Yale University

Department of Molecular, Cellular & Developmental Biology
PO Box 208104
New Haven, CT 06520-8104
(203) 432-3895 office
(203) 432-3854 fax
stephen.dellaporta@yale.edu

Thursday, October 10, 2002

Pleased find enclosed the resubmission of our original white paper titled "A Case for Sequencing the Trichoplax Genome" which received a moderate priority rating. We hope that this resubmission will provide a venue for discussion of points arising in the previous Council's review.

Just to highlight the changes in this resubmission:

- We have addressed the concern of the panel over whether *Trichoplax* should be the first lower metazoan sequenced by comparing the case for *Trichoplax* with that of traditional lower metazoan model organisms.
- We have also reworked our discussion of the expected utility of the *Trichoplax* genome to emphasize those points of immediate utility.
- A reevaluation of the existing data on the phylogenetic position of the *Trichoplax* has been added to a new section called "The Phylogenetic Position of *Trichoplax* and the Lower Metazoan Alternatives".
- We include our confirmation of the original estimates of the size of the Trichoplax genome to be approximately 50 Mb, the smallest known animal genome.
- Finally, we have added our progress in developing this system in the past 3 months, in particular the availability of new clones, the observation of cleaving embryos, and the preparation of a fosmid library.

Thank you for your consideration,

Respectfully submitted,
The Trichoplax Consortium

A Case for Sequencing the *Trichoplax* Genome

Principal Investigators of the *Trichoplax* Genome Consortium:

Stephen L. Dellaporta[σ], Leo W. Buss[σ], Maria A. Moreno[σ], and Bernd Schierwater[ψ], Yale University
George Weinstock[β], Baylor College of Medicine Human Genome Sequencing Center
Peter Holland, Oxford University (Collaborator)

[σ]Contact information: Yale University, Dept. MCD Biology and Ecology & Evolution, New Haven, CT 06520-8104. (203) 432-3895. Email: stephen.dellaporta@yale.edu, leo.buss@yale.edu, and maria.moreno@yale.edu
[β] Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030. (713) 798-6539 Email: gwstock@bcm.tmc.edu
[ψ] Affliliate Yale Scientist. (Primary address: Tieraerztliche Hochschule Hannover, Div. Ecology & Evolution, Hannover, Germany) Email: bernd.schierwater@ecolevol.de

Rationale

Comparative genomics is perhaps today's most important engine for the generation of new hypothesis-driven biological science. The degree to which this will remain the case is dependent upon the choice of new sequencing projects that "fill the gaps" between those organisms whose sequences have already been characterized. Perhaps no more compelling a gap remains to be filled than that of the Lower Metazoa, those organisms that lie at the base of the animal tree.

A group of four animal phyla, collectively known as the Lower Metazoa, lie between the Kingdom Fungi and the higher animals (Figure 1). The genome of an organism in this phylogenetic position will be required:

- To determine the gene and proteome content of the simplest known animals.
- To provide a genomics platform in a simple multicellular system. While yeast has proven the utility of simple model systems to human biology, we have heretofore not had a simple *multicellular* system. Lower metazoans are the simplest multicellular animals, hence provide the simplest platform for analysis of those biological processes unique to multicellular organisms, including cell-cell communication, signaling, cellular differentiation, development and patterning.
- To launch an experimental functional genomics platform for the study of the ancestral function of any animal-specific gene, RNA, or protein.
- To permit functional genomic approaches to problems of aging by providing, for the first time, genomic data and corresponding technology for an organism capable of indefinite asexual proliferation.
- To determine which genes are uniquely animal.
- To establish robust ancestor-descendant relationships between animal genes by providing, for the first time, a common ancestor of ecdysozoan (e.g., flies, nematodes) and deuterostome (e.g., human, mice, zebrafish, ascidian, sea urchin) genes to "root" the tree.
- To ascertain the limits of synteny between fungi (yeast) and higher metazoans on one hand and between ecdysozoans (flies and nematodes) and deuterostomes (human, zebrafish, ascidian) on the other.
- To determine whether genes absent in ecdysozoans, but present in humans represent gene losses in the Ecdysozoa or the genuine acquisition of novel function in Deuterostoma.

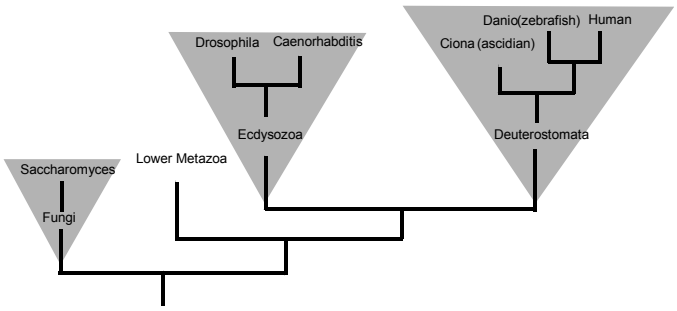


Figure 1. Phylogenetic Relationships among Major Metazoan Groups and Closest Sister Taxa

Four phyla comprise the Lower Metazoa: Placozoa, Porifera (sponges), Ctenophora (comb jellies), and Cnidaria (e.g., corals and jellyfish). 18S rDNA sequence data have unambiguous established that the Lower Metazoa are the sister group to the *all* higher animals [1-10]. Of the Lower Metazoa phyla, the Placozoa merit particular attention in that its sole representative, *Trichoplax adhaerens* possesses the simplest construction of any animal and it is an easily maintained experimental organism [11-27]. *Most, notably, Trichoplax has the smallest genome of any animal yet measured (50 Mb)[28-30], only 10X that of the Escherchia coli and less than 2% the size of the human genome.* Eight-to-ten times sequence coverage of the entire *Trichoplax* genome represents a project that can be completed in just a few months time at comparatively modest expense, but with a realistic expectation of a disproportionately large impact on our understanding of animal genome organization and evolution. As a model system, the simplest known animal with the smallest animal genome, the placozoan *Trichoplax adhaerens*, is highly suited as a cornerstone organism for understanding the underlying principles of animal organization.

The Biology of *Trichoplax*

Trichoplax is the sole representative of the Phylum Placozoa [14, 25, 31]. As Figure 2 illustrates, the organism bears a superficial resemblance to a giant amoeba. It is, however, a proper multicellular animal with four distinct cell types and differentiated dorsal and ventral epithelia [14, 17, 18, 21, 23, 32-35]. It is, without question, anatomically the simplest free-living (i.e., not parasitic) animal known.

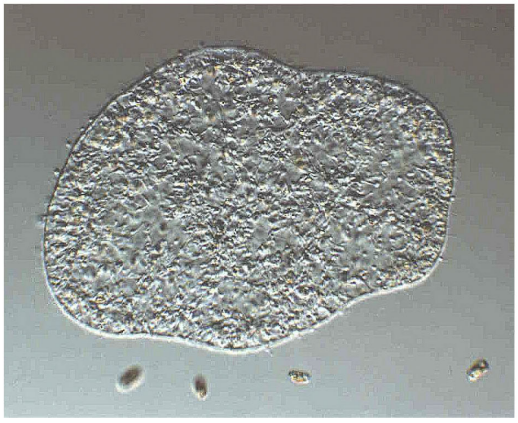


Figure 2. Morphology of Trichoplax

Trichoplax appears as a flat sheet of cells consisting of two epithelial layers, which sandwich a multi-nucleate syncytium. Four cell types exist: monciliated dorsal and ventral epithelia cells, ventral gland cells and the syncytial fiber cells (Figure 3). Nerves, sensory cells and muscles are absent, as are any 'tissues' or organs. The epithelia of *Trichoplax* lack a basal membrane; cells are connected by belt desmonsomes. Lipid inclusions, called 'shiny spheres', are regularly distributed over the dorsal epithelia.

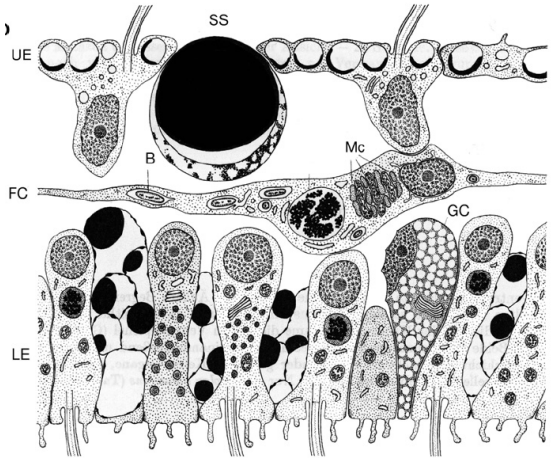


Figure 3. Histology of *Trichoplax*

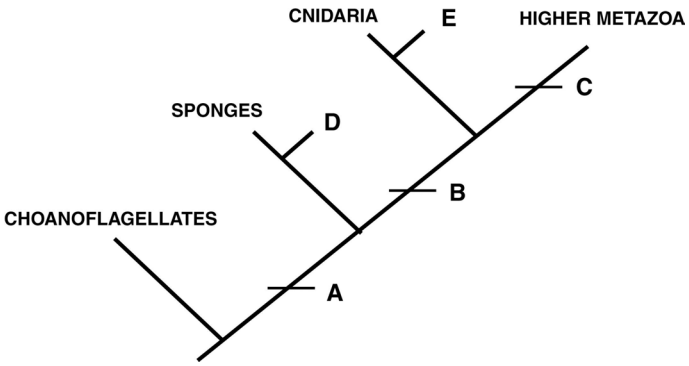
Diagrammatic cross section of *Trichoplax*: UE-upper epithelium, LE-lower epithelium, PC-contractile fiber cell, GC-gland cell, SS-shiny sphere, Mc-mitochondrial complex, B-bacterium in ER. (Modified after Grell & Ruthmann 1991)

Trichoplax is an exclusively marine organism distributed worldwide in tropical waters and it has been routinely collected from nearshore habitats, particularly mangrove communities [22]. In the laboratory, it can be maintained on a diversity of food sources, ranging from cryptomonad alga (e.g., *Rhodomonas*) to cynaobacteria (e.g., *Phormidium*)[36]. To feed, *Trichoplax* literally climbs atop its food using the ventral surface as a temporary extraorganismal gastric cavity [13, 33, 35]; digestion is both extracellular and by phagocytosis [19, 37]. The animals, when not feeding, are active and motile. Movement is effected by ventral ciliation [35] and by the fiber cell layer [38, 39]. The latter can be clearly as the cell layers are transparent. The animal lacks any polarity in its movement.

Reproduction is asexual and may take one of two forms [15]. Asexual reproduction by fission occurs under favorable environmental conditions, whereas small spherical colonies develop when conditions deteriorate [40]. Spherical 'swarmers' contain both cell layers and act as an asexual dispersive stage [41]. Sexual reproduction is suspected, but not been unambiguously established [11, 16]. Since our June submission, we have collected and established in permanent culture a number of clones from different geographic locations. Using these animals have succeeded in generating cleaving embryos. While these embryos arrested at the 16-cell stage, we are optimistic that experimentation with environmental conditions will permit us to complete the life cycle. If so, the full power of classical genetics will be available in this organism.

The haploid chromosome number is 6 and the total DNA content is 0.08 pg [28, 29, 42]. Genome size estimates, first published in 1977, were based on Fuelgen staining. We have recently verified these numbers based on the frequency of two single copy genes in several different genomic libraries. Two homeobox-containing proteins, *Trox 2* and *Mnx*, were confirmed to occur in single-copy by genomic Southern analysis, and their frequency assessed in libraries of known insert size and recombinant frequency. Specifically, we get 8-10X coverage in libraries of ca. 28,000 recombinants with an insert size of 15-20 kb, yielding a genome size estimate ca. 54MB, a value close to the 50MB figure obtained by Feulgen staining, confirming that *Trichoplax* possesses the smallest known animal genome, about half the size of *Caenorhabditis elegans* (97Mb).

The Phylogenetic Position of *Trichoplax* and the Lower Metazoan Alternatives



The ancestor-descendant relationships of the lower metazoans are summarized in Figure 4. *All* modern analyses, based largely on small and large subunit rDNA sequences, agree that the out-group to the animals are a 'protist' group known as chaonoflagelletes, that sponges are sister groups to the Cnidaria (corals, jellyfish, *Hydra*), and that the Lower Metazoa are sister to *all* higher animals (flies, worms, urchins, humans)[1-10]. Two matters remain unresolved in these analyses: the position of the Phylum Ctenophora and the position of the Phyla Placozoa. Depending upon the method of analysis (i.e., which algorithms are used, which taxa are compared, etc), the Ctenophora appear in either position B or C in Figure 5 and the Placozoa in positions A, B, or C [3, 5, 7, 8, 10, 43-45].

The extreme morphological simplicity of the *Trichoplax* has lead many researchers to regard *Trichoplax* as the basal-most animal, that is, in position A [12, 21, 46-48]. Morphological simplicity has long been presumed to be ancestral, since reductions in morphological complexity are typical only in cases of parasitism and/or miniaturization, whereas *Trichoplax* is a free-living organism. The two possible positions for *Trichoplax if its simplicity were secondarily reduced,* are denoted as positions D and E in Figure 5. Secondary reduction from the sponges, Position D, has never been supported in any published rDNA tree, nor has ever been entertained on morphological grounds. Secondary reduction from the Cnidaria, Position E, has recently been rendered unlikely based on mtDNA configuration (i.e., many cnidarians have linear mtDNA, whereas *Trichoplax* has circular tDNA) and by studies of 16S ribosomal RNA secondary structure [45].

Trichoplax, then, appears to be either the simplest animal (Position A)[12, 21, 46-48], the simplest animal other than sponges (Position B)[5, 7, 8, 43], or the closest common ancestor to all higher animals (Position C)[3, 10]. The question remains unresolved, not because of a shortage of *Trichoplax* data (our Consortium has completed over 1.5 Mb of sequence in initial feasibility experiments) but because of a dearth of information on comparative data from the relevant sister taxa. In specific response to the panel's suggestion that we use our preliminary results to resolve this question, we emphasize that doing so will require the isolation of paralogs of each of these genes from a number of different lower metazoan taxa. We have not had the opportunity to do so, but funding by the Human Frontiers Science Program will permit us to do just this.

Regardless of the outcome of this analysis, the need for a lower metazoan genome sequence remains a priority for the scientific community. Far more germane than the question of the precise position of *Trichoplax* within the Lower Metazoa is the question of whether it should be the first lower metazoan to be sequenced. Consider the alternatives. No representative of the sponges or the ctenophores are routinely maintained in laboratory culture as tractable experimental models. Excluding *Trichoplax*, only cnidarians are easily maintained. There are, indeed, several tractable cnidarian models with long research traditions, most notably the hydroids *Hydractinia* and *Hydra*. Each of these animals are substantially more complex than *Trichoplax*, having over a dozen cell types, and most importantly, both have genomes of considerable larger size (1.2 Gb for *Hydractinia,* and 1.6 Gb for *Hydra*).

In contrast, *Trichoplax* represents the simplest organized animal, consisting of only 4 somatic cell types. Its genome size of 5 X 10^7 bp, a mere 1.6% of humans, makes this the most cost-efficient selection among all lower metazoan organisms. All *Trichoplax* DNA, genomic libraries and protocols for shotgun sequencing are prepared and ready to go. An 8-10X coverage of *Trichoplax* raw genome sequence could be realistically completed in a few months time, about the same time that it would take to complete just one of fifteen fungal species on NIH's high priority list, and in just two weeks if the sequencing capacity of the BCCM were committed solely to this project. The effort required to sequence any of the traditional lower metazoan models is roughly 100 times that of *Trichoplax*.

Utility of the *Trichoplax* Genome Sequence

The principal benefits from the *Trichoplax* genome are those that derive from its unique phylogenetic position, its compact genome and its simple body plan. Unlike yeast, a unicellular organism, or higher metazoans, like flies and worms, with scores of cell types organized into tens of organs, *Trichoplax* represents the simplest extant multicellular animal. As such, we may expect that within the genome of *Trichoplax* lie insights to the fundamentals of many animal-specific processes important to the human condition including differentiation, cell-cell communication, signaling and pattern formation.

Comparative animal genomics: The finding that diverse higher animals, from fruitflies to nematodes to humans, share many of the same regulatory genes comprises some of the most significant scientific discoveries of the twentieth century. As more genomes are sequenced and gene function understood in these higher, or 'triploblast', model organisms, the number of structural and functional similarities continues to grow; the list now include the control of eye formation, body axis pattering, and the development of heart, nervous systems, gut and muscle. Whilst truly remarkable, however, these finding says little about the *origin* of the characters being studied. Until the more primitive 'diploblast' animals are analyzed in comparable detail, conclusions regarding the origins of genome organization, complexity and regulation will remain speculative.

As just one illustrative example, the study of homeobox genes has opened new avenues of research aimed at building an explanatory platform for understanding the diversity of animal life forms (e.g. [49]. The Antp superclass genes have only been described in animals and Hox class, at least, are definitely absent from plants and fungi. These genes are well known for their roles in axial patterning (e.g. marking position along the head-to-tail body axis in triploblasts) and because of their striking stereotypic genomic arrangement into gene clusters (organizational units). These units seem vital for correct spatial and temporal gene regulation, thereby forming a link between genome organization and gene function. It has been suggested by several authors that the origin of Hox genes was pivotal in allowing the emergence of animal multicellularity, since these gene clusters provide a genetic mechanism whereby embryos can partition information both spatially and temporally along a developing sheet of cells (e.g. Slack et al, 1993).

Once the *Trichoplax* genome sequence becomes available, it will be possible to undertake an exhaustive comparative study of the repertoire and expression of regulatory genes including homeobox genes in the simplest known animal. Genome data will determine whether clusters of homeobox genes exist illuminating the elusive relationship between gene clustering and animal body plans. In this way, *Trichoplax* genome sequence will reveal which of these gene classes are most ancient, which were originally clustered in animal genomes, putative ancestral roles, and how genomic clustering of homeobox genes evolved in concert with cellular and developmental complexity. Such illustrative examples of the utility of *Trichoplax* genome sequence can be made for several other areas of genome biology including structural features, such as telomeres and centromeres, epigenetics such as DNA methylation and chromatin structure, and information flow to RNA and the proteome.

Functional studies of human genes of anonymous function: Genes of unknown function comprise a considerable fraction of the human genome. *Any* such gene that is specific to metazoans (i.e., absent in yeast), but lost in the ecdysozoan lineage (absent in *Caenorhabditis* and *Drosophila*) must now be investigated in a higher, typically vertebrate, animal model. As the simplest animal, the *Trichoplax* genome is expected to share a large fraction of genes common to humans but lacking in non-animal genomes. While the genome sequence of yeast and other lower eukaryotes will continue to provide key insights into the function of many human proteins and processes, these systems will have limited value as compared to *Trichoplax* for animal-specific gene investigations.

Understanding the ancestral conditions: Evolution, in Jacob's words, is akin to tinkering; innovation comes only out of pre-existing materials. A corollary of this fact is less widely appreciated. Since innovation comes from pre-existing materials, the ancestral condition will constrain a systems future development. It is for this reason that an understanding of the organization of a suitably complex system requires an appreciation of its history. *Trichoplax* lacks nerves, but contains RFamide [50]. *Trichoplax* has an epithelial organization, but lacks a basement membrane. *Trichoplax* lacks muscles, yet the F-actin rich lamelliopodia of fiber cells comprise an effective locomotory system. *Trichoplax* possesses a dorsal-ventral axis but lacks an anterior-posterior axis. It may be expected the content and organization of this genome will yield insights into the animal communication and coordination before nerves, animal epithelia before true tissues, animal movement before muscles, and position information prior to an A-P axis.

Asexuality and gerontology: All animals that have been sequenced to date are organisms that a population biologist would call aclonal; that is, they reproduce by sex alone and have a fixed upper limit to their lifespans. *Trichoplax* is a clonal organism; it can and routinely does reproduce asexually. The clone we propose to sequence has an asexual generation time of about a day and has been maintained in continuous vegetative growth in the laboratory for >20 years. As far as we know, it can be maintained indefinitely. The question of how animals that have indefinite asexual lifespans achieve this end is of obvious interest to gerontologists and to cancer biologists. Do they have novel mechanisms of DNA repair? How do they evade oxidative damage? Are their telomeres unique? What are the characteristics that permit a stem cell population to remain mitotically active? These and any number of other questions become tractable once a genome is available for an organism capable of asexual reproduction.

Informing our understanding of the dawn of animal life: The *Trichoplax* sequence will provide the tools necessary to rapidly assess whether *Trichoplax* is rightly regarded, as classical invertebrate zoologists have contended, as the “mother of all metazoans”. Even if *Trichoplax* does not prove to be the basal-most animal, virtually any other phylogenetic position (for example, sister to the bilaterians or sister to the Eumetazoa) will inform our understanding of mode of life of early animals.

Facilitating research in other lower metazoan organisms: Research in lower metazoan model systems is often limited by the availability of genomic resources. As but one example of the potential for immediate impact, consider the example of graft rejection. Most lower metazoans display systems of allorecognition. Indeed, many believe that the evolutionary origins of the MHC lie in lower metazoan allorecognition. Molecular markers spanning a lower metazoan allorecognition complex exist and a walk is underway to characterize the interval in one cnidarian. The *Trichoplax* genome may well allow detection of long-range synteny in this region and the immediate characterization of an invertebrate allorecognition complex. Comparable impacts may be expected in other areas of active research using lower metazoans (e.g., natural products from sponges and cnidarians, novel fluorescent molecules from cnidarians, etc.).

Marine model and environmental genomics: The overwhelming majority of eukaryotic organisms sequenced to date have been terrestrial organisms (the pufferfish being the sole exception. The rapid environmental changes the oceans are now undergoing call for a sensitive experimental system wherein environmental changes can be readily detected. The small size of *Trichoplax*, its compact genome, its large surface-area to volume ratio, and its habit of respiring by diffusion all suggest that the development of expression array chips for *Trichoplax* will bring environmental genomics to the oceans.

Available Biological & Genomic Resources

Trichoplax are maintained in continuous laboratory culture and are available for public distribution. One such culture (BD) is propagated asexually in large-scale for genomic applications. Several additional isolates have been collected from marine environments at several locations worldwide and are likewise available for public distribution.

A constant source of high molecular weight (>100 Mb) DNA and high-quality RNA is available from the publicly available BD isolate. A small insert (2-3 kb) plasmid library with 300,000 clones, representing about 15X genome coverage, has been constructed at BCM. In a pilot genome sequencing experiment, approximately 1500 sequences were generated (examples are summarized in Table 1) with an average of 700 phred20 bases per sequence and 85% success rate per reaction, well above most standards. From a total read length of 1.4 Mb, the genome of *Trichoplax* is estimated to have a GC composition of 36%. Most importantly, a large fraction (over 70%) of these preliminary reads gave signficant hit to genes, with the majority showing greatest similarity to animal genes (Table 1). Clearly, the *Trichoplax* genome sequence will be rich in both gene content and density. These initial results provide further justification for the genome sequence of *Trichoplax*.

Table 1. Example BLAST Results from *Trichoplax* Genome Pilot Sequencing with highest E values

Homolog	Organism	*Score (bits)*	*E Value*
zinc finger, C3HC4 type	Caenorhabditis elegans	111	9e-24
mitogen inducible 2	Homo sapiens	80.4	2e-14
chondroitin sulfate proteoglycan 6	Homo sapiens	87.8	1e-16
hypothetical protein DKFZp434N0735	Homo sapiens	88.5	8e-17
fibropellin III	Heliocidaris erythrogramma	54.5	2e-22
EGF homolog	Strongylocentrotus purpuratus	54.8	3e-20
ribosomal protein L35A	Spodoptera frugiperda	70.2	3e-11
nitric-oxide synthase, brain	Rattus norvegicus	82.9	4e-15
hypothetical protein FLJ21432	Homo sapiens	56.6	1e-12
cystic fibrosis transmembrane conductance regulator	Rattus norvegicus	64.3	1e-12
peroxisomal fatty acyl-coA oxidase	Homo sapiens	71.3	5e-25
glutamate decarboxylase	Homo sapiens	68.1	8e-11
zinc finger protein 162	Mus musculus	99.7	9e-25

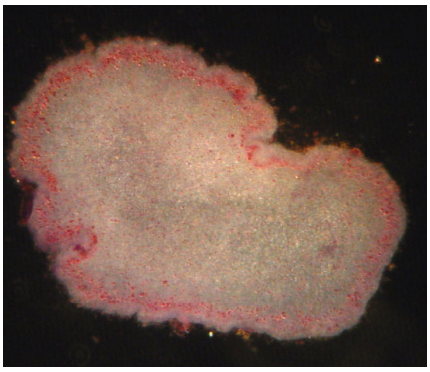
In addition, Yale has constructed several genomic libraries (8-10X coverage) and cDNA libraries in phage and plasmids. An arrayed fosmid [43] (30-40 kb inserts with 10X coverage) and a full-length cDNA library will be completed by September 2002. A modest EST project is currently underway in Hannover, Germany and Reading, UK. The EST project will switch to the sequencing of full-length cDNA clones from Yale for transcript mapping on annotated genome sequence.

Recent technical advances include the ability to perform whole mount *in situ* hybridization (Figure 5), immunolocalization (not shown) and gene functional studies with GFP and RNAi. Because large numbers of animals can be processed simultaneously in 96- or 384-well plates, it is possible to

systematically localize RNA from a large number of genes. For example, a high-throughput localization of the entire set of *Trichoplax* transcription factors genes would be feasible once genome sequence becomes available.

Figure 5. Whole Mount In Situ Hybridization

Trichoplax showing the pattern (red staining) of Cnox2 mRNA expression. Cnox2 is a member of the Antp homeobox gene superfamily.



Trichoplax Communities

The *Trichoplax* genome will have immediate impact on two communities and is expected to generate a third:

1. Bioinformatics/Comparative Genomics Community: While it is conventional to think in terms of a group of researchers devoted to a particular organism as the relevant community, the growth of comparative genomics and data mining as a research style effectively constitutes a new community. The phylogenetic position of *Trichoplax* alone will be sufficient to spawn new analyses, such as those outlined in the rationale and undoubtedly others we have yet to anticipate. We anticipate the *Trichoplax* genome sequence to become the standard basal group in the analysis of metazoan genomes, genes and biological processes.

2. Lower Metazoan Community: In addition to the *Trichoplax*, the lower metazoans include representatives of three phyla: the sponges, ctenophores and cnidarians. There exists a pressing need for lower metazoan genomic information. This need spawned a special session at the 9th International Workshop on Hydroid Development (2001) and a follow-up meeting was held at the Developmental Biology Center at the University of California, Irvine in July 2002. A *Hydra* EST project is now underway. *Hydra*, however, has a large (1600 MB), AT rich (71%), genome and proposals for whole genome sequencing have yet to be advanced. Appreciation of the larger need for a lower metazoan model system has led to the scheduling of a special session of the Society of Integrative and Comparative Biology meeting in January 2003 entitled "Model Systems for Basal Metazoans". Members of the *Trichoplax* Consortium will be active participants in this dialogue. While the *Trichoplax* genome will not exhaust enthusiasm for further lower metazoan genomics, it is surely the only genome that can be rapidly sequenced today. The availability of the *Trichoplax* sequence will surely facilitate gene discovery in other lower metazoans and for many problems will immediately provide an alternative lower metazoan model.

3. Establishment of a Trichoplax Community: Trichoplax has long been regarded as an enigmatic animal, receiving relatively little attention outside of an occasional study by invertebrate zoologists. This is unfortunate, as the animal is an eminently tractable laboratory organism. Recent funding of the *International Trichoplax Consortium* from the Human Frontiers Science Program indicates widespread scientific interest in the development of this system. The members of the *ITC* are committed to the development of *Trichoplax* as a new basal metazoan model system with swift distribution of biological and informatic resources to the international scientific community. As outlined in this proposal,

standard clones, culturing techniques, biological resources and molecular techniques have been and are continuing to be developed by this group. The members of this group expect that the availability of the *Trichoplax* resources will spawn the rapid development of *Trichoplax* as the prime experimental organism of choice for lower metazoan studies. Genome sequence will greatly accelerate the adoption of this system by the scientific community.

Experimental Strategies for *Trichoplax* Genome Sequencing

The methodology for whole genome shotgun sequencing of *Trichoplax* will be based on strategies and protocols currently employed at the BCM-HGSC. The *Trichoplax* genome is ca.50 MB. Because of the small size and novelty of this genome, it would be most informative to produce a finished sequence.

The approach being used at the BCM-HGSC for *Drosophila pseudoobscura* and microbial genomes will be used for *Trichoplax* as well. This involves WGS sequencing to 8x coverage, which will require about 1 million reads (80% pass rate and 500 phred20 bases per read). This is likely to be reduced since the genome may be smaller than 50 Mb and read lengths and pass rates are always improving. A small insert genomic library has been prepared and no difficulties were encountered, so there does not seem to be any unexpected issues with cloning genomic DNA from *Trichoplax*. In addition to the WGS reads from a small insert library, end sequencing of the fosmid library will be performed to aid in assembly. About 7500 fosmids will be end sequenced, representing about 5x clone coverage and about one half of the existing library.

Sequences will be assembled using the ATLAS assembler, developed for the rat genome project at the BCM-HGSC. The ATLAS assembler has been adapted for use in WGS projects such as this where there is no BAC sequencing. The draft sequence produced by the ATLAS assembly will be finished using the standard BCM-HGSC finishing pipeline. Although this has been built around BAC-based finishing for the human, it has been applied to microbial WGS finishing, analogous to the *Trichoplax* strategy. The sequencing of a large number of fosmid ends will allow the finishing to be localized to fosmids, however, analogous to the human genome.

The 1 million reads required is about 2 weeks of sequencing at the BCM-HGSC. The rate of finishing at the BCM-HGSC for human has been about 18 Mb per month, so *Trichoplax* could be completed in 3 months if the entire finishing capacity were directed at the project. The cost to complete the assembled draft (8X coverage) genome sequence is estimated at $2.0M. Once the draft genome sequence is complete, we estimate an additional cost $2.0M ($0.04 per bp) would be required to prepare the finished genome sequence. Because *Trichoplax* is so distantly related to other animals with sequenced genomes, a finished genome sequence would be most desirable and informative.

It would also be desirable to sequence full-length cDNAs, both for annotation purposes as well as for studies of gene expression networks. The BCM-HGSC is at a capacity of approximately 1000 cDNA sequences per month, although most of this is directed at human and (in the future) rat sequencing. Nevertheless, it would be desirable to direct some of this capacity at *Trichoplax* cDNAs. We request an additional $200K for this project.

As with other projects, annotation of the *Trichoplax* genome would be community oriented and involve researchers studying *Trichoplax* and related organisms, experts in specific areas (e.g. gene families such as regulatory proteins or transport systems), as well as bioinformaticians with tools and expertise at whole genome comparisons and other large-scale analyses. There is currently much activity in the area of distributed annotation systems as well as technology for displaying annotated genomes and data distribution. One example under development at the BCM-HGSC is the Genboree genome browser

(http://www.genboree.org) but there are others and by the time this project is completed we expect this area to have advanced and good solutions for these needs will be available. Yale University desires to play a major role in the annotation processes, including identifying and hiring postdoctoral personnel, trained at the BCM-HGSC on Genboree and bioinformatics.

References Cited

1. Aleshin, V.V., Vladychenskaya, N.S., Kedrova, O.S., Milyutina, I.A., Petrov, N.B. (1995) Phylogeny of invertebrates deduced from 18S rRNA comparisons. *Mol. Biol.* 29:843-855.
2. Wainright, P.O., Hinkle, G., Sogin, M. L., Stickel, S. K. (1993) Monophyletic origins of the Metazoa: an evolutionary link with fungi. Science. 260:340-342.
3. Collins, A.G. (1998) Evaluating multiple alternative hypotheses for the origin of Bilateria: An analysis of 18S rRNA molecular evidence. Proc Nat Acad Sci USA. 95:15458-15463.
4. Christen, R., Ratto, A., Baroin, A., Perasso, R., Grell, K. G., Adoutte, A. (1991) An analysis of the origin of metazoans, using comparisons of partial sequences of the 28S RNA, reveals an early emergence of triploblasts. EMBO J. 10:499-503.
5. Podar, M., Haddock, S.H., Sogin, M.L., and Harbison, G.R. (2001) A molecular phylogenetic framework for the phylum Ctenophora using 18S rRNA genes. Mol Phylogenet Evol. 21:218-30.
6. Schierwater, B. and Desalle, R. (2001) Current problems with the zootype and the early evolution of Hox genes. J Exp Zool. 291:169-74.
7. Cavalier-Smith, T., Allsopp, M., Chao, E., Boury-Esnault, N., and Vacelet, J. (1996) Sponge phylogeny, animal monophyly, and the origin of the nervous system. Can. J. Zool. 74:2031-2045.
8. Kim, J., Kim, W., and Cunningham, C.W. (1999) A new perspective on lower metazoan relationships from 18S rDNA sequences. Mol Biol Evol. 16:423-7.
9. Medina, M., Collins, A.G., Silberman, J.D., and Sogin, M.L. (2001) Evaluating hypotheses of basal animal phylogeny using complete sequences of large and small subunit rRNA. Proc Natl Acad Sci U S A. 98:9707-12.
10. Pawlowski, J., Montoya-Burgos, J.-I., Fahrni, J.F., Wüest, J., Zaninetti, L. (1996) Origin of the Mesozoa inferred from 18S rRNA gene sequences. *Mol. Biol. Evol.* 13:1128-1132.
11. Grell, K.G. (1971) Embryonalentwicklung bei *TrichopIax adhaerens.* F. E. Schulze. Naturwissenschaften. 58:570.
12. Grell, K.G. (1971) *TrichopIax adhaerens* F. E. Schulze und die Entstehung der Metazoen. Naturwiss. Rundschau. 24:160-161.
13. Grell, K.G. (1972) Eibildung und Furchung von *Trichoplax adhaerens* F. E. Schulze (Placozoa). Z. Morphol. Tiere. 73:297-314.
14. Grell, K.G. (1973) *Trichoplax adhaerens* (Placozoa). Bewegung und Organisation. Encyclopaedla Cinematographica Inst wiss.Film, Göttingen. Film E 1918
15. Grell, K.G. (1973) *Trichop1ax adhaerens* (Placozoa). Vermehrung. Encyciopaedla Cinematographica. Inst. wiss. Film, Göttingen. Film E1919
16. Grell, K.G. (1973) *Trichoplax adhaerens* (Placozoa). Eizellen und Furchungsstadien. Encyclopaedla Cinematographica Inst. wiss. Film, Göttingen. Film E 1920
17. Grell, K.G., Ruthmann, A. (1991) Placozoa in Microscopic Anatomy of Invertebrates. 2 Placozoa, Porifera, Cnidaria, and Ctenophora. Zool. Zhurn. 66:13-27.
18. Ivanav, D.L., Malakhov, V. V., Prilepsky, G. V., Tzetlin, A. (1982) Fine morphology and ultrastructure of a primitive multicellular organism *Trichoplax* sp.. 2. Ultrastructure of adult individuals. Zool. Zhurn. 61:645-652.
19. Ivanov, A.Y. (1973) *Trichoplax adhaerens*, a Phagocytella-like animal. Zool. Zhurn. 52:1117-1131.
20. Ivanov, D.L., Malakhov, V. V., Tzetlin, A. B. (1980) New findings (Discovery) of a primitive multicellular organism *Trichoplax sp.* Zool. Zhurn. 59:1735-1739.
21. Ivanov, D.L., Malakhov, V. V., Tzetlin, A. B. (1980) Fine morphology and ultrastructure of a primitive multicellular organism *TrichopIax sp.* 1. Morphology of adults and vagrants by the data of scanning electronmicroscopy. Zool. Zhurn. 59:1765-1767.
22. Pearse, V.B. (1989) Stalking the wild Placozoan: Biogeographiy and ecology of *Trichoplax* in the pacific. American Zoologist. 29:175A.

23. Rassat, J., A Ruthmann (1979) *Trichoplax adhaerens* F. E. Schulze (Placozoa) in the Scanning Electron Microscope. Zoomorphology. 93:59-72.
24. Schubotz, H. (1912) Ist *Trichoplax* die umgewandelte Planula einer Hydromeduse? Zool. Anz. 39: 582-585.
25. Schulze, F.E. (1891) Über *Trichoplax adhaerens*. Physik. Abh. Kgl. Akad. Wiss. Berlin.1-23.
26. Stiasny, G. (1903) Einige histologische Details über *Trichoplax adhaerens*. Z. Wiss. Zool. 75:430-436.
27. Schulze, F.E., Grell K. G. (1984) A new culture method for *Trichoplax adhaerens*. Zeitschrift für Naturforschung. C:1072.
28. Ruthmann, A., Wenderoth, H. (1975) Der DNA-Gehalt der Zellen bei dem primitiven Metazoon *Trichoplax adhaerens* F. F. Schulze. Cytobiologie. 10:421-431.
29. Ruthmann, A. (1977) Cell differentiation, DNA content and chromosomes of *TrichopIax adhaerens* F. E.. Schulze. Cytobiologie. 15:58-64.
30. Ruthmann, A., Grell, K. G, Benwitz, G. (1981) DNA content and fragmentation of the Egg Nucleus of *TrichopIax adhaerens.* Zeitschrift für Naturforschung. 36C:564-567.
31. Schulze, F.E. (1883) *Trichoplax adhaerens,* nov. gen., nov. spec. Zool. Anz. 6:92-97.
32. Grell, K.G. (1971) Über den Ursprung der Metazoen. Mikrokosmos. 60:97-102.
33. Grell, K.G., Benwitz, G. (1971) Die Ultrastruktur von *Trichoplax adhaerens F. E.* Schulze. Cytobiol. 4:216-240.
34. Grell, K.G., Benwitz, G. (1981) Ergänzende Untersuchungen zur Ultrastruktur von *TrichopIax adhaerens* F. E. Schulze (Placozoa). Zoomorphol. 98:47-67.
35. Ruthmann, A., Behrendt, G., Wahl, R. (1986) The ventral epithelium of *Trichoplax adhaerens* (Placozoa): Cytoskeletal structures, cell contacts and endocytosis. Zoomorphology. 106:115-122.
36. Grell, K.G. (1983) Ein neues Kulturverfahren für *Trichoplax adhaerens* F. E. Schulze. Z. Naturforsch. 38c:1072.
37. Wenderoth, H. (1986) Transepithelial cytophagy by *Trichoplax adhaerens* F. E. Schulze (Placozoa) feeding on yeast. Z. Naturforsch. 41C:343-347.
38. Kuhl, W., Kuhl, G. (1963) Bewegungsphysiologische Untersuchungen an *Trichoplax adhaerens* F. E. Schulze. Zool. Anz. Suppl. 26:460-469.
39. Kuhl, W., Kuhl, G (1966) Untersuchungen über das Bewegungsverhalten von *Trichoplax adhaerens* F. E. Schulze (Zeittransfomation: Zeitraffung). Z. Morph. Ökol. Tiere. 56:417-435.
40. Thiemann, M., Ruthmann, A. (1991) Alternative modes of asexual reproduction in *Trichoplax adhaerens* (Placozoa). Zoomorphol. 110:165-174.
41. Thiemann, M., Ruthmann, A. (1990) Spherical forms of *Trichoplax adhaerens* (Placozoa). Zoomorphol. 110:37-45.
42. Birstein, V.J. (1989) On the karyotype of *Trichaplax sp.* (Placozoa). Biol. Zentralbl. 108:63-67.
43. Collins, A.G. (2002) Phylogeny of Medusozoa and the evolution of cnidarian life cycle. J. Evol. Biol. 15:418-432.
44. Schierwater, B., Dellaporta, S., and DeSalle, R. (2002) Is the evolution of Cnox-2 Hox/ParaHox genes "multicolored" and "polygenealogical?" Mol Phylogenet Evol. 24:374.
45. Ender, A. and Schierwater, B. (2002) Placozoa are not derived cnidarians: Evidence from molecular morphology. Mol. Biol. Evol. *in press*
46. Grell, K.G., *Trichaplax adhaerens and the origin of Metazoa.* in *Origine dei Grandi Phyla dei Metazol. Acc. Naz. Lincei, Gonvegno intern*. 1981. p. 107-121.
47. Syed, T. and Schierwater, B. (2002) The evolution of the Placozoa: a new morphological model. Senckenbergiana lethaea. 82:259-270.
48. Syed, T. and Schierwater, B. (2002) *Trichoplax adhaerens*: a phylum discovered, forgotten and resdiscovered. View et Milieu. *in press*
49. Deutsch, J., Le Guyader, H. The neuronal zootype. An hypothesis. Comptes Rendus de l Academie des Sciences - Serie Iii, Sciences de la Vie. 321:713-719.
50. Schuchert, P. (1993) *Trichoplax adhaerens* (Phylum Placozoa) has cells that react with antibodies against the neuropeptide RFamide. Acta Zoologica. 74:115-117.