radiolabeled probes following conventional Southern blotting protocols. Positive clones were picked from 348-well plates and pre-cultured in 3 ml 2 x YT medium at 37°C over night. BAC DNA was isolated as described above. For restriction digestion fingerprints, 1 µg BAC DNA was digested with 10 U *HindIII*, *XbaI* and *NotI* at 37° C for 5 h, and loaded on a 0.7 % agarose pulsefield gel. DNA bands were transferred on a N+ nylon membrane and hybridized with the radiolabeled specific probe. Selected clones from fingerprints were sequenced by primer-walking directly on complete BAC clones using the Amersham Bioscience MegaBACE 1000 capillary sequencer. For BAC-end sequences, the LICOR plate sequencer was used.

7.2.12.3 Fluorescence in situ hybridization (FISH)

FISH was carried out on *Hydra* chromosomes, using probes labeled by random primer labeling with biotin according to the manufacturer's (Roche) protocol. In situ hybridization was performed as described by Schwarzacher and Heslop-Harrison (Schwarzacher, 2003) with some modifications. Chromosome preparations were treated with 100 µg/ml *RNaseA* (Sigma) for 50 min at 37°C in humid chamber, washed two times in 2 x SSC (5 min each) at 37°C, incubated in 0.01 % *pepsin* in 0.01 N HCl for 10 min at 37°C, washed in 1 x PBS for 1 min at RT and in 2 x SSC for 5 min at 37°C, dehydrated through an ethanol series (ice cold, 70 %, 90 % and 99 %, 2 min each) and

finally, dried. After pretreatment, preparation were mounted using frames for in situ hybridization (Peqlab) at 40°C with 90 µl pre-denatured (5 min at 96°C) hybridization solution, containing 150–200 ng of labeled genomic DNA, 25 % formamide, 4 × SSC, 10 % (w/v) dextran sulfate, 0.15 % (w/v) sodium dodecyl sulfate (SDS) and 2 µg salmon-sperm DNA. The slides were placed in a prewarming histological table and denaturated for 5 min at 68–69°C. The temperature was gradually reduced to 37–40°C, and the chromosome slides were incubated for 42–44 h at 37°C. Following hybridization, the slides were washed three times in 2 x SSC (5 min each) at 37°C, three times in 0.5 x SSC (3 min each) at 43°C, incubated in detection buffer (4 x SSC/ 0.1-0.2 % Tween) for 2 min at 37°C and blocked in 2 % (w/v) BSA/ 4 x SSC/ 0.2 % Tween for 25 min at 37°. Probes were detected with 13.5 µg/ml avidin or streptavidin, conjugated to FITC (Sigma). Detection reaction was performed in 2 % BSA/ 4 x SSC/ 0.1 % Tween for 1 h at 37°C. Slides were washed three times in 4 x SSC/ 0.1 % Tween (5 min each) at 37°C and rinsed in 1 x PBS at 37°C. Chromosomes were contrasted with 1 µg/ml DAPI and mounted in an antifade solution containing 60 % Glycerol in 1 x PBS and 5 % DABCO (Sigma).

8 MATERIALS

8.1 Servers, Software and Web resources

8.1.1 Servers

System	Description
Servers	DELL PowerEdge 1800, Dual Xeon 3 GhZ, 2 Gb RAM, 1x 73 Gb + 2x 146 Gb SAS HDDs
	Dual Xeon 2,8 GhZ, 2 Gb RAM, 2x 250 Gb S-ATA HDDs
NAS	Lacie Ethernet Disk, 2 Tb S-ATA storage

8.1.2 Software

Microsoft Windows OS	
BioEdit 7.0.5.3	http://www.mbio.ncsu.edu/BioEdit/bioedit.html
eSeq v.3	LI-COR Inc., Michigan Technology University
MegaBACE Sequence Analyzer 3.0	Amersham Biosciences, 2001.
Mega 3.1	http://www.megasoftware.net/index.html
seqtools 8.4	http://www.seqtools.dk
Unix based OS	
AAT	http://genome.cs.mtu.edu/aat/aat.html
AMOS / minimus	http://amos.sourceforge.net/
Apache webserver 2.2.4	http://httpd.apache.org/
CLC Workbench 3.2	http://www.clcbio.com
clview	http://www.tigr.org/tdb/tgi/software/
EMBOSS 4.1	http://emboss.sourceforge.net/
ESTScan 2-2.1	http://estscan.sourceforge.net/
HMMer 2.3.2	http://hmmer.janelia.org/
Mr. Bayes 3.1.2	http://mrbayes.csit.fsu.edu/
MUMmer 3.18	http://mummer.sourceforge.net/
MySQL 4.1	http://www.mysql.com/
ncbi toolbox	http://www.ncbi.nlm.nih.gov/IEB/ToolBox/index.
openSputnik 1.0.2	http://sourceforge.net/projects/opensputnik
Pasa	http://pasa.sourceforge.net/
PostgreSQL 7.4.16	http://www.postgresql.org/
seqclean	http://www.tigr.org/tdb/tgi/software/
tgicl	http://www.tigr.org/tdb/tgi/software/
TreePuzzle 5.2	http://www.tree-puzzle.de/
wwwBlast	http://www.ncbi.nlm.nih.gov/BLAST/download.s

8.1.3 Web resources

Name	URL
BCM Genome browser	http://www.hgsc.bcm.tmc.edu/projects/
Consite	http://mordor.cgb.ki.se/cgi-bin/CONSITE/consite
DDBJ	http://www.ddbj.nig.ac.jp/Welcome.html.en
EBI	http://www.ebi.ac.uk/
ENSEMBL	http://trace.ensembl.org/
FindModel	http://hcv.lanl.gov/content/hcv-
GenScan	http://genes.mit.edu/GENSCAN.html
GO Browser AmiGO	http://amigo.geneontology.org/cgi-bin/amigo/go.cgi
Graphical codon usage analyzer	http://gcua.schoedl.de/
Hydrabase	http://www.hydrabase.org
JGI Genome Browsers	http://www.jgi.doe.gov/
NCBI	http://www.ncbi.nlm.nih.gov/
PFAM	http://www.sanger.ac.uk/Software/Pfam/
PRAline	http://zeus.cs.vu.nl/programs/pralinewww/
Repbase	http://www.girinst.org/repbase/index.html
SignalP 3.0	http://www.cbs.dtu.dk/services/SignalP/
SMART	http://smart.embl-heidelberg.de/
Stellabase	http://www.stellabase.org
Superfamily	http://supfam.mrc-lmb.cam.ac.uk/SUPERFAMILY/
TFsearch	http://www.cbrc.jp/research/db/TFSEARCH.html
TMHMM 2.0	http://www.cbs.dtu.dk/services/TMHMM/
Transfac 7.0	http://www.gene-
Vista tools	http://genome.lbl.gov/vista/index.shtml

8.2 Media, buffers and solutions

Name	Ingredients
2xYT medium	16 g Tryptone, 10 g Yeast extract, 5 g NaCl, add H_2O to 1 l, autoclave
Denaturation/ Transfer buffer	0,4 M NaOH; 1 M NaCl
Denhardt´s (50 x)	1 % Polyvinylpyrrolidone, 1 % Ficoll, 1 % BSA fraction V, filter sterile, -20°C
Dilution buffer	Amersham Biosciences
DNA loading buffer (6x)	50 % Glycerol, 10 mM EDTA pH 8, 0.1 % SDS, 0.025 % Bromphenolblue, 0.025 % Xylencyanol
Freezing medium	10 g Tryptone, 5 g Yeast extract, 20 g NaCl, 6.3 g K_2HPO_4, 1.8 g KH_2PO_4, 0.5 g Sodium-Citrate, 98.8 g $MgSO_4$, 0.9 g $(NH_4)_2SO_4$, 51.2 ml 86% Glycerol, add ddH_2O to 1 l, autoclave
Hybridization buffer (Southern blots)	6 x SSC, 5 x Denhardt´s, 0.5 % SDS, sonicated salmon sperm (conc.?)
Hybridization buffer (BAC library)	5 x SSPE, 5 x Denhardt´s, 0.2 % SDS, sonicated salmon sperm
LB^{+amp} medium	20 g LB Broth Base, 15 g Bactoagar, add H_20 to 1 l, autoclave, add 1:1000 ampicillin stock (100 mg/ml)
M-solution	1 mM $CaCl_2$, 1 mM NaCl, 0,1 mM $MgSO_4$, 0,1 mM KCl, 1 mM Tris-HCl, pH 7.8
SOC medium	10 ml SOB medium, 100 µl 2 M glucose
SOB medium	20 g Bactotryptone, 5 g Yeast extract, 0.58 g NaCl, 0.19 g KCl in 1 l, 100 µl $MgCl_2$, 10mM $MgSO_4$
SSC (20 x)	3 M NaCl, 0.3 M Sodium-citrate, pH 7.0
SSPE (20 x)	3 M NaCl, 0.2 M Na_2HPO_4, 0.02 M EDTA, pH 7.4
TAE (50 x)	2 M Tris-HCl, 0.05 M EDTA, pH 8.0
TBE long run buffer (10 x)	1340 mM Tris-HCl, pH 7.5; 450 mM boric acid; 25 mM EDTA, pH 8.3 – 8.7
TE	10 mM Tris-HCl, pH 7.5, 1 mM EDTA pH 8.0
Washing buffer 1	2 x SSC, 0.1 % SDS
Washing buffer 2	1 x SSC, 0.1 % SDS
Washing buffer 3	0.2 x SSC, 0.1 % SDS

8.3 Chemicals, Antibiotics, Enzymes and Kits

Product	Manufacturer
Chemicals	
α-p32 dCTP / dATP	GE Healthcare
Acetic acid	Roth
Agarose	Roth
Agarose (InCert)	Sigma
Agar-Agar	Roth
Ammonium acetate	Roth
APS	Roth
Bactoagar	Roth
Bactotryptone	Roth
Boric acid	Roth
$CaCl_2$	Sigma
Chlorophorm	Roth
Dimethylformamide	Merck
DNA size standarts (100 bp, 1 kbp)	Fermentas
dNTPs (100 mM)	Fermentas
Ethanol	Roth
EDTA	Sigma
Ethidium bromide (50 ng/ml)	Merck
Formaldehyde	Merck
Formamide	Roth
Glucose	Roth
Glycerol	Roth
HCl	Roth
IPTG	Sigma
Isopropanol	Roth
Low range marker	NEB
Methanol	Roth
$MgCl_2$	Roth
$MgSO_4$	Roth
Na_2HPO_4	Roth
NaH_2PO_4	Roth
Paraformaldehyde	Merck
Phenol	Roth
PMSF	Sigma

RapidGEL-XL 40% concentrate	USB
Sea salt (Ocean Zac plus)	Zoo Zajac
Sephadex G50	Pharmacia
Sodium acetate	Roth
Sodium azide	Roth
Sodium chloride	Roth
Sodium citrate	Roth
Sodium hydroxide	Roth
Sodium pyruvate	Roth
Sodium larylsarcosine	Roth
Sonicated salmon sperm DNA	Invitrogen
Spermine / Spermidine	Sigma
TEMED	Merck
Tris base	Roth
Tris HCl	Roth
Tryptone	Roth
Urea	Roth
XGal	Sigma
Yeast extract	Roth

Antibiotics	
Ampicillin	Fluka
Chloramphenicol	Fluka

Enzymes	
Taq polymerase	GE Healthcare
Platinum Taq polymerase	Invitrogen
Platinum Taq polymerase high fidelity	Invitrogen
T4 DNA Ligase	NEB
Klenow fragment	Fermentas
Proteinase K	Fermentas
SAP (alkaline phosphatase)	Fermentas
Hind III	Fermentas
EcoRI	Fermentas
BamHI	Fermentas
Not I	Fermentas

Kits	
DYEnamic ET Terminator Cycle Sequencing Kit	Amersham Biosciences
SequiTherm EXCEL™ II DNA Sequencing Ki	Epicentre
QIAfilter Plasmid Midi Kit	Qiagen
CopyControl™ BAC Cloning Kits	Epicentre
BACMAX™ DNA Purification Kit	Epicentre
GELase™ Agarose Gel-Digesting Preparation	Epicentre
NucleoSpin Plasmid quick pure	Macherey & Nagel
NucleoSpin Extract II	Macherey & Nagel
pGEM®-T Easy Vector	Promega
Microcon YM-50 columns	Millipore

8.4 Laboratory machines and devices

PCR Thermocycler	
Cyclone gradient	peqLab
Primus 96 *plus*	MWG-Biotech
Primus 25	MWG-Biotech
Gelelectrophoresis	
various AGE chambers	BioRAD
CHEF-3 PFGE system	BioRAD
Incubators / Shakers	
HIS25	Grant Boekel
KS1 (rotator)	Edmund Bühler
Mini 10	Thermo Hybaid
Thermo-Incubator	Heraeus Instruments
Thermomixer compact	Eppendorf
Thermomixer Certomat H	B. Braun Biotech
Ultratemp 2000	Sternkopf
UV devices	
Imaging-system	Biorad
ImaGo compact imaging system	B+L Systems
UV-Stratalinker 1800	Stratagene

Electroporation devices	
Gene Pulser II	Biorad
Pulse Controller II	Biorad
Centrifuges	
Centrifuge 5415 D	Eppendorf
Centrifuge 5417 R (cooling centrifuge)	Eppendorf
Minifuge RF	Heraeus Instruments
Labofuge 1	Heraeus Instruments
Sequencer	
LICOR DNA Analyzer Gene Read IR 4200	LiCOR Biotech
LiCOR DNA Analyzer 4300	LiCOR Biotech
MegaBACE 1000 capillary sequencer	Molecular Dynamics
Other material	
384 well microtiter plates	Genetix
96 well microtiter plates	MJ Research
Seal-Mats for 96 well plates	BioZym
Q-trays	Genetix
Electroporation cuvettes	Peqlab
Imager plate eraser	Raytest
BioPhotometer	Eppendorf
Nylon membrane	Amersham
Chromatographic paper	Whatman
Milli-Q Academic System	Millipore
Phospho Imager FLA-5000	FUJI
Phospho Imagingplates	FUJI
VARIOKLAV autoclave Typ 400 EV	H+P Labortechnik GmbH
pH-Meter pH 211	Hanna Instruments
Vortex Genie 2	Scientific Industries
Plastic labware	Sarstedt, Eydam, Eppendorf
Dialysis membranes	Millipore
Dialysis tubes	Medicell Int. Ltd.
Xray films (Hyperfilm)	Amersham
BAC filters (Performa II)	Genetix

8.5 Oligonucleotides

LSU/28s primers:

28SrRNA_F 5'-GCTAAGCTTTGACGAGTAGG-3',

28SrRNA_R 5'-CTGCCACAAGCCAGTTATC-3';

SSU/18s primers:

18SrRNA_F 5'-GATCCTGCCAGTAGTCATATG-3',

18SrRNA_R 5'-GAGTCAAATTAAGCCGCAGG-3';

16s primers:

16SrRNA_F 5'-GGATGCAGTAACTCTGACTG-3',

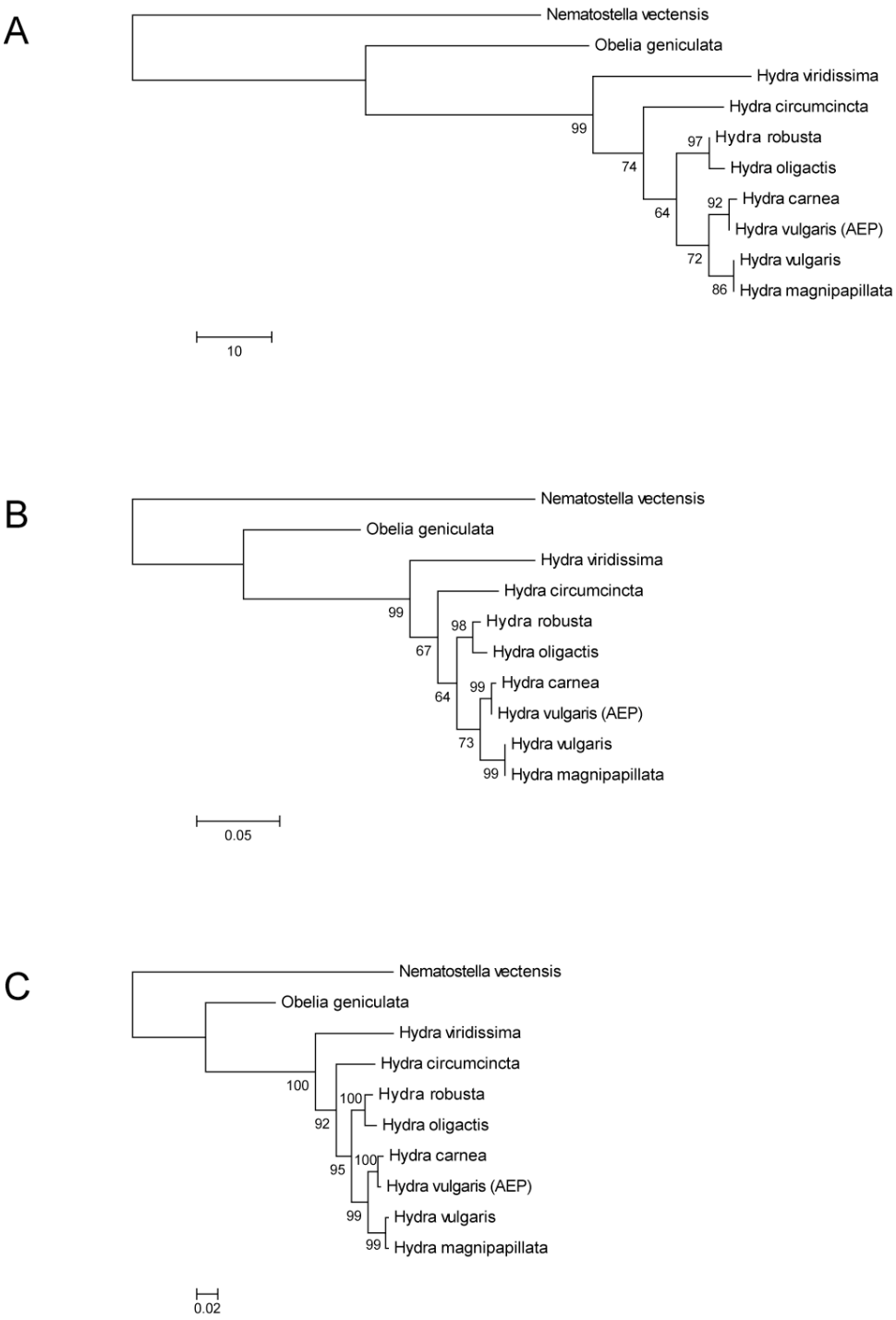
16SrRNA_R 5'-CCTGTTATCCCTAAGGTAGC-3';

CO1 primers:

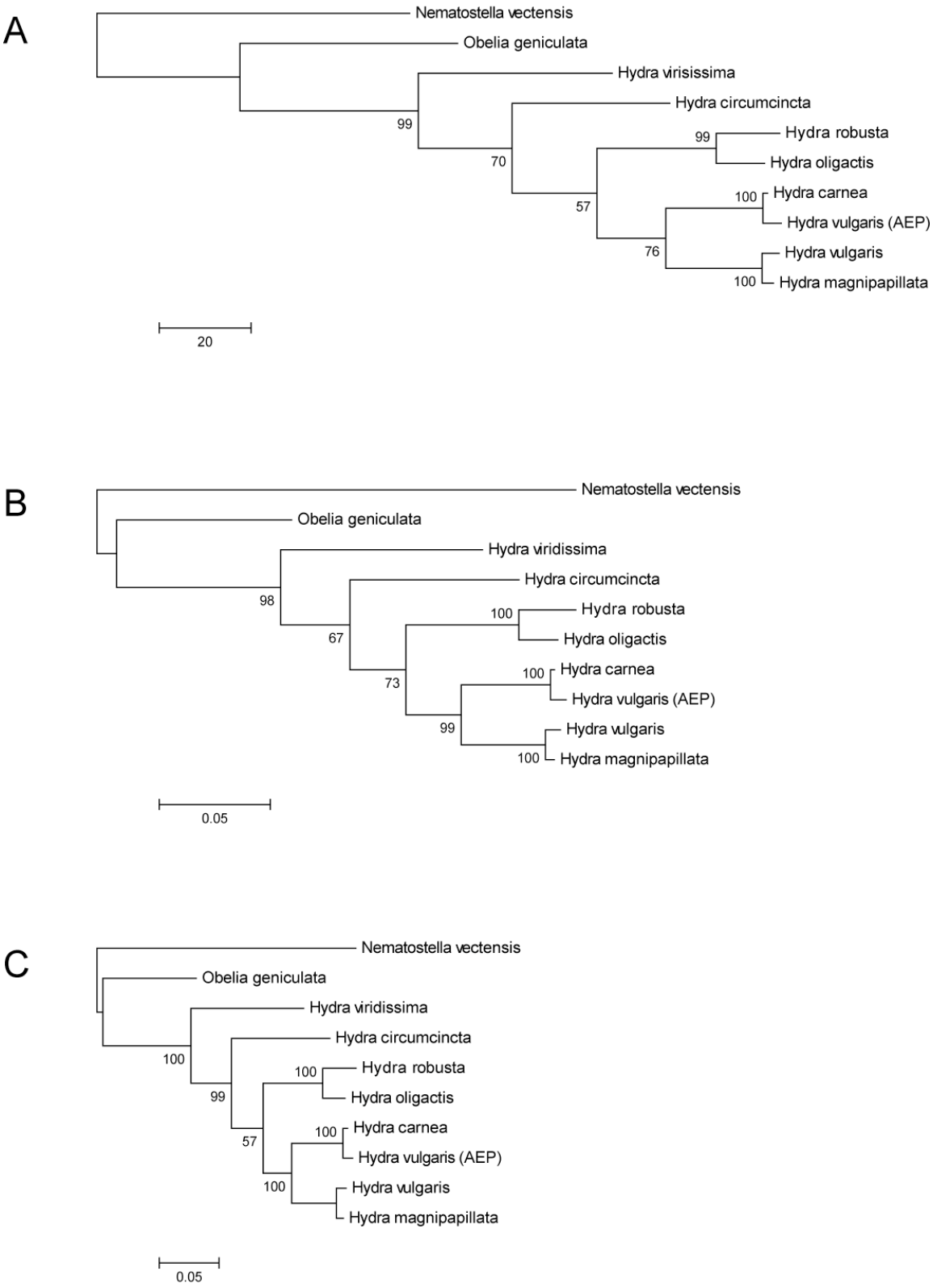
CO1_F 5'-GGATGCAGTAACTCTGACTG-3',

CO1_R 5'-CTATCAGTTAGTAGCATAGTTAT-3'.

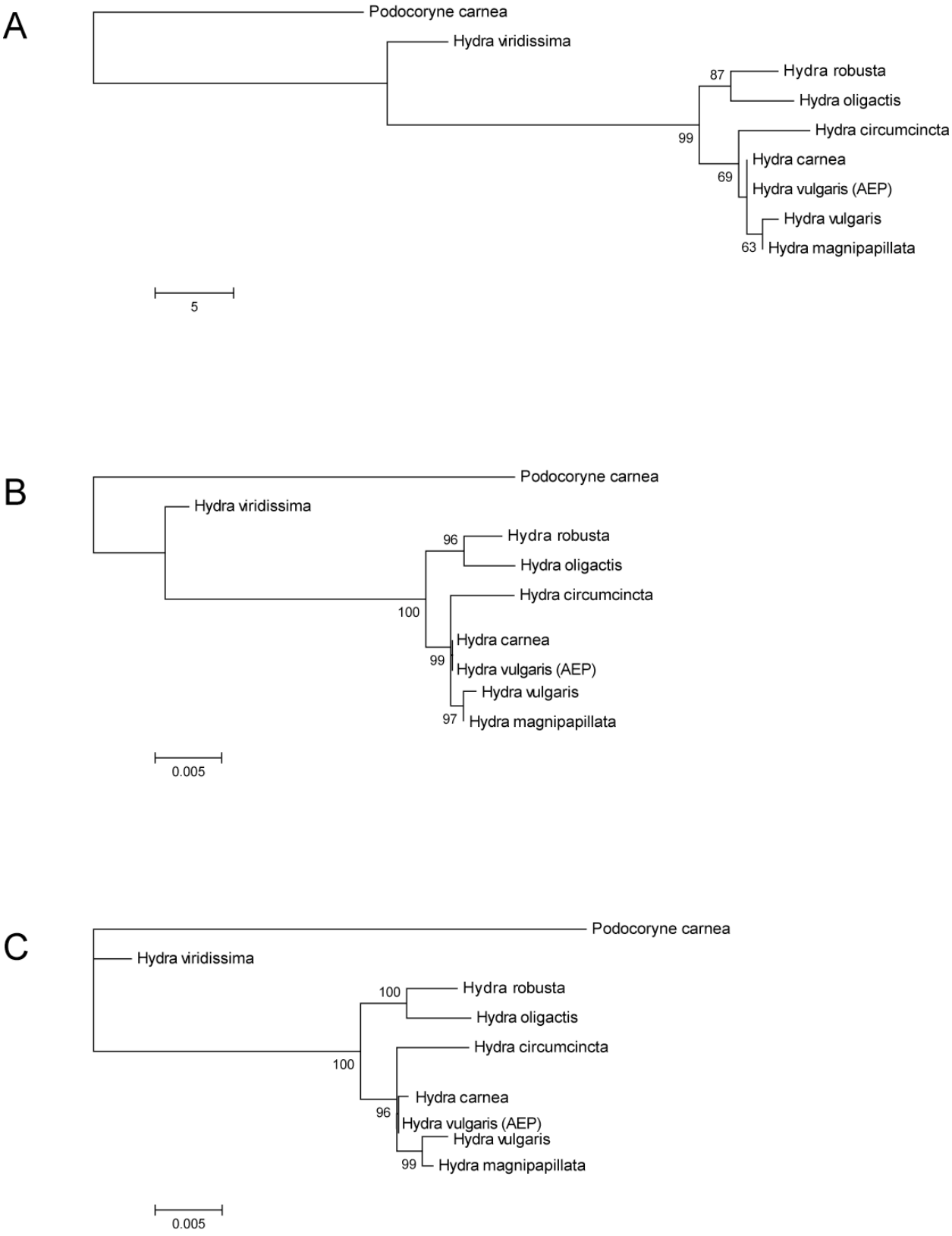
9 APPENDIX



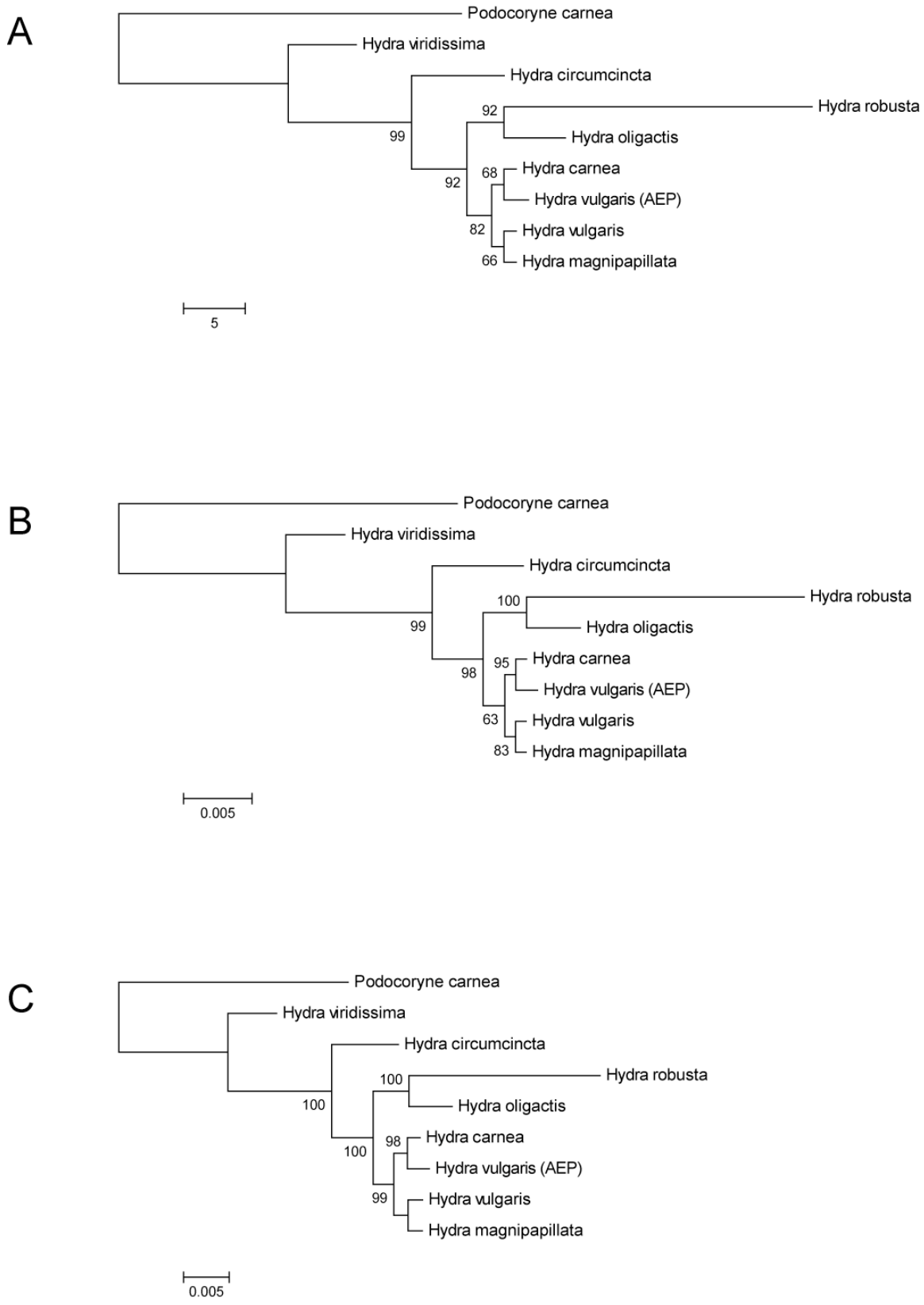
Appendix Figure 1: *Phylogenetic trees resulting of A) maximum parsimony (MP), B) maximum likelihood (ML) and C) Bayesian inference (BI) analyses for the mitochondrial 16s rRNA gene. Bootstrap values for MP and ML as well as Bayesian posterior probabilities are given at the nodes. Branch lengths are scaled to the corresponding indicator bars that display substitutions per site. Figure taken from Hemmrich et al., 2006.*



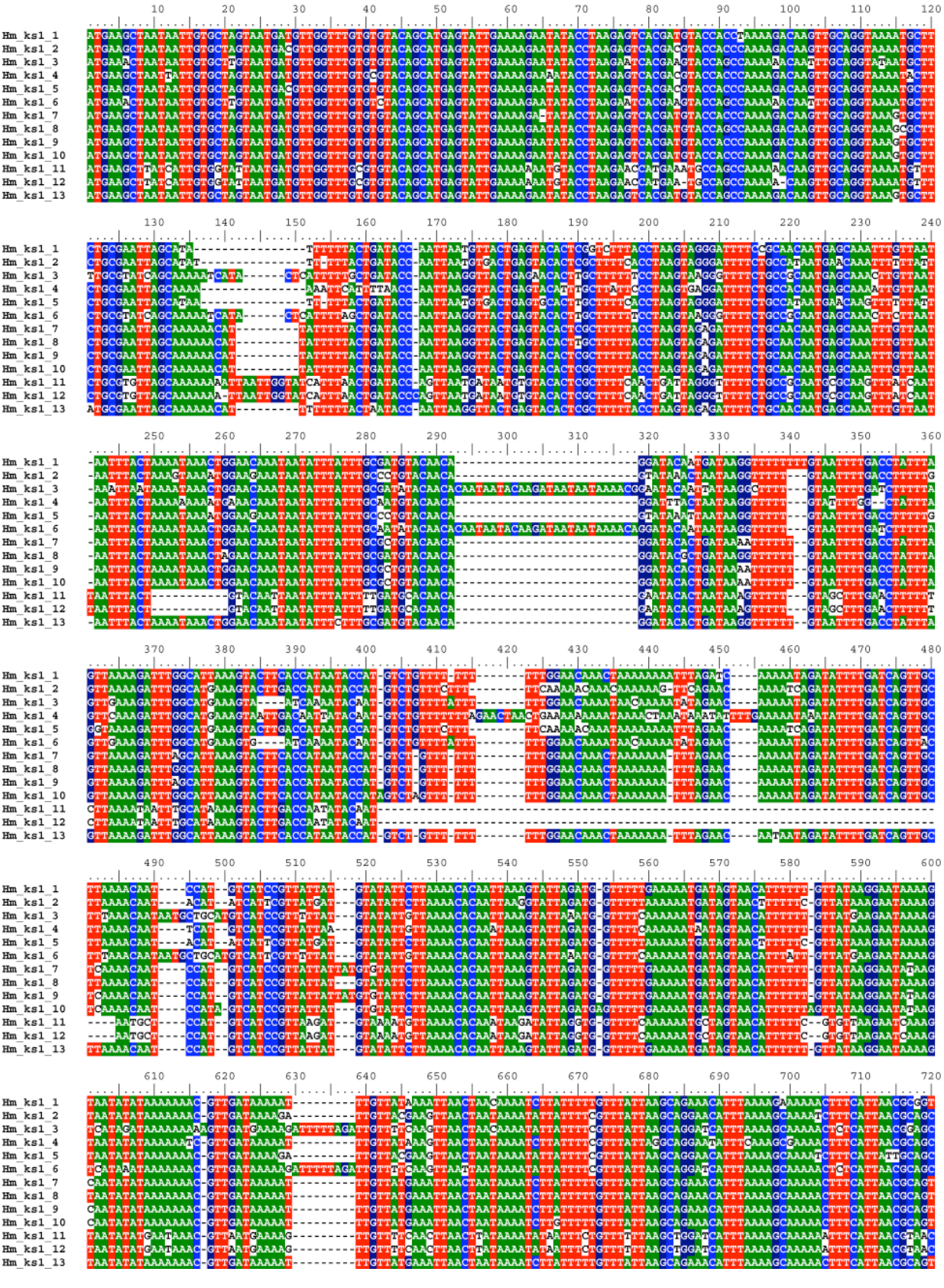
Appendix Figure 2: *Phylogenetic trees resulting of A) maximum parsimony (MP), B) maximum likelihood (ML) and C) Bayesian inference (BI) analyses for the mitochondrial cytochrome oxidase 1 (CO1) gene. Bootstrap values for MP and ML as well as Bayesian posterior probabilities are given at the nodes. Branch lengths are scaled to the corresponding indicator bars that display substitutions per site. Figure taken from Hemmrich et al., 2006.*

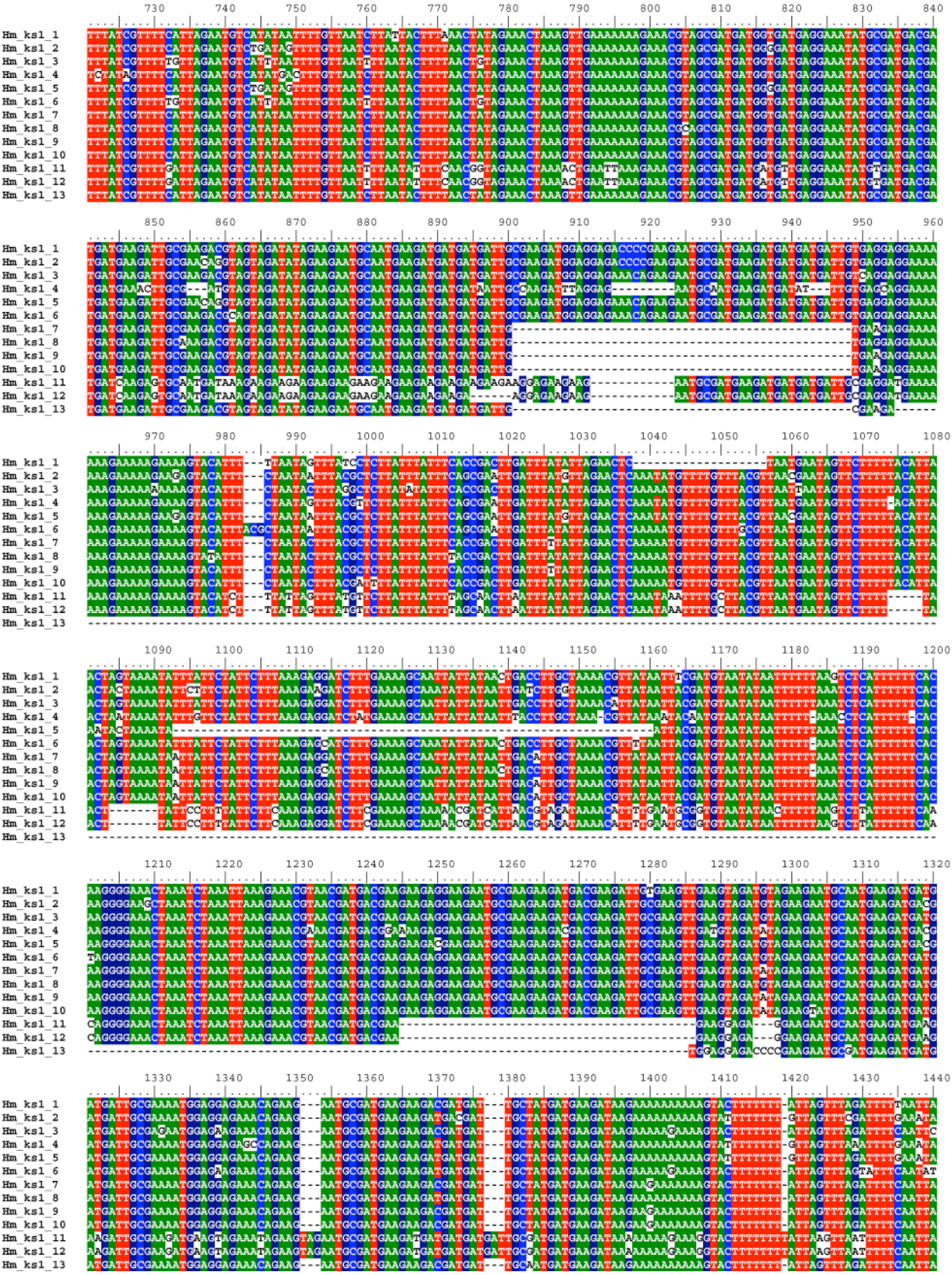


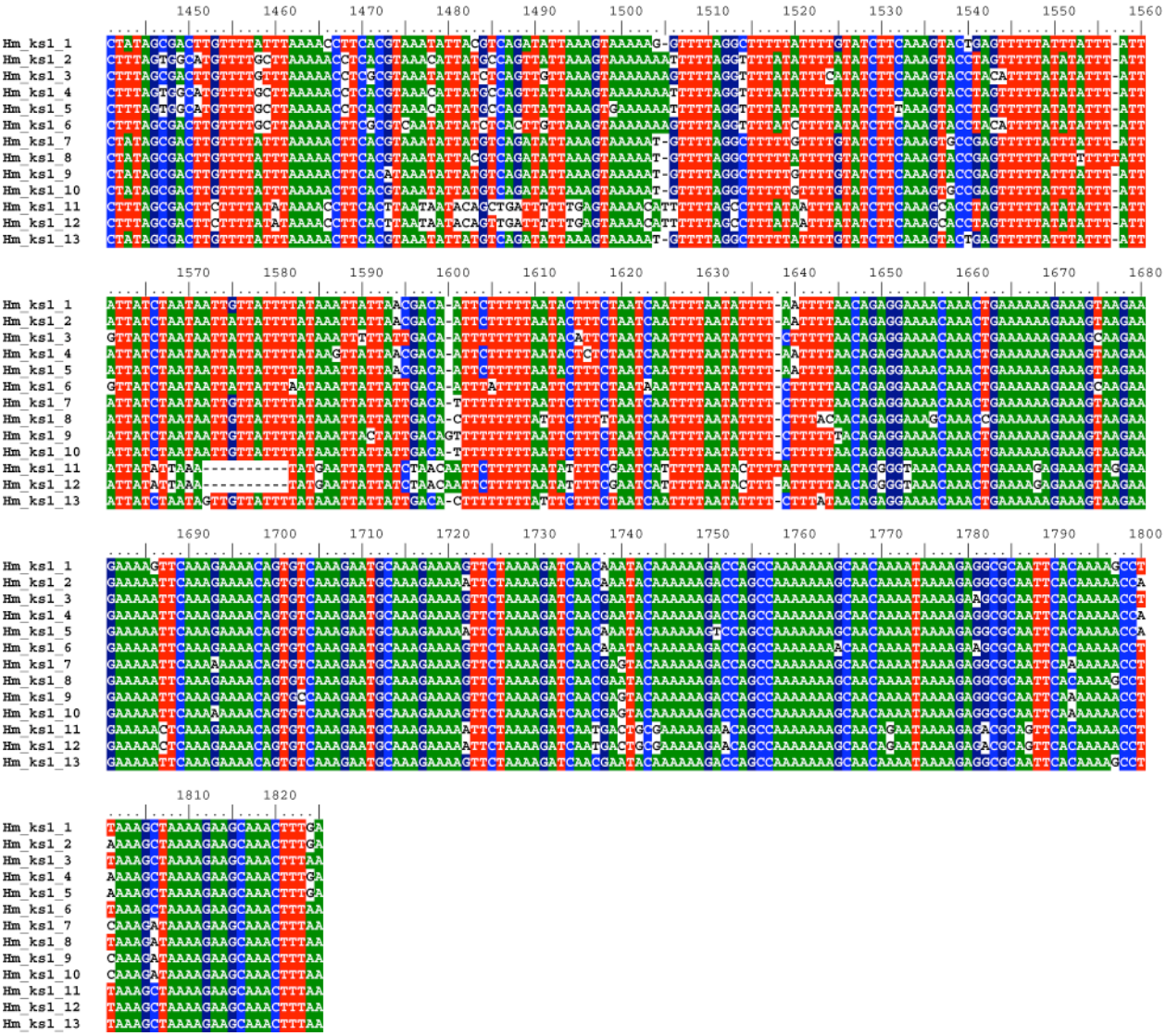
Appendix Figure 3: *Phylogenetic trees resulting of A) maximum parsimony (MP), B) maximum likelihood (ML) and C) Bayesian inference (BI) analyses for the nuclear 18s rRNA gene. Bootstrap values for MP and ML as well as Bayesian posterior probabilities are given at the nodes. Branch lengths are scaled to the corresponding indicator bars that display substitutions per site. Figure taken from Hemmrich et al., 2006.*



Appendix Figure 4: *Phylogenetic trees resulting of A) maximum parsimony (MP), B) maximum likelihood (ML) and C) Bayesian inference (BI) analyses for the nuclear 28s rRNA gene. Bootstrap values for MP and ML as well as Bayesian posterior probabilities are given at the nodes. Branch lengths are scaled to the corresponding indicator. Figure taken from Hemmrich et al., 2006.*


Hm_ksl_1
Hm_ksl_2
Hm_ksl_3
Hm_ksl_4
Hm_ksl_5
Hm_ksl_6
Hm_ksl_7
Hm_ksl_8
Hm_ksl_9
Hm_ksl_10
Hm_ksl_11
Hm_ksl_12
Hm_ksl_13


730 740 750 760 770 780 790 800 810 820 830 840
Hm_ksl_1
Hm_ksl_2
Hm_ksl_3
Hm_ksl_4
Hm_ksl_5
Hm_ksl_6
Hm_ksl_7
Hm_ksl_8
Hm_ksl_9
Hm_ksl_10
Hm_ksl_11
Hm_ksl_12
Hm_ksl_13
850 860 870 880 890 900 910 920 930 940 950 960
970 980 990 1000 1010 1020 1030 1040 1050 1060 1070 1080
1090 1100 1110 1120 1130 1140 1150 1160 1170 1180 1190 1200
1210 1220 1230 1240 1250 1260 1270 1280 1290 1300 1310 1320
1330 1340 1350 1360 1370 1380 1390 1400 1410 1420 1430 1440



Appendix Figure 5: MAFFT alignment of 13 Hydra magnipapillata ks1 genes. *Included in the alignment are coding sequences plus introns. The alignment served as basis for the NJ analysis of ks1 relationships.*

Database naming convention for *www.compagen.org*		
<*database type*_*species code*_*date of construction*>		
database type:		**Abrev.:**
Expressed sequence tags		**dbEST**
Whole genome shotgun		**dbWGS**
Cap3 assembled sequences		**dbCAP3**
ESTScan predicted UniGenes		**dbUNI**
Predicted peptides		**dbPEP**
Gene models		**dbGMOD**
Whole genome assemblies		**dbASM**
species code:	**Common name:**	
Hydra magnpapillata	Freshwater polyp	**HMAG**
Hydra magnipapillata sf-1		**HMAGsf1**
Hydra vulgaris		**HVUL**
Hydra AEP		**HAEP**
Hydra viridissima		**HVIR**
Nematostella vectensis	Starlet sea anemone	**NVEC**
Acropora millepora	Coral	**AMIL**
Aropora palmata	Elkhorn coral	**APAL**
Porites lobata	Lobe coral	**PLOB**
Hydractinia echinata	Colonial hydroid	**HECH**
Montastrea faveolata	Coral	**MFAV**
Biomphalaria glabrata	Bloodfluke	**BGLA**
Aplysia californica	California sea hare	**ACAL**
Daphnia pulex	Water flea	**DPUL**
Daphnia magna	n.a.	**DMAG**
Litopenaeus vannamei	Pacific white shrimp	**LVAN**
Penaeus monodon	Black tiger shrimp	**PMON**
Strongylocntrotus purpuratus	Sea urchin	**SPUR**
Petromyzon marinus	Sea lamprey	**PMAR**
Monosiga brevicollis	Choanoflagellate	**MBRE**
Monosiga ovata	n.a.	**MOVA**
Trichoplax adherens	Placozoan	**TADH**
Reniera sp.	Demosponge	**RENI**
Molgula tectiformis	Tunicate	**MTEC**
Branchiostoma floridae	Florida lancelet	**BFLO**
date of construction:		**YYMMDD**

Appendix Table 1: *Overview of the "Compagen" database naming convention*

Additional sequences identified but not further investigated:	**Accn.#**
***Nematostella* TIR-domain encoding sequences:**	
predicted protein encoding 2 TIR domains	gnl|ti|613621229
predicted protein encoding 1 TIR domain and 2 ARM domains	gnl|ti|595419898
predicted protein encoding 1 TIR, 2 TM, 1 ANK and 3 Ig domains	gnl|ti|571936680
***Acropora palmata* TIR-domain encoding sequences:**	
predicted protein similar to *Nematostella* IL-1R like protein	gnl|ti|824028928
	gnl|ti|824031090
***Nematostella* MAC/PF domain encoding sequences:**	
predicted protein encoding a putative homolog of the spondin gene (MAC/PF and Spondin domain)	gnl|ti|557738010

Appendix Table 2: *Accession numbers for additional sequences identified within the database searches that were not further characterized in the present study. Table taken from Miller and Hemmrich et al., 2007.*

10 LIST OF PUBLICATIONS:

Printed or "in press":

Miller DJ, **Hemmrich G**, Ball EE, Hayward DC, Khalturin K, Funayama N, Agata K, Bosch TC. The innate immune repertoire in Cnidaria – ancestral complexity and stochastic gene loss. *Genome Biol. 2007 Apr 16;8(4):R59 [Epub ahead of print]*

Hemmrich G, Anokhin B, Zacharias H, Bosch TC. Molecular phylogenetics in *Hydra*, a classical model in evolutionary developmental biology. *Mol Phylogenet Evol. 2006*; in press.

Kurn U, Sommer F, **Hemmrich G**, Bosch TC, Khalturin K. Allorecognition in urochordates: Identification of a highly variable complement receptor-like protein expressed in follicle cells of *Ciona*. *Dev Comp Immunol. 2007; 31(4):360-71*.

Augustin R, Franke A, Khalturin K, Kiko R, Siebert S, **Hemmrich G**, Bosch TC. Dickkopf related genes are components of the positional value gradient in *Hydra*. *Dev Biol. 2006; 296(1):62-70.*

Genikhovich G, Kurn U, **Hemmrich G**, Bosch TC. Discovery of genes expressed in *Hydra* embryogenesis. *Dev Biol. 2006; 289(2):466-81*.

Submitted or in preparation:

Khalturin K, Anton-Erxleben F, Milde S, Plötz C; Wittlieb J, **Hemmrich G**, Bosch TC. Transgenic stem cells in *Hydra* reveal an early evolutionary origin for key elements controlling self-renewal and differentiation. Submitted.

Miller DJ, **Hemmrich G**, Bosch TC. The evolution of immunity - A low life perspective. Submitted.

Hemmrich G, Anokhin B, Khalturin K, Bosch TC. Characterization of the taxon-specific ks1-gene family in the basal metazoan *Hydra*. In preparation.

Bosch TC, Zill H, Augustin R, **Hemmrich G**, Kremling S, Röhlk C, Anton-Erxleben F, Leippe M, Stanisak M, Grötzinger J, Jung S, Podschun R, Bartels J, Harder J, Schröder JM. Uncovering ancestral innate immunity: the simple metazoan *Hydra* uses epithelial cells for host defense. In preparation.

Bosch TC, **Hemmrich G**, Khalturin K, Milde S. Differentiation of a taxon-specific cell type requires taxon-specific genes. In preparation.

Khalturin K, Anton-Erxleben F, Wittlieb J , **Hemmrich G**, Bosch TC. Molecular control of species-specific morphologies at the basis of animal evolution: Identification of genes controlling morphological variation in *Hydra*. In preparation.

11 DECLARATION

I confirm that this is my own work and the use of all material from other sources has been properly and fully acknowledged.

Hiermit versichere ich, dass ich die vorliegende Dissertation eigenständig verfasst und keine anderen als die angegebenen Hilfmittel und Quellen benutzt habe.

Teile dieser Arbeit wurden bereits veröffentlicht oder zur Publikation eingereicht (S. 119)

Kiel, den

_____

12 ACKNOWLEDGEMENTS

I am grateful to all members of the Bosch laboratory for collaboration, discussion, patience and a positive working atmosphere.

Special thanks to Prof. Thomas C.G. Bosch, who supervised my work. Thank you for advice, discussion, constructive criticism, support and a lot of freedom.

I also want to thank people, who collaborated in several research projects: Boris Anokhin, David Miller, Kiyokazu Agata and Noriko Funayama.

A big THANK YOU goes to my two families and to my girlfriend. To them I dedicate this thesis.

Kiel, 04/2007.