A Case for Sequencing the *Trichoplax* Genome

**Principal Investigators of the *Trichoplax* Genome Consortium:**

Stephen L. Dellaporta\(^\sigma\), Leo W. Buss and Bernd Schierwater\(^\gamma\), Yale University
George Weinstock\(^\beta\), Baylor College of Medicine Human Genome Sequencing Center
Peter Holland\(^\xi\), Oxford University (Collaborator)

\(^\sigma\)Contact information: Yale University, Dept. MCD Biology and Ecology & Evolution, New Haven, CT 06520-8104. (203) 432-3895. Email: stephen.dellaporta@yale.edu and leo.buss@yale.edu
\(^\beta\) Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030. (713) 798-6539 Email: gwstock@bcm.tmc.edu
\(^\gamma\) Affiliate Yale Scientist. (Primary address: Tierärztliche 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 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 identify 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,
- To launch an experimental platform for the study of the ancestral function for any animal-specific gene, RNA, protein or biological process.
Four phyla comprise the Lower Metazoa: Placozoa, Porifera (sponges), Ctenophora (comb jellies), and Cnidaria (e.g., corals and jellyfish). While the phylogenetic relationships amongst these taxa are uncertain, the group is unambiguously established to be a sister taxa to the higher animals [1-7]. The Placozoa merit particular attention in that its sole representative, *Trichoplax adhaerens* possesses the simplest construction of any animal [8-23] and it is an easily maintained experimental organism [24]. Most notably, *Trichoplax* has the smallest genome of any animal yet measured (less than 50 Mb)[25-27], only 10X that of the *Escherichia coli* and less than 1% the size of the human genome. Six-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 new 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 principles of animal complexity.

**The Biology of Trichoplax**

*Trichoplax* is the sole representative of the Phylum Placozoa [11, 22, 28]. 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 [11, 14, 15, 18, 20, 29-32]. It is, without question, anatomically the simplest free-living (i.e., not parasitic) animal known.

*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...
Trichoplax is an exclusively marine organism distributed worldwide in tropical waters and it has been routinely collected from nearshore habitats, particularly mangrove communities [19]. In the laboratory, it can be maintained on a diversity of food sources, ranging from cryptomonad alga (e.g., *Rhodomonas*) to cyanobacteria (e.g., *Phormidium*)[33]. To feed, *Trichoplax* literally climbs atop its food using the ventral surface as a temporary extraorganismal gastric cavity [34]; digestion is both extracellular and by phagocytosis [16, 35]. The animals, when not feeding, are active and motile. Movement is effected by ventral ciliation [32] and by the fiber cell layer [36, 37]. 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 [12]. Asexual reproduction by fission occurs under favorable environmental conditions, whereas small spherical colonies develop when conditions deteriorate [38]. Spherical 'swarmers' contain both cell layers and act as an asexual dispersive stage [39]. Sexual reproduction is suspected, but has not been unambiguously established [8, 13]. The haploid chromosome number is 6 and the total DNA content is 0.08 pg [25, 26, 40].

The extreme morphological simplicity of the *Trichoplax* has lead many researchers to regard *Trichoplax* as the basal-most animal [9, 18, 41]. Morphological simplicity is 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. Several alternative phylogenetic positions, however, are possible. Two that have held some prominence are that sponges are the basal-most animals followed by a clade with *Trichoplax* sister to all animals excluding sponges or all animals excluding sponges and ctenophores. The second common alternative is to regard *Trichoplax* as the nearest common ancestor to all higher metazoans, that is the bilaterians (or triploblasts). Clearly other relationships can be imagined. Attempts to resolve these relationships by sequence comparisons of small and large subunit rDNAs have been inconclusive [42].
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 currently being established in the laboratory. Additional isolates will be available for public distribution as soon as permanent cultures have been established.

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 phred 20 bases per sequence and 85% success rate per reaction, well above most standards. From a total read length of 1.4 megabases, the genome of *Trichoplax* is estimated contain a GC composition of 36%.

Table 1. Example BLAST Results from *Trichoplax* Genome Pilot Sequencing

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

In addition, Yale has constructed several genomic libraries (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 4) and immunolocalization (not shown) of *Trichoplax*. 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 4. 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. Researchers in each of these phyla have their own organizations sponsoring periodic meetings. While sponges and ctenophores have not yet to be established as tractable laboratory systems, several cnidarians, particularly the hydroids *Hydra* and *Hydractinia*, have rich traditions as experimental animals.

Nevertheless, 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 is scheduled for hydroid workers at the Developmental Biology Center at the University of California, Irvine in July 2002. A *Hydra* EST project is now underway, as is the production of a cosmid libraries. *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.

Utility of the Trichoplax Genome Sequence

The principal benefits from the Trichoplax genome are those that derive from its unique phylogenetic position and the most important of these have been summarized in the rationale. We expect, however, that the availability of this sequence will spawn the development of a new research community centered on this organism and that a variety of unique features of the organism will permit progress in several domains. We speculate freely here on the future we imagine.

1. 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.

2. 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 [44]. *Trichoplax* has an epithelial organization, but lacks a basement membrane. *Trichoplax* lacks muscles, yet the F-actin rich lamellipodia 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.

3. **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.

4. **Functional studies of genes of unknown function:** A considerable fraction of the human genome is genes of unknown function. Genes that are 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. The availability of a simple metazoan model will provide an experimental platform for the study of genes of unknown function.

5. **Drug testing:** *Trichoplax* has a variety of advantages as an experimental organism. Its asexual generation time is on the order of a day, it can be maintained in clonal homogeneity, it can be grown in small volumes, and its large surface area to volume ratio facilitates uptake of small molecules. The characteristics are akin to what one expects in a protist or fungal system, yet *Trichoplax* provides these advantages in an animal model.

6. **Wound healing/ regeneration:** The animals whose genomes have been sequenced all have limited capacities for wound healing and regeneration. *Trichoplax* rapidly closes experimentally inflicted wounds and can regenerate entire organisms from fragments. Indeed, *Trichoplax* can be dissociated into single cells and the fragments will reassemble into a viable organism just as is the case for sponges [45]. Cell adhesion, directed morphogenetic movement, dedifferentiation, wound healing and regeneration may all be readily addressed in this system.

7. **Natural products:** The search for pharmaceuticals from natural products is ongoing, with most major taxa having been surveyed for biologically active compounds. *Trichoplax*, as the sole representative of its phylum, is a novel source. As a slow moving,
free-living organism without obvious defenses, it seems certain that *Trichoplax* must produce and concentrate metabolites that deter predators.

8. *Marine model and environmental genomics:* All of the organisms sequenced to date have been terrestrial organisms. While an ascidian project and plans for an urchin project are well advanced, we can readily expect that the *Trichoplax* genome along will permit us to understand the complement of genes associated with the marine habit. Moreover, 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.

9. *Innate immune systems:* The innate immune system of invertebrates has led to new appreciation of elements of the vertebrate immune system (e.g., the characterization of *Toll*-like receptors). Many attributes of the vertebrate system, however, are unknown in ecdysozoans (e.g., fruit flies and nematodes), but are widely distributed in Lower Metazoa. As one example, graft rejection does not occur in ecdysozoans, yet allorecognition is a prominent feature of the ecology of lower metazoans and of vertebrates. Indeed, many believe that the evolutionary origins of the MHC lie in lower metazoan historecognition. 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.

10. *Genome Organization & Content:*

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 patterning, 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. [46]). The *Antp* superclass genes have only been described in animals and, at least for the Hox class, they 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 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.

**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 less than 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. This involves WGS sequencing to 8x coverage, which will require about 1 million reads (80% pass rate and 500 Phred 20 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 $2.5M. Once the draft genome sequence is complete, we estimate an additional cost $2.5M ($0.05 per bp) would be required to prepare the finished genome sequence.

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 $1M 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**