HHS Public Access

Author manuscript

Nature. Author manuscript; available in PMC 2015 June 24.

Published in final edited form as:

Nature. 2010 March 25; 464(7288): 592–596. doi:10.1038/nature08830.

The dynamic genome of *Hydra*

A full list of authors and affiliations appears at the end of the article.

Abstract

The freshwater cnidarian *Hydra* was first described in 1702[1] and has been the object of study for 300 years. Experimental studies of *Hydra* between 1736 and 1744 culminated in the discovery of asexual reproduction of an animal by budding, the first description of regeneration in an animal, and successful transplantation of tissue between animals[2]. Today, *Hydra* is an important model for studies of axial patterning[3], stem cell biology[4] and regeneration[5]. Here we report the genome of *Hydra magnipapillata* and compare it to the genomes of the anthozoan *Nematostella vectensis*[6] and other animals. The *Hydra* genome has been shaped by bursts of transposable element expansion, horizontal gene transfer, *trans*-splicing, and simplification of gene structure and gene content that parallel simplification of the *Hydra* life cycle. We also report the sequence of the genome of a novel bacterium stably associated with *H. magnipapillata*. Comparisons of the *Hydra* genome to the genomes of other animals shed light on the evolution of epithelia, contractile tissues, developmentally regulated transcription factors, the Spemann–Mangold organizer, pluripotency genes and the neuromuscular junction.

The genomic basis of cnidarian evolution has so far been viewed from the perspective of an anthozoan, the sea anemone *Nematostella vectensis*[6]. *Hydra* is a medusozoan that diverged from anthozoans at least 540 millions year ago. Features of *Hydra* and *Nematostella* are

Correspondence and requests for materials should be addressed to R.E.S. (resteele@uci.edu) or D.S.R. (dsrokhsar@gmail.com).
†Present addresses: Department of Cell and Developmental Biology, John Innes Centre, Norwich NR4 7UH, UK (P.A.W.); Institute of Human Genetics, University of Heidelberg, D-69120 Heidelberg, Germany (A.-K.G.); Center for Bioinformatics and Computational Biology, National Institute of General Medical Sciences, Bethesda, Maryland 20892-6200, USA (K.A.R.); Department of Ecology and Evolutionary Biology, Rice University, Houston, Texas 77251-1892, USA (N.H.P.); Ochadai Academic Production, Ochanomizu University, Ohtsuka, Bunkyo, 1128610 Tokyo, Japan (A.O.).
*These authors contributed equally to this work.
‡Deceased.

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Author Contributions J.A.C., E.F.K., O.S., H.R.B., C.N.D., D.S.R. and R.E.S. directed the project and wrote the manuscript. E.F.K., K.A.R., R.L.S. and J.C.V. directed genome sequencing and assembly at JCVI. J.B., D.B., K.D., C.P., N.S., G.G.S., L.D.V. and B.W. were responsible for library construction, sequence production and genome assembly at JCVI. D.S.R., J.A.C., O.S., T.M., D.M.G., U.H., T.K., S.E.P., S.S. and N.H.P. carried out genome assembly and gene annotation at UC Berkeley. Construction of cDNA libraries and analysis of ESTs was carried out by H.R.B., R.E.S., D.F.K., S.E.H., L.G., D.L., L.L., J.P., B.B., P.A.W., T.F., C.N.-F., T.G., J.S.H., E.H., S.H., M.H., K.I., A.O., T.T., T.C.G.B., K.K., G.H., A.F., R.A., S.F., T.W.H., C.G., P.G.B., B.B., Y.N., S.O., H.W. and D.E.M. T.W., T.R., P.G.B., C.E.D., P.T. and C.N.D. carried out analysis of the *Curvibacter* genome and HGT candidate genes. A.M.H., B.M.W., O.S. and K.J.P. carried out the microRNA analyses. U.T., B.H., A.W., P.R.H.S., X.Z., R.A., M.-K.E., A.-K.G., W.S., T.F. and A.B. carried out analyses of genes involved in various biological processes. J.A.C., E.F.K. and O.S. are joint first authors.

Two different assemblies of the *Hydra magnipapillata* strain 105 genome were generated and deposited in GenBank under accession numbers ABRM00000000 and ACZU00000000. The *Curvibacter* sp. genome sequence has been deposited in GenBank under accession numbers FN543101, FN543102, FN543103, FN543104, FN543105, FN543106, FN543107 and FN543108.

The authors declare no competing financial interests.

compared in Supplementary Table 1. We generated draft assemblies of the *Hydra magnipapillata* genome using a whole-genome shotgun approach (Supplementary Information sections 1–3 and Supplementary Figs 1–3). The *Hydra* genome is (A+T)-rich (71% A+T), and includes ~57% transposable elements (see below). Although the sequenced strain reproduces clonally in the laboratory by asexual budding, it is diploid with substantial heterozygosity (~0.7% single nucleotide polymorphism between alleles), which we find is distributed along the genome as expected if it were drawn from a randomly mating population (Supplementary Information section 3). These features complicate shotgun sequencing and assembly. Two complementary assemblies (CA and RP) were generated (Supplementary Information section 3) and deposited in GenBank. The CA assembly (1.5 gigabases (Gb)) has contig and scaffold N50 values of 12.8 kilobases (kb) and 63.4 kb, respectively. The RP assembly (1.0 Gb) has a contig N50 length of 9.7 kb and a scaffold N50 length of 92.5 kb. The CA assembly gives an estimated non-redundant genome size of 1.05 Gb. The RP assembly gives an estimated non-redundant genome size of 0.9Gb (see Supplementary Information section 3 for a discussion of genome size calculations). For analysis, we chose the assembly that minimized sequence redundancy owing to the separate assembly of haplotypes (see Supplementary Information section 3 for further discussion). Approximately 99% of known *Hydra* genes are found in both assemblies, attesting to their completeness with respect to protein-coding genes.

Although the present *Hydra* assembly is too fragmented for a chromosome-scale analysis, we found evidence for synteny with other metazoans. Of the 33 longest gene-rich *Hydra* scaffolds (that is, those containing genes from at least 10 *Hydra*/*Nematostella* orthologue groups), 15 (45%) were significantly enriched ($P < 0.01$) for genes from specific eumetazoan linkage groups[6], indicating that vestiges of the ancestral eumetazoan genome organization persist in *Hydra*. This is in contrast to the highly diverged genomes of *Drosophila* and *Caenorhabditis elegans*, which show no synteny with other metazoans by these methods.

We estimate that the *Hydra* genome contains ~20,000 bona fide protein-coding genes (excluding transposable elements), based on expressed sequence tags (ESTs), homology and *ab initio* gene prediction (Supplementary Information section 6). The amino acid substitution rate in the *Hydra* lineage is enhanced relative to the *Nematostella* lineage; the sequence divergence between a *Hydra* peptide and itshuman orthologue is typically greater than the sequence divergence between *Nematostella* and human (Supplementary Information section 8 and Supplementary Fig. 4) as expected based on the longer branch leading to *Hydra* in peptide-based phylogenies[6]. Similarly, the rate of intron loss has been higher in the *Hydra* lineage; we find that 22%(126 out of 575) of the introns shared by *Nematostella* and human in well-aligned coding regions have been lost in *Hydra*. Conversely, only 6% (28 out of 476) of the introns shared by *Hydra* and human are absent in *Nematostella*.

Transposable elements make up ~57% of the *Hydra* genome and represent over 500 different families (Supplementary Information section 9). The most abundant element, comprising ~15% of the genome (Fig. 1 and Supplementary Table 3), is a non-long-terminal-repeat (non-LTR) retroelement of the chicken repeat 1 (CR1) family. To our knowledge, elements of this family are more abundant in the *Hydra* genome than in any

other sequenced animal genome (in comparison, the CR1 family occupies only ~1%of the *Nematostella* assembly and 3% of the chicken assembly). This retrotransposon is still active in *Hydra*, as indicated by its representation in 105 ESTs. We also found 789 cases of intronless genes that were derived recently from multi-exon genes, most probably through retrotransposition. DNA transposons (predominantly 'cut-and-paste' elements of the mariner, Transib and hAT (hobo-*Ac*-Tam3) types) occupy ~20% of both the *Hydra* and *Nematostella* genomes, and are also active in *Hydra* based on the presence of ESTs.

Timing of transposable element activity using sequence divergence of extant copies reveals at least three periods of element expansion (at ~5%, ~20% and ~40% nucleotide substitutions; Fig. 1 and Supplementary Figs 5 and 6). In marked contrast, comparable expansions are absent from the *Nematostella* genome (Supplementary Fig. 7). Most individual *Hydra* transposable element families show discrete bursts of expansion (Fig. 1b, c) that are possibly associated with population bottlenecks[7]. The correspondence between speciation times in the genus *Hydra* and the timing of transposon activity may have been associated with the approximately threefold increase in genome size (Fig. 1a) in *H. magnipapillata*, *H. vulgaris* and *H. oligactis* relative to *H. viridissima* (380 megabases (Mb))[8].

Addition of short RNA leader sequences to the 5′ ends of messenger RNAs by *trans*-splicing occurs in a subset of metazoans and unicellular eukaryotes[9]. Transcripts from at least one-third of EST-supported genes in *Hydra* undergo *trans*-spliced leader addition (Supplementary Information section 10). *Hydra* has multiple spliced leader genes (Supplementary Table 9), and a given transcript may be *trans*-spliced with several different spliced leaders. Notably, *trans*-splicing is absent from *Nematostella* (Supplementary Information section 10). It now seems likely that *trans*-splicing has evolved multiple times independently[9].

Trans-splicing occurs in *Hydra viridissima* (N. A. Stover and R.E.S., unpublished data; GenBank accession number DQ092354) and in several other hydrozoans (Supplementary Table 10), and may be an ancestral feature of the class. Spliced leader addition gives a eukaryotic cell the opportunity to combine genes into operons, the multi-cistronic transcripts of which can be resolved into individual mRNAs by *trans*-splicing. We found 32 potential *Hydra* operons (Supplementary Information section 10, Supplementary Table 11 and Supplementary Fig. 9), but no obvious evidence for functional relationships between genes in these operons.

Bacteria are stably associated with *Hydra*[10]. Electron micrographs reveal bacterial cells underneath the glycocalyx, the coat that overlies the apical surface of the ectodermal epithelial layer of *Hydra* (Supplementary Fig. 10). Our assembly yielded eight large putative bacterial scaffolds as evidenced by: (1) high G+C content (in contrast to the low G+C content of the *Hydra* genome); (2) no high-copy repeat sequences typical of *Hydra* scaffolds; and (3) closely spaced single-exon open reading frames with best hits to bacterial genes (Supplementary Information section 11, Supplementary Fig. 11 and Supplementary Table 12). These scaffolds span a total of 4Mb encoding 3,782 single-exon genes and represent an estimated 98% of the bacterial chromosome. Phylogenetic analysis of 16S

rRNA (Supplementary Fig. 12) and conserved clusters of orthologous groups of proteins (COGs) indicate that this bacterium is a novel *Curvibacter* species belonging to the family Comamonadaceae (order Burkholderiales)[11]. About 60% of annotated *Curvibacter* sp. genes have an orthologue in another species of Comamonadaceae (Supplementary Table 13). Notably, the *Curvibacter* sp. genome encodes nine different ABC sugar transporters, compared to only one or two in other species of Comamonadaceae (Supplementary Table 14), possibly reflecting an adaptation to life in association with *Hydra*.

Non-metazoan genes among cnidarian ESTs have been reported previously[12], and we have now found further examples of such genes in the *Hydra* genome assembly. These genes are candidates for horizontal gene transfer (HGT) (Supplementary Information section 12). Seventy-one *Hydra* gene models showed closer relationships to bacterial genes than to metazoan genes based on sequence similarity and phylogenetic analysis (Supplementary Table 15). Of these, 51 have no blast hits to other metazoans, except in a few cases to *Nematostella*. Potential donors of these HGT candidates are widely distributed among different bacterial phyla (Supplementary Table 15) and show no enrichment for close relatives of *Curvibacter*. Approximately 70% of the HGT candidates have EST support, and transcripts from 30% of the genes have spliced leaders, indicating unambiguously that they are derived from *Hydra* and not from associated bacteria (Supplementary Table 15). The HGT candidates generally have fewer introns than *Hydra* genes and nearly one-half are single-exon genes (Supplementary Fig. 14), as expected if they were relatively recently acquired by *Hydra*. A number of the HGT candidates encode sugar-modifying enzymes. Three genes encode enzymes in the branch of the bacterial lipopolysaccharide synthesis pathway that leads to formation of the activated heptose precursor of the lipopolysaccharide inner core (Supplementary Fig. 13). This pathway could modify endogenous glycoproteins or proteoglycans in *Hydra*.

We also identified 90 transposable elements that were potentially horizontally transferred into the *Hydra* genome. These elements have expanded recently (less than 10% nucleotide divergence from their consensus) and have no older copies in the genome. The most frequent element class consists of hAT transposons with 34 different families, although all major classes of transposable element (DNA transposon, LTR and non-LTR elements) are represented. Transposable elements have been shown previously to be horizontally transferred inmetazoans[13].

We identified 51 unique non-tRNA/non-rRNA transcripts that correspond to putative non-coding RNA genes based on 454 sequencing of short transcripts from *Hydra* (Supplementary Information section 13 and Supplementary Table 16). At least 17 of these are microRNAs (miRNAs), compared to 40 identified miRNAs in *Nematostella*[14]. Surprisingly, only a single miRNA gene in the available data sets, *miR-2022*, is common to both cnidarian species.

Hox and ParaHox gene families arose from a megacluster that included a number of other homeobox genes (for example, NK genes)[15]. With the exception of engrailed, descendants of all of the classes of homeobox genes in the megacluster are found in *Nematostella*[16,17]. *Hydra* is missing a substantial fraction of megacluster descendants[16], indicating secondary

loss. For example, the *eve* and *emx* genes are absent from *Hydra*, although they are present in *Nematostella* and several hydrozoans (Supplementary Table 17). The loss of these genes from *Hydra* is therefore recent in relation to the diversification of hydrozoans. These genes are expressed in a cell-type-specific manner in larvae and adults of *Nematostella*[17] and *Hydractinia*[18]; it is intriguing that the loss of these genes correlates with the absence of a larval stage in *Hydra* (Supplementary Table 17). The absence of these genes in *Hydra* indicates that despite their near-universal presence in animals, it is possible to construct a metazoan without either of them. In addition to the loss of *emx* and *eve* genes, *Hydra* has undergone several other marked gene losses; for example, it lacks fluorescent protein genes and key circadian rhythm genes (Supplementary Information section 14).

All major bilaterian signalling pathways, including Wnt, transforming growth factor-β, Hedgehog, receptor tyrosine kinase and Notch, are present in *Hydra* and *Nematostella*. An important signalling centre in *Hydra* is the head organizer, which uses the Wnt signalling pathway to establish positional values along the body column[19,20]. The head organizer, which is located at the apical tip of the adult polyp, is derived from the gastrula blastopore in cnidarians. A transplanted head organizer has the capacity to induce axis formation[21], similar to the Spemann–Mangold organizer in *Xenopus*. Orthologues of a number of genes known to act in the Spemann–Mangold organizer in *Xenopus* are present in the *Hydra* and *Nematostella* genomes. Moreover, several of the secreted signalling molecules and transcription factors encoded by these genes are expressed specifically in the *Hydra* head organizer and the blastopore organizer in the *Nematostella* gastrula (Supplementary Information section 15 and Supplementary Table 18). Thus, the *Hydra* head organizer and the *Xenopus* Spemann–Mangold organizer may share common descent from an organizer in the ancestor of cnidarians and bilaterians.

The extracellular portions of two *Hydra* receptor tyrosine kinases[22,23] contain a novel protein domain, sweet tooth (SWT). The SWT domain is also present in ESTs from the hydrozoan *Clytia*, but is absent from all other sequenced genomes, including that of *Nematostella* (Supplementary Fig. 15). SWT is among the most abundant protein domains encoded in the *Hydra* genome. The SWT domain is present in one or more copies in predicted secreted proteins. Given its presence in receptors and secreted proteins, we deduce that the SWT domain defines a large, diverse and novel set of signalling proteins.

Hydra contains a pluripotent stem cell type that gives rise to germ cells, nerve cells, nematocytes and secretory cells[4]. Of the five genes that have been shown to induce pluripotency in differentiated somatic cells of mammals (*Myc*, *Nanog*, *Klf4*, *Oct4* and *Sox2*)[24], homologues of three (*Nanog*, *Klf4* and *Oct4*) are clearly not present in the *Hydra* genome. *Hydra* has four *Myc* homologues. There are two members of the Sox B group in *Hydra*. The Sox B group includes *Sox2*, but the evolutionary relationship between vertebrate *Sox2* genes and *Hydra* Sox B genes is not clear[25]. We conclude that the stem cell genetic network in *Hydra* probably has an evolutionary origin independent from the network used in mammalian stem cells. Studies of diverse cnidarians support this scenario (see Supplementary Information section 14 for details).

Hydra's shape is formed by epitheliomuscular cells, a cell type unique to cnidarians. A survey of genes that encode muscle structural and regulatory proteins in *Hydra* and *Nematostella* reveals a conserved eumetazoan core actin-myosin contractile machinery shared with bilaterians (Supplementary Table 19). Both cnidarians, however, lack crucial, specific regulators associated with vertebrate striated (troponin complex) or smooth muscles (caldesmon), indicating that these specializations arose after the cnidarian–bilaterian split. *Hydra* also shows secondary simplifications relative to *Nematostella*, which has a greater degree of muscle-cell-type specialization, including specialized retractor muscle cells. *Hydra* lacks several components of the dystroglycan complex (α/ε-sarcoglycan and β-sarcoglycan, α/β-dystroglycan and γ-syntrophin), which may lead to a less robust tethering of actin to the cell membrane than in *Nematostella*. Similarly, the absence of a bona fide myosin light chain kinase and phosphatase in *Hydra* indicates a divergence or loss of regulation by myosin regulatory light chain phosphorylation. The greater degree of muscle-cell-type specialization in *Nematostella* is also mirrored in the higher number of myosin light chain genes in this species. Thus, even among cnidarians, we see substantial variation in muscle-associated components superimposed on the eumetazoan core, with the *Hydra* muscular system representing a secondary simplification from a more complex cnidarian ancestor.

Ultrastructural studies show that nerve cells in *Hydra* form synapses on contractile epitheliomuscular cells (Fig. 2a), and that these synapses contain dense core vesicles, paramembranous densities and cleft filaments[26] similar to canonical neuromuscular junctions in bilaterians. Several components of the bilaterian neuromuscular junction (choline transporter, nicotinic acetylcholine receptor) are encoded in the *Hydra* genome (Supplementary Information section 16 and Supplementary Table 20) and their expression is consistent with a role in neuromuscular signalling (Supplementary Figs 16 and 17). Other components, however, are found only in a possibly primitive form (putative carnitine acetyltransferases that lack the diagnostic residues for choline selectivity), and some components are absent (the vesicular acetylcholine transporter; Fig. 2b). Together, these data indicate that a canonical bilaterian neuromuscular junction was probably not present in the last common ancestor of cnidarians and bilaterians. *Hydra* is known to use neuropeptides for the control of behaviour[27], and these may be contained in the dense-core vesicles seen at *Hydra* synapses.

In *Hydra* and *Nematostella*, epitheliomuscular cells have an apical junctional belt in the form of a septate junction, clear apical–basal polarity, and hemidesmosome-like contact sites with the extracellular matrix (mesoglea) on their basal surface (Fig. 3a). The *Hydra* and *Nematostella* genomes encode almost all of the proteins known from bilaterians to be involved in the establishment of cell–cell and cell–substrate contacts (Fig. 3b and Supplementary Fig. 18). This indicates that the common cnidarian–bilaterian ancestor possessed a genetic inventory for the formation of all types of eumetazoan cell–cell and cell–substrate junctions. The presence of innexin genes in the *Hydra*[28] and *Nematostella* genomes (Fig. 3b and Supplementary Fig. 19) combined with the lack of connexin genes in non-chordate genomes clearly support the view that innexin-based gap junctions are an ancestral eumetazoan feature, and that gap junctions formed by connexins[29] arose later in

animal evolution. Similarly, the lack of occludin genes in cnidarians and other non-chordates (Fig. 4) indicates that occludins and their function in tight junction formation first arose in the deuterostome lineage.

Although some gene families associated with cell–cell and cell– substrate interactions are also found in placozoans, demosponges and choanoflagellates, it is important to note that there are cell-adhesion-associated protein domains specific to cnidarians and bilaterians. For example, *Hydra* and *Nematostella* have classic cadherins exhibiting a highly conserved, bilaterian-type cytoplasmic (CCD) domain(Fig. 3b and Supplementary Fig. 18) that is able to interact with β- and p120/δ-catenin (Supplementary Information section 17). So far, only one sponge cadherin gene that encodes a cytoplasmic domain with weak similarity to the eumetazoan CCD domain has been detected.

The sequencing of the *Hydra* genome has revealed unexpected relationships between the genetic makeup of the animal and its biology. The genes encoding the proteins that form epithelial junctions in bilaterians are present in *Hydra* yet there are obvious differences in structures of the junctional complexes. Despite the morphological similarity of neuromuscular junctions in bilaterians and *Hydra*, several of the key genes required to make this junction in bilaterians are absent from *Hydra*. *Hydra* has a complete set of muscle genes but lacks mesoderm and forms muscles only in epithelial cells. Most of the genes required for stem cell pluripotency in mammals are absent from *Hydra*, yet *Hydra* has a multipotent stem cell system that functions similarly to stem cell systems in bilaterians. The availability of the *Hydra* genome sequence and methods to manipulate it[30] provide an opportunity to understand how this remarkable animal evolved.

METHODS SUMMARY

The genome of *Hydramagnipapillata* strain 105was sequenced at the J. CraigVenter Institute using the whole genome shotgun approach. Two different assemblies were generated and deposited in GenBank (accession numbers ABRM00000000 and ACZU00000000). Complementary DNA libraries were prepared using standard methods and ESTs were generated at the National Institute of Genetics (Mishima, Japan) and the Genome Sequencing Center(Washington University, St Louis). ESTs have been deposited in the dbEST database at the National Center for Biotechnology Information. The *Curvibacter* sp. genome sequence has been deposited in GenBank (accession numbers FN543101, FN543102, FN543103, FN543104, FN543105, FN543106, FN543107 and FN543108).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Authors

Jarrod A. Chapman[1,*], Ewen F. Kirkness[2,*], Oleg Simakov[3,4,*], Steven E. Hampson[5,‡], Therese Mitros[4], Therese Weinmaier[6], Thomas Rattei[6], Prakash G. Balasubramanian[3], Jon Borman[2], Dana Busam[2], Kathryn Disbennett[2], Cynthia Pfannkoch[2], Nadezhda Sumin[2], Granger G. Sutton[2], Lakshmi Devi Viswanathan[2],

Brian Walenz[2], David M. Goodstein[1], Uffe Hellsten[1], Takeshi Kawashima[4], Simon E. Prochnik[1], Nicholas H. Putnam[1,4,†], Shengquiang Shu[1], Bruce Blumberg[7,8], Catherine E. Dana[8,9], Lydia Gee[7,8], Dennis F. Kibler[5], Lee Law[7,8], Dirk Lindgens[7,8], Daniel E. Martinez[10], Jisong Peng[7,8], Philip A. Wigge[11,†], Bianca Bertulat[3], Corina Guder[3], Yukio Nakamura[3], Suat Ozbek[3], Hiroshi Watanabe[3], Konstantin Khalturin[12], Georg Hemmrich[12], André Franke[12], René Augustin[12], Sebastian Fraune[12], Eisuke Hayakawa[13], Shiho Hayakawa[13], Mamiko Hirose[13], Jung Shan Hwang[13], Kazuho Ikeo[13], Chiemi Nishimiya-Fujisawa[13], Atshushi Ogura[13,†], Toshio Takahashi[14], Patrick R. H. Steinmetz[15], Xiaoming Zhang[16], Roland Aufschnaiter[17], Marie-Kristin Eder[17], Anne-Kathrin Gorny[17,†], Willi Salvenmoser[17], Alysha M. Heimberg[18], Benjamin M. Wheeler[19], Kevin J. Peterson[18], Angelika Böttger[20], Patrick Tischler[6], Alexander Wolf[20], Takashi Gojobori[13], Karin A. Remington[2,†], Robert L. Strausberg[2], J. Craig Venter[2], Ulrich Technau[15], Bert Hobmayer[17], Thomas C. G. Bosch[12], Thomas W. Holstein[3], Toshitaka Fujisawa[13], Hans R. Bode[7,8], Charles N. David[20], Daniel S. Rokhsar[1,4], and Robert E. Steele[8,9]

Affiliations

[1]US Department of Energy Joint Genome Institute, Walnut Creek, California 94598, USA

[2]The J. Craig Venter Institute, Rockville, Maryland 20850, USA

[3]Institute of Zoology, Department of Molecular Evolution and Genomics, University of Heidelberg, D-69120 Heidelberg, Germany

[4]Center for Integrative Genomics, Department of Molecular and Cell Biology, University of California, Berkeley, California 94720, USA

[5]Department of Computer Science, University of California, Irvine, California 92697-3435, USA

[6]Department of Genome-Oriented Bioinformatics, Technische Universität München, D-85354 Freising, Germany

[7]Department of Developmental and Cell Biology, University of California, Irvine, California 92697-2275, USA

[8]Developmental Biology Center, University of California, Irvine, California 92697-2275, USA

[9]Department of Biological Chemistry, University of California, Irvine, California 92697-1700, USA

[10]Department of Biology, Pomona College, Claremont, California 91711, USA

[11]The Salk Institute, La Jolla, California 92037, USA

[12]Zoologisches Institüt, Christian-Albrechts-University, D-24098 Kiel, Germany

[13]National Institute of Genetics, Yata 1, 111, Mishima 411-8540, Japan

[14]Suntory Institute for Bioorganic Research, Osaka 618-8503, Japan

[15]Department of Molecular Evolution and Development, University of Vienna, A-1090 Vienna, Austria

[16]Department of Anatomy and Cell Biology, The University of Kansas Medical Center, Kansas City, Kansas 66160, USA

[17]Institute of Zoology and Center for Molecular Biosciences, University of Innsbruck, A-6020 Innsbruck, Austria

[18]Department of Biological Sciences, Dartmouth College, Hanover, New Hampshire 03755, USA

[19]Department of Computer Science, North Carolina State University, Raleigh, North Carolina 27695, USA

[20]Department of Biology II, Ludwig-Maximilians-University, D-82152 Planegg-Martinsried, Germany

Acknowledgements

We are grateful to S. Clifton, R. Wilson and the EST sequencing group at the Genome Sequencing Center at the Washington University School of Medicine for their efforts in generating the *Hydra* ESTs and to the National Science Foundation for its support of the *Hydra* EST project (grant number IBN-0120591). Funding for the sequencing of the *Hydra* genome was provided by the National Human Genome Research Institute. We thank J. Gerhart who, as co-chair of the National Human Genome Research Institute Working Group on Comparative Genome Evolution, advocated sequencing of the *Hydra* genome. The septate junction electron micrograph in Fig. 3 was provided by G. Rieger. K.J.P. thanks N. Margulis for technical assistance and the NSF for support (grant number DEB-0716960). U.T. was supported by the Austrian Science Fund and the Norwegian Research Council. B.H. was supported by Austrian Science Fund grants FWF P16685 and FWF P20734. D.S.R. was supported by R. Melmon and the Gordon and Betty Moore Foundation. Work at the US Department of Energy Joint Genome Institute was supported by the Office of Science of the US Department of Energy under Contract No. DE-AC02-05CH11231. T.F. and T.G. were supported by Grants-In-Aid for Scientific Research from the Ministry of Education, Culture, Sports, Science, and Technology of Japan. T.C.G.B. was supported by grants from the Deutsche Forschungsgemeinschaft (DFG SFB617-A1) and from the DFG Cluster of Excellence programs ‘The Future Ocean’ and ‘Inflammation at Interfaces’. T.W.H. was supported by grants from the Deutsche Forschungsgemeinschaft including SFB488-A12 and the DFG Cluster of Excellence program ‘CellNetworks’. Support for H.W. was provided by the TOYOBO Biotechnology Foundation and the Alexander von Humboldt Foundation. Support for A.B., B.B., C.G., C.N.D., H.W., P.G.B., S.O., T.F. (Mercator Professor) and Y.N. was provided by the Deutsche Forschungsgemeinschaft.

References

1. van Leeuwenhoek, A. The Collected Letters of Antoni van Leeuwenhoek. Palm, LC., editor. Vol. XIV. Swets and Zeitlinger; 1996. p. 169-173.
2. Trembley, A. Mémoires, pour servir à l’historie d’un genre de polypes d’eau douce, à bras en forme de cornes. Jean and Herman Verbeek; 1744.
3. Bode, PM.; Bode, HR. Pattern Formation, A Primer in Developmental Biology. Malacinski, GM.; Bryant, SV., editors. Macmillan; 1984. p. 213-241.
4. David CN, Murphy S. Characterization of interstitial stem cells in *Hydra* by cloning. Dev. Biol. 1977; 58:372–383. [PubMed: 328331]
5. Holstein TW, Hobmayer E, Technau U. Cnidarians: an evolutionarily conserved model system for regeneration? Dev. Dyn. 2003; 226:257–267. [PubMed: 12557204]
6. Putnam NH, et al. Sea anemone genome reveals ancestral eumetazoan gene repertoire and genomic organization. Science. 2007; 317:86–94. [PubMed: 17615350]
7. Lynch M. The evolution of genetic networks by non-adaptive processes. Nature Genet. 2007; 8:803–813.

8. Zacharias H, Anokhin B, Khalturin K, Bosch TCG. Genome sizes and chromosomes in the basal metazoan *Hydra*. Zoology. 2004; 107:219–227. [PubMed: 16351940]
9. Douris V, Telford MJ, Averof M. Evidence for multiple independent origins of trans-splicing in Metazoa. Mol. Biol. Evol. 2009 Nov 25.
10. Fraune S, Bosch TC. Long-term maintenance of species-specific bacterial microbiota in the basal metazoan *Hydra*. Proc. Natl Acad. Sci. USA. 2007; 104:13146–13151. [PubMed: 17664430]
11. Ding L, Yokota A. Proposals of *Curvibacter gracilis* gen. nov., sp. nov. and *Herbaspirillum putei* sp. nov. for bacterial strains isolated from well water and reclassification of [*Pseudomonas*] *huttiensis*, [*Pseudomonas*] *lanceolata*, [*Aquaspirillum*] *delicatum* and [*Aquaspirillum*] *autotrophicum* as *Herbaspirillum huttiense* comb. nov., *Curvibacter lanceolatus* comb. nov., *Curvibacter delicatus* comb. nov. and *Herbaspirillum autotrophicum* comb. nov. Int. J. Syst. Evol. Microbiol. 2004; 54:2223–2230. [PubMed: 15545462]
12. Technau U, et al. Maintenance of ancestral complexity and non-metazoan genes in two basal cnidarians. Trends Genet. 2005; 21:633–639. [PubMed: 16226338]
13. Pace JK II, Gilbert C, Clark MS, Feschotte C. Repeated horizontal transfer of a DNA transposon in mammals and other tetrapods. Proc. Natl Acad. Sci. USA. 2008; 105:17023–17028. [PubMed: 18936483]
14. Grimson A, et al. Early origins and evolution of microRNAs and Piwi-interacting RNAs in animals. Nature. 2008; 455:1193–1197. [PubMed: 18830242]
15. Butts T, Holland PW, Ferrier DE. The urbilaterian Super-Hox cluster. Trends Genet. 2008; 24:259–262. [PubMed: 18472178]
16. Chourrout D, et al. Minimal ProtoHox cluster inferred from bilaterian and cnidarian Hox complements. Nature. 2006; 442:684–687. [PubMed: 16900199]
17. Ryan JF, et al. Pre-bilaterian origins of the Hox cluster and the Hox code: evidence from the sea anemone, *Nematostella vectensis*. PLoS One. 2007; 2:e153. [PubMed: 17252055]
18. Mokady O, Dick MH, Lackschewitz D, Schierwater B, Buss LW. Over one-half billion years of head conservation? Expression of an ems class gene in *Hydractinia symbiolongicarpus* (Cnidaria: Hydrozoa). Proc. Natl Acad. Sci. USA. 1998; 95:3673–3678. [PubMed: 9520424]
19. Hobmayer B, et al. WNT signalling molecules act in axis formation in the diploblastic metazoan *Hydra*. Nature. 2000; 407:186–189. [PubMed: 11001056]
20. Broun M, Gee L, Reinhardt B, Bode HR. Formation of the head organizer in hydra involves the canonical Wnt pathway. Development. 2005; 132:2907–2916. [PubMed: 15930119]
21. Browne E. The production of new hydranths in *Hydra* by the insertion of small grafts. J. Exp. Zool. 1909; 7:1–23.
22. Bridge DM, Stover NA, Steele RE. 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. 2000; 220:253–262. [PubMed: 10753514]
23. Reidling JC, Miller MA, Steele RE. Sweet Tooth, a novel receptor protein-tyrosine kinase with C-type lectin-like extracellular domains. J. Biol. Chem. 2000; 275:10323–10330. [PubMed: 10744720]
24. Feng B, Ng JH, Heng JC, Ng HH. Molecules that promote or enhance reprogramming of somatic cells to induced pluripotent stem cells. Cell Stem Cell. 2009; 4:301–312. [PubMed: 19341620]
25. Jager M, Queinnec E, Houliston E, Manuel M. Expansion of the SOX gene family predated the emergence of the Bilateria. Mol. Phylogenet. Evol. 2006; 39:468–477. [PubMed: 16439161]
26. Westfall JA, Yamataka S, Enos PD. Ultrastructural evidence of polarized synapses in the nerve net of *Hydra*. J. Cell Biol. 1971; 51:318–323. [PubMed: 5111879]
27. Takahashi T, et al. Systematic isolation of peptide signal molecules regulating development in hydra: LWamide and PW families. Proc. Natl Acad. Sci. USA. 1997; 94:1241–1246. [PubMed: 9037037]
28. Alexopoulos H, et al. Evolution of gap junctions: the missing link? Curr. Biol. 2004; 14:R879–R880. [PubMed: 15498476]
29. Yen MR, Saier MH Jr. Gap junctional proteins of animals: the innexin/pannexin superfamily. Prog. Biophys. Mol. Biol. 2007; 94:5–14. [PubMed: 17507077]

30. Wittlieb J, Khalturin K, Lohmann JU, Anton-Erxleben F, Bosch TCG. Transgenic *Hydra* allow *in vivo* tracking of individual stem cells during morphogenesis. Proc. Natl Acad. Sci. USA. 2006; 103:6208–6211. [PubMed: 16556723]

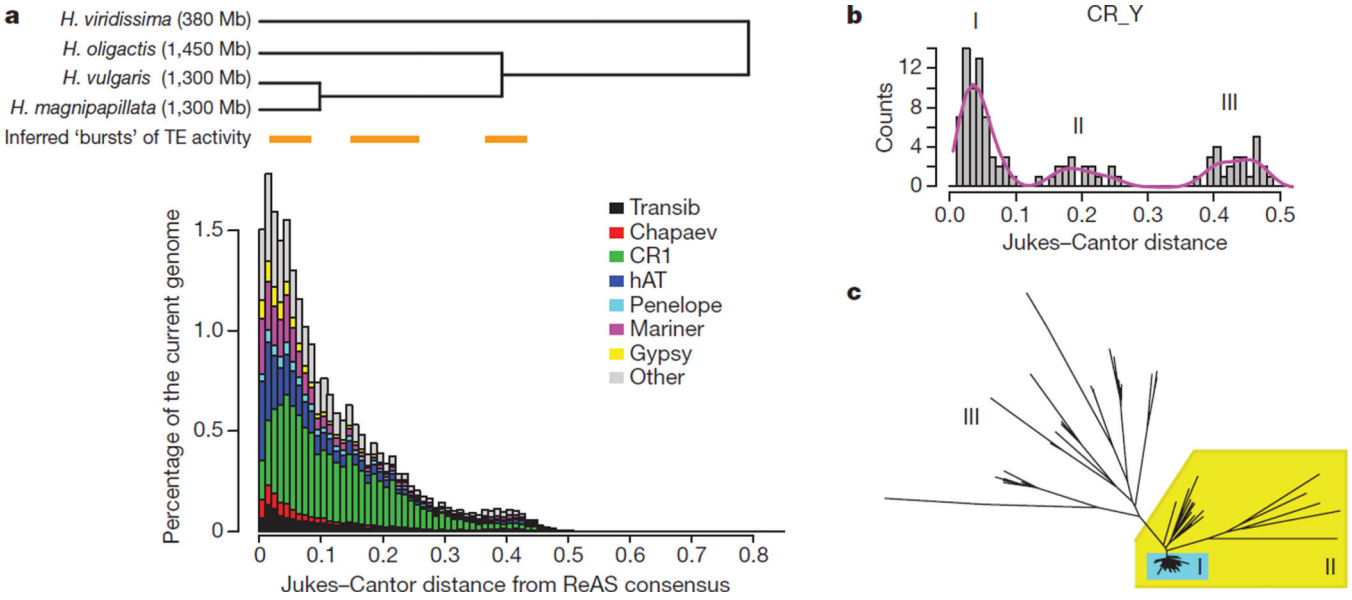

Figure 1. Dynamics of transposable element expansion in *Hydra* reveals several periods of transposon activity

a, The top panel shows phylogenetic relationships between four *Hydra* species based on ESTs (using Nei-Gojobori synonymous substitution rates; see Supplementary Fig. 8). The bottom panel shows the fraction of the genome that is occupied by a specific repeat class at a given divergence from the repeat consensus generated by the ReAS (recovery of ancestral sequences) algorithm (see Supplementary Information section 9). Substitution levels are corrected for multiple substitutions using the Jukes–Cantor formula $K = -3/4\ln(1-i4/3)$, where i is per cent dissimilarity on the nucleotide level from the repeat consensus. This substitution level for transposons is equivalent to Nei-Gojobori synonymous substitution rates in the ESTs. Three element expansions are inferred, the most distinct are the most ancient at ~0.4 and the most recent at 0.05 divergence levels. The middle expansion at about ~0.2 is not well synchronized and is more clearly seen for individual element classes in Supplementary Figs 5 and 6. **b, c**, Example of periods of activity of a single *Hydra* CR1 retrotransposon family (**b**) and the maximum likelihood phylogeny of the family (**c**).

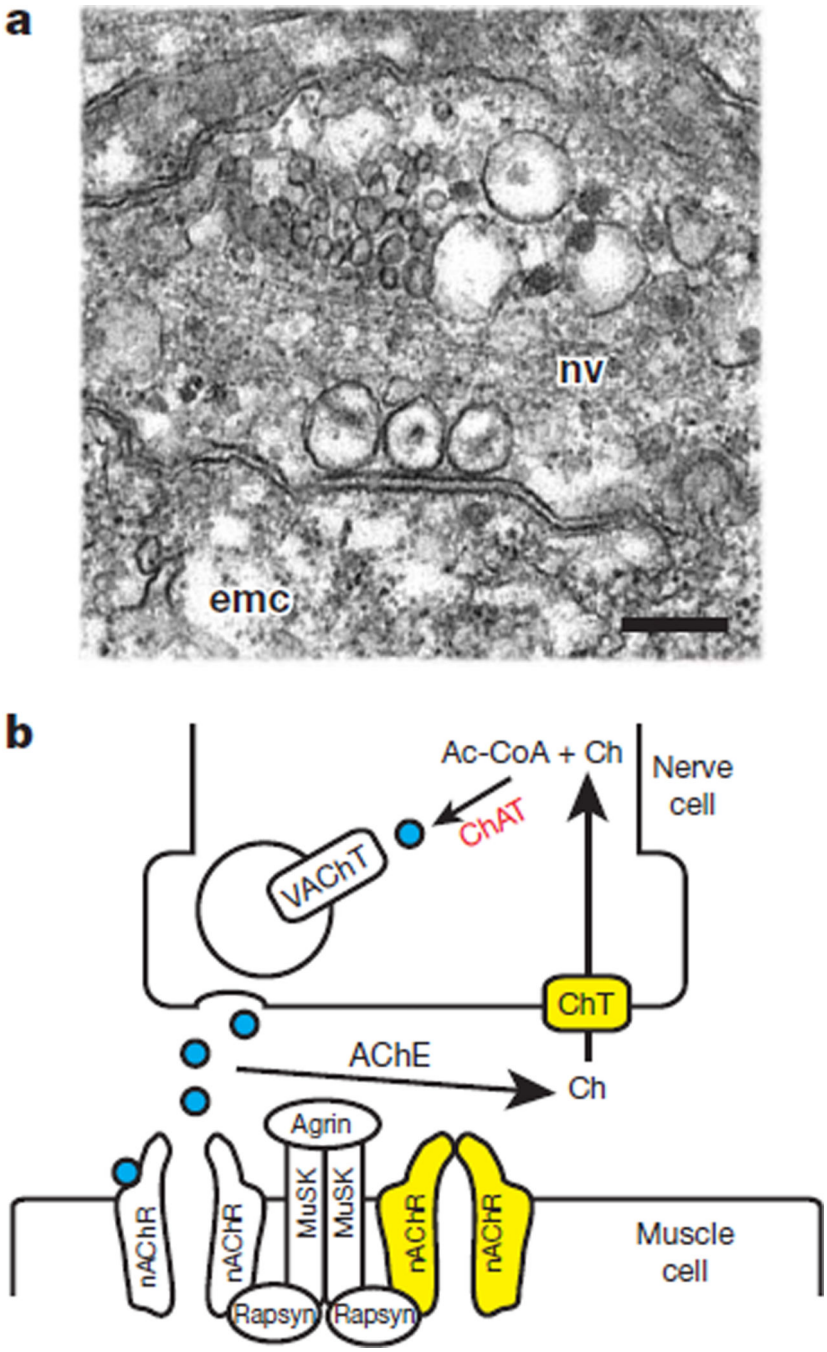

Figure 2. The neuromuscular junction in *Hydra*

a, Electron micrograph of a nerve synapsing on a *Hydra* epitheliomuscular cell. emc, epitheliomuscular cell; nv, nerve cell. Three vesicles are located in the nerve cell at the site of contact with the epitheliomuscular cell. Scale bar, 200 nm. **b**, Schematic diagram of a canonical neuromuscular junction. Yellow indicates presence in *Hydra*. Choline acetyltransferase (ChAT) is shown in red because it is not clear whether *Hydra* has an enzyme that prefers choline (Ch) as a substrate. Acetylcholine (ACh) molecules are shown as blue circles. The nicotinic acetylcholine receptor (nAChR) is shown in the open state with

acetylcholine bound (left), and in the closed state in the absence of bound acetylcholine (right). AChE, acetylcholinesterase; ChT, choline transporter; MuSK, muscle-specific kinase; VAChT, vesicular acetylcholine transporter.

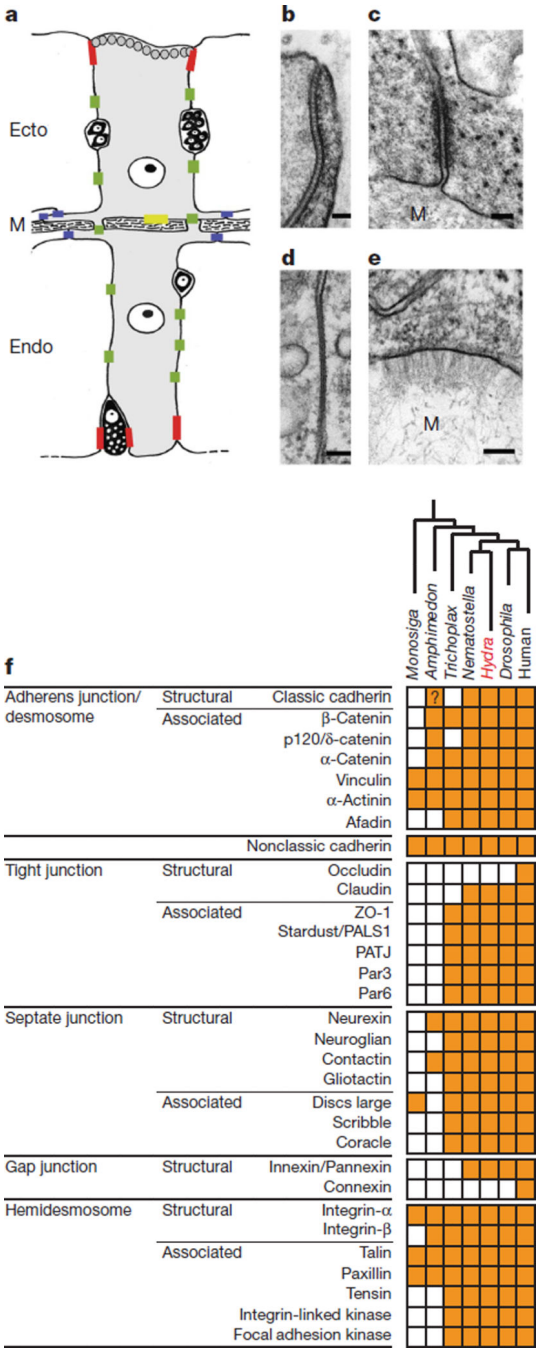

Figure 3. *Hydra* cell junctions

a, Schematic diagram of the positions of cell–cell and cell–matrix contacts in *Hydra* epitheliomuscular cells. Septate junction, red; gap junctions, green; spot desmosomes, blue; hemidesmosome-like cell–matrix contact, yellow. Ecto, ectodermal cell; Endo, endodermal cell; M, mesoglea. For simplicity the nervous system has been omitted. **b–e**, Electron micrographs of cell–cell and cell–matrix contacts in *Hydra*. **b**, Apical septate junction. **c**, Spot desmosome between basal muscle processes. **d**, Gap junction in the lateral cell membrane. **e**, Hemidesmosome-like cell–mesoglea contact site. Scale bars in **b–e** indicate

100 nm. **f**, Phylogenetic distribution of cell–cell and cell–substrate contact proteins. A filled box indicates the presence of an orthologue from the corresponding protein family as identified by SMART/Pfam analysis or conserved cysteine patterns. See Supplementary Information section 17 and Supplementary Table 21 for details.