Proposal for Construction of an *Aplysia californica* BAC Library

James J. Russo[1], Jingyue Ju[1], S. Puthanveetti[1], Monica Medina[2], Douglas L. Crawford[3], Patrick J. Walsh[4], Bruce Birren[5], Eric S. Lander[5], Pieter DeJong[6], Eric R. Kandel[1], and Leonid L. Moroz[7]
Aplysia californica sequencing consortium*

[1]Columbia University, [2]Joint Genome Institute, [3]The Center for Marine Genomics & [4]National Resource for *Aplysia* at the University of Miami, [5]Whitehead Institute & MIT Center for Genome Research, [6]BACPAC Resource Center at Childrens Hospital Oakland Research Institute, and [7]University of Florida

Contact Information: [1]Ph.(212)851-5165; Fax(212)851-5215; jjr4@columbia.edu (J.J.Russo); [1]Ph.(212)851-5172; Fax (212)851-5176; dj222@columbia.edu (J.Ju); [1]Ph.(212) 543-5204; Fax (212)543-5474; erk5@columbia.edu (E.R.Kandel); or [7]Ph.(904)461-4020; Fax (904)461-4054; moroz@whitney.ufl.edu (L.L.Moroz).

Summary: This proposal to obtain a 10x coverage genomic BAC library for *Aplysia californica* is an addendum to the accompanying request for deep draft sequencing of the *Aplysia* genome by the same consortium of investigators. The marine opisthobranch mollusc *A. californica* is a powerful experimental system in cellular, molecular and behavioral neuroscience because of the distinctive organization of its nervous system, which makes it advantageous for cellular analysis of a variety of behaviors and learning and memory. It is also a member of the Lophotrochozoa superclade of bilaterian animals with currently limited genomic information though it is the largest domain of the animal kingdom, and a unique model for developmental and evolutionary studies. The rationale for obtaining a BAC library is twofold. First, it will enhance the ability to assemble the whole genome sequence by taking advantage of the BAC library to generate scaffolds through paired-end reads. Second, it will provide an urgently needed resource for *Aplysia* investigators who will identify and utilize members of this large-insert clone library containing genes of their particular interest for further molecular studies, sequence closure (BAC-based sequencing), *in situ* experiments associating these BACs with specific chromosomes, and, most importantly, targeting novel regulatory regions of genes involved in basic neural functions.

1. The importance of *Aplysia* to biomedical or biological research:

1.1. *Aplysia* as a key model for comparative genomics. *Aplysia californica* and related opisthobranchs are free-living representatives of Mollusca (class Gastropoda), the second largest animal phylum (after Arthropods) 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). Notably, Mollusca belong to the Lophotrochozoa - one of three newly established superclades of bilaterian animals, and one that is underrepresented with regard to the establishment of genomic resources and molecular information that can support modern systematic, phylogenetic, evolutionary and developmental studies in this important animal taxon.

1.2. *Aplysia* as an emerging model for developmental biology. 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. The diverse life history strategies adopted by many species, combined with extreme body plan modifications, make molluscs an ideal group with which 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 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, exibit a high level of homoplasy (convergence, parallel evolution and reversals) in many morphologic traits. Some of this homoplasy is most likely due to differences in regulation of developmental networks. Genomic information from *Aplysia* will catalyze applications of molecular methods to study developmental mechanisms in the entire molluscan clade; it will help explain how gene expression controls major events in spiralian development, torsion-detorsion processes, neurogenesis, network organization and maturation.

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

1.3. *Aplysia* has provided fundamental insights into the basic organization of neuronal functions. The importance of *Aplysia* to modern neuroscience can not be overestimated. 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 the current neurogeneticist.

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 are studied better in *Aplysia* than in *Drosophila*, *C. elegans* and vertebrates. In a larger sense *Aplysia* is synergistic to these systems.

The distinction of *Aplysia* as a neurobiological 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, hundreds of neurons have been uniquely identified at the single cell level and have been 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 mediating circuitry. (7) Some of these behaviors can be modified by different forms of learning. As a result, one can gain access to genomic mechanisms of basic neuronal functions and to study these mechanisms in real physiological time with single-neuron resolution. No other models today can provide similar single-neuron resolution and its coupling to molecular mechanisms.

As a result, significant breakthroughs in our 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 and the control of feeding.

Finally, *Aplysia* provides a highly tractable paradigm 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 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 mostly 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.

1.4. *Aplysia* as a basic model for evolutionary and ecological functional genomics. The unique features of *Aplysia* and related gastropods will make it possible to study the genomic basis for 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 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, populational 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. Given the fact that the majority of molecular studies have been carried out in *Aplysia*, it should continue to serve as the reference mollusc for such studies. The BAC library requested herein is the first step toward a large-scale genomic multidisciplinary program with the *Aplysia* genome and functional genomic studies on the horizon. None of the representatives of this largest and most diverse phylum has sufficient genome resources linked to well-established biological, developmental and neurobiological background information; establishing such a genomic underpinning for *Aplysia* will provide unique opportunities for integrative and system biology.

2. Uses to which the BAC library would be put, in addition to genomic sequencing: As can be seen below the *Aplysia* research community is unified and enthusiastic about the needs for access to genomic resources. Practically every one of about 100 laboratories is performing molecular research directed to cloning and identification of regulatory regions of many neuronal genes. Our estimate is that more than 400 genes are currently being targeted by key players in development, learning, memory, neurological disorders, etc. As one can see from the more than 20 letters we obtained last week (available upon request), the advent of new and comprehensive *Aplysia* sequence data should stimulate and expand this research community. They will be anxious to have as accurate a sequence as possible as soon as possible, and the use of BAC end sequences for creating scaffolds and verifying assemblies, is crucial to achieving this goal in an organism the size of *Aplysia* (~1.9 Gb). Many of these investigators are also primed to take immediate advantage of the requested BAC clone resource, which will be available to them for conducting a number of experiments, including mapping, gene sequencing, and potentially comparative studies with vertebrate models. It is our prognosis that within the next couple of years the BAC library will be used for acquisition of information about the organization and regulatory regions of several hundred *Aplysia* genes, if not more. Fundamental aspects of the genomic bases of evolutionary and neurobiological processes will be experimentally approached with access to BAC libraries.

3. The size of the research community that could potentially use the BAC library and the community's interest in and support for having a BAC library: About one hundred active laboratories work on *Aplysia* in the U.S. and elsewhere. These scientists provide the required infrastructure and a solid foundation for most biological aspects of the organism. This is reflected in more than 3,500 publications (1965-2003) dealing with *A. californica* alone and over 16,000 for all 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. *Aplysia* investigators work in such distinct fields as 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). 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. Finally, there are more than two dozen international malacological societies 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 most successful of all animal phyla.

4. Whether the organism will be, or has been, proposed to NHGRI or another publicly funded agency for BAC-based genomic sequencing and the status of that request: To our knowledge, there is no proposal for BAC library construction from *Aplysia californica* or any related species from the entire subclass of opisthobranch molluscs. In parallel with this BAC library proposal, we have submitted a white paper to sequence the genome of *Aplysia.*

5. Other genomic resources that are available that will complement this resource: Members of the consortium at Columbia University and the University of Florida have received funding and recognition by NGHRI as a Center of Excellence in Genomic Science, to establish genomic approaches for studying the genomic basis of neuronal diversity and plasticity, including such technological innovations as rapid methods of DNA sequencing, nanoarrays for gene expression, and the use of subcellular detection systems for mRNA localization. All of these approaches are being developed utilizing, indeed taking advantage of, the giant neurons of *Aplysia*. In addition, the consortium has created and will expand their current set of over 100,000 sequenced ESTs (including ~35,000 unique sequences) from this organism's central nervous system, emphasizing individual ganglia, giant neurons, and even cellular components such as neural cell bodies or processes. Even before detailed sequence is available, or if the genome sequencing is not funded, it will still be possible to take advantage of these ESTs, which have revealed many signaling molecules and pathways previously suspected but never demonstrated in *Aplysia*, to select BACs. These BACs will permit complete gene sequences to be generated, particularly in cases where RACE methods prove intractable. Furthermore, they can be used to identify promoter elements through trapping or directed sequencing efforts. While in many cases, the whole genome draft sequence being requested in conjunction with this proposal will be complete and accurate through the length of genes of interest, there will often be sequence gaps that may include not only coding sequence but critical regulatory elements (promoters, enhancers, silencers, and the like); the BAC resource will be an excellent tool for filling in these gaps. Moreover, the amplification of BAC segments, each derived from single molecules, will be useful for obtaining information on extended haplotypes that are displayed by one copy of the

genome; the whole genome sequences, by definition, will contain a mixture of alleles, that can only be disambiguated within the stretch of a single sequence read. As other resources we also have more than 40 high quality cDNA libraries from different larval stages, CNS, individual ganglia, and identified neurons as well as from pure neuronal processes. We generated six normalized libraries from different sources, constructed two types of cDNA microarrays and maintain two neuronal transcriptome databases at the University of Florida and Columbia University. Over 3,000 novel transcripts from our datasets were identified using blastx and their functional annotation is currently in progress.

6. The strain of the organism proposed and rationale for *Aplysia* selection: The summary of features supporting the rationale is presented in the Table 1.

Table 1. **The rationale for selection of *Aplysia californica* as a target for a BAC library construction**

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° 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
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 processes
Behavior:	Well-characterized
Development:	Well-described
Genome size:	1.8 Gb (est. 30% repetitive)
Haploid chromosome #:	17
cDNA/EST resources:	>100,000 (~35,000 unique sequences) from the CNS, characterized individual neurons, neuronal processes, other specific tissues/life stages
Gene expression:	Microarrays, *in situ* hybridization, GFP-constructs
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 paradigm for cellular and systemic 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, developmental, memory and disease model
Databases:	EST databases, nervous system databases
Resources:	NIH supported National Resource for *Aplysia*
Cultivation:	Yes, animals can be bred in captivity; inbred lines can be generated
Number of Labs:	ca. 100 (*Aplysia*); >200 related molluscan species
Number of investigators:	> 1500
Genomic sequencing:	An accompanying white paper requesting 8x draft sequence based on plasmid, fosmid and BAC shotgun reads

Most importantly, *Aplysia* nation-wide facilities are providing reliable sources for all life stages of the organism, and inbred lines can be generated and used for a diversity of genomic applications including BAC library construction. The National Resource Center for *Aplysia* (see below) supplies exclusively *A. californica* to the research community—it is the species used by essentially all researchers worldwide—hence the decision to sequence this species.

7. The size of the genome: The *Aplysia* genome is about 1.8-2 x 10^9 bp in size and is distributed among 17 haploid chromosomes (Lasek & Dower, 1971; Hinegardner, 1974).

8. The availability of a source of DNA for construction of the BAC library (evidence of its quality for this purpose): DNA will be obtained from sperm from a single *Aplysia* specimen. The organisms are easily reared in the laboratory from fertilized ova to mature adults. 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, including one of the authors of this proposal (PJW). This Center supplies over 20,000 cultured *Aplysia* each year at all developmental stages to the research communities throughout the world. Though there are no isogenic lines as yet for this species, and normally specimens are collected in the wild as brood stock and outbred, it is very important to note that it is possible to inbreed F1 from these wild populations and subsequent progeny of *Aplysia* out to approximately 7 generations before inbreeding depression is observed. Thus, we are confident that several generations of inbreeding can be performed to decrease polymorphism in animals used for library production, EST and genomic sequencing. Dr. Walsh assures us

that he will be able to supply the needed amount of DNA (several hundred micrograms from a single such partially inbred organism) that will be necessary to prepare this BAC library as well as the Fosmid and plasmid libraries for the genome sequencing effort. But even if for some reason that proves difficult, it will be possible to collect enough DNA from inbred siblings that should have relatively homogeneous genomes over much of their length. Protocols exist in the Kandel and Moroz laboratories and at the Columbia Genome Center for isolating large fragments of DNA for BAC preparation; in fact, we have sent DNA in agarose plugs to Dr. DeJong in the past for other genomes and used these BACs for physical mapping and shotgun sequencing.

9. Specifications for the library (*e.g.*, library depth, BAC insert size) and supporting scientific rationale for these specifications: There are no special requirements for the BAC vector; standard methods used at CHORI by Dr. DeJong's group to obtain an approximate 10-fold coverage library of the 1.9 Gb genome (127,000 BACs of 150 kb average insert size) will suffice. Based on their experience, it is more difficult to obtain high quality (random and stable) BAC libraries for genomes with especially high AT content. The AT content of our full coding- and 3'-non-coding sequence enriched EST collection is about 60% and will most likely be only slightly higher for the genome as a whole; therefore BAC cloning should be routine. Preferably two half-libraries should be prepared utilizing distinct restriction enzymes, to decrease the chances that some regions of the genome will be uncloneable due to unusually large or small distances between restriction sites.

10. Time frame in which the library is needed: It will be desirable to have the BAC library as soon as possible to support the needs of the *Aplysia* research community and within half a year we would like to analyze it for candidates of neuronal, developmental and memory related genes at the laboratories of the authors of the proposal. If the *Aplysia* genome sequencing is approved, BAC end sequencing will be required for aid in assembly. Given the current and expected future speed of shotgun sequencing at MIT WI-CGC and other genome centers, the whole process could be accomplished within just a few months. However, we would like to stress that given the usefulness of the BAC library to the *Aplysia,* gastropod, mollusc, and neurobiology communities, and their readiness to take advantage of it, even if the sequencing has to await other projects assigned higher priority, the earlier the library can be generated, the better.

11. Other support that is available or has been requested for the construction of the desired library: No other support is available to us for BAC library construction. We have federally funded grants dealing with specific neurobiological questions and development of novel technologies in the fields of neuroscience and genomics. As far as we are aware, there has been no prior request to make an *Aplysia* BAC library, nor is there currently one available.

12. The need for an additional BAC library if one or more already exists: No library available at present.

13. Any other relevant information: If required, we would be happy to provide letters of support from major laboratories of the *Aplysia* research community and answer any questions the committee has.

References:

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.

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

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

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

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

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.

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.