Proposal for the Sequencing of a New Target Genome:
White Paper for a Planarian Genome Project

The *Schmidtea mediterranea* Sequencing Consortium
Represented by the authors of this document: A. Sánchez Alvarado[A], P. W. Reddien, P. A. Newmark, Chad Nusbaum[B]

[A]Contact Information: Ph. 801-581-3548; Fax 801-585-5171; sanchez@neuro.utah.edu

In consultation with

[B]The Whitehead Institute Center for Genome Research

Summary. Freshwater planarians are best known for their ability to regenerate complete animals from tiny fragments of their bodies. This regenerative ability is based upon totipotent, somatic stem cells present in the adult planarian. Planarians are free-living representatives of the Platyhelminthes, an understudied yet evolutionarily and clinically important phylum. Planarians are easily cultured in the laboratory and serve as useful experimental models for addressing many fundamental problems in biology that cannot be studied in *Drosophila* or *C. elegans*, including: wound healing, regeneration, somatic stem cells, and tissue homeostasis. *Schmidtea mediterranea* serves as a model planarian species: it is a stable diploid (2n=8); inbred lines are available for sequencing; 4,500 unique ESTs have already been sequenced; and the genome size of $4.8x10^8$ is approximately half that of other common planarians. A planarian genome project will provide a critical resource for developing this invertebrate model system by facilitating: gene identification, microarray experiments, comparative genomics, RNAi screens, genetic screens, and the identification of promoter sequences. With such tools available, the planarian will serve as a complementary model invertebrate, allowing functional genomic analyses relevant to studies of metazoan evolution, developmental biology, and human health.

Table of Contents

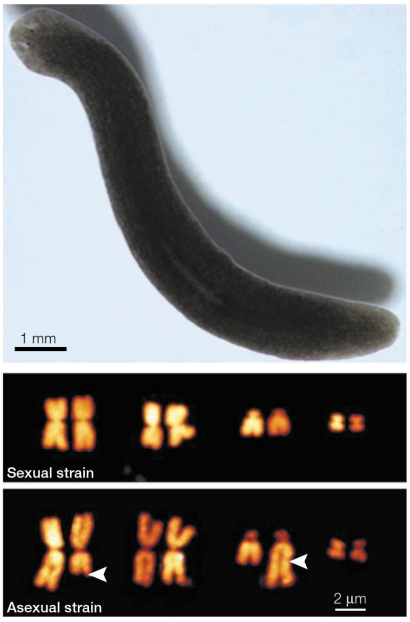

Figure 1. The planarian *S. mediterranea*. Top, Individual asexual animal. Bottom, Metaphase chromosome of diploid sexual and asexual strains; asexual reproduction is associated with a translocation (arrowheads).

Overview: **the planarian *Schmidtea mediterranea* as a model organism**

Planarians are free-living representatives of the phylum Platyhelminthes, a group of some 50,000 species of flatworms. Flatworms are among the simplest bilaterally symmetric animals: they are acoelomates, yet they possess derivatives of all three germ layers organized into complex organ systems. Thus, Platyhelminthes have been thought to occupy an important position in Metazoan evolution[1-4]. Current models place the Platyhelminthes in a large assemblage of protostome invertebrates, known as the Lophotrochozoa[5], a sister group to the Ecdysozoa (to which insects and nematodes belong).

Planarians are best known for their capacity to regenerate complete individuals from minuscule body parts[6,7], as well as for their ability to "de-grow" when starved[8-13]. Such extraordinary plasticity in the adult is in direct contrast to the rigidity displayed by currently used invertebrate models such as *Caenorhabditis elegans* and *Drosophila melanogaster*. The difference lies in a population of adult somatic stem cells, called neoblasts, that are distributed throughout the planarian body. Neoblasts are the only mitotically active cells in planarians[14], and their division progeny generate the 30-40 different cell types found in these organisms. In intact planarians these stem cells replace cells lost to normal physiological turnover; whereas, in amputated animals, they give rise to the regeneration blastema, the structure in which missing tissues are regenerated.

Until the mid-20th Century, planarians were a key model for studying development and regeneration. Yet, as attention shifted towards animals amenable to classical genetic analysis, the use of planarians declined[15]. Recently, however, the successful introduction of cell[14,16], molecular[17], and RNAi techniques[18] in planarians, along with heightened interest in stem-cell biology[19] and the plasticity of the differentiated state[20], has re-kindled interest in these fascinating organisms[21,22]. Sequencing the 4.8 x 10^8 bp genome[23] of the sexual, diploid planarian *S. mediterranea* will provide a vital resource for the development of a unique model to study metazoan evolution, regeneration, and the regulation of pluripotentiality. Mechanistic insights into these basic biological problems will have deep and obvious implications for the improvement of human health.

A. *Specific biological rationales for the utility of new sequence data*. Comparative genomics has become an important strategy for identifying candidate genes and regulatory elements underpinning complex biological problems. However, detailed mechanistic understanding of biological problems can best be obtained through experimentation. Therefore, it is vital to obtain genome sequences from organisms in which gene function can be studied and manipulated. The planarian *S. mediterranea* has emerged as a promising model in which to functionally address a number of long-standing and fundamental problems relevant to human health and biology that are not readily studied in well-established models such as *Drosophila* and *C. elegans*.

1. Improving human health: *Regenerative processes*. The human body has considerable capacity for repairing and replacing some of its tissues, especially those with high rates of cell turnover such as blood and epithelia. However, all human tissues do not share such regenerative properties. In fact, most debilitating conditions afflicting *Homo sapiens* in the developed world result from a failure of many tissues to repair themselves after injury. Spinal cord injury, for example, affects at least 250,000 Americans[24], and stroke and cardiac infarct are responsible for the death of nearly one million people per year in the U.S. alone, more than the next six leading causes of death combined[25]. The ability of planarians to regenerate all tissues after amputation provides unique opportunities to identify molecular mechanisms allowing muscle and neurons to regenerate. A systematic, molecular dissection of these processes in the planarian *S. mediterranea* is currently underway using RNA-mediated interference and DNA microarrays[26]. The *S. mediterranea* genome project (SmedGP) is essential for efficient gene prospecting and for identifying DNA regulatory regions that choreograph gene expression patterns driving planarian wound healing and regeneration. Through functional and comparative genomics, these studies should assist in the rational design of protocols for optimizing tissue repair in *H. sapiens*.

Stem cell disorders. In humans, a large number of malignancies exist that are caused by defects in stem cell maintenance, proliferation, and differentiation[27]. These range from aniridia, leukemias, and myelodysplastic syndromes to teratocarcinomas, epitheliomas, and neuroblastomas. Understanding the mechanisms that regulate stem cell function will have profound consequences for the design of effective therapies to treat these and many other human diseases. Our limited understanding of stem cell biology is mirrored by the relative inefficiency of clinically available stem cell-based therapies. For instance, only 30-45% of children treated for neuroblastomas by autologous stem cell rescue (ASCR) after myeloablative chemoradiotherapy achieve long-term, disease-free survival[28-30]. Therefore, understanding how stem cell proliferation, pluripotentiality, and differentiation are

regulated *in vivo* will improve the efficiency of ASCR and other stem cell-based therapies. Given that all metazoans depend on stem cells for their survival, it is likely that the mechanisms regulating stem cell biology have been conserved throughout evolution. Sequence information derived from the SmedGP, along with loss-of-function and population dynamic studies of the planarian neoblasts[31] will help identify key players involved in stem cell maintenance, proliferation, and differentiation. This information should prove invaluable to the study of stem cells as causes of human disease and for improving and devising stem cell-based therapies.

Infectious diseases. In addition to free-living forms, the phylum Platyhelminthes includes parasitic organisms, such as cestodes (*Taenia solium*) and digenetic trematodes (*Schistosoma mansoni* and *Schistosoma japonicum*), which are responsible for inflicting disabling and often life-threatening diseases upon nearly 300 million people throughout the world[32]. Further, it is estimated that about 5 million Americans (including military, diplomatic personnel, residents of foreign countries, and travelers) each year are at risk of contracting a tropical disease[33-36]. Because the life cycles of these parasites may involve as many as three different hosts, as well as both sexual and asexual reproductive strategies[37,38], experimental manipulation of such animals in the laboratory is difficult. The SmedGP, combined with the ease with which *S. mediterranea* can be maintained and manipulated experimentally in the laboratory, will allow functional genomic studies aimed at identifying platyhelminth-specific genes (absent from the human genome) that are necessary for flatworm survival. Thus, the SmedGP will streamline the identification of specific pharmacological targets for the prevention and treatment of helminthic infections.

2. Informing human biology: Experimental exploration of the many unique aspects of planarian biology described below should significantly enhance our understanding of human biology. The SmedGP will greatly facilitate study of these topics by providing a complete listing of genes and promoter sequences essential for the utilization of available tools (e.g., RNAi and genetic screens, microarrays, *in situ* hybridizations) and the development of new methodologies and resources (e.g., transgenesis, mutation mapping, gene knockouts).

Somatic Stem Cells. A major obstacle hampering progress in the understanding of stem cell regulation lies in the difficulty of accessing and studying these cells *in vivo*. This barrier is further compounded by the limitations of *in vitro* culture systems, which are unable to emulate the microenvironments in which stem cells reside and that are known to provide critical regulatory signals for their proliferation and differentiation[39]. Given the complexity of vertebrate adult somatic stem cell populations and their relative inaccessibility to *in vivo* molecular analyses, studies of somatic stem cells should benefit from analyses of simpler animal models. Study of *Drosophila* and *C. elegans* has provided invaluable contributions to our understanding of genes and pathways involved in a variety of human diseases. However, most somatic tissues in *Drosophila* and *C. elegans* are post-mitotic and, thus, are devoid of adult somatic stem cells. In contrast, *S. mediterranea* possesses an experimentally accessible population of stem cells that play a role in tissue maintenance and regeneration. The ability to label neoblasts with BrdU[14], purify them by fluorescence activated cell sorting[40], and silence gene expression by RNAi[41] should be very useful in identifying and functionally testing the mechanisms regulating stem cell activities. Sequence information derived from the SmedGP will be invaluable in the identification of control regions involved not only in maintaining the undifferentiated state of stem cells, but also in regulating their differentiation potential.

Germ cells. Germ cells represent an intriguing example of highly differentiated cells that maintain their totipotentiality and ability to reproduce themselves indefinitely. In the best-studied invertebrates (*Drosophila* and *C. elegans*) and many vertebrates, germ cells are specified early in embryogenesis by maternally supplied, localized cytoplasmic determinants. In mammals, however, germ cells are specified later in embryogenesis in a process that requires inductive interactions. The mechanisms that link these two apparently disparate modes of germ cell specification remain unclear. Unlike *Drosophila* and *C. elegans*, and more akin to mammals, planarians do not segregate their germ cell lineage during early embryogenesis; rather, germ cells are formed from stem cells post-embryonically. The biology of *S. mediterranea* is well suited for investigating the mechanisms that specify germ cell fate. Two strains of *S. mediterranea* exist: sexual worms that reproduce as cross-fertilizing hermaphrodites[42] and asexual worms that reproduce strictly by transverse fission. The sexual strain has a normal karyotype, whereas the asexual strain harbors a chromosomal translocation and is unable to differentiate germ cells. Furthermore, planarian stem cells can be eliminated by irradiation[43], resulting in adults devoid of germ line. Stem cells can be purified and transplanted into irradiated animals, effectively replacing the germline of one strain with that of another and allowing experiments testing the site (neoblast or soma) of gene function[44]. The SmedGP will facilitate analysis of the mechanisms by which cell-cell interactions specify germ cell fates, and should help

define pathways that have been conserved between this simple invertebrate and higher organisms, including humans.

Tissue/organ differentiation. During embryogenesis, coordinated morphogenetic movements are responsible for the formation and distribution of organs. In many cases multiple organogenic events take place simultaneously; thus it is difficult to identify the molecular mechanisms specific to each event and the extent of interdependency that may exist between them. Replacing the loss of a specific organ in an adult requires access and deployment of the specific developmental programs involved in the ontogeny of the missing body part(s). Thus, regeneration presents us with a unique opportunity to study tissue differentiation and organogenesis in isolation from other embryonic events. Genome sequence information, combined with functional studies in *S. mediterranea* should help identify the mechanisms by which tissues and organs are formed.

***Tissue plasticity*.** Differentiated and regenerating tissues in planarian fragments display a striking ability to reset their positional values to produce a normal, properly proportioned animal. For instance, a small fragment removed from the right flank of a planarian is capable of re-specifying its body midline to regain bilateral symmetry, while simultaneously preserving anteroposterior and dorsoventral polarities and resetting these axes to their appropriate positional values[7,45]. Such plasticity illustrates the enormous capacity possessed by planarians to maintain and regulate form and function. Relevant to human biology, understanding how this plasticity is dynamically regulated in the adult tissues of planarians should help identify factors that could be manipulated in adult human tissues to facilitate repair and/or replacement of tissues damaged by injury, degenerative disorders, and aging.

Tissue Homeostasis. The replacement of differentiated cells is a major challenge metazoans must face[46,47]. Humans, for example, must replace an estimated 10 billion cells every day[48]. Despite the importance of tissue homeostatic processes to human biology and health, relatively little is known about how the genome controls adult tissue homeostasis. Thus, numerous questions remain unanswered, including: How are cells in intact organs specified to die during turnover? How does a cell's death trigger its subsequent replacement by stem cells? How do organ systems maintain their order and function while in a state of cell flux? How do animals control and coordinate the size and cell number of multiple organ systems? How are developmental genes re-expressed in the adult during tissue homeostasis? Adult planarians constantly replace their tissues during normal cell turnover and regulate their size based upon metabolic conditions[13,49-51] through the proliferation and differentiation of the totipotent neoblasts. By contrast, other invertebrate model systems, such as *Drosophila* and *C. elegans*, have a more determined mode of development and utilize stem cells in the adult primarily for the regulation of the germline[52,53]. Planarians provide a unique and experimentally tractable system for studying such homeostatic processes: all tissues are regulated in the adult, tissue turnover is robust and rapid (as little as 7-10 days[14]), and developmental genes can be silenced by RNAi in adult animals[18].

***Aging and senescence*.** The use of invertebrate models to study aging has contributed significantly to our understanding of the mechanisms that regulate lifespan[54]. Studies in most invertebrates are limited by the fact that the adult soma is largely or entirely post-mitotic[52]. For example, mechanisms of aging uncovered in *C. elegans* are restricted to those affecting the lifespan and function of non-dividing adult cells[55]. In *S. mediterranea* however, adult tissues are constantly replaced, allowing the study of tissue aging during normal cell turnover. Asexual metazoans may be immortal, akin to the germ line of sexual metazoans[56,57]. Individual animals in clonal lines of some planarian species replicating by fission have been maintained for over 15 years and may indeed prove immortal[45]. In the sexual *S. mediterranea* strain somatic neoblasts may serve as an immortal cell source for producing germ as well as somatic cells in adults, and these animals do in fact have extremely long lives (greater than three years[58]), likely due to the ability of the neoblasts to replace aging cells. By developing assays to measure the age of cells or tissues[59,60] it should be possible to identify genes that: 1) regulate the rate at which existing cells are lost through cell death; and 2) that promote longevity by controlling cell replacement. These genes should define mechanisms by which cells determine their age and are specified for replacement. The long lifespan of planarians, coupled with the utilization of stem cells for the replacement of aged tissues, should provide a unique model for understanding aging in more complex metazoans, including humans.

***Behavior and learning*.** The evolutionary conservation of mechanisms that control nervous system development and function has been a notable discovery from studies of invertebrate models[61,62]. For example, behavioral studies in *Drosophila* have uncovered the mechanisms by which mammals regulate circadian rhythms[63] and studies in the mollusk *Aplysia* have identified genes and mechanisms by which animals, including humans, learn

and store memories[64,65]. Planarians display classical conditioning and memory formation, a phenomenon that has been reproduced in thousands of high school science fair projects[66-68]. The robust and complex behavioral responses of planarians to stimuli, combined with modern methodologies for studying gene function (genome sequence, RNAi, microarrays, *in situs*) should allow for the identification of mechanisms underlying how the planarian nervous system can control learning, memory, social behavior, chemotaxis, rheotaxis, and geotropism. Given the degree of conservation between planarian and human genes (see sections **A.3-4** below), understanding how the *S. mediterranea* genome controls behavior should provide insight into how learning and behavior are controlled in humans.

3. Informing the human sequence. *In vivo* experimental investigation at the molecular and cellular levels is easier in *S. mediterranea* than in higher organisms because of the rapid growth rate and clonal homogeneity of laboratory strains[42]. The ability to inhibit planarian genes by RNAi provides a rapid assay for testing the functions of novel genes. Furthermore, in contrast to *C. elegans*, the adult planarian nervous system is not refractory to RNAi[41], allowing the functions of neuron-specific genes to be tested rapidly. When comparative BLASTx analyses between ~3,000 *S. mediterranea* ESTs and the orfeomes of *C. elegans*, *D. melanogaster*, and *Homo sapiens* are performed, the predicted proteins of 124 *S. mediterranea* ESTs have significant similarity only to proteins encoded by the human orfeome[17]. Of these, 63 are similar to human proteins of unknown function. RNAi experiments, combined with whole-mount *in situ* hybridization data to analyze tissue-specificity, will help determine the function of these novel human proteins that have been lost from both the *Drosphila* and *C. elegans* genomes. The recent identification and functional analysis of *nou-darake*[41], a planarian ortholog of a human FGF receptor-like gene not found in the genomes of *C. elegans* or *Drosophila*, illustrates the untapped potential of planarians to provide insight into the human sequence.

4. Providing a better connection between human and non-human sequences. At this time, the only complete genome sequences from the protostome invertebrates are those of *Drosophila* and *C.elegans*. These species are much more closely related than originally thought: recent molecular studies indicate that they are both members of the ecdysozoa, a large group of molting invertebrates[69]. This incomplete phylogenetic representation can lead to incorrect interpretation of the conservation and evolution of gene families. For example, the identification of some genes present in the human genome, but absent from *Drosophila* and *C. elegans*, has led to the proposal that these genes arose as a result of direct horizontal gene transfer (HGT) between bacteria and vertebrates[70]. However, identification of *S. mediterranea* homologues of thymidine phosphorylase/endothelial cell growth factor 1(BLASTx $E=5x10^{-30}$), acyl-CoA dehydrogenase (BLASTx $E=2x10^{-21}$), epoxide hydrolase (BLASTx $E=5x10^{-29}$), and formiminotransferase cyclodeaminase (BLASTx $E=4x10^{-42}$) suggests that these loci are not shared by bacteria and vertebrates via HGT, but rather by descent through common ancestry[71,72]. This example clearly illustrates that even limited sequencing of *S. mediterranea* DNA can deepen our understanding of the human sequence. Since planarians reside in a phylogenetically separate position from *C. elegans* and *Drosophila* (see section **A.8**), sequencing the genome of *S. mediterranea* will provide critical information on genome evolution by filling in a major gap in the representation of complete invertebrate genome sequences (see Appendix, letters by Drs. Carroll, Giribet, and Ambros).

5. Expanding understanding of biological processes relevant to human health: ***Cancer***. Many types of cancers result from stem cell proliferation gone awry[27]. During planarian regeneration, the wound epithelium triggers proliferation of stem cells in the vicinity of the wound; this proliferation is halted when the appropriate amount of tissue has been regenerated[73]. The identification of mechanisms underlying this precise control of cell proliferation during regeneration will illuminate how cell division can be controlled in adult animals and identify candidate mechanisms that may be misregulated in human cancers.

Neurobiology. The planarian central nervous system consists of a bilobed brain and two longitudinal ventral nerve tracts connected by commissural neurons. When planarians suffer an injury that severs their nerve tracts, nerve cells can be repaired. Quite remarkably, when planarians are decapitated they can completely regenerate a new brain. Importantly, new neurons and axons that replace damaged tissues can functionally integrate with the pre-existing nervous system. Humans, in contrast, have very limited capacity for the replacement of injured nervous tissue; in the US alone, spine injuries and dysfunctions affect 250,000-400,000 people[24]. The SmedGP will provide a unique opportunity for identifying genes that control nervous system repair, which could serve as pharmacologic targets for the enhancement of regenerative capacities.

6. Additional surrogate systems for human experimentation. Ethical/political concerns have resulted in limitations on research using human embryonic stem cells. Such considerations do not apply to studies of the planarian stem cells. Because the mechanisms for regulating stem cell maintenance, proliferation, and differentiation are fundamental for the propagation of species, it is likely that the mechanisms controlling stem cell populations have been highly conserved. In fact, orthologues of the *Drosophila* gene, *piwi*, have been implicated in stem cell maintenance in divergent species ranging from plants and *C. elegans*, to planarians and humans[17,74]. The accessible stem cell population of *S. mediterranea,* coupled with RNAi and the SmedGP, will provide a surrogate system and an excellent model for analyzing the basic mechanisms that control stem cell behavior (see Appendix, letters by Drs. Spradling, Gage, Kimble, and McLaren).

7. Facilitating the ability to do experiments. The SmedGP will be invaluable for the development of planarians as experimental models to probe fundamental problems of human biology not readily accessible in current model organisms. The genome sequence will provide a wealth of information and resources that cannot be obtained by cDNA sequencing alone.

• Comparative genomics between *S. mediterranea* and other metazoans will facilitate predictions about how specific genes and/or pathways control biological processes. The capacity for experimental manipulation of gene function in planarians will allow the testing of such predictions.

• The SmedGP will identify the complete set of planarian genes. This will include alternatively spliced gene products and genes that are not easily found in EST collections (e.g., chemoreceptors[75]).

• Small non-coding RNAs form a surprisingly large family of genes (e.g., >100 in *C. elegans* [76,77]) that can only be found and studied using a complete genome sequence (see Appendix, letter by Dr. Ambros). The SmedGP will allow the identification of such RNAs and point to their possible involvement in key aspects of planarian biology, such as stem cell regulation and regeneration.

• The SmedGP will identify shared regulatory elements in genes found to have similar expression patterns through large-scale *in situ* hybridization screens[17].

• The genome sequence will be essential for the design and interpretation of comprehensive microarray experiments.

• The SmedGP will provide promoter sequences for developing transgenesis methodologies and GFP lines for labeling stem cells and other tissues for lineage and phenotypic analyses.

• The genome sequence will aid in the development of strategies for gene knock-outs (either using PCR screening of random genomic deletions requiring genomic sequence[78] or homologous recombination) and gene replacement.

• Genome sequence will greatly facilitate forward genetic strategies by allowing the development of polymorphism maps to molecularly identify genes disrupted by mutations[79,80] or to clone genes disrupted by transposons[81,82]. Such genetic strategies are possible because sexual *S. mediterranea* are cross-fertilizing hermaphrodites with a generation time between one and two months[83] that can be "selfed" by amputating and crossing regenerated clones.

8. Expanding our understanding of evolutionary processes.

***Phyletic relationships*.** Flatworms are among the simplest bilaterally symmetrical animals and have served a key role in the construction of metazoan phylogenies due to their unique body plan. They possess three germ layers organized into complex organ systems, yet they lack a coelom (mesoderm-lined body cavity) and their digestive system lacks an anus[37]. Thus, Platyhelminthes have been thought to occupy an important position in Metazoan evolution[1-4]. Recent phylogenetic models have radically revised our views of the relationships between the protostome invertebrates. The protostomes are now divided into two large groups: the Ecdysozoa (including arthropods and nematodes) and the Lophotrochozoa (including annelids, mollusks, and flatworms). At this time, all of the complete genome sequences from protostomes have been obtained from the Ecdysozoa (*C. elegans* and *Drosophila*); thus, the Lophotrochozoa are currently under-represented in current genomic databases. Because of the evolutionary distance of flatworms from those animals with currently available genomes, the genome sequence of *S. mediterranea* will fill a large gap in the metazoan lineage by providing a critical resource for improving the resolution of phyletic relationships.

Evolution of human genes. Are all of the genes found in humans, but not in the ecdysozoa, deuterostome inventions? The SmedGP will identify genes conserved with other protostome taxa and man, but lost in the ecdysozoa (see section **A. 4** for specific examples). These data will have profound impact on our understanding of the evolution of genes, gene pathways, and biological processes.

Relationships between gene pathways. A complete *S. mediterranea* genome sequence will allow comparison of planarian genes with those of humans and model systems such as mouse, *Drosophila*, *C. elegans*, and zebrafish. Such comparisons, established evolutionary relationships between these model systems, and functional data for gene and pathway function are important factors for testing the universality and relationship to human biology of discoveries made in model systems. For instance, are stem cell processes primordial or derived in different systems? If conservation of stem cell regulation is found between the germline of *Drosophila* and the neoblasts of *S. mediterranea*, then depending upon their evolutionary relationship, the shared modes of stem cell regulation are likely to be shared with humans. In other words, if *Drosophila* and *S. mediterranea* were related within the same bilaterian clade (as current models suggest for *Drosophila* and *C. elegans*), then conserved functions may be due to evolution of the function strictly within that clade. But, if *Drosophila* and *S. mediterranea* are as closely related to each other as either is to humans, then conserved functions have a high probability of existing in man. Thus, the current position of *S. mediterranea* in a sister phylum to humans and to *Drosophila* and *C. elegans* places it in a key position for complementing studies in other invertebrate models for informing the universality of model organism discoveries.

B. Strategic issues in acquiring new sequence data

1. Demand for platyhelminth genome sequence. Besides planarian researchers, the primary communities that will directly benefit from the SmedGP include laboratories that study regeneration, stem cells, development, parasitism, metazoan evolution, platyhelminthes in general, and a much broader community interested in comparative genomics (see Appendix). In the past two years, the planarian community has published in journals such as *Nature*, *Nature Reviews Genetics*, *Development*, and *PNAS*, and the platyhelminth community at large has published nearly 400 manuscripts (PubMed), suggesting that data obtained from the SmedGP will be used productively by members of the planarian and the platyhelminth communities. In addition, keen interest in planarians by other biomedical disciplines is evidenced by the letters of support (see Appendix), and the number of researchers from the US and abroad (~35 per day and over 4,000 since October 2002) consulting the *Schmidtea mediterranea* database (SmedDb; http://planaria.neuro.utah.edu)[17]. Such interest is bound to expand with a genome project since SmedGP will aid in making the biology and experimental tools of planarians more commonplace to fundamental, biomedical, and applied research. For instance:

a. Stem cell researchers: The identification of genes and gene interactions responsible for regulating the planarian neoblasts will have applications to stem cells in vertebrates, including humans. Having a complete inventory of *S. mediterranea* genes and the ability to rapidly test their function by RNAi will aid the efforts of stem cell biologists to better understand the regulation of cellular pluripotentiality, and provide target molecules for regenerative medicine in mammals (see Appendix, letters by Drs. Spradling, Gage, Kimble, and McLaren).
b. Developmental Biologists: The SmedGP will allow the identification of the planarian homologs to genes known to play key roles in the development of other model organisms. Many genes in common model systems such as *Drosophila*, *C. elegans*, zebrafish, and mice confer embryonic lethality when mutated, precluding their study at later developmental stages. Planarians are known to express many developmentally relevant genes as adults[84]. Since RNAi can be used to disrupt gene function in intact and regenerating adults[18], it is possible to perturb genes after embryonic development has proceeded normally, effectively overcoming embryonic lethality problems. This provides a unique opportunity for developmental biologists to study the function of key developmental regulators at later stages such as during morphogenesis and organogenesis (see Appendix, letters by Drs. Gall, Brown, and Ambros).
c. Parasitologists: There is a critical need for a non-parasitic model capable of complementing at the molecular level studies of the parasitic platyhelminthes. This fact has considerably complicated the molecular study of diseases such as schistosomiasis and thus hampered the development of therapies for more than 300 million human beings afflicted by helminthic diseases. The SmedGP will produce new molecular data that should help identify and functionally assess platyhelminth-specific genes. This will

allow the design of novel therapeutic strategies and be of great usefulness to researchers engaged in the study of helminthic parasites (see Appendix, letter by Dr. Day).

d. Evolutionary Biologists: The SmedGP will provide an invaluable resource for comparative genomics and molecular evolutionary studies since no genome sequence is yet available for any members of the Lophotrochozoa. We predict that the SmedGP sequencing data will be widely anticipated by taxonomists and evolutionary developmental biologists engaged in understanding the evolution of body plans and developmental processes (see Appendix, letters by Drs. Carroll and Giribet).

Finally, the SmedGP will have a tremendous impact on the emerging planarian community. The availability of *S. mediterranea* ESTs, microarrays, and RNAi, coupled with the sequencing of its genome, is likely to lead to an influx of talented young and established investigators (e.g., see appendix, letter by Dr. Vale) to the field to utilize and expand upon the resources generated by the SmedGP.

2. Suitability of *Schmidtea mediterranea* for experimentation:

Table 1. Key Features of *S. mediterranea*

Biological Traits	
Haploid Chromosome #	4
Genome Size	$4.8x10^8$
Adult Animal Size	
Asexual biotype	1-8 mm
Sexual biotype	1-3 cm
Cellular Organization	Multicellular
Germ layers	3 (triploblastic)
Ploidy	Stable diploid
Generation Time	
Asexual biotype	10 days
Sexual biotype	1 month
Cultivation	Freshwater, aquatic animal; easy to rear, expand, and breed in the lab
Genetic Resources/Tools	
cDNA/EST resources	~5,000 genes
Gene Inactivation	RNAi
Gene Expression	Microarrays and whole-mount *in situ* hybridizations
Immunocytology	20 cell types distinguishable by antibodies
Loss-of-function phenotypes	Growing rapidly. Presently, nearly 20
Spontaneous mutants	3 thus far in sexual strain
Special Strengths	Asexual and sexual clonal lines developed
	Inbred line available
	"Self crossing" by amputation, regeneration, and breeding
	Regeneration and stem cell animal model
	Ease of transplantation and surgical manipulations
	Ability to carry out large-scale gene inactivation screens with RNAi
	Ability to study embryogenesis and adult regeneration in same species
	Purification of stem cells by FACS
	Tissue culture of neurons
	Extensive classical literature
Weaknesses	Transient transgenesis only; stable transgenesis expected in near future
	Stem cell culture poorly developed
Databases	EST resource, gene expression data and antibody markers resource **http://planaria.neuro.utah.edu**
Number of labs	Ca. 30 in planarians; nearly 400 in platyhelminthes

3. The rationale for the complete sequence of the organism. The wealth of information stored in the planarian genome is reflected by this organism's remarkable biological properties. Given the powerful tools available in planarians for post-genomic analyses, a full genome sequence will provide methodologies, resources, and strategies not provided by cDNAs alone that will help fully exploit the properties of *S. mediterranea* to address a broad range of biological and human health-related problems. Among these are:

a. **Genomic prospecting** of DNA regulatory regions responsible for staging the gene expression choreography driving planarian wound healing and regeneration.

b. Mapping and identifying **genes for complex biological traits** such as **stem cell maintenance**, **tissue homeostasis**, **aging** and **senescence**, **behavior** and **learning** (including regulatory genes normally expressed at low copy number and thus difficult to obtain by other approaches).

c. Insights into **developmental processes** and **germline regulation**.
d. Increase our understanding of the **evolutionary relationships** between different phyla.
e. Identification of **platyhelminth-specific genes** to develop vaccines and/or treatments for helminthic diseases.
f. Improve the **identification and functional characterization of orthologs of disease genes and genes of unknown function in humans**.

4. Cost and readiness. The *S. mediterranea* genome is estimated to have 480 Mb, i.e., less than 16% the size of the human and mouse genomes. The Whitehead Institute Center for Genome Research (WI-CGR) has expressed strong interest in pursuing the SmedGP (see Appendix, letter by Dr. C. Nusbaum). An inbred line of the diploid sexual strain of *S. mediterranea* will be used to generate genomic DNA for sequencing. Based on the sequence data obtained from the *S. mediterranea* EST project[17], we estimate an AT content of 40% for the coding regions, and thus anticipate no significant difficulties in cloning, sequencing, and the eventual assembly of the genome. In addition, the sequencing of nearly 10,000 cDNAs of *S. mediterranea* did not identify large numbers of retrotransposons, and Southern blot analyses have failed to detect large numbers of transposable elements[85]. Therefore, the combination of a diploid genome with a notable absence of large numbers of transposable elements and expressed retrotransposons should make the assembly of contiguous clones and chromosome walking possible in *S. mediterranea*.

The following strategy will be used to sequence and assemble the *S. mediterranea* genome:

a) Whole-genome shotgun (WGS) sequencing. The WI-CGR has extensive experience with WGS sequencing and has successfully assembled genomes ranging from 5.8 Mb to 2.7 Gb. Standard WGS procedures at the WI-CGR involve sequence read production from sheared random whole genome inserts of 4 Kb (high copy) and 10 Kb (low copy) plasmids, as well as 40 Kb Fosmids and 50kb "Jumping" clones. For the *S. mediterranea* project, this corresponds to a total of approximately 7.5 million attempted reads to deliver 10x coverage of the genome. We assume a pass rate of 84% and an average Phred 20 read length of 770 b for plasmids and jumping clones, and 75% and 680 Phred 20 read length for Fosmids. This sequence will provide a combined physical coverage of approximately 69.6 x. Details are shown in Table 2.

Sequence and template coverage of planned whole genome assembly: Table 2

	# clones	#seq. reads	Seq. coverage	Physical coverage
4 kb Plasmids	2,225,000	4,450,000	6.0 x	13.1 x
10 kb Plasmids	1,113,000	2,226,000	3.0 x	23.2 x
40 kb Fosmids	235,000	471,000	0.5 x	19.6 x
50 kb Jumping	186,000	371,000	0.5 x	13.7 x
Totals	3,759,000	7,518,000	10.0 x	69.6 x

b) WGS assembly. WGS sequencing reads will be assembled with Whitehead's Arachne assembly tool[86,87]. Arachne was specifically designed for assembly of large genomes with significant repeat content. The algorithm takes advantage of subclone read pairing and size information in addition to sequence overlap data. By using different library types and insert sizes, we minimize the effects of cloning bias and maximize the efficiency of the Arachne hierarchical linking algorithm to resolve repeat regions and generate long range linking in the assembly process. Arachne has been extremely effective in assembling at least 15 genomes ranging from 0.8 Mb to 2.7 Gb in size, including genomes with up to ~40% repeat content. For example, the 7x Arachne assembly of the 2.7 Gb mouse genome resulted in an assembly consisting of 89 ultracontigs placed on the 20 mouse chromosomes and covering ~96% of the genome, with an length weighted average (N50) scaffold size of 16 Mb[88]. Further, 95% of the bases in the assembly have a calculated base quality of 40 or greater, which approaches the standard of finished sequence. Since the repeat content of its genome appears modest, and the proposed sequence coverage higher, we anticipate the *S. mediterranea* assembly will have very long range scaffold contiguity and very high base quality.

c) Closure of the genomic sequence. The knowledge gained from performing closure and finishing at large scale at WI-CGR has produced new, streamlined approaches to these activities. Expert software now directs much of the work required to close gaps in a fully automated fashion, evaluating each sequence gap, matching it with the appropriate laboratory workflow and evaluating gap closure success. All laboratory work is performed by robots. Finishing is now routinely performed using directed primer walks off Fosmid templates. The nearly 19.6x Fosmid

template coverage of the *S. mediterranea* genome is predicted to cover at least 97% of the genome, and will be stored in just over 600 384-well plates, vastly simplifying logistics of finishing. It will be straightforward to finish selected regions of the genome, but significant finishing can be performed of the entire genome for less than 10% of the cost of the shotgun sequencing.

d) Access to the *S. mediterranea* genome data. Genome data will be released in accordance with NHGRI rules. All traces will be submitted to the NCBI trace repository. In addition, the completed genome will be sent to GenBank. Genomes in progress at the WI-CGR are served on the Whitehead Institute's website with BLAST servers, downloadable data, and integration of existing datasets such as ESTs and physical maps. In order to distribute and display sequence and associated genomic information, the *S. mediterranea* sequencing consortium is organizing an *S. mediterranea* Genome Database to be hosted by the University of Utah Center for High Performance Computing (see Appendix, letters by Drs. Facelli and Gesteland). The *S. mediterranea* community will work closely with WI-CGR and the Center for High Performance Computing to make the data as user-friendly as possible to the scientific community at large.

Sequencing an entire 480 Mb genome requires approximately 7.5 million attempted reads for a 10x assembly. The WI-CGR operates the leading low cost facility in genome sequencing, and continues to aggressively pursue driving down costs. As a result, sequencing costs will be significantly lower at the time of initiation of sequencing than they are at present, so costs will be negotiated at that time with the NHGRI. The above plan was developed in close consultation with WI-CGR. Discussions on SmedGP started in October 2002 and culminated in February 2002.

5. Other (partial) sources of funding.
NIH NIGMS award RO-1 GM057260-07 for microarray and automated *in situ* hybridizations (ASA)
NSF award IBN-0237825 for sexual strain ESTs (PAN)
Riken award for the construction of a physical map for *Schmidtea mediterranea* (Dr. Kiyokazu Agata)

Letters of Support: Roster (see Appendix 1)

The following individuals have written in support of the SmedGP as nationally and internationally recognized experts, representatives of specific communities, or both.
C. Nusbaum, Co-director, Genome Sequencing and Analysis, Whitehead Institute Center for Genome Research; **R. F. Gesteland**, Vice President for Research, Distinguished Professor, Human Genetics, University of Utah; **J. Facelli**, Director, Center for High Performance Computing, University of Utah, and Consortium member; **K. Agata**, leading planarian biologist, Group Director, RIKEN Center for Developmental Biology, Kobe, Japan, and Consortium Member; **V. Ambros**, leading developmental biologist and non-coding RNA expert, Professor of Genetics, Dartmouth Medical School; **D. D. Brown**, leading developmental biologist, Staff Member, Carnegie Institution of Washington, Dept. of Embryology; **S. B. Carroll**, leading evolutionary and developmental biologist, Investigator HHMI, Professor of Genetics and Molecular Biology, University of Wisconsin-Madison; **T. Day**, leading helminthic parasitologist, Assistant Professor and Chair Neuroscience Program, Iowa State University of Science and Technology; **F. H. Gage**, leading stem cell and regeneration biologist, Adler Professor, Laboratory of Genetics, The Salk Institute; **J. G. Gall**, leading cell and developmental biologist, Staff Member, Carnegie Institution of Washington, Dept. of Embryology, American Cancer Society Professor of Developmental Genetics; **G. Giribet**, leading molecular taxonomist, Assistant Professor of Biology and Curator of Invertebrates, Museum of Comparative Zoology, Harvard University; **J. Kimble**, leading developmental and stem cell biologist, Investigator HHMI, Professor of Genetics and Molecular Biology, University of Wisconsin-Madison; **A. McLaren**, leading germ and stem cell biologist, The Wellcome Trust/Cancer Research UK Institute of Cancer and Developmental Biology, University of Cambridge, United Kingdom; **A. Spradling**, leading developmental and stem cell biologist, Director, Dept. of Embryology, Carnegie Institution of Washington, Investigator HHMI, co-PI, Berkeley Drosophila Genome Center; **R. Vale**, leading biochemist, Investigator HHMI, Professor of Cellular and Molecular Pharmacology, University of California in San Francisco.

Institutional Affiliations of White Paper Authors: A. Sánchez Alvarado, Dept. of Neurobiology & Anatomy, University of Utah School of Medicine; P. W. Reddien, Dept. of Neurobiology & Anatomy, University of Utah School of Medicine; P. A. Newmark, Dept. of Cell & Structural Biology, University of Illinois at Urbana-Champaign; Chad Nusbaum, Whitehead Institute Center for Genome Research.

References Cited

1. Adoutte, A., Balavoine, G., Lartillot, N. & deRosa, R. Animal evolution. The end of the intermediate taxa? *Trends Genet* **15**, 104-8 (1999).
2. Henry, J. Q., Martindale, M. Q. & Boyer, B. C. The unique developmental program of the acoel flatworm, Neochildia fusca. *Dev Biol* **220**, 285-95. (2000).
3. Willmer, P. *Invertebrate Relationships* (Cambridge University Press, New York, 1994).
4. Tyler, S. in *Interrelationships of the Platyhelminthes* (eds. Littlewood, T. J. & Bray, R. A.) 3-12 (Taylor and Francis, London, New York, 2001).
5. Balavoine, G. The early emergence of platyhelminths is contradicted by the agreement between 18S rRNA and Hox genes data. *C R Acad Sci III* **320**, 83-94 (1997).
6. Randolph, H. Observations and experiments on regeneration in planarians. *Arch. Entw. Mech. Org.* **5**, 352-372 (1897).
7. Morgan, T. H. Experimental studies of the regeneration of *Planaria maculata*. *Arch. Entw. Mech. Org.* **7**, 364-397 (1898).
8. Schultz, E. Über Reduktionen. I. Über Hungerserscheinungen bei *Planaria lactea*. *Arch. Entwm.* **18**, 555-577 (1904).
9. Lillie, F. R. A comparison of the power of regeneration in three genera of planarians, viz. *Planaria*, *Phagocata*, and *Dendrocoelum*. *Science* **13** (1901).
10. Berninger, J. Über die Einwirkung des Hungers auf Planarien. *Zool. Jahrb.* **30**, 181-216 (1911).
11. Child, C. M. Studies on the dynamics of morphogenesis and inheritance in experimental reproduction. III. The formation of new zooids in planaria and other forms. *J Exp Zool* **11**, 220-280 (1911).
12. Abeloos, M. Recherches expérimentales sur la croissance et la régénération chez les planaires. *Bull. Biol.* **1**, 1-140 (1930).
13. Oviedo, N. J., Newmark, P. A. & Sánchez Alvarado, A. Allometric scaling and proportion regulation in the freshwater planarian *Schmidtea mediterranea*. *Dev Dyn* **226**, 326-33 (2003).
14. Newmark, P. & Sánchez Alvarado, A. Bromodeoxyuridine specifically labels the regenerative stem cells of planarians. *Dev Biol* **220**, 142-53 (2000).
15. Mitman, G. & Fausto Sterling, A. in *The Right Tools for the Job: at Work in 20th-Century Life Sciences* (Princenton University Press, Princeton, New Jersey, 1992).
16. Robb, S. M. C. & Sánchez Alvarado, A. Identification of immunological reagents for use in the study of freshwater planarians by means of whole-mount immunofluorescence and confocal microscopy. *genesis* **32**, 293–298 (2002).
17. Sánchez Alvarado, A., Newmark, P., Robb, S. M. C. & Juste, R. The *Schmidtea mediterranea* database as a molecular resource for studying platyhelminthes, stem cells and regeneration. *Development* **129**, 5659-5665 (2002).
18. Sánchez Alvarado, A. & Newmark, P. A. Double-stranded RNA specifically disrupts gene expression during planarian regeneration. *Proc. Natl. Acad. Sci. USA* **96**, 5049-5054 (1999).
19. Spradling, A., Drummond-Barbosa, D. & Kai, T. Stem cells find their niche. *Nature* **414**, 98-104 (2001).
20. Wilmut, I., Schnieke, A., McWhir, J., Kind, A. & Campbell, K. Viable offspring derived from fetal and adult mammalian cells. *Nature* **385**, 810-3 (1997).
21. Brown, K. in *HHMI Bulletin* 28-32 (2002).
22. Pearson, H. The regeneration gap. *Nature* **414**, 388-390 (2001).
23. Agata. Personal communication.
24. http://www.spinalcord.uab.edu/show.asp?durki=21446.
25. http://www.ninds.nih.gov/health_and_medical/disorders/sci.htm.
26. ASA & PWR. unpublished results.
27. Reya, T., Morrison, S. J., Clarke, M. F. & Weissman, I. L. Stem cells, cancer, and cancer stem cells. *Nature* **414**, 105-11 (2001).
28. Graham Pole, J., Casper, J. & Elfenbein, G., et al. High-dose chemoradiotherapy supported by marrow infusions for advanced neuroblastoma: A Pediatric Oncology Group Study. *J Clin Oncol* **9**, 152-158 (1991).

29. Stram, D., Matthay, K., M, O. L. & al, e. Consolidation chemoradiotherapy and autologous bone marrow transplantation versus continued chemotherapy for metastatic neuroblastoma: A report of two concurrent Children's Cancer Group studies. *J Clin Oncol* **14**, 2417-2426 (1996).
30. Matthay, K., Harris, R. & Reynolds, C., et al. Improved event-free survival for autologous bone marrow transplantation versus chemotherapy in neuroblastoma: A phase III randomized Children's Cancer Group Study. *Proc Amer Soc Clin Oncol* **17**, 525a. (1998).
31. ASA & PAN. Unpublished results.
32. http://www.who.int/inf-fs/en/fact115.html.
33. Congress, U. S. (Office of Technology Assessment, Washington, DC, 1985).
34. CDC. Acute Schistosomiasis with Transverse Myelitis in American Students Returning from Kenya. *Morbidity and Mortality Weekly Report* **33**, 445-7 (1984).
35. CDC. Acute Schistosomiasis in U.S. Travelers Returning from Africa. *Morbidity and Mortality Weekly Report* **39**, 141-142,147-148 (1990).
36. CDC. Schistosomiasis in U.S. Peace Corps Volunteers-Malawi, 1992. *Morbidity and Mortality Weekly Report* **42** (1993).
37. Hyman, L. H. *The Invertebrates: Platyhelminthes and Rhynchocoela the Acoelomate Bilateria* (McGraw-Hill Book Company Inc., New York, 1951).
38. Brusca, R. C. & Brusca, G. J. *Invertebrates* (Sinauer Associates, Sunderland, MA, 1990).
39. Spradling, A., Drummond-Barbosa, D. & Kai, T. Stem cells find their niche. *Nature* **414**, 98-104 (2001).
40. Agata, K. & ASA. Unpublished results.
41. Cebrià, F. et al. FGFR-related gene nou-darake restricts brain tissues to the head region of planarians. *Nature* **419**, 620-4 (2002).
42. Newmark, P. A. & Sánchez Alvarado, A. Not your father's planarian: a classic model enters the era of functional genomics. *Nature Reviews Genetics* **3**, 210-219 (2002).
43. Bardeen, C. R. & Baetjer, F. H. The inhibitive action of the Roentgen rays on regeneration in planarians. *J. Exp. Zool.* **1**, 191-195 (1904).
44. Baguñà, J., Saló, E. & Auladell, C. Regeneration and pattern formation in planarians. III. Evidence that neoblasts are totipotent stem cells and the source of blastema cells. *Development* **107**, 77- 86 (1989).
45. Brøndsted, H. V. *Planarian Regeneration* (Pergamon Press, London, 1969).
46. Kerr, J. F., Wyllie, A. H. & Currie, A. R. Apoptosis: a basic biological phenomenon with wide-ranging implications in tissue kinetics. *Br J Cancer* **26**, 239-257 (1972).
47. Hall, P. A. & Watt, F. M. Stem cells: The generation and maintenance of cellular diversity. *Development* **106**, 619-633 (1989).
48. Heemels, M. T. Apoptosis. *Nature* **407**, 769 (2000).
49. Lillie, F. R. Some notes on regeneration and regulation in planarians. *Amer. Nat.* **34**, 173-177 (1900).
50. Baguñà, J. & Romero, R. Quantitative analysis of cell types during growth, degrowth and regeneration in the planarians Dugesia mediterranea and Dugesia tigrina. *Hydrobiologia* **84**, 181-194 (1981).
51. Baguñà, J. et al. in *Experimental Embryology in Aquatic Plants and Animals* (ed. Marthy, H.-J.) 129-162 (Plenum Press, New York, 1990).
52. Sulston, J. E. & Horvitz, H. R. Post-embryonic cell lineages of the nematode, *Caenorhabditis elegans. Dev. Biol.* **56**, 110-156 (1977).
53. Lin, H. The stem-cell niche theory: lessons from flies. *Nat Rev Genet* **3**, 931-40 (2002).
54. Tissenbaum, H. A. & Guarente, L. Model organisms as a guide to mammalian aging. *Dev Cell* **2**, 9-19 (2002).
55. Herndon, L. A. et al. Stochastic and genetic factors influence tissue-specific decline in ageing *C. elegans*. *Nature* **419**, 808-814 (2002).
56. Weismann, A. *The Germ-Plasm. A Theory of Heredity*. (Charles Scribner's Sons, New York, 1893).
57. Martínez, D. E. Mortality patterns suggest lack of senescence in hydra. *Experimental Gerontology* **33, No. 3**, 217 (1998).
58. ASA. Unpublished observations.
59. Child, C. M. *Patterns and Problems of Development* (University of Chicago Press, Chicago, 1941).
60. Braeckman, B. P., Houthoofd, K., De Vreese, A. & Vanfleteren, J. R. Assaying metabolic activity in ageing *Caenorhabditis elegans*. *Mech Ageing Dev* **123**, 105-119 (2002).

61. Bargmann, C. I. Neurobiology of the *Caenorhabditis elegans* genome. *Science* **282**, 2028-33. (1998).
62. Sokolowski, M. B. *Drosophila*: Genetics meets behaviour. *Nat Rev Genet* **2**, 879-890 (2001).
63. Panda, S., Hogenesch, J. B. & Kay, S. A. Circadian rhythms from flies to human. *Nature* **417**, 329-335 (2002).
64. Carew, T. J., Walters, E. T. & Kandel, E. R. Classical conditioning in a simple withdrawal reflex in *Aplysia californica*. *Journal of Neuroscience* **1**, 1426-1437 (1981).
65. Kandel, E. R. The molecular biology of memory storage: a dialogue between genes and synapses. *Science* **294**, 1030-1038 (2001).
66. Griffard, C. D. & Peirce, J. T. Conditioned discrimination in the planarian. *Science* **144**, 1472-1473 (1964).
67. Block, R. & McConnell, J. Classically conditioned discrimination in the planarian, Dugesia dorotocephala. *Nature* **215**, 1465-6 (1967).
68. McConnell, J. Learning in invertebrates. *Ann Rev Physiol* **28**, 107-136 (1966).
69. Aguinaldo, A. et al. Evidence for a clade of nematodes, arthropods and other moulting animals. *Nature* **387**, 489-93 (1997).
70. Lander, E. S. et al. Initial sequencing and analysis of the human genome. *Nature* **409**, 860-921 (2001).
71. Kyrpides, N. C. & Olsen, G. J. Archaeal and bacterial hyperthermophiles: horizontal gene exchange or common ancestry? *Trends Genet* **15**, 298-9. (1999).
72. Stanhope, M. J. et al. Phylogenetic analyses do not support horizontal gene transfers from bacteria to vertebrates. *Nature* **411**, 940-4. (2001).
73. Sánchez Alvarado, A. Regeneration in the Metazoans: Why does it happen? *BioEssays* **22**, 578-590 (2000).
74. Benfey, P. N. Stem cells: A tale of two kingdoms. *Curr Biol* **9**, R171-2. (1999).
75. Robertson, H. M. The large srh family of chemoreceptor genes in *Caenorhabditis* nematodes reveals processes of genome evolution involving large duplications and deletions and intron gains and losses. *Genome Res* **10**, 192-203 (2000).
76. Lau, N. C., Lim, L. P., Weinstein, E. G. & Bartel, D. P. An abundant class of tiny RNAs with probable regulatory roles in *Caenorhabditis elegans*. *Science* **294**, 858-62 (2001).
77. Lee, R. C. & Ambros, V. An extensive class of small RNAs in *Caenorhabditis elegans*. *Science* **294**, 862-4 (2001).
78. Jansen, G., Hazendonk, E., Thijssen, K. L. & Plasterk, R. H. Reverse genetics by chemical mutagenesis in *Caenorhabditis elegans*. *Nat Genet* **17**, 119-21 (1997).
79. Wicks, S. R., Yeh, R. T., Gish, W. R., Waterston, R. H. & Plasterk, R. H. Rapid gene mapping in *Caenorhabditis elegans* using a high density polymorphism map. *Nat Genet* **28**, 160-4 (2001).
80. Peichel, C. L. et al. The genetic architecture of divergence between threespine stickleback species. *Nature* **414**, 901-5 (2001).
81. Cooley, L., Kelley, R. & Spradling, A. Insertional mutagenesis of the *Drosophila* genome with single P elements. *Science* **239**, 1121-8 (1988).
82. Bessereau, J. L. et al. Mobilization of a *Drosophila* transposon in the *Caenorhabditis elegans* germ line. *Nature* **413**, 70-4 (2001).
83. ASA. Unpublished.
84. Umesono, Y., Watanabe, K. & Agata, K. A planarian *orthopedia* homolog is specifically expressed in the branch region of both the mature and regenerating brain. *Develop. Growth Differ.* **39**, 723-727 (1997).
85. Garcia-Fernandez, J. et al. High copy number of highly similar mariner-like transposons in planarian (Platyhelminthe): evidence for a trans-phyla horizontal transfer. *Mol Biol Evol* **12**, 421-31 (1995).
86. Jaffe, D. B. et al. Whole-genome sequence assembly for Mammalian genomes: arachne 2. *Genome Res* **13**, 91-96 (2003).
87. Batzoglou, S. et al. ARACHNE: a whole-genome shotgun assembler. *Genome Res* **12**, 177-89 (2002).
88. Mouse Genome Sequencing Consortium. Initial sequencing and comparative analysis of the mouse genome. *Nature* **420**, 520-562 (2002).