Re: Sequencing the *Aplysia* Genome
FOR GENOME SEQUENCING SUBMISSION: **submission_requests@mail.nih.gov**

Dear Dr. Felsenfeld and Committee members,

We are resubmitting as an attachment a white paper requesting an 8-fold coverage draft genome sequence for the opisthobranch mollusc, *Aplysia californica*. As described in the proposal, *Aplysia* is a major model organism for studies in cellular, molecular and behavioral neuroscience, integrative physiology and system biology. It is also ideally placed taxonomically to serve as an important organism for comparative genomics. Access to the genome sequence will provide a stimulus to the already large and dynamic *Aplysia* community to carry out previously intractable studies in the areas of metazoan evolution, developmental biology, neuroscience and human health. That they are anxious for this is attested to by the more than three dozen major laboratories working with *Aplysia* and related molluscs, who have sent supporting letters or agreed to have their names attached to this document.

This proposal was written in collaboration with Drs. Lander and Birren from the sequencing center (WI-CGR) at MIT, and was vetted by major investigators in neurobiology, genomics, and related model systems, who have also sent letters of support.

Specifically, the revised white paper addresses four issues that arose during the previous review process, namely

(1) **The utility of whole genome sequence** (versus ESTs or cDNAs) for answering the most important questions for the community (**pages 8-9**). In particular, we stress the fact that the growing value of *Aplysia* as a unique and experimentally tractable model system for functional genomics and neurobiology would be significantly compromised by the lack of genomic information. The access to a whole set of genes in the organism, their regulatory regions, and their intron-exon organization is crucial for applying existing molecular approaches and taking full advantage of the largest (up to 1 mm) somatic cells in the animal kingdom. It provides the long-desired opportunity to bring genomic sciences into the unexplored realm of the living neuron, an opportunity to link activities of gene products within cellular compartments to systems biology and the behavior of the whole organism.

(2) **The strength of the organism community** and its ability to usefully exploit a full genome sequence (**pages 3-4**). More than 100 laboratories and NIH supported *Aplysia* resources and centers provide the required tools and infrastructure for the community to immediately explore the functions of individual genes as well as their orchestrated activities at the level of the single cell with nearly real-time resolution. As a result, many fundamental problems in functional genomics, cell biology and neuroscience can be analyzed more effectively in *Aplysia* than in *Drosophila*, *C. elegans* and vertebrates.

(3) Accepting the obvious needs of comparative genomics for an additional lophotrochozoan genome (and molluscan genome in particular) we consider that a choice may have to be made between *Aplysia* and another mollusk under consideration by this process (especially those with smaller genomes). To this end, we added a section (*Aplysia* amongst other Molluscs; **pages 1-2**) as well as **Figs. 1 & 2** (Appendix) that stress the fact that *Aplysia* is

probably the best-studied mollusk so far. Due to the value of *Aplysia* as a widely accepted and important model system for biomedical sciences, the *Aplysia* research community is best suited to instantly explore the utility of the genome.

(4) Finally, it might appear that *Aplysia* is a somewhat derived gastropod, and a more basal mollusk may provide more insight into questions related to genome evolution. As stated by Brusca & Brusca (2003), 'the Mollusca is such a diverse phylum, and so many taxa below the class level are apparently artificial (i.e. polyphyletic or paraphyletic), that efforts to trace their evolutionary history have often led to frustration." In view of the fact that the phylogenetic relationships among even different molluscan classes are not well established (in part due to the lack of sufficient molecular and genomic data) four major molluscan classes (including Gastropoda) can be conservatively collapsed into one clade with Gastropoda as the largest group. Two apparently more basal molluscan classes are significantly less accessible for experimentalists. Thus, exploration of their genomes will also be tempered by the lack of required tools and background information that would be difficult to obtain within a reasonably short time (*e.g.*, it took nearly 40 years to map *Aplysia* networks).

On balance, we propose that *Aplysia* should be one of the high priority sequencing targets as it will allow the exploration of a lophotrochozoan genome in great detail and will have immediate and major impact on research in the biomedical sciences including major breakthroughs in the understanding of the basic principles of neuronal organization and memory mechanisms.

Please also note that we have included two sets of Appendix Material:

(a) a list of scientists providing letters of support; a summary table demonstrating the features of *Aplysia* that make it an ideal experimental model; a table with information on our current EST libraries; lists of major societies, meetings, databases and laboratories that have an interest in *Aplysia* biology; and text references; and

(b) a separate PDF file containing the actual letters of support.

The first appendix currently is part of the main document file, but if you prefer, we can separate it.

We thank you in advance for your guidance, and please let us know if there is any other information you need.

On the behalf of the *Aplysia* genome consortium,

Leonid Moroz, Ph.D. & Eric R. Kandel, M.D.

Contact information:
Leonid Moroz, Ph.D.
University of Florida
Whitney Laboratory for Marine Bioscience, UF Brain Institute,
Department of Neuroscience & Bio-Nano Center,
9505 Ocean Shore Blvd.

St. Augustine, FL 32080
Phone: (904) 461-4020; Fax: (904) 461-4052; Email: moroz@whitney.ufl.edu

Eric R. Kandel, M.D.
Columbia University
Center for Neurobiology and Behavior
Senior Investigator, Howard Hughes Medical Institute
1051 Riverside Drive
New York, NY 10032
(212) 543-5204 tel; (212) 543-5474 fax; Email: erk5@columbia.edu

Sequencing the *Aplysia* Genome: a model for single cell, real-time and comparative genomics*

Leonid L. Moroz[1], Jingyue Ju[2], James J. Russo[2], S. Puthanveetti[2], Andrea Kohn[1], Monica Medina[3], Patrick J. Walsh[4], Bruce Birren[5], Eric S. Lander[5], and Eric R. Kandel[2,6]

[1]University of Florida, [2]Columbia University, [3]Joint Genome Institute, [4]National Resource for *Aplysia* at the University of Miami, [5]Whitehead Institute & MIT Center for Genome Research, and [6]HHMI

Contact Information: [1]Ph. (904) 461-4020; Fax: (904) 461-4052; moroz@whitney.ufl.edu (L.L. Moroz); [2]Ph.(212)851-5172; Fax: (212) 851-5176;dj222@columbia.edu (J.Ju); [2]Ph.(212) 543-5204; Fax: (212) 543-5474; erk5@columbia.edu (E.R. Kandel).

Summary

The marine opisthobranch mollusc *Aplysia californica* is a powerful experimental system in cellular, molecular and behavioral neuroscience as well as cell and evolutionary biology because of the distinctive organization of its nervous system, which makes it advantageous for cellular and comparative analysis of a variety of behaviors and learning and memory. *Aplysia*'s large neurons allow examination of neuronal architecture of instinctive and learned behaviors at the level of single characterized cells and defined cellular compartments (*e.g.*, synapses). As a result, many fundamental problems in neurobiology can be analyzed more effectively in *Aplysia* than in *Drosophila*, *C. elegans* and vertebrates. In a larger sense work on *Aplysia* is synergistic with these other experimental systems. We propose to sequence the genome of *Aplysia* both to gain access to genomic mechanisms of basic neuronal and other functions and to study these mechanisms in real physiological time with single-neuron resolution.

The distinction of *Aplysia* as a neurobiological and cellular model system is due to the following: (1) Its nervous system has a relatively small number of nerve cells. (2) Many of these cells are large (sometimes gigantic, up to 1 mm in diameter). (3) As a result of their size, pigmentation, and position in the nervous system, many cells can easily be uniquely identified at the single cell level and can be reliably linked to the animal's behavior. (4) The cells provide enough messenger RNA to generate cDNA libraries from single cells. (5) These neurons can be isolated and cultured *in vitro* and they form circuits which can be explored at molecular and cellular detail. (6) The animal generates a variety of behaviors many of which can be specified in terms of their underlying circuitry. (7) Some of these behaviors can be modified by different forms of learning.

Aplysia are easy to rear in the laboratory from fertilized ova to mature adults and it is possible to obtain an inbred stock of reproducing animals. In 1995, the NIH established a National Resource Center for *Aplysia* at the University of Miami to meet the growing needs of the biomedical community. This Center supplies over 20,000 cultured *Aplysia* at all developmental stages annually to the research community throughout the world. This community consists of more than 100 laboratories and over a thousand investigators with overall research budgets of tens of millions of dollars annually. In addition to its value to the neurobiological community, *Aplysia* is also of interest from a comparative biological perspective. *Aplysia* is a free-living representative of Mollusca, the second largest animal phylum (after Arthropoda). Members of this phylum have received relatively little genetic study, even though they are of considerable significance for evolutionary and developmental biology and for basic and applied biomedical studies. *Aplysia* has a stable diploid genome consisting of 17 haploid chromosomes with highly polyploid central neurons (n=120,000). The genome size of 1.8×10^9 base pairs is typical for many molluscan species .

Over 200,000 expressed sequence tags (ESTs) have already been sequenced, including more than 50,000 unique sequences representing several thousand genes, from the central nervous system of *Aplysia*. These ESTs have revealed many signaling molecules and pathways, including their key receptors, kinases and substrate proteins, but many rare yet key genes (*e.g.*, ionic channels) have escaped detection so far. To achieve a complete understanding of the biologically important genes and their regulatory elements, a complete genomic analysis is required. In turn, genomic analysis will be greatly facilitated by the available EST data, which provides many critical markers for annotation.

The *Aplysia* genome project will serve two major functions. First, it will be important for neuroscience by providing a critical resource for facilitating: (a) identification of all genes and regulatory regions of this valuable organism, (b) assembling microarrays for global expression studies including those based on single neurons and their processes, and (c) gene function analysis by RNA interference screens. Second, the *Aplysi*a genome will be important for comparative, developmental and evolutionary biology. *Aplysia* will serve as a complementary invertebrate/molluscan genomic model from the most diverse and second largest clade of bilaterally symmetrical animals. The *Aplysia* genome will promote studies of metazoan evolution, developmental biology, neuroscience, and human health, including identifying and validating therapeutic targets for human disorders that are related to aging, cancer, the central nervous system, as well as learning and memory mechanisms.

* A complete list of the laboratories of the *Aplysia* sequencing consortium is attached at the end of the document.

The mollusc *Aplysia* as a Key Model for Comparative Genomics. *A. californica* and related opisthobranchs are free-living representatives of Mollusca (class Gastropoda), the second largest animal phylum (after Arthropoda) with more than 100,000 extant marine, freshwater and land species that trace their origin to the Cambrian period. Because of their shell and chitinous radula (a major part of the feeding apparatus) this phylum has one of the best paleontological records. More than 500 million years of their evolutionary history has been reconstructed; 70,000 known well-preserved molluscan fossils provide both landmarks for calibration of molecular clocks and important resources for the needs of comparative and evolutionary genomics. For example, the origin of the opisthobranch molluscs can be traced back to the late Paleozoic era (ca. 350 Mya during the Carboniferous period). The suspected sister group to anaspideans (sea hare- and *Aplysia*-like ancestors) dates back to the Jurassic period (ca. 200 Mya). *Akera*, the most primitive sea hare, dates back to ca. 165 Mya and the first *Aplysia*-like records only appear in the Miocene era (ca. 25 Mya). Importantly, Mollusca belong to the Lophotrochozoa - one of three newly established superclades of bilaterian animals (Fig. 1). Although Lophotrochozoa comprise more than 15 animal phyla, none of the representatives of this largest and systematically most diverse domain of life has had its genome sequenced (Fig. 1). Therefore, sequencing the *Aplysia* genome will be a crucial landmark in the entire field of comparative genomics, a vital step in understanding the organization of genomes in all animals. In combination with the unique organization of *Aplysia*'s nervous system, access to the *Aplysia* genome will greatly inform not only neuroscience but biology in general. Genomic approaches will be combined with tools from modern neuroscience to probe the intracellular domains and signaling pathways of different identified cells with different behavioral functions. Thus, the accessibility to *Aplysia* sequences will open new opportunities to study the behaviors of highly differentiated cellular populations and elementary systems of neurons in real physiological time at a resolution unavailable with other animal models.

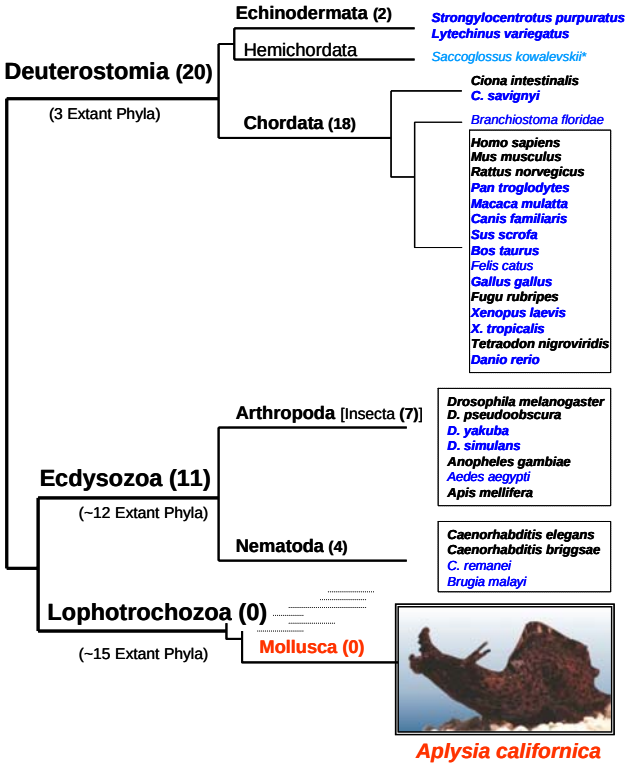

Fig. 1. *Aplysia* is a member of the largest and genomically least understood subdivision of bilaterian animals. This diagram shows the current status of genome sequencing from different Bilateria relevant to their systematic position and phylogeny. Bilaterian animals can be grouped into three major superclades where Lophotrochozoa is the largest clade composed of more than 15 animal phyla including Mollusca (*e.g.*, Brachiopoda, Annelida, Nemertina, Plathyhelmintes, Bryozoa, etc.). Despite their importance no genome has been sequenced so far from any representative of Lophotrochozoa. In contrast, there are 31 genomes (indicated as (N)) from representatives of two other superclades where sequencing is complete (black) or in progress (blue). The proposed sequencing of the *Aplysia* genome will therefore be a crucial landmark in the whole field of comparative genomics and will be an essential step toward understanding the organization of genomes in bilateral animals and molluscs in particular.

***Aplysia* among other Molluscs.** The rationale to select *Aplysia* as the next target for molluscan genome sequencing can be summarized as follows. *First, Aplysia* is probably the best-studied mollusc with about one hundred of laboratories worldwide using this organism for molecular, cellular, physiological, behavioral and developmental studies. It is an especially important model system for biomedical research in the fields of neuroscience, cell and comparative biology (see below). As a result, extensive background information about *Aplysia* has been accumulated for the last fifty years (>3,500 publications) providing all required resources, tools and experimental approaches for the needs of functional genomics and effective annotation of novel genes in this organism. There are only a few other molluscan species that might come close to *Aplysia* as model organisms (see Fig 1-2 in Appendix to compare *Aplysia* with other molluscan models). However, the genome sizes for some candidates are significantly larger (e.g. 4-5 Gb in Cephalopods), the species are not as widely available, or the existing research communities and resources are substantially smaller, limiting full-scale exploration of their genomes at this time.

Second, *Aplysia* ideally represents the class Gastropoda; the largest and systematically most diverse class of the phylum with deep evolutionary roots and an early appearance in the Cambrian period. Consequently, members of this class have a very long evolutionary history, enormous adaptive radiation, and well-tractable lineages, evolutionary and ecological trends supported by abundant paleontological data. Two classical taxonomical groups of Gastropoda (such as Prosobranchia and Pulmonata) are recognized as polyphyletic clades whereas the third subclass (Opisthobranchia, marine gastropods) appears to be a monophyletic group. Notably, many aspects of anatomy, physiology and development of *Aplysia* fully represent a canonical molluscan organization and gastropod organization in particular. In principle we agree that a potential selection of species from the more basal molluscan classes (Polyplacophora (chitons) and Monoplacophora) for genome sequencing may provide additional insights into questions related to genome evolution on a grand evolutionary scale (Fig. 2, Appendix). However, living monoplacophorans are virtually unreachable for experimentalists and were discovered as a living fossil only in 1957. Most aspects of chiton biology are too poorly understood to efficiently utilize their genome data within the next few years. In contrast, *Aplysia* offers a unique combination of characteristics making it a desirable model organism both for experimental and comparative biologists. It is a truly appealing molluscan candidate for genomic sequencing in the very near future. Furthermore, the largest somatic cells in the animal kingdom are found in the CNS of *Aplysia* (see the next section) and offer unrivaled opportunities for a cost-effective analysis of the genome's operation in real physiological time (on a scale ranging from seconds to days).

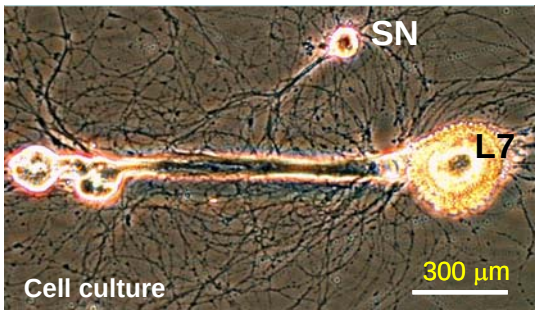

Fig. 2. Giant neurons in *Aplysia* are the largest somatic cells in the animal kingdom. They are tolerant to environmental stress and all aspects of their biology as well as network analysis can be studied for days in the CNS maintained in a simple media. The micrograph shows the simplest memory forming network reconstructed in cell culture. SN (sensory neuron) forms monosynaptic connections with L7 (motoneuron).

Fundamental Insights into the Basic Organization of Neuronal Functions. For more than 50 years, the relatively simple and thoroughly studied organism *Aplysia* has been a workhorse for a large segment of the neurobiology community, in the same way that the bacterium *E. coli* was to molecular geneticists, and *Drosophila* and *C. elegans* are to current geneticists and developmental biologists. There are several challenges that confront the neurobiological sciences in the genomic era. One of them is to try to understand how enormous neuronal diversity is generated in the nervous system (perhaps more than 2,000 different cell types in the mammalian brain) and how it is related to behavioral functions. Although the complexity of the brain appears to be overwhelming, many major fundamental questions can be addressed only at the level of single neurons. These problems are more tractable in *Aplysia* which has 20,000 central nerve cells. By contrast the mammalian brain has between 10^{11} and 10^{12} neurons and *Drosophila* has ~200,000 (quite small) neurons. Each of the 302 neurons from *C. elegans* have been identified individually, but they are too small (<5 μm) to be routinely used for global genomic profiling. By contrast, the neurons of *Aplysia* are giant (100-1000 μm), which allows study of gene expression profiles at the level of identified neurons. In addition, these polyploid neurons by themselves are of special interest for genomic science, since they represent unique multifunctional integrative centers with multiple genome copies, a feature that can be changed during development.

In the past *Aplysia* has proven to be a useful model organism because of the technical advantages it offered for the cellular study of behavior (Kandel, 1976, 1979, 2001). These advantages include the following:

(1) Behaviors are produced by a small number of cells. The nervous system of *Aplysia* contains only about 20,000 central nerve cells. These are distributed in about 10 central ganglia, each containing about 2,000 nerve cells. A given ganglion, such as the abdominal ganglion, controls several behaviors so that the number of cells committed to a single behavior may be as small as 100.

(2) Many of the nerve cells are gigantic (Fig. 2). Some of the cells, such as R2 and LP1, are 1000 μm in diameter and are highly polyploid (n>120,000). These are the largest somatic cells in the animal kingdom; only eggs are larger. Because of their size, one can extract individual cells, generate cDNA libraries from them, and inject DNA constructs and antibodies into them.

(3) Many of these large cells are uniquely identifiable in every member of the species and have well-defined homologs even in distantly related species so that comparative aspects of behavior can be studied at the cellular level.

(4) The identified nerve cells form precise connections with one another that are readily detectable electrophysiologically. These connections can be mapped on a cell-to-cell basis because one can record in the cell body large synaptic potentials produced by single presynaptic neuron.

(5) The neurons can be dissociated and placed in cell-culture (Fig. 2) where they form precise connections. As a result, components of a network can be reconstructed in cell culture and studied in great detail.

(6) In cell culture one can dissect out subcellular components of the neuron, and analyze them at the genomic level.

(7) One can map the behavior of the animal in terms of specific nerve cells and their interconnections.

(8) The animal has a surprisingly rich repertoire of learning capabilities. As a result, it is possible to work out the cellular and molecular mechanisms of different forms of learning and of both short- and long-term memory. Moreover, identical cells can be studied in every animal of the species in both naïve and trained animals, and the evolution of selected signaling mechanisms directly analyzed at the level of single neurons. Furthermore, it is possible to examine expression patterns in different parts of the same cell, for example, central (cell body) or peripheral (dendritic) regions during learning and memory.

(9) In the past *Aplysia* has been used primarily in the analysis of simple behaviors, but recently the *Aplysia* community has started to tackle more complex behaviors, such as feeding, walking, defensive inking, control of circulation, reproduction, social behaviors, etc. Furthermore, *Aplysia* has become a very useful model system to study mechanisms of development and aging, single cell proteomics and metabolomics, toxicology and drug discovery, as well as many aspects of population and marine biology, ecology and evolution. These problems are not only relevant to an understanding of specific regulatory systems but expand the applicability of this model organism for nearly all aspects of integrative physiology and system biology.

The Demand for Genomic Sequence Data, the Strength of the *Aplysia* Research Community, and its Ability to Usefully Exploit a Full Genome Sequence. About one hundred active laboratories work on *Aplysia* in the U.S. and elsewhere. These scientists provide the required research infrastructure for this organism. This is reflected in more than 3,500 publications (1965-2003) dealing with *A. californica* alone and over 16,000 for related species. Thus, among the available invertebrate models, after *C. elegans* and *Drosophila*, the molluscan research community is, along with possibly the sea urchin group, the most prominent. An *Aplysia* sequence database would benefit thousands of investigators working in the fields of neuroscience, comparative genomics, developmental and evolutionary biology, as well as in marine sciences, aquaculture and pharmaceutical sciences. NIH and NSF alone support at least 120 research grants directly or indirectly linked to research on *Aplysia* and related gastropods (calculated from NIH CRISP & NSF databases). Practically all of them require genomic information or have a research component and resources related to cloning and gene identification. The estimated total funding for molluscan based research grants at the NIH and NSF for the last fiscal year exceed $15M. This excludes parasitology research--molluscs are well-known secondary hosts for various diseases of human and animals.

Over the last forty years molluscan neurobiology has been an active frontier in neuroscience. The Society for Neuroscience's annual meeting attracts more than 30,000 attendees and always holds at least three sections with a significant representation of molluscan models. At this multidisciplinary meeting there are about 150 presentations a year (calculated from the SFN abstracts during the last three years) that reflect the current status and significance of *Aplysia* and related species in neurobiology. In addition, there are at least 2 or 3 international meetings dedicated to the neuroscience and the behavior of invertebrate models including molluscs. Finally, there are more than two dozen international malacological societies (summarized in Appendix), many with their own annual meetings and specific journals, which deal with all aspects of molluscan studies. Clearly, this situation reflects the biological and biomedical significance of the Mollusca as one of the largest and evolutionary one of the most successful animal phylum.

The molluscan research community is unified and enthusiastic about the need for the *Aplysia* genome sequence. This is reflected in the 36 letters of support obtained for this proposal from experts in this field and related areas of biological sciences (see the Appendix). As one can see from the attached letters the advent of new and comprehensive *Aplysia* sequence data should greatly stimulate and expand this research community.

The access to a full-genome *Aplysia* sequence will instantly lead to functional annotation of many hundreds of human orthologs relevant to all aspects of cell biology, neuroscience, development, morphogenesis, toxicology, drug discovery, and ecological and behavioral genomics. The role of non-coding regions and regulatory areas in the *Aplysia* genome will also be experimentally validated at nearly real time resolution.

Finally, the current demands of the *Aplysia* research community for highly representative microarrays can not be underestimated. Urgent demands exist to explore the orchestrated activity of thousands of genes at the level of single and functionally characterized cells in various physiological states. Such an approach extends the utility of

the *Aplysia* genome far beyond the specific goals of the existing *Aplysia* research community; it will provide experimentally testable paradigms to link precise molecular events of gene expression and regulation to systematic functions at all levels of multicellular organization in bilaterian animals: from cells to organism, from genomic networks to control of animal behaviors. A foundation will be established to extend genomic research on *Aplysia* to large-scale comparative surveys among various molluscan lineages and other Lophotrochozoan phyla leading to experimental analysis of multiple innovations (morphological, biochemical, developmental and behavioral) that are well documented within different gastropod families and orders.

***Aplysia* as an Emerging Model in Genomic Science.** The *Aplysia* genome is about 1.8 x 10^9 bp in size and is distributed among 17 haploid chromosomes (Lasek & Dower, 1971; Hinegardner, 1974). The *Aplysia* genome sequencing project will be carried out at the MIT genome center directed by Eric Lander, in collaboration with Jingyue Ju's laboratory at the Columbia Genome Center, and Eric Kandel's laboratory at the Center for Neurobiology and Behavior, both at Columbia University College of Physicians and Surgeons, and with Leonid Moroz's laboratory at the University of Florida. Genome annotation will be performed in conjunction with other *Aplysia* researchers throughout the U.S. and the world. In this way, we propose to further promote familiarity with the *Aplysia* genome to all members of the biological community.

As a first step in this direction, the Columbia Genome Center and the Whitney Laboratory at the University of Florida established a consortium in 2002 to initiate an *Aplysia* cDNA sequencing project with Incyte Genomics and Amersham Bioscience. We have constructed standard and normalized cDNA libraries from the CNS, individual ganglia, identified neurons and even pure neuronal processes of *Aplysia*. Out of over 200,000 ESTs, we have obtained more than 30,000 unique *Aplysia* sequences, which include many diverse splice forms and possible non-coding RNAs. Using these sequences, which represent a significant fraction of the total number of cDNAs expressed in the nervous system (Appendix Table 2), we have validated the applicability of microarrays and large scale gene expression profiling to single cell studies. For example we have estimated total RNA content in larger molluscan neurons (serotonergic MCC cells and cholinergic R2 cells) based on our ability to accomplish linear amplification of RNA from these individual cells, and in parallel, known amounts of RNA. We obtained 50-100 ng of total RNA is equivalent to about 1-2 ng of mRNA or ~ 10^9 mRNAs of average size per molluscan neuron. In contrast, a typical vertebrate neuron contains only 2-20 pg of total RNA (~0.05-0.5pg of mRNA or ~50,000-500,000 mRNAs) less than 0.1-0.01% of what found in a typical *Aplysia* neuron. The amount of material from a single *Aplysia* neuron was sufficient to carry out several individual microarray experiments.

On the basis of the above efforts, we have acquired funding for the development of innovative technologies to conduct genomic analyses on single living neurons. By coupling this technological project with the *Aplysia* genome sequence, we will obtain a comprehensive set of genetic tools, resources and information. This data set will be exported to the larger *Aplysia* community simultaneously with its acquisition, allowing investigators to utilize the information for their own studies in cellular and system biology. As a further step in that direction, the Miami Center for Marine Genomics (including the P. Walsh laboratory) joined the consortium in the spring of 2003, focusing on the development of cDNA libraries and EST sequencing from all developmental stages of *Aplysia*. The Miami Center houses the unique NIH supported *Aplysia* facility, a national and world resource serving the growing needs of the biomedical community. We expect that within the next few months we will generate an additional set of ~30,000 unique larval ESTs from normalized libraries of major developmental stages of the organism.

Major Outcome of the *Aplysia* Genome Project. The work with *Aplysia* has been essential not only for our understanding of the basic processes of neural functions, but also for highlighting cellular processes that are targets for drug discovery in virtually all areas of cell and system biology and toxicology. Genomic sequence information from *Aplysia* will be utilized by a large biomedical community with well developed tools and research programs. The *Aplysia* genome will be annotated and ready for applications in neuroscience, comparative genomics and biotechnology development.

I. Neuroscience and Cell Biology. Sequencing the *Aplysia* genome will allow the discovery of essentially every gene with its regulatory components. This will contribute greatly to the study of the genomic basis of network organization and behavior, mechanisms of neuronal specificity, cell-cell communication, and long-term plasticity. Taking advantage of the extensive background information already obtained for this organism, straightforward experiments can be designed to reveal a complete set of genes involved in learning and memory mechanisms, and synaptogenesis, as well as neuronal development, specialization, damage and degeneration. A hierarchical, step-by-step experimental approach (single synapse ↔ neuron ↔ network ↔ behavior), in combination with the genomic information, will link the integrative molecular machinery of a complex eukaryotic cell with the dynamic organization of a multicellular organism.

II. Comparative Genomics and Evolution. *Aplysia* serves as an important model molluscan species since it belongs to one of the basal branches of opisthobranch gastropods as well the Lophotrochozoa lineage of animals (see Fig. 1). For that reason, the *Aplysia* genome will serve as a valuable outgroup and a sister group to all ecdysozoans (including flies and nematodes). The *Aplysia* genomic sequence and organization will immediately benefit comparative studies since *Aplysia* is a member of the largest and least understood domain of multicellular animals and will help to answer fundamental questions about the design of animal genomes by comparative examination across great evolutionary distances. Hence, *Aplysia*'s evolutionary position makes it a prime target for genome sequencing.

III. Genomic technologies. Finally, the advantages of *Aplysia* as a model organism will push the frontiers of functional genomics and innovative technological development in genomic science. For example, using new technologies in cell and protein imaging, parallel studies can be performed in single living cells and complex interacting cell populations that constitute behaviorally meaningful networks. Also, there are ongoing efforts to study parallel gene expression using micro- and nano-chip technologies. Single cell SAGE using massively parallel sequencing approaches and imaging technologies such as molecular beacons are being developed that focus on real-time gene expression profiling from living cells and RNA trafficking in subcellular compartments. The importance of these technologies has been recently recognized by NHGRI in the newly established Center of Excellence in Genomic Science at Columbia University and the University of Florida.

In summary, *Aplysia* can bring genomic science to all levels of organization from defined subcellular domains in a single neuron to complex behaviors. The model will allow monitoring the global transcriptome at the subcellullar level, in real physiological time and with minimal damage and interference to cellular processes. Below we will briefly summarize additional specific biological rationales and strategic issues associated with the sequencing of the *Aplysia* genome. The list includes only a few of the many research opportunities that *Aplysia* offers as a model organism.

1. Improving human health: ***Age-Related Memory Loss and Mental Retardation.*** Significant breakthroughs in the understanding of cellular and molecular mechanisms of memory formation have come from studies on *Aplysia*. These processes have proven to be conserved among animals, and the significance of *Aplysia*'s contribution has become a textbook staple. As an example of the effects of such research, studies of learning in *Aplysia* have led to drugs now in clinical trial that can reverse memory loss in elderly people (http://www.memorypharma.com/news24.html). In addition, the study of learning in *Aplysia* has proven important in the study of motivation, arousal and the control of feeding and locomotory patterns.

The availability of the *Aplysia* genome sequence will allow the discovery of novel genes and their regulatory regions underlying learning and neuronal plasticity. This will lead to the establishment of new experimental models of neurological diseases and to the discovery of new therapeutic targets for human central nervous system disorders. As an example, our recent discovery and cloning of *Aplysia* orthologs of several genes underlying Alzheimer's disease (e.g. presenilin) and mental retardation (Fragile X Mental Retardation Protein or FMRP) has provided a unique opportunity to dissect the molecular pathways leading to changes in synaptic strength, neuronal plasticity, and dysfunction (*e.g.*, using RNAi) in memory-forming networks.

Taking into account the relatively short life cycle of *Aplysia californica* (240-300 days), the use of the proposed model opens unique possibilities for studying molecular and cellular mechanisms of memory formation and loss during aging at the level of single characterized neurons in real physiological time. Once identified in a simpler model system, the targeted mRNA orthologs and their encoded proteins can be analyzed in mammalian systems – an approach proven to be highly efficient in modern neuroscience. A similar case can be made for the use of *Aplysia* as a model system for neuroregeneration (*e.g.*, after stroke or spinal cord injury).

Cancer. The *Aplysia* nervous system and its component neurons grow continuously during the entire life cycle of the animal. The recent identification of tumor related genes might be associated with such permanent growth and present novel insights to study the functional role of these genes in network organization, plasticity and regeneration as well as to model related neurooncological diseases and pathological states.

Nutrition: Food, toxins and drug discovery. Gastropod, bivalve and cephalopod molluscs are important food sources in many regions of the world. However, the relevant molluscan diseases and hormonal controls on the organisms' reproduction and development are only beginning to be understood, with few identified hormones or pheromones. *Aplysia* genome sequencing will provide an essential bridge in this field and help target the relevant homologs in close and more distantly related molluscan species. In addition, many gastropods are well known as unique sources of toxins and biologically active compounds (e.g., *Conus*), and the availability of molluscan sequences can serve as an efficient resource in the search for prototypes in drug discovery. For example FMRF-amide, a widespread molluscan hormone, was found to be conserved through many phyla and FMRFa analogs

have been discovered in mammals, including humans. Most importantly, FMRFa and related peptides directly modulate nociceptive (pain-related) pathways and are involved in memory mechanisms. FMRFa is just one of more than 500 expected molluscan neuropeptides and hormones.

2. Informing human biology: *Signaling pathways.* Neurotransmitters, hormones and other signaling molecules are precisely embedded within specific neuronal pathways as well as tightly linked to the complex molecular machinery of every neuron in our brain. Moreover, there is a remarkable evolutionary conservation of transmitter mechanisms across species. Many of these signaling cascades and their genomic controls remain uncharacterized. It is not clear how these pathways are coordinated and developed. How many and what genes are regulated following specific activation of synaptic inputs or a given transmitter? What are their targets? The answers to these and other fundamental questions are crucial for the understanding of cellular and systems biology of all tissues in any organism, and for the development of new therapeutic drugs and strategies. The *Aplysia* genome project will immediately lead to direct experimental determination and analysis of global genomic responses of specific neurons in networks, including the changes that occur following transmitter-induced modulation of synaptic strength, network arousal or depression, development and regeneration. Armed with the entire genome sequence of *Aplysia*, we will then identify different isoforms of signaling proteins and determine their regulation, as well as identify downstream targets of different signaling pathways, their interactions and coordination. This approach will greatly facilitate proteomic identification of novel genes and post-translational modifications of many target proteins down to the single cell level. Furthermore, gene expression profiling in specific cells can be linked to large-scale metabolic profiling and quantification of major signaling molecules in the same cells using microanalytical approaches such as capillary electrophoresis and MALDI TOF mass spectrometry. Importantly, signaling pathways are expected to be different in non-neuronal cells, and the genome project will make it feasible to study these networks in other systems of *Aplysia*.

Polyploidy. The *nucleus* of large cells such as R2 may approach 500-1000 microns in diameter. Lasek and Dower in 1971 and Coggeshall in 1967 found that the DNA content of the nucleus of R2 increases with animal size from approximately 2,000 times the haploid amount in small animals to 75,000 times the haploid amount in large animals, indicating that the DNA has already undergone 15 duplications without intervening cell division. Moreover, the number of nuclei in the giant cell is increased proportionally with the DNA content of the nucleus. This phenomenon is not unique to *Aplysia*. Nuclear DNA in excess of diploid amounts has been found in all large neurons of the mammalian nervous system including motor neurons in the spinal cord and Purkinje cells in the cerebellum. The combined power of genomics and cell biology of *Aplysia* may help shed light on the unexplained biological significance and role of the widespread distribution of polyploid neurons in mammals.

3 & 4. Informing the human sequence and providing a better connection between the sequences of non-human organisms and humans. The identification of *Aplysia* orthologs of human genes (especially for hypothetical human proteins) will lead to direct experimental analysis of their functions in the nervous system. Due to the experimental advantages of its large cells, gene discovery and annotation in *Aplysia* is expected to be much easier than in other organisms. The unique opportunity to inject specific transcripts or probes into virtually any cell and monitor the physiological consequences in real time allows one to examine the role of non-coding RNAs and gene regulatory networks in neural functions. How many genes are expressed in a single cell? How many transcripts are transported to a specific synapse? How many genes are sufficient for certain integrative functions of neurons? These questions represent major challenges for our understanding of brain mechanisms since every cell in the nervous system may be a phenotypically unique cell type, expressing its own repertoire of the genome. The genomic basis of such diversity can in part be linked to the diversity of splice forms and regulatory mechanisms that are likely to be overrepresented in neuronal tissues. *Aplysia* will reveal the biological roles of this transcriptional diversity in heterogeneous cell populations and networks.

The significance of *Aplysia* for comparative genomics is outlined above and the availability of its sequences will clearly help in the understanding of genome organization in other major animal clades. Our EST collection revealed a number of unusual orthologs that are found in *Aplysia* and vertebrates, but have not been found in other invertebrates, suggesting their loss in evolution (*e.g.*, purinergic P2X receptors - Table 2, Appendix). On the other hand, *Aplysia* and human genes for hypothetical proteins were found to be highly expressed in specific neurons and neurites, implying neuronal functions. Some unusual but evolutionarily conservative substitutions within well-known gene sequences (*e.g.*, for NOS, NMDA-like receptors and FMRP) probably indicate novel regulatory mechanisms, specifically linked to their diverse expression patterns in the CNS.

5. Expanding the understanding of basic biological processes relevant to human health; Development and neurogenesis: Spiral cleavage (one of two basic cleavage patterns in animals) and schizocoely, both characteristic of *Aplysia*, are widely distributed across more than a dozen animal phyla and are critical for

studying the organization of animal body plans. Our EST screening has allowed us to identify many crucial developmental genes and potential pathways in *Aplysia* (*e.g.*, members of the HOX family, hedgehog, Wnt signaling, *etc.*), but many low abundant transcripts are still missing. The diverse life history strategies adopted by many species, combined with extreme body plan modifications, make molluscs an ideal group to investigate the evolutionary plasticity of the underlying mechanisms of development. For example, one of the most remarkable and dramatic steps taken during the course of molluscan evolution was the advent of gastropod torsion, a unique reorganization of the organ positions associated with the loss of some organs and topographical reorganization and varying degrees of centralization of the nervous system. Molluscan lineages, and opisthobranchs in particular, exhibit a high level of homoplasy (convergence, parallel evolution and reversals) in many morphological traits. Some of this homoplasy is most likely due to the differences in regulation of developmental networks. Genomic information from *Aplysia* will help explain how gene expression controls major events in spiralian development, torsion-detorsion processes, neurogenesis, network organization and maturation.

Aplysia metamorphosis. Although most of current knowledge about the mechanisms of this life history transition comes from vertebrates (*e.g*., amphibians) and ecdysozoans (*e.g*., *Drosophila*), new insights from the lophotrochozoan *Aplysia* will be highly informative not only due to its phylogenetic position but also because of its settlement in a marine rather than terrestrial or freshwater environment.

Adult neurogenesis. Another interesting feature of development in *Aplysia* is the increase in the number of cells in the course of post-embryonic growth. The best studied case is the increase in the number of neurons in the "bag" cell clusters in the abdominal ganglion. These cells are absent in juveniles prior to stage 13 but slowly increase to about 800 cells in mature animals. Since the bag cells' released hormones trigger egg laying, an increase in their number may be necessary to supply adequate hormones for the animal's reproductive phase. But the increases are not restricted to the bag cells. Other small cells also increase with the result that the total number of cells in the abdominal ganglion increases from about 1,000 in juveniles to about 2,500 to 3,000 in mature animals. These cells seem to arise from deep regions of the neuropil and from the ectoderm of the body wall providing potential insights into neurogenesis within the adult brain. Identification of those genes and their regulatory elements with altered expression patterns in the precursors to these cells, require identification of the complete *Aplysia* genetic repertoire.

6. Facilitating the ability to conduct experiments using *Aplysia* as a model: In principle, any gene can be introduced into any cell of *Aplysia* via transfection or microinjection. Many genetic studies can then be performed. One such application, gene inactivation, can be carried out by the injection of antisense RNAs into the cell or by RNA interference (Lee *et al*. 2001). In addition, GFP constructs have been designed, expressed and tracked in the intact CNS and *in vitro* for *Aplysia* (Kim et al., 1998, 2003). The abilities and limitations of *Aplysia* for genomic manipulations are summarized in Table 1 (Appendix).

7. Expanding the understanding of evolutionary processes. The *Aplysia* genome sequences will help to evaluate the newly proposed hypothesis of bilaterian phylogenetic relationships encompassing three major superclades (see Fig. 1). Molluscan sister groups will then be determined. Subsequent sequencing of key genes will expand the understanding of the major genomic events that took place in the lophotrochozoan lineage prior to the evolution of the great diversity of *Bauplans* (body plans) within this major clade of the animal kingdom. Among other molluscs, substantial sequence information (~5,000 ESTs, mainly blood cell-derived) is only available from the blood fluke *Biomphalaria glabrata*, a highly derived freshwater pulmonate species. The limpet *Lottia gigantea*, a specialized prosobranch species recently selected by DOE for genome sequencing due to its relatively small genome size, can be considered a favorable complementary organism for comparative genomics of molluscan-specific genes. However, because of the current lack of a *Lottia* research community and a lack of sufficient background information about this species, the experimental approaches required for functional annotation of limpet-specific genes are currently nonexistent and unlikely to be fully developed. For that reason, the availability of the genome of an experimentally favorable model organism such as *Aplysia* is required for annotation of molluscan genes and identification of the genomic basis for major transitions in the molluscan lineage.

Evolutionary and Ecological Functional Genomics. The unique features of *Aplysia* and related gastropods will make it possible to study the genomic basis of evolution of the nervous system at the level of individual homologous neurons. Many principal neurons identified in *Aplysia* have homologs even in distantly related species such as land and freshwater pulmonate molluscs. In certain cases these cells preserve their unique identity and functions, while in others they have different phenotypes and properties. Homologous networks and behaviors can also be modified following changes in animal ecology. Importantly, of all animal groups none have more diversity and variability of organization than molluscs. They live in all major ecological habitats from the deep ocean floor to land and freshwater environments. Ecological adaptations in feeding, reproductive or defensive

strategies can be observed both within the same species/genus and across different orders and subclasses. Yet, in many cases, their evolutionary trends can be tracked down to specific homologous networks and behaviors. Here, population and comparative genomics can merge with neuroscience and neuroethology in a field of integrative biology where origin and evolution of neuronal functions will be linked to the origin of species and major changes in their body plans. As an illustration of the initial efforts in this direction, the Moroz laboratory has generated several thousand ESTs from several important models in comparative neurobiology, including *Lymnaea*, *Tritonia*, *Pleurobranchaea* and *Clione*. Most of the ESTs were collected from identified neurons in the feeding network that have homologs in *Aplysia*. We have noted that changes in feeding strategies (*e.g*., from grazers to opportunistic predators) are associated with specific changes in neuronal signaling (*e.g*.,nitric oxide signaling). This will provide an ideal approach to study the evolution of genomic regulatory pathways in characterized neuronal networks.

Origin of higher neuronal functions. The identification of memory-relevant genes in *Aplysia* will provide unique prospects to assess the function of their homologs both in less and more centralized brains of related gastropods, as well as in the cephalopod brain; these have evolved following a well-documented process of detorsion and ganglionic fusion within molluscan groups. As a result, the link from simple to higher neuronal functions will be established with the molecular dissection of associated mechanisms. Therefore, the evolution of elementary cognitive functions can be traced across all major grades of nervous system organization starting from the mainly decentralized diffuse-like nervous system found in chitons (Polyplacophora) to the peak of neuronal organization (consisting of up to one billion neurons) reached by the octopi, cuttlefish and squid, which are able to perform highly complex and intelligent behaviors, comparable to those of some birds and mammals. Yet, they preserve basically the same neuronal organizational plan as *Aplysia*, with a fusion of all major ganglia in the central brain and mainly unipolar neurons sending their processes to the ganglionic neuropil.

Strategic Issues in Acquiring New Sequence Data

The suitability of *Aplysia* for experimentation and *Aplysia* research community: These considerations are addressed above and summarized in Table 1 of the Appendix.

The rationale for the complete sequence of the *Aplysia* genome: The complete sequence of the *Aplysia* genome would provide information about regulatory sequences that are often located in non-coding regions that can be great distances from the gene that they are acting on. Some biologically important genes are rarely expressed or are expressed at very low levels and thus often absent from EST libraries. In addition, a major value of this organism is its widespread use in neurobiology for studies involving behavior, learning, memory and arousal. Behaviors are not due to the action of a single cell or gene; rather they are generated by larger neural networks sometimes involving heterogeneous cell populations. Indeed, important changes occur subcellularly in different portions of the same neurons. To prepare and sequence cDNA libraries from every neuron or even every ganglion would be quite difficult. Single cell amplification from large cells with high concentrations of mRNA molecules is possible, but the technique is subject to high rates of bias and error introduction when carried out with smaller neurons with smaller mRNA stores. In contrast, genomic sequences can be obtained from any diploid tissue source in the organism. Once the genome sequence is available, it will be possible to conduct global expression assays or SAGE (Serial Analysis of Gene Expression) and determine subcellular localizations of mRNAs or proteins in any cell of interest.

The utility of the whole genome sequence for answering the most important questions for the community was briefly outlined in the previous sections. The coordinating committee that reviewed our first application believed that the case had not been made completely as to *why a whole genome sequence was needed for the uses discussed in the White Paper, as opposed to ESTs or cDNAs*. Now with more than 200,000 *Aplysia* ESTs, ~ 80 cDNA libraries and hundreds of full-length cDNAs in our hands we can define several major shortcomings of EST or even full-length cDNA sequences that can be immediately overcome by the availability of the genomic sequence.

***First* and most important**, gene transcripts, while indispensable for determining the expression patterns in cells and tissues, are by definition devoid of promoter and enhancer regions that regulate their activity. Taking into account the major advantage of *Aplysia* as a model organism for single-cell real time genomics, the lack of information about the regulatory regions is one of the major gaps in the field. Identification of such genetic elements would have to be accomplished gene by gene in the absence of an available underlying genomic sequence; with that sequence, such information would be much easier to achieve and could be done globally, for all genes at once. This is particularly true for identifying promoters, making rather straightforward assumptions about their typical distance from the transcriptional start site, and taking advantage of the ever-growing information contained in databases on their common motif structures. Another major area of interest is the identification of small interfering RNAs and other small non-protein-coding genes that are almost never found in

standard cDNA libraries; these molecules are increasingly being recognized as playing important roles in regulation during development, pathological and physiological processes. Since they are often complementary to segments of the genes they control, have palindromic or other known structural properties, it should be possible to identify candidates within genomic sequences.

The second problem with cDNA libraries is that, even after multiple rounds of normalization, it is difficult and sometimes impossible to obtain sequences for the rarer transcripts, many of which will turn out to play crucial regulatory roles. One would expect most of these to be expressed at functionally meaningful levels only at certain times or to be found mainly in specific cells of a tissue or in discrete cellular locations. Indeed, in our collection of more than 150,000 ESTs from normalized CNS libraries, we have been unable to find many of the expected potassium channels, amiloride sensitive Na channels, transcription factors and other regulatory low abundant genes (see examples in Table 2, Appendix). A genomic sequence, in contrast, is by definition fully normalized, making all the genes accessible with an essentially stochastic search, such as is accomplished by 8x coverage sequencing of random libraries.

The third disadvantage of working with cDNA libraries alone is methodological. For the most part, high representation cDNA libraries are made from the 3' ends of genes, taking advantage of their poly(A) tails, but resulting in clones that have a high proportion of non-coding sequence. While these regions might be taken advantage of in some cases for designing oligonucleotides for microarrays or tags for SAGE studies, often they contain genome-wide repeats, and in any case, they give no information about the encoded protein. If the 3'-UTR is nearly as long as or longer than the cDNA clone (a feature of many cloned *Aplysia* genes), the context is lost and it becomes essentially useless. This is probably the case in the existing *Aplysia* EST collections, where less than 10% of clones can be annotated at this stage. However, if a genomic sequence is available that overlaps with this sequence, and extends to the coding portion of the gene, the context will be restored, and the non-coding sequence might still be a useful tag.

Fourth, genomic sequences also immediately provide gene anatomy (exon-intron boundaries) and in combination with cDNA sequences, identify alternative splicing and other gene features. An enormous diversity of splice forms is a vital component of regulatory cascades discovered in nervous tissues and developmental stages of other organisms. It also implies a similar organization in *Aplysia*, but ESTs with relatively short error-prone reads or even extensive full-length cDNA sequencing, will not be an efficient way to identify and characterize the diversity of splice forms in specific cells and modulation of their expression at different functional stages.

A fifth downside to cDNA libraries is that in order to identify the majority of the organism's genes, one must prepare libraries not only from different tissues or cells (for the CNS this means cDNA libraries from every one of several thousand neuronal types and theoretically from every neuron or synapse), but also at different developmental stages, or following a variety of stresses or other stimuli. Then, each of these libraries has to be sequenced to a high degree of coverage (typically more than 8x, not a cost efficient approach by itself) with a low likelihood of identifying low abundance transcripts. As indicated earlier our screen of more than 80 *Aplysia* libraries resulted in the identification of 17 potassium channels (compared with ~50-90 K-channels identified in sequenced genomes of other invertebrates) and only one amiloride sensitive channel (compare with 10-30 orthologs in other sequenced genomes). These genes are highly cell specific and absolutely crucial for many cellular and systemic functions in any organism. Clearly, it would be more efficient to determine the complete collection of genes using a single shotgun strategy, and then take advantage of much more cost-effective and resourceful approaches such as global expression microarrays to reveal clusters of co-regulated genes.

Finally, even if it were possible to obtain a very complete set of ESTs (estimated at several million), as mentioned above, they are biased toward their 3' ends. At present, only ~17% of the sequences in *Aplysia* have an ORF > 150 amino acids long. One can envision various ways to obtain the remainder of the genes, ranging from inefficient (rounds of 5'-RACE on each clone) to difficult (using random oligonucleotides, adapter- and 5' cap-based strategies to obtain full-length cDNA libraries, followed by separation of the clones into size categories for transposon-based sequencing). Even if one were to carry out these studies just for the majority of CNS-expressed genes, obtaining full-length clones required for the current needs of the existing *Aplysia* neurobiological community (gene expression studies, proteomic analysis and construction of molecular probes) is unrealistic if the currently standard approaches and tools are used.

In summary, the importance of obtaining a whole genome sequence for *Aplysia californica* cannot be overstated. Until now, the hundreds of workers in the field have been hampered by not knowing the catalog of genes that are present in the organism and expressed in different tissues and cells. Despite the unique possibility of isolating mRNA's and preparing cDNA libraries from single well-defined *Aplysia* neurons, without knowing the organism's gene repertoire, this information has limited usefulness. It is possible to generate very large

numbers of EST sequences (in fact, the Columbia and University of Florida groups have done this), but we have strong evidence that this type of collection of sequences, or even an equivalent set of full-length sequences, will be insufficient for understanding cell and systems biology in *Aplysia* without the accompanying genome sequence as a blueprint. For the several dozen laboratories working on behavior and neurobiology, a major goal would be to generate a standardized and highly representative microarray containing the full set of *Aplysia*'s genes; this can only be accomplished by genomic sequencing.

Sequencing strategy and cost: The *Aplysia* genome size is estimated to be ~1.8 Gb. Experience and insights gained from sequencing a number of large genomes, including mouse, human, chimp, *Tetraodon* and *Ciona*, using both clone based and whole genome methods lead us to propose a whole genome shotgun strategy for the *Aplysia* genome. An assembly comprised of ~8-fold whole genome shotgun coverage will provide the community with rapid access to most of the *Aplysia* genome sequence and support gene discovery and comparative analyses. Inclusion of coverage as read-pairs from large insert clones (Fosmids and/or BACs) (see Table 1) will provide long-range continuity to the assembly to facilitate comparative studies with other genome sequences. The plasmid and Fosmid libraries will be generated using randomly sheared *Aplysia* DNA. By performing the sequencing at a large, efficient sequence production facility such as the WI-CGR, we expect to reap the benefits of the continued decrease in sequencing costs. At current costs, the estimated budget to generate this 8X coverage of the *Aplysia* genome is on the order of $20M.

Table 1. Proposed whole genome shotgun sequencing of the *Aplysia* genome.

Clone type	Insert size	No. of reads	Seq coverage	Phys coverage
plasmid	4 kb	20M	5.8x	21x
plasmid	10 kb	6.1M	1.78x	26x
Fosmid	40 kb	1.0M	0.29x	10.5x
BAC	130 kb	0.3M	0.09x	10.2x
TOTALS	---	27.4M	7.96x	48x

These calculations assume a genome size of 1.9Gb, an average sequence read length of 630 bp (phred 20) and an average pass rate of 88%. The physical coverage calculation assumes that all clones contribute read pairs.

All libraries (including BAC) will come from the same sperm sample and the same individual of an inbred line (F5 or F7 generation) to minimize potential effect of polymorphism (below). At 8x coverage (~7x phred 20 base coverage in the actual assembly), well over 90% of *Aplysia* genes will be partially sequenced.

Sequence readiness of *Aplysia californica* genome. There are two features of any genome that are helpful to know before undertaking a sequencing project of such scale: (1) What percent of the genome is composed of repetitive sequence elements, and how many different repeats exist? (2) Are the populations used in experimental studies homogeneous with regard to their genome sequences? In other words, how many polymorphic or heterozygous positions would one expect to find? In a carefully done complex study, it was demonstrated that repetitive DNA appears to be relatively densely interspersed with single copy sequences in *A. californica* (about 300 bases of repeat every thousand bases), and that this organization is quantitatively similar to that of the sea urchin *S. purpuratus* (Angerer et al., 1975). Genome sequencing for this species is already underway.

The second issue is especially important in marine organisms like *Aplysia*; it is critical both for obtaining efficient assembly and to assure that the genome sequence obtained is useful for investigators independent of which they collect their specimens. Fortunately, most of the worldwide supply of *Aplysia* currently comes from a relatively homogenous population of animals bred for more than 8 years in controlled aquaculture at the Miami NIH *Aplysia* Center. To avoid potential difficulties in genome assembly, we will utilize DNA from sperm of a single inbred *Aplysia* (F5 or F7) from these facilities to prepare the complete shotgun libraries in advance. Excess sperm will be stored if, for some reason, the initial supply of DNA is insufficient.

At the NCRR National Resource for *Aplysia*, animals are bred in captivity with great ease. The normal procedure for producing animals for the community at large is to obtain brood stock from wild populations and breed them. After a mating life, a brood stock is partially replaced. Nonetheless, we started to custom breed *Aplysia* and it was experimentally established that one can obtain several (up to seven) generations before inbreeding depression starts, and this small number of generations will decrease the overall variation among individuals. Thus, we are confident in our ability to obtain inbred animals with decreased polymorphism levels, thereby abetting sequence assembly. At present, F1-F3 progeny of genotyped parents are available and we continue inbreeding specifically for this project; by controlling diet and water temperature to achieve a lower age of sexual maturity, we expect to have F5 animals at the time funding becomes available (Jan. 2005) or F7 animals later in spring 2005.

We are using novel (single-copy DNA loci including a microsatelite and an exon coding sequence of the gene FMRFamide) and the established polymorphic microsatellite loci (Medina & Walsh 2000) to monitor the progress of the current inbreeding program. Genomic DNA was isolated from inbred animals and wild animals (6-12 animals per generation). A minimum number of PCR cycles was used to obtain and clone the markers mentioned above. Three individual clones were sequenced from each animal and all the sequences were aligned and analyzed for the presence of polymorphic sites. Our initial studies on F2 inbreed animals showed a 78% decrease in single nucleotide polymorphism (SNP) for the coding region per 500 bp and 15% decrease in SNP for the microsatelite region per 500 bp. Deletion/insertion polymorphisms (DIPs) were only observed in the microsatelite region and the F2 inbred animals showed a 45% decrease per 500 bp. The analysis of F3 animals and other loci in the F1-F3 progeny is in progress. We assume that it is possible to obtain the seventh generation (F7) in the sib-mating *Aplysia* program. Methods based on Mendelian genetics indicate that 73.4% of the genome will be homozygous by descent in F7 of a sib-mating program making sequence assembly most feasible. Thus, we will select a specimen and collect sperm from a line of F5-F7 inbred animals for use in this the project.

In addition, we would like to note that the level of natural polymorphism has been estimated in five natural populations of *Aplysia* (Medina & Walsh, 2000) on the Pacific Cost of the Southern California and Mexico. This study suggests that there was no geographic genetic population subdivision in 177 individuals examined. The genetic homogeneity found in *A. californica* is concordant with its relatively long pelagic larval stages (~35 days), their ability for oceanic long-distant dispersal and high level of gene flow within the current relatively restricted *Aplysia* areole. Thus, field-collected specimens seem to be part of a panmictic metapopulation and will likely be genetically similar. To determine polymorphism rates on a larger scale, we took advantage of our *Aplysia* EST collection obtained using partially sequenced cDNA libraries from several animals, both naturally collected and raised in culture. Our analysis of all the different ESTs in the library, admittedly a biased coding sequence-enriched set, indicates that polymorphism density in *Aplysia* is slightly higher than that obtained for coding regions of the human genome (about one SNP per 300 bases). As a final point, since one of the main purposes of the genome sequence is to identify genes of interest, individual investigators will be responsible for confirming sequences of cloned genes, and the significance of any variable bases, in their specific studies.

Methods: We propose that the majority of the sequencing, along with high throughput automated annotation be done at WI-CGR. In addition, they or a commercial entity (*e.g.*, Agencourt) could be contracted to store and dispense clones to the community at large, using their large-scale robotic freezer retrieval systems. The Center of Excellence in Genome Science at Columbia University and University of Florida, along with the larger *Aplysia* community, will be assigned the task of higher level annotation through the "jamboree" approach. Finally, the Genome Centers at Columbia, the Univeristy of Florida and the University of Miami will be responsible for further cDNA sequencing using the previously established successful approaches to generate neuronal ESTs. The methods for DNA purification, PCR, sequencing and assembly by the WI-CGR at MIT are well known and robust. The same methods will be used for this project as are used for other high throughput eukaryotic genome sequencing projects (e.g. see details of the program ARACHNE developed for the assembly of large (mammalian-size) genomes: Batzoglou *et al*, 2002; Jaffe *et al*., 2003).

Are there other (partial) sources of funding available or being sought for this sequencing project? *None.*

Other ongoing or pending projects: *BACs.* There are no sequences of large clones such as BACs for *Aplysia* but we have obtained NHGRI funding to construct the BAC libraries with an NIH request to preserve the same source of material for future genome sequencing. We will use these funds to construct BAC libraries from sperm of F5 (or F7) when it will be available in January – March 2005.

Using the Paracel GeneMatcher computer, along with a suite of gene identification and phylogenetic software, we will assemble and continue annotation of the existing EST collection and the nearly complete gene structure will be determined for some *Aplysia* genes. Further information on the most important genes will derive from our Center of Excellence in Genome Science in which nanotechnology-based methods are being developed using *Aplysia* for rapid sequencing, unicellular and subcellular expression (nanoarray) studies, and intracellular tracking of RNAs within neurons using powerful photochemical systems. Comparative clustering of the results of these studies will highlight the sets of genes involved in the most important signaling and behavioral pathways. Along with the draft genome sequence, this will allow us to further annotate their genomic context.

Finally, we want to emphasize that NIH already funds the *Aplysia* hatchery and resource facilities. If the *Aplysia* genome is sequenced, the demand for organisms will increase as more laboratories switch to address scientific questions with this species, as has been the case for other genome projects. In this case the supply is already in place since RSMAS could easily increment their yields without much delay or complication.

Appendix 1:

Letters of Support: Roster (36 letters are currently available)

The following individuals have written in support of the Aplysia genome project, as prominent scientists with a general perspective, representatives of specific communities, or both. We have included the letters in Appendix 2 (attached separately).

H. Robert Horvitz, Prof. of Biology and Investigator, Howard Hughes Medical Institute, Dept of Biol. & the McGovern Instit. for Brain Research at MIT – leading researcher in genomics and neurobiology, and an expert in the *C.elegans* field; G. Rubin, Vice-President, HHMI and Director of HHMI's Janelia Farm Research Campus, Prof. of Genetics and Development at UC Berkeley – a leading genome research scientist and an expert in *Drosophila* genomics; G. Somero, Prof. of Marine Sciences, Director, Hopkins Marine Station, Stanford Univ. – a leading scientist in the field of comparative physiology, biochemistry, ecological physiology and genomic approaches to non-traditional model organisms; G. Robinson, Prof. of Integrative Biology & Entomol, Univ. Illinois – a leading researcher in the honey bee genome project, expert in the sociobiology, genetics and neurobiology of insects; I. Meinertzhagen, Prof. of Neurobiology, Neuroscience Instit., Dalhousie Univ. – a leading researcher in the neurobiology of ascidians and insects, one of the key researchers in the *Ciona* genome project; D. Willows, Prof. of Biology, Director of Friday Harbor Marine Laboratories, Univ. Washington – a leading scientist in comparative neurobiology of various marine models including *Tritonia* (a species related to *Aplysia*); A. Selverston, Prof. of Neuroscience, UC San Diego – a leading researcher in neuronal networks analysis and comparative neurobiology of arthropod and molluscan models; L. Kaczmarek, Prof. of Pharmacology and Cellular & Molecular Physiology, Yale Univ. – a leading researcher in ion channels and related signaling in mammalian models and *Aplysia*; J. Byrne, Prof./ Chairman Dept. of Neurobiol. & Anatomy, and Center for Neurobiology of Learning and Memory, Univ. Texas Houston – a leading *Aplysia* neuroscientist, an expert in molecular and system mechanisms of learning and memory, neuronal and genomic networks, computational biology; G. Orlovski, Prof. Neurophysiology, Karolinska Institute, *Nobel Institute for Neurophysiology*, Sweden – a leading expert in system organization of brain and neuronal pattern generators in mammals and molluscan models; K. Weiss, Prof./Chair Dept. of Physiol. & Biophysics, Mt. Sinai Medical Center – a leading *Aplysia* neurobiologist, an expert in network and behavior, molecular biology of signaling systems and motivational mechanisms; T. Carew, Prof./Chair Dept. of Neurobiology & Behavior, UC Irvine – a leading expert in learning and memory mechanisms, neurodevelopment and behavioral neuroscience, as well as the neurobiology of *Aplysia*; R. Gillette, Prof. Integrative & Molec. Physiology, Univ. Illinois – a leading researcher in comparative neurobiology of molluscan models including *Pleurobranchaea* (a species related to *Aplysia*); G. Smit, Prof./Chair. Center for Neurogenomics and Cognitive Research, Univ. Amsterdam – a leading researcher in neurogenomics and proteomics of mammalian models and *Lymnaea* (a species related to *Aplysia*); E. Walters, Prof. of Physiology, Univ. Texas – a leading *Aplysia* neuroscientist, an expert in mechanisms of memory and neuronal injury; P. Forscher, Prof. of Molecular, Cellular and Developmental Biology, Dept. Molec., Cellular and Developmental Biology, Yale Univ. – a leading researcher in cell biology of neurons and the cytoskeleton; J. Jacklet, Prof. of Neurocience,State Univ. New York – a leading *Aplysia* neurobiologist, an expert in circadian clocks; M. Greenberg, Woods Hole Marine Institute and Univ. of Florida – an expert in invertebrate biology, pharmacology and bioactive peptides, a leading researcher in molluscan biology; J. Sweedler, Prof. of Chemisty, Univ. Illinois – a leading researcher in single-neuron microanalysis, neuroproteomics and biotechnology development; R. Croll, Prof. of Biology, Dalhousie Univ. – a leading molluscan neuroscientist and expert in developmental biology; S. Benner, Prof. of Chemistry, Univ. Florida – a leading researcher in biotechnology and genomics technology development, bioinformatics and evolutionary biology; A. Eskin, Prof. of Biology, Univ, Houston – a leading *Aplysia* biologist and neuroscientist; P. Katz., Prof of Biology, Georgia State Univ. – a leading comparative neurobiologist, an expert in molluscan and arthropod models; A. Susswein, Prof. of Biolgy, Israel – a leading *Aplysia* biologist and behavioral scientist; R. Chase, Prof. of Biology, McGill Univ. – a leading molluscan neurobiologist, M. Miller, Prof./Director Institute of Neurophysiology Univ. Puerto Rico – a leading comparative neurobiologist of *Aplysia* and related models; K.Martin, Prof. of Psychiatry and Biol. Chem., UCLA, CA - a leading *Aplysia* neurogeneticist, and expert in synaptic plasticity in molluscan and mammalian models; K. Lukowiak, Prof. of Physiology, Univ.Calgary – a leading researcher in neurobiology of *Aplysia*, *Lymnaea* and related species, an expert in learning and neuronal plasticity; G. Nagle, Prof of Neuroscience, Univ. Texas, Medical Branch – *Aplysia* molecular and system neurobiologist, an expert in molluscan development, hormonal and pheromone systems; W. Sossin, Prof. of Neurology, McGill Univ. – *Aplysia* molecular biologist and neuroscientist; P. Benjamin, Prof of Neuroscience, Director Sussex Center for Neuroscience, UK - a leading researcher in network organization, an expert in molluscan models related to *Aplysia*; D. Glanzman, LA, Prof. of Physiological Science and Neurobiology, UCLA – a leading *Aplysia* neurophysiologist, an expert in learning mechanisms; T. Abrams, Prof. of Pharmacol., Univ. Maryland – a researcher of molecular mechanisms of synaptic plasticity in *Aplysia* N. Syed, Prof. of Physiology, Univ. Calgary – a leading molluscan neuroscientist, an expert in reconstruction of neuronal networks *in vitro* and molecular mechanisms of synaptogenesis; G. Cottrell, Prof. of Physiology, Univ. St. Andrews – a leading molluscan neurobiologist and comparative pharmacologist, an expert in pharmacology, biophysics and molecular biology of ionic channels and receptors of many vertebrate and molluscan models including *Aplysia*, Neil Magoski, Asst Prof., Queens Univ. - *Aplysia* neurobiologist, specialist in ionic channels and their regulation; M. Hadfield, Prof. of Zoology, Director of Kewalo Marine Laboratory, Univ. Hawaii, HI – a leading expert in developmental biology of marine invertebrates and opisthobranch molluscs, mechanisms of metamorphosis and comparative biology of sensory systems

We have also discussed the project and been in contact with more than 50 scientists from the *Aplysia* research community, as well as leading researchers in fields of cellular and system neuroscience, comparative physiology, toxicology, pharmacology, developmental and evolutionary biology, biotechnologies and genomics, who expressed both enthusiasm and support for the *Aplysia* genome sequencing efforts. Additional letters are available upon request.

Suitability of *Aplysia californica* for experimentation

Taxonomy	Superclade: Lophotrochozoa* Phylum: Mollusca Class: Gastropoda (Opisthobranchia, Anaspidea) *No representative genome of Lophotrophozoa have been sequenced

Table 1. **Summary of features**

Adult animal size (length):	up to 50-60 cm [permanent growth during the life cycle]
Animal weight:	up to 3.5 kg; fertilized egg is ~ 100 µm
Life cycle:	240-300 days, embryonic period 8-11 days at 25^{o} C.
Reproductive age:	2-3 months (diet dependent)
Number of eggs:	up to 100 per egg capsule and 1,500,000 -3,000,000 per egg mass
Cellular organization:	Complex multicellular
Cultivation:	Yes, animals can be bred in captivity; inbred lines can be generated
Central Nervous System:	10 paired ganglia (some fused) with ~15,000-20,000 neurons
Neurons:	Many giant (up to 1 mm), polyploid, surface location, easily identified;
Networks:	Hundreds of defined synaptic connections in all major networks; about 40-60% of central neurons are linked to animal's behaviors including learning and memory
Behavior:	Well-characterized
Development:	Well-described
Genome size:	1.8 Gb (est. 30% repetitive)
Haploid chromosome #:	17
Genetic resources/tools	
cDNA/EST resources:	>200,000 ESTs (~50,000 unique sequences) from the CNS, characterized individual neurons, neuronal processes, other specific tissues/life stages
Gene expression:	Microarrays, in situ hybridization, GFP-constructs
BAC libraries:	Will be constructed as a part of this project (supported by NIH)
Gene inactivation:	RNAi, antisense RNAs or oligonucleotides
Special Strengths:	**-** Single neuron genomics with subcellular resolution & real-time measurements from identified cells and synapses **-** General biological, evolutionary and developmental importance; **-** Unique advantages for cellular and system neuroscience **-** Comprehensive knowledge of *Aplysia* biology and behavior **-** Detailed mapping in the brain with single-neuron resolution **-** Cell and tissue culture, including identified neurons and reconstruction of operational networks in cell culture **-** Behavioral, aging, regeneration, developmental, memory and disease model
Databases:	**-** EST databases, nervous system databases
Resources:	**-** NIH supported National Resource for *Aplysia*
Number of Labs:	- ca. 100 (*Aplysia*); ca. >200 related molluscan species
Number of investigators:	> 1500
Weaknesses:	Transient transgenesis only; weak population and classic genetics

Table. 2. **An illustrative set of *Aplysia*'s transcripts identified in our EST collection**

Gene Orthologs	*Aplysia* ESTs	*Drosophila* genome	*C. elegans* genome	Human genome
Protein Kinases	**~260**	**240**	**454**	**500**
Protein Phosphatases	**~70**	**66**	**106**	**100**
Ionotropic Glu Receptors	**14**	**11**	**10**	**15**
Ionotropic ACh receptors	**15**	**12**	**56**	**17**
Ca-channels	**10**	**9**	**12**	**38**
Na-channels	**3**	**4**	**4**	**11**
K- channels**	**17**	**42**	**91***	**~50***
Amiloride-sensitive Na**	**1**	**24**	**27**	**11**
CNG-channels**	**4**	**9**	**9**	**22**
Cadherins/Protocadherins	**16**	**17**	**16**	**113**
Synaptotagmin	**8-10**	**3**	**3**	**10**
Semaphorin	**4**	**6**	**2**	**2**
Fragile X MRP	**1**	**1**	**1**	**2**
Plexin	**1**	**2**	**0**	**9**
P2X purinoreceptors[#]	**2**	**0**	**0**	**10**

These estimates are preliminary since for many transcripts the functional annotation has not been performed. We specifically identified glutamate (Glu) receptors using different molecular approaches; their occurrence based solely on the existing EST collection is 8.

** - Apparently underrepresented (low abundance) transcripts in the *Aplysia* EST collection. These and many other families of rare or cell specific transcripts (e.g. transcription factors, gap junction proteins or olfactory receptors) are represented by only 1-10% of their expected numbers in the *Aplysia* genome (based on comparison with other invertebrate models with sequenced genomes).

[#] - These ATP-type receptors were identified in *Aplysia* and vertebrate genomes only; they were not found in any other invertebrate species (including *C.elegans* and *Drosophila*). In addition, in the *Aplysia* EST collection, we have identified more than 220 human orthologs that were also not found in the sequenced genomes of *C. elegans* and *D. melanogaster* and likely were lost in these lineages but present in a common bilaterian ancestor.

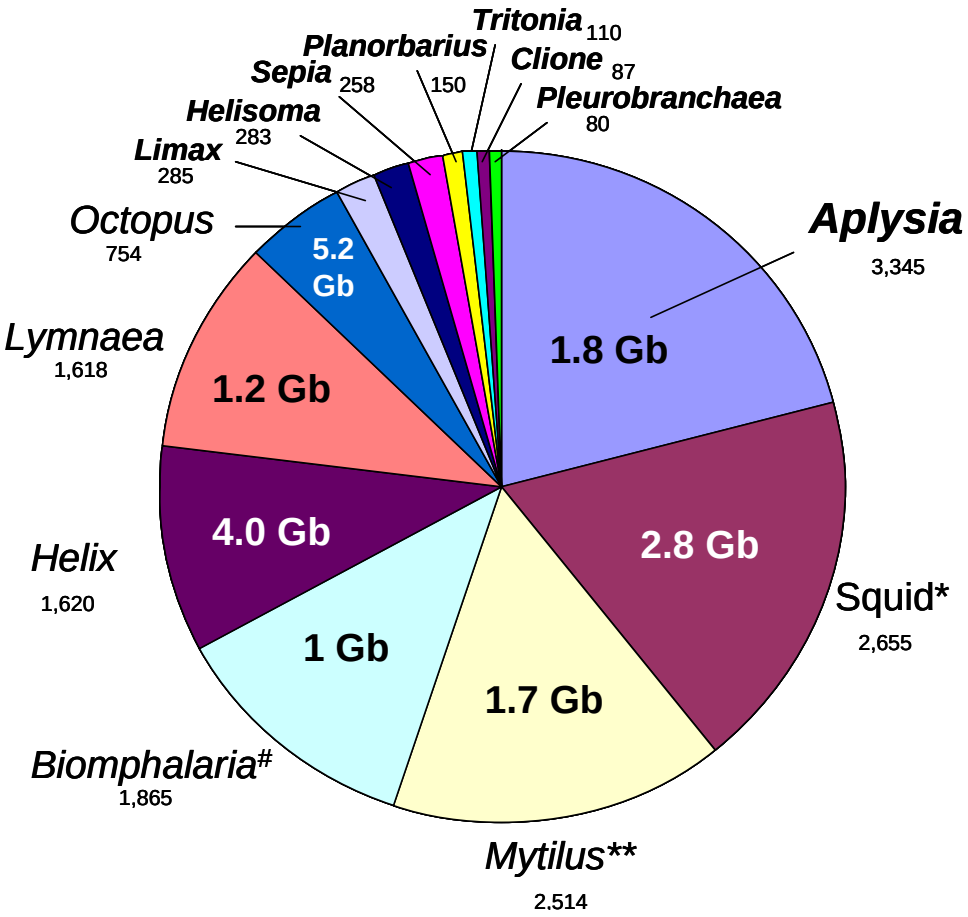

Fig. 1. Experimental molluscan models: number of publications in the fields of neuroscience and biology (PubMed data, October 10, 2003) and estimated genome sizes for selected molluscan models (Gb). Please note that the number of publications for squid (*Loligo*), *Mytilus* and *Biomphalaria* include closely related species of these genuses. Mean genome size for molluscs: 1.8pg ± 0.07. Smallest molluscan genome size: 0.43pg in *Lottia gigantea*, the owl limpet. *Lottia* was recently selected as a new genome target by DOE. There are only 3 publications in PubMed (1966-2004) and 24 in AFSA and Oceanic Abstracts for *Lottia* and several scientists working in the US.

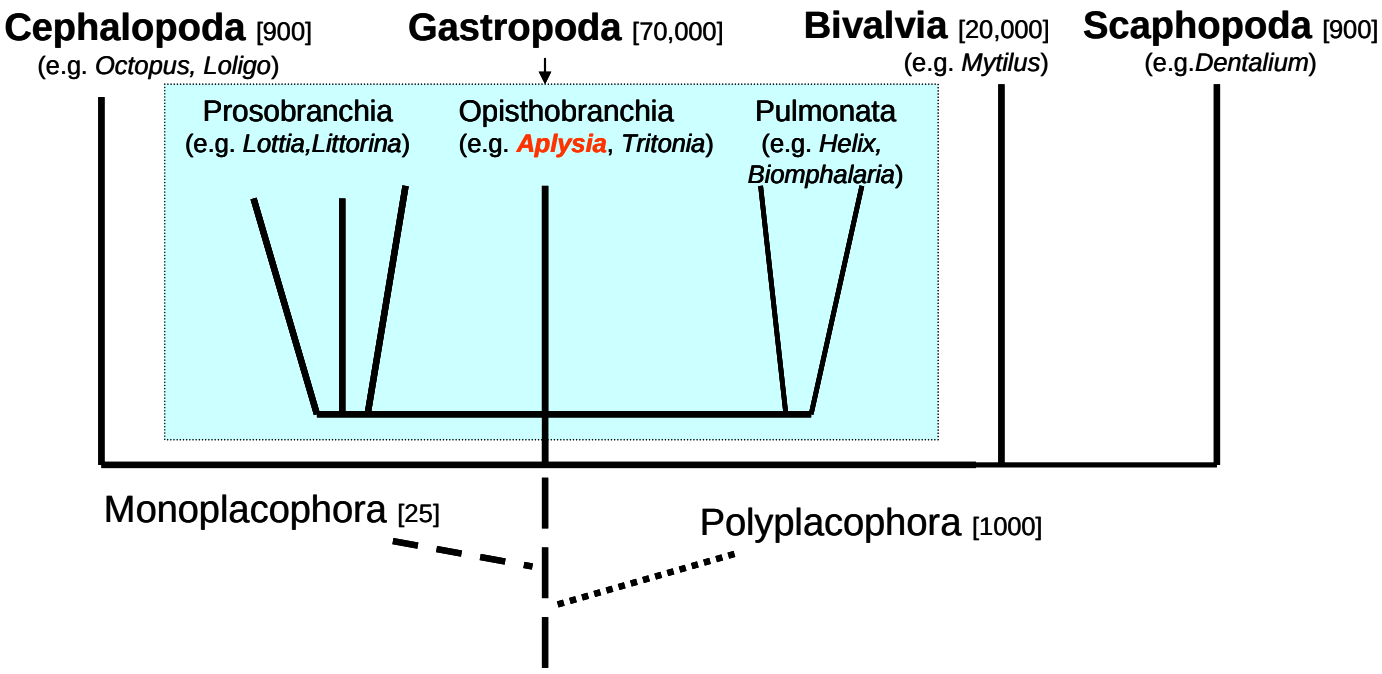

Fig. 2. Six Major Classes of the Phylum Mollusca. The numbers in brackets indicate number of described species (Brusca, Brusca, 2003). Aplacophora [~370 species] is a less investigated class (probably a basal lineage with Polyplacophora) and not shown here. As stated by Brusca & Brusca (2003), 'the Mollusca is such a diverse phylum, and so many taxa below the class level are apparently artificial (i.e. polyphyletic or paraphyletic), that efforts to trace their evolutionary history have often led to frustration." In view of the fact that the phylogenetic relationships among the different molluscan classes are not well established (in part due to the lack of sufficient molecular and genomic data) we conservatively collapse four major classes into one clade. For the same reason, we do not mark a precise branching point for two apparently more basal molluscan classes. Clearly, Gastropoda is the major class among molluscs with the best characterized fossil records and established research community. It is second (after insects) in number of extant species but the most diverse among all animals in their range of morphological forms (with widest grades of complexity known among bilaterians) and ecological adaptations. Selected members of this class such as *Aplysia* have the largest somatic cells in the animal kingdom and present unique experimental models for single cell and functional genomics.

References

Angerer, RC, Davidson EH, Britten RJ. (1975). DNA sequence organization in the mollusc *Aplysia californica*. Cell 6:29-39.

Bartsch, D., Casadio, A., Karl, K. A., Serodio, P., and Kandel, E. R. (1998) CREB1 encodes a nuclear activator, a repressor, and a cytoplasmic modulator that form a regulatory unit critical for long-term facilitation. *Cell*. 2:211-23.

Batzoglou, S., Jaffe, D.B., Stanley, K., Butler, K., Gnerre, S., Mauceli, E., Berger, B., Mesirov, J.P. and Lander, E.S. (2002). "ARACHNE: a whole-genome shotgun assembler." *Genome Res* 12(1): 177-89.

Brembs, B., F. D. Lorenzetti, et al. (2002). "Operant reward learning in *Aplysia*: neuronal correlates and mechanisms." *Science* 296(5573): 1706-9.

Brusca, R.C., Brusca, GJ. (2003). Invertebrates. Sinauer Associates, Inc., Publ. 936p

Bailey, C. H., Kaang, B. K., Chen, M., Martin, K. C., Lim, C. S., Casadio, A., and Kandel, E. R. (1997) Mutation in the phosphorylation sites of MAP kinase blocks learning-related internalization of apCAM in *Aplysia* sensory neurons. *Neuron.* 6:913-24.

Cash D, Carew TJ. (1989) A quantitative analysis of the development of the central nervous system in juvenile *Aplysia californica*. J Neurobiol. Jan;20(1):25-47.

Casadio, A., Martin, K. C., Giustetto, M., Zhu, H., Chen, M., Bartsch, D., Bailey, C. H., and Kandel, E. R. (1999) A transient, neuron-wide form of CREB-mediated long-term facilitation can be stabilized at specific synapses by local protein synthesis. *Cell.* 2:221-37.

Chain, D. G., Schwartz, J. H., and Hegde, A. N. (1999) Ubiquitin-mediated proteolysis in learning and memory. *Mol Neurobiol*. 2-3:125-42.

Chase, R. (2002). Behavior and its neuronal control in gastropod molluscs. Oxford Univ. Press, 314 p.

Coggeshall, RE. (1967) A light and electron microscope study of the abdominal ganglion of *Aplysia* californica. *J Neurophysiol*.6:1263-87

Cohen L, Hopp HP, Wu JY, Xiao C, London J. (1989). Optical measurement of action potential activity in invertebrate ganglia. *Annu Rev Physiol*.;51:527-41.

Davidson EH. (1990) How embryos work: a comparative view of diverse modes of cell fate specification. *Development*. Mar;108(3):365-89

Fang, X., Li, J. J., Perlette, J., Tan, W., and Wang, K. (2000) Molecular beacons: novel fluorescent probes. *Anal Chem*. 23:747A-753A.

Giusti AF, Hinman VF, Degnan SM, Degnan BM, Morse DE. (2000) Expression of a Scr/Hox5 gene in the larval central nervous system of the gastropod *Haliotis*, a non-segmented spiralian lophotrochozoan. Evol Dev. Sep-Oct;2(5):294-302.

Guan, Z., Kim, J. H., Lomvardas, S., Holick, K., Xu, S., Kandel, E. R., and Schwartz, J. H. (2003) p38 MAP kinase mediates both short-term and long-term synaptic depression in *Aplysia. J Neurosci*. 19:7317-25

Guan, Z., Giustetto, M., Lomvardas, S., Kim, J. H., Miniaci, M. C., Schwartz, J. H., Thanos, D., and Kandel, E. R. (2002) Integration of long-term-memory-related synaptic plasticity involves bidirectional regulation of gene expression and chromatin structure. *Cell*. 4:483-93.

Lespinet O, Nederbragt AJ, Cassan M, Dictus WJ, van Loon AE, Adoutte A. (2002) Characterization of two snail genes in the gastropod mollusc *Patella vulgata*. Implications for understanding the ancestral function of the snail-related genes in Bilateria. *Dev Genes Evol. May*; 212(4):186-95.

Jacob, M. H. (1984) Neurogenesis in *Aplysia californica* resembles nervous system formation in vertebrates. *J Neurosci.* 5:1225-39.

Jaffe, D. B., Butler, J., Gnerre, S., Mauceli, E., Lindblad-Toh, K., Mesirov, J.P., Zody, M.C. and Lander, E.S. (2003). "Whole-genome sequence assembly for mammalian genomes: Arachne 2." *Genome Res* 13(1): 91-6.

Hart M. (2000). Phylogenetic analyses of mode of larval development. *Semin Cell Dev Biol*. Dec;11(6):411-8.

Hegde, A. N., Inokuchi, K., Pei, W., Casadio, A., Ghirardi, M., Chain, D. G., Martin, K. C., Kandel, E. R., and Schwartz, J. H. (1997) Ubiquitin C-terminal hydrolase is an immediate-early gene essential for long-term facilitation in *Aplysia. Cell.* 1:115-26.

Hinegardner, R. (1974). Cellular DNA content of the Mollusca. *Comparative Biochemistry and Physiology* **47A**: 447-460.

Hinman VF, Degnan BM. (2002). Mox homeobox expression in muscle lineage of the gastropod *Haliotis asinina*: evidence for a conserved role in bilaterian myogenesis. *Dev Genes Evol*. Apr;212(3):141-4.

Kandel, E.R. (1976) Cellular basis of behavior. W.H. Freman & Co., San Francisco, 727p.

Kandel, E.R. (1979) Behavioral Biology of *Aplysia*, W.H. Freeman and Company. 463p.

Kandel, E.R. (2001) The molecular biology of memory storage: a dialogue between genes and synapses. *Science*. **294**, 1030-1038.

Kandel ER, Kriegstein A, Schacher S. 1980. Development of the central nervous system of *Aplysia* in terms of the differentiation of its specific identifiable cells. *Neuroscience*.5(12):2033-63.

Kim, H. K. and B. K. Kaang (1998). "Truncated green fluorescent protein mutants and their expression in *Aplysia* neurons." *Brain Res Bull* 47(1):35-41.

Kim, J. H., H. Udo, et al. (2003). "Presynaptic activation of silent synapses and growth of new synapses contribute to intermediate and long-term facilitation in *Aplysia*." *Neuron* 40(1): 151-65.

Kriegstein AR. 1977. Development of the nervous system of *Aplysia californica*. *Proc Natl Acad Sci U S* A. Jan;74(1):375-8.

Kriegstein AR. 1977. Stages in the post-hatching development of *Aplysia californica*. *J Exp Zool.* Feb;199(2):275-88.

Kriegstein AR, Castellucci V, Kandel ER. 1974 Metamorphosis of *Aplysia californica* in laboratory culture. *Proc Natl Acad* Sci USA. 71(9):3654-8.

Lasek, R.J. and W.J. Dower (1971). *Aplysia californica*: analysis of nuclear DNA in individual nuclei of giant neurons. *Science* **172**: 278-180.

Lee, J. A., H. K. Kim, et al. (2001). "Overexpression of and RNA interference with the CCAAT enhancer-binding protein on long-term facilitation of *Aplysia* sensory to motor synapses." *Learn Mem* 8(4): 220-6.

Manning, G., Whyte, D. B., Martinez, R., Hunter, T., and Sudarsanam S. (2002) The protein kinase complement of the human genome. *Science.* 298:1912-34.

Martin, K.C, Barad. M., and Kandel, E.R. (2000) Local protein synthesis and its role in synapse-specific plasticity. *Curr. Opin. Neurobiol.* 10:587-592.

Medina M and PJ Walsh (2000) Comparison of four Mendelian loci of the California sea hare (*Aplysia californica*) from populations of the coast of California and the Sea of Cortez. *Mar Biotechnol* 2:449-455.

Morrison, D. K., Murakami, M. S., and Cleghon, V. (2000) Protein kinases and phosphatases in the *Drosophila* genome. *J Cell Biol*. 2:F57-62.

Moroz, L.L., Gillette, R. and Sweedler, J.V. (1999). Single-cell analyses of nitrergic neurons in simpler nervous systems. *J. Exp. Biol.* **202**:333-341.

Muller, U.and Carew, T. J. (1998) Serotonin induces temporally and mechanistically distinct phases of persistent PKA activity in *Aplysia* sensory neurons. *Neuron*. 6:1423-34.

Page L.R.. Gastropod ontogenetic torsion: Developmental remnants of an ancient evolutionary change in body plan. (2003) *J Exp Zool*., 15;297B(1):11-26.

Purcell, A. L., Sharma, S. K., Bagnall, M. W., Sutton, M. A., and Carew, T. J. (2003) Activation of a tyrosine kinase-MAPK cascade enhances the induction of long-term synaptic facilitation and long-term memory in *Aplysia. Neuron.* 3:473-84.

Sahly, I., H. Erez, et al. (2003). "Effective expression of the green fluorescent fusion proteins in cultured *Aplysia* neurons." *J Neurosci Methods* 126(2): 111-7.

Sugita, S., Baxter, D. A., and Byrne, J. H. (1997) Modulation of a cAMP/protein kinase A cascade by protein kinase C in sensory neurons of *Aplysia. J Neurosci*. 19:7237-44.

Tong, A. K., Li, Z., Jones, G. S., Russo, J. J., and Ju, J. (2001) Combinatorial fluorescence energy transfer tags for multiplex biological assays. *Nat Biotechnol.* 8:756-9.

Major recent and upcoming scientific meetings:

1. Society for Neuroscience Meeting, New Orleans, USA; 7-12 November, 2003: 142 presentations on *Aplysia* and related species

2. 10th Symposium on Invertebrate Neurobiology, Tihany, Hungary, July 5-9, 2003

3. Behaviour and Neurophysiology of Molluscs; Surrey, United Kingdom; 16-17 April; 2004

4. The 15th World Congress of Malacology, Perth, Western Australia, 11-16 July 2004

5. 70th Annual Meeting of the American Malacological Society, Sanibel Island, Florida, July 31 - August 4, 2004.

6. Society for Neuroscience Meeting, San Diego, October 23-27, 2004:

Databases:

1. *Aplysia* Resource NIH facilities: http://www.rsmas.miami.edu/groups/sea-hares/

2. Malacological societies: http://manandmollusc.net/links_mala.html
http://www.il-st-acad-sci.org/malacol.html#societiesi
http://erato.acnatsci.org/ams/
http://www.sunderland.ac.uk/~es0mda/msl1.shtml
http://www.amonline.net.au/invertebrates/mal/malsoc/

3. List of of Malacological Meetings: http://erato.acnatsci.org/ams/pdfs/symposia.pdf

4. SFN: http://web.sfn.org/

5. *Aplysia* Database Project of Identified Cell Anatomy, Physiology, and Behavior: http://mollusc.med.cornell.edu

6. Aplysia EST databases (under construction): http://genomics3.biotech.ufl.edu:8080/bq/blastquest.jsp

The list of major laboratories of the *Aplysia* research community (PIs are listed in an alphabetical order). The list is primarily based on permanent users of *Aplysia* NIH research facilities and active laboratories in the field supporting the goals of the current white paper; a more extended list is available upon request)

Thomas Abrams, University of Maryland, MD
Ted Abel, Univ Pennsylvania, PA
Richard Ambron, Columbia University, NY
James Apland, US Army Med. Research Inst., TX
Craig Bailey, Columbia University, NY
Steven Blackband, University of Florida, FL
Vladimir Brezina, Mt. Sinai Medical Center, NY
James Blankenship, University of Texas, Galveston, TX
John Byrne, University of Texas Houston, TX
Thomas Carew, University of California, Irvine, CA
Hillel Chiel, Case Western Reserve University, OH
Gregory Clark, University of Utah, UT
Elizabeth Cropper, Mt. Sinai Medical Center, NY
Roger Croll, Dalhousie University, Canada
Larry Cohen, Yale University
Charles Derby, Georgia State University, GA
Stephen DeWeerth, Georgia Institute of Technology, GA
Frederic Doussau, Neurotransmission et Sec. Neroendoc., France

Robert Elson, University of California, San Diego, CA
Arnold Eskin, University of Houston, TX
Michael Ferragamo, Gustavus Adolphus College, MN
Thomas Fischer, Wayne State University, MI
Steven Fredman, Meharry Medical College, TN
Paul Forscher, Yale University
William Frost, Chicago Med School, IL
Michael Geusz, Bowling Green State University, OH
Daniel Gibson III, Worcester Polytechnic Institute, MA
Rhanor Gillette, University of Illinois, IL
David Glanzman, University of California, CA
Daniel Goldberg, Columbia University, NY
Jon Jacklet State, University of New York, NY
Yiming Liu Jackson, State University, MS
Ken Lukowiak, Univeristy of Calgary, Canada
Robert Hawkins, Columbia University, NY
Ashok Hegde, Wake Forest University Medical Center, NC
Robert Hill, University of Rhode Island, RI
Benahim Hochner, Hebrew University of Jerusalem, Israel
Bong-S, Kaan, Seoul Natl. University, South Korea
Leonard Kaczmarek, Yale University
Eric Kandel, Howard Hughes Medical Institute, NY
Paul Katz, Georgia State University, GA
Mark Klein, UCLA
Bernard Leib, Institure fuer Zoologie, Germany
Neil Magoski, Queens University, Canada
Kelsey Martin, University of California, Los Angeles, CA
Earl Mayeri, University of California, San Francisco, CA
Mark Miller, University of Puerto Rico, PR
Leonid Moroz, University of Florida, FL
Kermit Murray, Louisiana State University, LA
Gregg Nagle, University of Texas, Medical Branch, Galveston, TX
Dwight Nelson, University of Saint Thomas, MN
Sherry Painter, University of Texas, Medical Branch, Galveston, TX
Terence Peters University of California, San Diego, CA
Samuel Schacher, Columbia University, NY
James Schwartz, Columbia University, NY
Allen Selverston, University of California, San Diego, CA
Wayne Sossin, Montreal Neurological Institute, Canada
Micha Spira, Hebrew University of Jerusalem, Israel
John Stein, Brown University, RI
Abraham Susswein, Bar-Ilan Univ, Israel
Jonathan Sweedler, Beckman Institute, IL
Joseph Thornton, Columbia University, NY
Stuart Thompson, Stanford University, CA
Steven Treistman, Univ of Mass. Medical School, MA
Ferdinand S. Vilim, Mt. Sinai Medical Center, NY
Edgar T. Walters, University of Texas, TX
Klaudiusz Weiss, Mt. Sinai Medical Center, NY
William Wright, Chapman University, CA
Jian-Young Wu, Georgetown University, DC
Yalan Zhang, Yale School of Medicine, CT