Sequencing the genome of threespine sticklebacks (*Gasterosteus aculeatus*)

David Kingsley, Ph.D.
Professor of Developmental Biology
HHMI and Stanford University
Stanford University School of Medicine
Stanford, CA 94305-5329
kingsley@cmgm.stanford.edu
(650) 725-5954

Introduction:

Rapid progress in genome biology has now provided nearly complete sequence coverage for several different organisms. Comparative analysis suggests many fundamental pathways and gene networks are conserved between organisms. And yet, the morphology, behavior, physiology, disease susceptibility, and lifespan of different organisms are obviously and profoundly different. What are the genetic mechanisms that control these differences between species? Have the unique characteristics of humans and other animals been created by changes in gene number, by alterations in the functional attributes of particular proteins, or by diversification of the regulatory mechanisms that control where and when genes are normally expressed? Comparative sequence of different genomes can enumerate the number of genes in different organisms, the types of protein-coding differences, and the divergence in noncoding DNA. Sequence analysis alone however, cannot tell us how which of the tens of millions of sequence changes seen between typical vertebrate genomes are actually responsible for the unique characteristics of different animals.

If different life forms could be crossed, it would be possible to use genome-wide linkage analysis to study both the number and locations of the genetic changes that create specific functional differences between the original starting parents. Early theoretical work suggested that evolution was likely to proceed by a very large number of genetic changes of very small effect [1]. If this were true, it would be very difficult to identify the number and location of DNA sequence changes that control evolutionary differences between groups. However, more recent studies suggest that adaptive evolution is likely to occur by a mixture of genetic changes, some of which account for a substantial fraction of the total variance in evolutionary traits [2]. In invertebrate systems, it has been possible to use genetic crosses to examine the genetic architecture of both morphological and physiological differences between different subspecies of plants and insects. Many of these studies have identified particular chromosome regions that control a substantial fraction of variation in specific traits [3]. In a few cases, it has been possible to identify specific genes responsible for particular differences [4-6].

Much less is known in higher animals, despite great interest in the mechanisms that underlie the unique characteristics that have evolved in the vertebrate lineage. Vertebrates provide some of the most dramatic examples of morphological and physiological change during evolution. Classic examples include the formation of distinctive forms of Darwin's finches on the Galapagos Islands, the generation of large species flocks of cichlid fish in the great lakes of East Africa, and the rise and diversification of many different mammalian groups in the Tertiary period, including humans. It has been proposed that the diversity of vertebrate groups may depend upon unusual features of the vertebrate genome, including diversification of gene function following successive rounds of either genome or chromosome segment duplication that have occurred in the vertebrate lineage [7,8]. However, formal genetic studies of species differences in vertebrates have been severely limited by the impracticality of crossing and raising large numbers of animals, and the absence of molecular genetic linkage maps for all but a small handful of laboratory and domesticated animals. A recent review of genetic studies of evolutionary differences in naturally occurring species identified more than 20 previous studies in plants and insects [3], but no systematic studies based on crosses between naturally occurring vertebrate species. As a result, even in the genome sequence era, we still do not know how many genetic changes are required to evolve new traits in vertebrates, which particular genes are involved, and whether changes in gene number, protein sequence, or regulatory information are responsible for specific traits that have evolved in different groups.

The threespine stickleback *Gasterosteus aculeatus* offers a unique biological opportunity for detailed study of the genomic and genetic basis of species differences in vertebrates. This small marine fish has undergone one of the most recent and dramatic adaptive radiations on earth [9]. Sticklebacks normally live in the ocean but migrate into freshwater streams and lakes every

spring to breed. At the end of the last Ice Age, widespread melting of glaciers led to dramatic changes in sea level and land elevation. As a result of this global climate change, tens of thousands of new freshwater lakes and streams were created in formerly ice-covered regions throughout the Northern hemisphere. Ocean sticklebacks colonized many of these newly created lakes and streams, and in many cases became isolated in new environments following the end of widespread melting and subsequent land elevation. These newly established populations have diverged over the course of only 10 to 15,000 years in response to different ecological conditions in each lake and stream, including large differences water temperature, depth, color, and salinity; food sources; predators; day length; and seasonal stability of different environments. Thousands of evolutionary experiments throughout the Northern Hemisphere have since given rise to new populations of sticklebacks with marked changes in body size, body shape, feeding specializations, size and pattern of skeletal structures, presence or absence of defensive armor, salinity tolerance, temperature preference, parasite resistance, lifespan, and behavior. These differences are as large as those normally seen between different species or genuses of animals, and the divergent stickleback types were originally classified as over 40 different species. Although many of the specialized forms are known be reproductively isolated even when in contact with each other (a formal definition of species), the reproductive barriers between forms are largely either behavior or mechanical. As a result, fully viable and fertile F1 and F2 hybrids between most forms can be generated using laboratory matings or in vitro fertilization. The ability to generate crosses between widely different forms provides an unprecedented opportunity to use formal genetic analysis to study the number and location of genomic that underlie evolution of natural species differences in a vertebrate system [9].

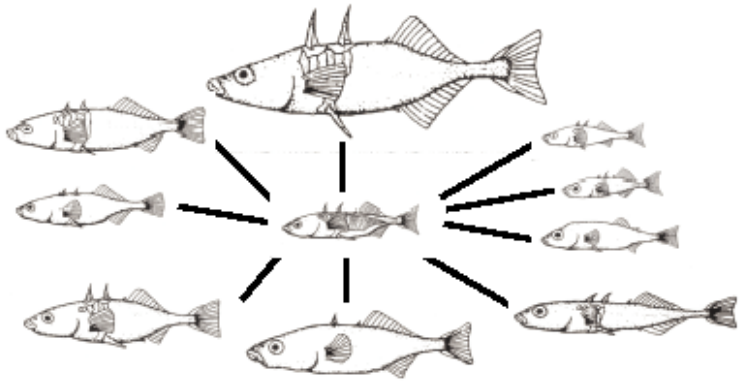

Figure 1: Morphological diversity of sticklebacks. Very different looking fish have radiated from marine ancestors (center). Note alterations in body size, feeding morphology, number of dorsal spines (zero to 3), pattern and number of lateral plates (0 to 38), and pelvic fin development (completely absent in some populations). Many of the derived freshwater populations are found in postglacial lakes and streams that formed only 10,000-15,000 years ago. (Modified from Bell and Foster, 1994).

Stickleback biology has already been extensively characterized. The abundance, wide distribution, ease of collecting, and interesting behavioral and physical characteristics of sticklebacks have made them a favorite research organism for most of the last century. Tinbergen's pioneering work on the reproductive behavior of male and female sticklebacks formed an important basis for his Nobel Prize in Physiology and Medicine in 1973. The diverse stickleback research community has produced one of the largest research literatures for any vertebrate model organism, including more than 2,000 research papers, and several full-length textbooks on the morphology, distribution, life-history, and ecology of different stickleback populations [9-13].

Sticklebacks have also attracted great interest from geneticists interested in the biological basis of species differences in vertebrates. Several groups have carried out genetic crosses between fish from different locations [14-20]. These experiments show that many of the important evolutionary differences between fish populations are likely to be controlled by relatively small numbers of genes, providing a unique opportunity to study the molecular basis of evolutionary divergence in vertebrates. A large collection of microsatellite markers has recently been

developed for the fish, and assembled into a genome-wide linkage map for sticklebacks [18]. Using this map, specific chromosome regions have been mapped that control dramatic changes in morphological characters between fish populations [18-22]. Recently the molecular basis of one of these morphological traits has been traced to cis-acting regulatory changes in a specific developmental regulator [20]. These studies have shown that major morphological transformations in the vertebrate skeleton may be controlled by a relatively small number of genes, and have also identified the type of gene and the type of mutation (coding or regulatory) that have led to morphological divergence. A similar combination of genetic and genomic studies can now be applied to many of the other dramatic differences between stickleback populations. These studies should ultimately make it possible to answer some of the most longstanding and fundamental questions about the type of DNA sequence alterations that underlie the appearance of new characters in natural populations, and whether evolution is constrained to use particular mechanisms over and over again to reach a particular state in independent populations around the world.

A. *Specific biological rationales for the new sequence data*

A1. Utility of the stickleback system for studying complex traits in natural populations

Sticklebacks offer unique experimental advantages for genetic studies of species differences and quantitative traits in naturally occurring organisms. The extreme recency of the stickleback radiation in post-glacial lakes greatly increases the chances that the underlying genetic bases of many traits are relatively simple. Ten thousand years since the end of the last Ice Age corresponds to approximately 10,000 generations of stickleback breeding. This is orders of magnitude less than the tens of millions of generations that separate most mammal groups. Importantly, the number of breeding generations underlying the stickleback radiation is roughly comparable to the estimated 10,000 generations of human breeding that have occurred since the most recent common ancestor of human mtDNA originated in Africa [23]. Further study of the adaptive radiation of sticklebacks may thus reveal genetic mechanisms that are broadly relevant to the evolution of complex traits and physiological differences in other groups that have recently migrated and adapted to many different environments, including humans. Unlike most mammals, sticklebacks are abundant, easy to collect in the wild, and easy to raise in large numbers in the laboratory. Their small size, relatively rapid generation time (six months in the laboratory), and large clutch size makes it possible to raise thousands of progeny for high-resolution genetic mapping. This is a particularly important advantage when studying the genetic basis of complex traits, since it facilitates the recovery of offspring carrying different combinations of all the genes that may be contributing to a particular character.

A2, A5: Utility of sticklebacks for increasing our understanding of basic developmental, physiological, and neurological processes relevant to human health. Many of the differences that have evolved in different stickleback populations involve the same physiological systems that play an important role in human health and disease. For example, the process of bone formation varies tremendously in different stickleback groups, including large variations in jaw size and total tooth number, normal development or complete absence of the pelvic hindfin, and up to 35 fold differences in number of bony plates along the side of the fish [9]. Understanding the genetic basis of variation in bone formation in natural populations is likely to reveal fundamental new information relevant to skeletal diseases in humans, including limb and tooth abnormalities, variation in total bone density, and susceptibility to diseases like osteoporosis.

The colonization of different environments around the world has led to profound physiological changes in different fish populations. Sticklebacks have adapted to water with salinities from fresh to marine, temperatures from 4 to 30 degrees; and pHs from 3 to 10 [9]. Fish from different environments show marked preference for particular temperatures [24], and can either migrate freely between oceans and streams, or have completely lost their ability to survive in high salt environments [25-27]. Further study of such naturally occurring differences should inform studies of

general metabolism and salt handling physiology in other animals, including genetic differences in metabolism, obesity, and susceptibility to hypertension in humans.

Stickleback populations show marked differences in susceptibility to parasites, with some populations having evolved resistance to particular parasites that are common in their immediate environment[28,29]. Active research in this area has established simple protocols for exposure of sticklebacks and monitoring the physiological impact of parasite infection [30,31]. Genetic crosses between resistant and susceptible populations, in combination with further molecular studies of immune functions in the fish [28], will help elucidate fundamental mechanisms of host-parasite interactions, and the genetic basis of achieving resistance to widespread pathogens.

Sticklebacks also differ greatly in total lifespan, with some populations characteristically surviving for a single year in the wild, and some populations living up to six times longer [32,33]. Further studies of short-lived and long-lived populations may provide important new insights into the genetic and physiological basis of aging in many different animals, including humans.

Finally, the biological basis of instinct and complex behavior is one of the most challenging frontiers of current biomedical research. Sticklebacks have long been one of the major model systems for studying complex behavior in vertebrates. The Nobel-Prize wining work of Niko Tinbergen has provided some of our most fundamental ideas on how complex behavior patterns can emerge from a series of separate motivational states, visual stimuli, and responses to environmental cues [34]. The robust and interesting behaviors shown by different sticklebacks, and the ease of studying them in the laboratory, has led to extensive characterization of different behavioral traits in stickleback populations around the world. Fish from different locations reproducibly orient to different cues [25,35], build different kinds of nests on different substrates [25,36], prefer different kinds of mates [37], court mates in different ways [38], explore or hide when presented with novel environments, have instinctive fear or disregard for particular predators[39,40]; vary in aggressiveness [41]; and show an extended period of nest building and parental care for offspring, or scatter and abandon eggs immediately after fertilization[42]. Further genetic and molecular studies of these profound behavioral differences should provide new insights into fundamental mechanisms of sensory perception, signal processing, and motivational states in vertebrates. This research is broadly relevant to many important areas of human behavior and human health policy, including aggression, novelty seeking and drug abuse, reproductive behavior, and parental investment or abandonment of offspring.

A3. Utility of the threespine stickleback for informing the human sequence, and identifying the functions of specific sequence features in the human genome. Genomics provides an embarrassment of riches for thinking about the possible molecular basis of unique traits in humans and other animals, including changes in total gene number, expansion and contraction of specific gene families, changes in amino acid sequences of proteins, changes in splicing patterns, and changes in regulatory information that controls where and when genes are expressed. However, we currently do not know which of the tens to hundreds of millions of sequence differences between humans and other animals are actually responsible for the unique adaptations of humans or other groups. Even in an era where complete sequences are being generated for many different animals, we may never be able to identify the actual base pair differences responsible for particular traits, if we do not have additional methods to sort through the enormous number of sequence changes seen in any large-scale genome project.

Sticklebacks provide a unique opportunity to cross different populations, raise large number of progeny, and use genome-wide linkage mapping to determine the number and location of the chromosome regions that have the greatest functional impact on any trait being studied. This in turn, will make it possible to answer several of the most fundamental and longstanding questions in genomics and genetics of higher animals, including:

1) How many separate chromosome regions are required to produce substantial phenotypic change in natural populations?
2) What types of genes underlie specific morphological, physiological, and behavioral differences?
3) What kind of DNA sequence changes lead to the appearance of unique traits (coding or regulatory alterations, point mutations versus larger rearrangements, rare new mutations or selection on standing variation already present in populations)?

Because the genomic and genetic tools in sticklebacks can be applied to a large range of different traits, we expect the research on this organism to reveal important general principles that are applicable to evolution of many different characters in vertebrates, including biomedically relevant traits in humans.

A4. Providing a better connection between the sequences of non-human organisms and the human sequence. The primary rationale for sequencing sticklebacks is their unique experimental utility for studying the genetic and genomic basis of complex traits in natural populations, not their unique phylogenetic position. However, just as sequence from a range of different mammals has helped define islands of sequence conservation that correspond to coding and regulatory regions [43,44], the stickleback sequence should help annotate important genomic sequences conserved both among different fish groups, and between fish and mammals. Sticklebacks are spiny-rayed teleosts whose last common ancestors with pufferfish and zebrafish are thought to have lived at least 55 million and 125 million years ago, respectively [45]. Comparisons between stickleback and zebrafish genome sequence thus cover approximately the same phylogenetic distance as comparisons between mice and man, and should have comparable utility for identifying key coding and regulatory regions in teleost fish, (one of the most diverse and economically important groups of vertebrates, including one of the major model systems for vertebrate biomedical genetics). The genome size of sticklebacks is intermediate between pufferfish and zebrafish, (about twice that of fugu, and one half that of zebrafish) [46]. Comparison of all three genomes will help identify the detailed mechanisms of expansion and contraction of genome size during vertebrate evolution, and the reproducibility of gene duplication and gene loss pathways in separate lineages [8,47].

A6. Facilitating the development of new monitoring systems for environmental contaminants relevant to human health and disease. Sticklebacks are widely distributed in lakes and streams throughout industrial countries. As ubiquitous native organisms that occur in natural environments near most industrialized centers, they represent an ideal sentinel organism for monitoring the possible presence of environmental contaminants of concern to human health and disease [48]. Previous work has already shown that female sticklebacks begin to produce the male glue protein spiggin when exposed to compounds with androgen activity. Enzyme-linked immunosorbent assay for spiggin production in female sticklebacks is currently one of the only available quantifiable bioassays for detecting the presence of androgenic compounds in natural environments [49]. Development of similar assays for other environmental contaminants is currently limited by the years required to clone interesting biomarkers de novo from sticklebacks. A stickleback genome sequence would make it possible to screen for many other genes whose expression changes in response to other environmental contaminants, facilitating the further development of the fish as a native sentinel organism for environmental monitoring in many different countries**.**

A7. Enhancing the ability to carry out positional cloning experiments. Previous studies have developed a genome-wide linkage map of threespine sticklebacks. Using this map, researchers have already identified major chromosome regions that control many of the interesting phenotypic differences between natural populations of sticklebacks [18-22]. One of the most important conclusions from these studies is that real evolutionary differences between natural populations can indeed be mapped to particular chromosome regions. In the best cases, it has

even been possible to move from these initial candidate regions to the actual genes responsible for major morphological differences in stickleback populations, and to study whether evolution proceeds by changes in coding or regulatory regions of these genes [20].

Although exciting recent progress has clearly show the potential of the system for identifying the genetic and genomic basis of evolutionary change, further progress is being slowed by the laborious process of cloning each candidate region, screening for candidate genes, and sequencing in many different intervals. The situation is directly analogous to the state of human genetics prior to the human genome project, when individual labs spent years developing the local genomic resources required to clone a single trait, and had to repeat this process each time they began studies in a new area. Development of a stickleback genome sequence will make it possible for the research community to identify candidate genes quickly for many different traits, and to evaluate those candidates rapidly be comparing sequence and expression in different populations. Whole genome sequence will also make it possible to apply entirely new approaches to the study of stickleback genetics, including the use of linkage disequilibrium in natural populations. Such methods are likely to be particularly powerful given the structure and history of isolated stickleback populations, and will further expand the number of interesting differences that can be traced to particular chromosome regions, genes, and DNA sequence changes.

A8. Expanding our understanding of evolutionary processes in general. We have already touched on the important role that sticklebacks can play in determining the number and type of genomic changes required to evolve new traits in natural populations. One of the most important features of this system is that sticklebacks also offer an opportunity to study whether evolution is constrained to use particular mechanisms over and over again to reach a particular endpoint. The thousands of different glacial lakes and streams in the Northern Hemisphere represent thousands of independent evolutionary experiments. Previous mitochondrial sequencing studies have shown that sticklebacks in different lakes have unique mitochondrial haplotypes that are independent of the degree of morphological similarity between the fish [50]. These results suggest that many of the characteristic stickleback morphologies have evolved repeatedly in different locations, presumably in response to similar ecological conditions occurring at different sites (deep water lakes, shallow water lakes, particular constellations of predators and food sources, etc).

Genetic mapping and genetic complementation crosses between different stickleback populations provides and elegant method to test whether evolution uses the same or different genes to evolve similar characters in different populations. Complementation crosses suggest that the same major locus controls bony patterning in several different populations [15,16,19,21]. In addition, a major locus controlling pelvic hindfin reduction maps to the same major chromosome region in pelvic-reduced fish from both Alaska and British Columbia [19,20]. These intriguing results need to be extended to many other traits. However, they already suggest that evolution may not have hundreds of ways to reach a given endpoint. Instead, particular genes and mechanisms appear to be used repeatedly during parallel evolution, suggesting the presence of important developmental, genomic, or environmental constraints on how natural populations evolve under selection in the wild.

We currently have no idea why evolution may be constrained to use particular genetic mechanisms to evolve a particular character. Particular loci may be more or less prone to mutation or rearrangement, have regulatory or coding sequences that are more amenable to specific changes, have different degrees of specificity or pleiotropy when mutations do occur, have larger or smaller phenotypic effects under selection, or may differ in their natural population frequency prior to selection. Cloning and identification of the loci responsible for repeated evolution of stickleback populations should help answer these fundamental questions,

and provide new insight into the genomic and developmental constraints that underlie the appearance of new characters during vertebrate evolution.

B. *Strategic issues in acquiring new sequence data:*

B1. The demand for the new sequence data. What is the size of the research community that will use it? What is the community's enthusiasm for having the sequence? Sticklebacks are one of the most extensively studied non-mammalian vertebrate model organisms. Decades of previous work have characterized all aspects of the biology of the fish, from ecology, morphology, physiology, endocrinology, parasitology, toxicology, behavior, development, and evolution [9-11]. More than a hundred labs in 22 countries around the world have published primary research papers on sticklebacks in the last 10 years.

Many new labs have also been attracted to this area recently, including researchers with extensive experience in other major vertebrate model organisms, including classical mouse genetics (David Kingsley, Katie Peichel; Stanford University); vertebrate embryology (Cheryl Tickle); and zebrafish genetics and development (Charles Kimmel, John Postlethwaite; Eugene OR). With widespread support of reviewers working in all the major vertebrate and invertebrate model organisms, NIH recently established a new Center of Excellence in Genomic Science (CEGS) at Stanford, dedicated in large part to developing additional genomic resources (markers, BAC libraries, linkage maps) that could be used for further study of the stickleback system. To facilitate further spread of molecular methods throughout the field, an intensive summer laboratory course in stickleback molecular genetics is also now being taught each summer as Stanford, modeled after the longstanding courses in phage, yeast, and zebrafish genetics taught at Cold Spring Harbor and Woods Hole (see http://cegs.stanford.edu). The stickleback course has already begun attracting additional researchers in the field, including graduate students, postdocs, and principal investigators who are moving from C. *elegans*, *Drosophila*, or zebrafish studies into stickleback research.

International stickleback meetings are scheduled every 3 to 4 years to stimulate further exchange of information between labs. At the Fourth International Stickleback Conference held in summer of 2003, a general discussion was held to assess interest in a stickleback genome project. The entire audience of 61 researchers from 13 countries unanimously endorsed the goal of obtaining the stickleback genome sequence, because of the crucial role this sequence could play in accelerating research projects in many different fields.

B2. The suitability of the organism for experimentation. Sticklebacks are abundant, hardy, easy to collect in the wild with minnow traps or seine nets, and easy to grow in inexpensive aquaria setups in the laboratory. They typically come into breeding condition once a year in the wild, but can also be brought into reproductive condition artificially in the lab by appropriate manipulation of temperature and light cycle. Under optimal laboratory conditions, the generation time from fertile adult to fertile adult is approximately 6 months [9].

No inbred strains are available, in part because the emphasis in the field has always been on the study of natural variation present in wild populations, rather than on unusual characters that may develop with continued artificial selection and inbreeding in the laboratory. However, the wild populations provide an incredible diversity of morphological, physiological, and behavioral traits that have evolved in different environments. Many of these traits are maintained when fish from different locations are raised in the laboratory, and genetic experiments have shown that the differences can be mapped to particular chromosome regions. The thousands of papers already written on unique characteristics of different fish around the world make it possible for researchers to pick almost any character of interest, and quickly find populations appropriate for further genetic studies.

Sticklebacks normally mate using external fertilization, and the young develop outside mothers in nests tended by male sticklebacks. External development and the large transparent stickleback embryo facilitate the collecting of material for developmental studies. Simple aerated beakers can substitute for nests in the laboratory, making it simple to monitor the progress of clutches and crosses.

Although many of the natural populations show phenotypic differences as large as those between different genuses of fish, sticklebacks from different locations can be easily crossed using artificial fertilization. Methods for stripping eggs and combining them with sperm for crosses are easily performed in both the field and laboratory, facilitating the genetic analysis of many different population differences. Clutch sizes are large (100 to 200 eggs), and pairs of fish can be mated repeatedly, making it possible to recover very large families for genetic mapping and QTL analysis [18-20].

Finally, techniques have recently been developed for both gene transfer and gene inactivation experiments in sticklebacks. Foreign DNA can be injected into fertilized fish eggs, resulting in appropriate tissue specific expression in transgenic animals, and germ-line transmission of the introduced constructs [51]. These methods make it possible to study the functions of both coding and regulatory information in genomic regions of interest. Gene transfer methods will also make it possible to measure the phenotypic impact of transferring particular genes between populations, and to test whether particular characters can be transferred by introducing specific genes. Morpholinos can also be injected into fertilized stickleback eggs, making it possible to reduce or inactivate the function of a given target gene during early development [52]. This method has already been used to study the role of duplicate genes in stickleback hindbrain development [53]. Because stickleback eggs are large, all of the methods required for transgenic and morpholino experiments can be carried out with relatively simple injection equipment and dissecting microscopes. At the Stanford Summer course on Stickleback molecular genetics last summer, groups with no previous experience in embryology or microinjection were able to make transgenic sticklebacks on their first day of injecting.

B3. Rationale for the complete sequence of the organism. Previous studies have already shown the utility of sticklebacks for genetic analysis of many different complex traits that have evolved in natural populations. This work has also shown that the chromosome regions controlling different traits are widely distributed among almost all stickleback linkage groups [18-22,54]. A genome-wide effort will facilitate molecular analysis of many different traits simultaneously. Given the large number of different morphological, physiological, and behavioral differences that can be studied in the fish, a genome-wide sequence is the most efficient way to facilitate the molecular analysis of many different characters.

Whole genome information is particularly important for this effort, rather than a concentration only on ESTs or coding regions. The ultimate goal of the research is to determine the detailed type of genomic changes that underlie evolution of new characters in natural populations. To evaluate the contribution of both coding and regulatory information, it is essential to have both types of regions available for detailed analysis. This has already become a rate-limiting step in current studies. For example, recent work has shown that a major morphological transformation in the stickleback skeleton (presence or absence of hindfins) is controlled in large part by genetic changes in a known developmental control gene [20]. Null mutations in that gene cause lethality in experimental animals, due to the pleiotropic role of the regulator in multiple tissues. In sticklebacks, the coding region of the gene is identical in different populations. However, some populations have a cis-acting change in gene regulation that alters the expression pattern in some tissues but not others. This type of regulatory alteration may be a common way of avoiding the detrimental effects that would accompany mutations in key developmental regulators [55-57]. However, such mutations are very difficult to study in the absence of detailed genomic information for characterizing the cis-acting regulatory sequences that surround a gene of

interest. Stickleback genome sequence will make it possible to identify the actual site and nature of such changes, and to compare how the same gene may be hit in different ways when similar traits evolve independently in different lakes and streams around the world [19,58].

B4. Size of the genome, sequence characteristics, and choice of populations for sequencing. The 675 megabase stickleback genome [46] is quite compact for vertebrates, less than one quarter the size of a typical mammalian genome. Sticklebacks have 21 cytologically visible chromosomes[59]. Although no sexually dimorphic chromosomes are known, genetic studies suggest the presence of an evolving X-Y chromosome system, with males as the heterogametic sex [54]. Sample sequencing has been carried out on 21 stickleback BAC clones during the course of preliminary genome characterization projects at Stanford, resulting in a total of 4.19 Mb of sequence (2.54 Mb of which is currently finished) [60]. Based on this initial analysis of many different widely distributed chromosome regions, (including 4 clones that map to the evolving sex chromosomes), sticklebacks have an average GC content of 42% (range 41.6 to 48.5%), a simple sequence content of 1.75% (range 0.5 to 6.8%), and a low complexity sequence content of 2.63% (range 0.8 to 8.7%). The percentage of low complexity sequence in sticklebacks is approximately half that seen in a comparable sample of zebrafish BAC clones (based on 3.94 Mb of zebrafish sequence analyzed by identical methods at Stanford). The higher average GC content of sticklebacks, and the lower percentage of simple sequence content, has made genomic sequence from stickleback clones substantially easier to assemble than zebrafish clones in our experience.

The choice of populations to sequence is driven by a combination of the evolutionary history of the fish, and the specific populations for which genetic and genomic resources are already readily available. Most existing freshwater populations are thought to be derived from anadromous populations that migrate each year between salt and fresh water [9]. Thus a sequence assembly from a representative anadromous population should be obtained to sample the presumed ancestral state for many different traits. Conversely, many freshwater populations have evolved a similar set of phenotypic characteristics over and over again, including loss of armor plates, reduction of pelvic and dorsal spines, increased body depth, and changes in jaws, teeth and gill raker as fish adapt to the different food sources and predators present in near-shore, freshwater environments. A second sequence obtained from one of the highly divergent freshwater stickleback populations would make it possible to carry out a genome-wide comparison of the sequence changes that underlie evolution of many different characters that have evolved repeatedly during adaptive radiation of sticklebacks.

We recommend that a representative anadromous sequence be obtained from a single male fish from the Little Campbell River, British Columbia. This population of migratory fish was the basis of an extensive classical study characterizing a whole range of morphological, physiological, and behavioral differences between anadromous and freshwater animals[25]. The same ocean going population has also been used to set up multiple recent crosses between ocean and freshwater derived forms. The results of these genetic mapping studies will make it possible to compare genome sequence information with the actual location of particular chromosome regions that are most important for controlling specific characters. Several fish from the Little Campbell River were trapped this summer and used to make high molecular weight DNA preps from single individual males. These preparations have been screened by Pieter DeJong's lab at Children's Health Organization Research Institute in Oakland, California, and the one with highest yield is currently being used to construct a high coverage BAC library from a single individual. The original fish samples have also been saved, and the remaining tissues are available for constructing additional libraries for DNA shotgun analysis.

We recommend that sequence from a representative freshwater derived population be obtained from a single male fish of the benthic population from Paxton Lake, British Columbia. The Paxton benthic population shows simultaneous development of many typical characteristics of

derived freshwater fish. It has also already been the subject of extensive ecological and genetic studies [17,61-64], including the generation of more than 2600 F2 fish for detailed genetic mapping of chromosome regions that underlie evolutionary divergence [20]. In addition, Pieter DeJong's group has already generated high coverage BAC libraries from single Paxton Benthic individuals [65]. One male was used to generate an approximately 10X coverage BAC library using EcoR1 partial digests, and a second male was used to generate a 10X coverage BAC library using MboI partial digests. Both libraries, have large average insert sizes (155 to 165 kb), and are currently available for public distribution from BACPAC resources at Oakland. We preserved the original fish used to prepare these libraries, and additional tissue is therefore available for making shotgun-sequencing libraries from the same individual.

The optimal strategy for whole genome sequencing should be worked out in consultation with the major genome centers, and should be guided by their extensive previous experience with sequencing and assembly of medium-sized genomes. The primary biological goal should be to obtain a high quality long-range assembly covering essentially the complete stickleback genome (95% to 100% sequence coverage). The sequencing depth and quality should be chosen to achieve an average error rate of approximately 1 in 10,000 base pairs. Preliminary studies of several genomic regions suggest that the average polymorphism rate in stickleback populations is approximately 1 in 500 base pairs [60]. By setting the genome sequencing depth to obtain an error rate approximately 1/20th of the polymorphism rate, 95% of detected changes in the final sequence assembly should correspond to real polymorphisms or causative mutations, rather than sequencing errors. That level of accuracy will be sufficient to identify large number of SNPs for genetic mapping and linkage disequilibrium studies. It will also make it possible to catalog most the candidate DNA sequence changes that occur in the genomic intervals known to control evolutionary divergence between a representative ancestral and derived form, and to design functional tests to decide which of the sequence changes are responsible for particular traits.

B5. Other projects to support and enhance a stickleback genome project. The Stanford CEGS grant will continue to work in parallel to develop high quality genetic and physical maps of the stickleback genome. The genetic and physical maps will make it possible to integrate the stickleback whole-genome sequence information with higher order mapping information needed to identify the genetic and molecular basis of evolutionary traits in different populations. The center has already collaborated with Marco Marra's group in Vancouver to develop a first generation physical map of the stickleback genome, based on fingerprints of 100,000 clones from an initial BAC library[65] (see http://cegs.stanford.edu/Physical_map.jsp and http://www.bcgsc.ca/lab/mapping/data). Because this library was made from a large number of pooled individuals, the Stanford Center will work with the Vancouver group to develop similar BAC fingerprint physical maps based on the new BAC libraries prepared from the individual Paxton Benthic and Little Campbell males chosen for DNA sequencing. This physical mapping effort, in combination with end sequence reads from the individual BAC clones in the libraries, will make it possible to tie stickleback whole genome shotgun sequence assemblies to larger clone contigs. Microsatellite and SNP markers within all of the largest clone contigs will also be genetically mapped on a stickleback linkage cross of 94 F2 animals, establishing the position of all the major sequence assemblies on the genetic linkage map of sticklebacks. These physical and genetic mapping efforts projects will help identify large-scale synteny relationships between stickleback and other vertebrate genomes. They will also make it possible to move rapidly from genetic mapping studies of many different traits, to relevant BAC clone assemblies, to sequence contigs, and to individual sequence differences between marine and freshwater populations. This type of combined genetic and genomics effort will provide the tools needed for a detailed functional study of the number and type of mutations that control evolution of new traits in natural populations. We expect the results of this work to be a fundamental advance in our understanding of how new traits evolve in vertebrates, and how the unique morphological, physiological, and behavioral characteristics of different animals are actually encoded in the genome.

References

1. Fisher, R. A. *The genetical theory of natural selection* (Oxford University Press, Oxford, 1930).
2. Orr, H. A. & Coyne, J. A. The genetics of adaptation - a reassessment. *American Naturalist* **140**, 725-742 (1992).
3. Orr, H. A. The genetics of species differences. *Trends in Ecology & Evolution* **16**, 343-350 (2001).
4. Stern, D. L. A role of Ultrabithorax in morphological differences between Drosophila species. *Nature* **396**, 463-6 (1998).
5. Sucena, E. & Stern, D. L. Divergence of larval morphology between Drosophila sechellia and its sibling species caused by cis-regulatory evolution of ovo/shaven-baby. *Proc Natl Acad Sci U S A* **97**, 4530-4 (2000).
6. Kopp, A., Duncan, I., Godt, D. & Carroll, S. B. Genetic control and evolution of sexually dimorphic characters in Drosophila. *Nature* **408**, 553-9 (2000).
7. Holland, P. W., Garcia-Fernandez, J., Williams, N. A. & Sidow, A. Gene duplications and the origins of vertebrate development. *Dev Suppl*, 125-33 (1994).
8. Amores, A., Force, A., Yan, Y. L., Joly, L., Amemiya, C., Fritz, A., Ho, R. K., Langeland, J., Prince, V., Wang, Y. L., Westerfield, M., Ekker, M. & Postlethwait, J. H. Zebrafish hox clusters and vertebrate genome evolution. *Science* **282**, 1711-4 (1998).
9. Bell, M. A. & Foster, S. A. (eds.) *The evolutionary biology of the threespine stickleback* (Oxford University Press, Oxford, 1994).
10. Wootton, R. J. *The biology of the sticklebacks* (Academic Press, London, 1976).
11. Wootton, R. J. *A functional biology of sticklebacks* (University of California Press, Berkeley and Los Angeles, 1984).
12. Coad, B. W. A bibliography of sticklebacks. *Syllogeus* **35**, 1-42 (1981).
13. Keivany, Y. http://www.geocities.com/CapeCanaveral/Hall/1345/stickbibl.html. (2002).
14. Münzing, J. Biologie, variabilitat und genetik von *Gasterosteus aculeatus* L. (Pisces) untersuchugen im elbegebiet. *Int Rev Ges Hydrobiol* **44**, 317-381 (1959).
15. Hagen, D. W. & Gilberts, L. G. Genetics of plate morphs in freshwater threespine sticklebacks. *Heredity* **31**, 75-84 (1973).
16. Avise, J. C. Genetics of plate morphology in an unusual population of threespine sticklebacks (*Gasterosteus aculeatus*). *Genetical Research* **27**, 33-46 (1976).
17. Hatfield, T. Genetic divergence in adaptive characters between sympatric species of stickleback. *American Naturalist* **149**, 1009-1030 (1997).
18. Peichel, C. L., Nereng, K. S., Ohgi, K. A., Cole, B. L., Colosimo, P. F., Buerkle, C. A., Schluter, D. & Kingsley, D. M. The genetic architecture of divergence between threespine stickleback species. *Nature* **414**, 901-5 (2001).
19. Cresko, W. A., Amores, A., Wilson, C., Murphy, J., Currey, M., Phillips, P., Bell, M. A., Kimmel, C. & Postlethwait, J. H. The genetic basis of recurrent evolution: Armor loss in Alaskan populations of threespine stickleback, *Gasterosteus aculeatus*. (submitted).
20. Shapiro, M. D., Marks, M. E., Peichel, C. L., Blackman, B. K., Nereng, K. S., Schluter, D. & Kingsley, D. M. Genetic and developmental basis of evolutionary pelvic reduction in threespine sticklebacks. (submitted) (2003).
21. Colosimo, P. F., Peichel, C. L., Hosemann, K. E., Nereng, K. S., Blackman, B. K., Shapiro, M. D., Schluter, D. & Kingsley, D. M. Molecular genetics of lateral plate

morphs in threespine sticklebacks. *Fourth International Conference on Stickleback Behavior and Evolution (*Stromstad, Sweden, 2003)
22. Miller, C., Peichel, C. L., Colosimo, P. F., Shapiro, M. D., Blackman, B. K., Nereng, K. S., Schluter, D. & Kingsley, D. M. Genetic and developmental basis of phayngeal divergence in sticklebacks. *Fourth International Conference on Stickleback Behaviour and Evolution (*Stromstad, Sweden, 2003)
23. Ingman, M., Kaessmann, H., Paabo, S. & Gyllensten, U. Mitochondrial genome variation and the origin of modern humans. *Nature* **408**, 708-13 (2000).
24. Roed, K. H. Temperature preference of the 3-spined stickleback, *Gasterosteus aculeatus* L. (Pisces), collected at different seasons. *Sarsia* **64**, 137-141 (1979).
25. Hagen, D. W. Isolating mechanism in threespine sticklebacks (Gasterosteus). *Journal of the Fisheries Research Board of Canada* **24**, 1637-& (1967).
26. McPhail, J. D. Predation and the evolution of a stickleback (*Gasterosteus*). *J. Fish. Res. Bd Can* **26**, 3183-3208 (1969).
27. Guderly, H. E. in *The Evolutionary Biology of the Threespine Stickleback* (eds. Bell, M. A. & Foster, S. A.) 85-113 (Oxford University Press, New York, 1994).
28. Wegner, K. M., Reusch, T. B. H. & Kalbe, M. Multiple parasites are driving major histocompatibility complex polymorphism in the wild. *Journal of Evolutionary Biology* **16**, 224-232 (2003).
29. Kalbe, M., Wegner, K. M., Reusch, T. B. H. & Kurtz, J. Habitat-specific parasite resistance in three-spined sticklebacks. *Fourth International Conference on Stickleback Behaviour and Evolution (*Stromstad, Sweden, 2003)
30. Barber, I. A non-invasive morphometric technique for estimating cestode plerocercoid burden in small freshwater fish. *Journal of Fish Biology* **51**, 654-658 (1997).
31. Arnott, S. A., Barber, I. & Huntingford, F. A. Parasite-associated growth enhancement in a fish-cestode system. *Proceedings of the Royal Society of London Series B-Biological Sciences* **267**, 657-663 (2000).
32. Reimchen, T. E. Extended longevity in a large-bodied stickleback, Gasterosteus, population. *Canadian Field-Naturalist* **106**, 122-125 (1992).
33. Baker, J. A. in *The Evolutionary Biology of the Threespine Stickleback* (eds. Bell, M. A. & Foster, S. A.) 144-187 (Oxford University Press, New York, 1994).
34. Pickren, W. E. An elusive honor - psychology, behavior, and the Nobel Prize. *American Psychologist* **58**, 721-722 (2003).
35. Odling-Smee, L. & Braithwaite, V. A. The influence of habitat stability on landmark use during spatial learning in the three-spined stickleback. *Animal Behaviour* **65**, 701-707 (2003).
36. Macdonald, J. F., Bekkers, J., Macisaac, S. M. & Blouw, D. M. Intertidal breeding and aerial development of embryos of a stickleback fish (Gasterosteus). *Behaviour* **132**, 1183-1206 (1995).
37. Boughman, J. W. Divergent sexual selection enhances reproductive isolation in sticklebacks. *Nature* **411**, 944-8 (2001).
38. Kitano, J., Mori, S. & Peichel, C. L. Genetic basis for variation in male courtship behaviours of threespine sticklebacks. *Fourth International Conference on Stickleback Behaviour and Evolution (*Stromstad, Sweden, 2003)

39. Giles, N. & Huntingford, F. A. Predation risk and inter-population variation in antipredator behavior in the 3-spined stickleback, *Gasterosteus aculeatus* L. *Animal Behaviour* **32**, 264-275 (1984).
40. Huntingford, F. A. & Wright, P. J. Inherited population differences in avoidance-conditioning in 3-spined sticklebacks, *Gasterosteus aculeatus*. *Behaviour* **122**, 264-273 (1992).
41. Huntingford, F. A. Further evidence for an association between lateral scute number and aggressiveness in the threespine stickleback, *Gasterosteus aculeatus*. *Copeia*, 717-720 (1981).
42. Blouw, D. M. Evolution of offspring desertion in a stickleback fish. *Ecoscience* **3**, 18-24 (1996).
43. Cooper, G. M., Brudno, M., Green, E. D., Batzoglou, S. & Sidow, A. Quantitative estimates of sequence divergence for comparative analyses of mammalian genomes. *Genome Res* **13**, 813-20 (2003).
44. Thomas, J. W., Touchman, J. W., Blakesley, R. W., Bouffard, G. G., Beckstrom-Sternberg, S. M., Margulies, E. H. et al. Comparative analyses of multi-species sequences from targeted genomic regions. *Nature* **424**, 788-93 (2003).
45. Lauder, G. V. & Liem, K. F. The evolution and interrelationships of the actinopterygian fishes. *Bull. Mus. Comp. Zool.* **150**, 95-197 (1983).
46. Hinegard, R. Evolution of cellular DNA content in teleost fishes. *American Naturalist* **102**, 517-523 (1968).
47. Force, A., Lynch, M., Pickett, F. B., Amores, A., Yan, Y. L. & Postlethwait, J. Preservation of duplicate genes by complementary, degenerative mutations. *Genetics* **151**, 1531-45 (1999).
48. Katsiadaki, I., Scott, A. P. & Mayer, I. The potential of the three-spined stickleback (Gasterosteus aculeatus L.) as a combined biomarker for oestrogens and androgens in European waters. *Marine Environmental Research* **54**, 725-728 (2002).
49. Katsiadaki, I., Scott, A. P., Hurst, M. R., Matthiessen, P. & Mayer, I. Detection of environmental androgens: A novel method based on enzyme-linked immunosorbent assay of spiggin, the stickleback (Gasterosteus aculeatus) glue protein. *Environmental Toxicology and Chemistry* **21**, 1946-1954 (2002).
50. Taylor, E. B. & McPhail, J. D. Evolutionary history of an adaptive radiation in species pairs of threespine sticklebacks (Gasterosteus): insights from mitochondrial DNA. *Biol. J. Linnean Soc* **66**, 271-291 (1999).
51. Hosemann, K. E., Colosimo, P. F. & Kingsley, D. M. A simple and efficient microinjection protocol for making transgenic sticklebacks. *Behavior*, (in press) (2004).
52. Nasevicius, A. & Ekker, S. C. Effective targeted gene 'knockdown' in zebrafish. *Nat Genet* **26**, 216-20 (2000).
53. Cresko, W. A., Amores, A., Yan, Y. L., Boone, J., Baltus, D., Singer, A. & Postlethwait, J. H. Different ways to make a hindbrain: functional divergence of duplicated Krox20 genes across stickleback, pufferfish, and zebrafish. *Fourth International Conference on Stickleback Behaviour and Evolution (*Stromstad, Sweden, 2003)
54. Peichel, C. L. & Kingsley, D. M. The genetic basis of sex determiantion in sticklebacks. (in preparation).
55. Carroll, S. B. Endless forms: the evolution of gene regulation and morphological diversity. *Cell* **101**, 577-80 (2000).

56. Stern, D. L. Evolutionary biology. The problem of variation. *Nature* **408**, 529, 531 (2000).
57. Tautz, D. Evolution of transcriptional regulation. *Current Opinion in Genetics & Development* **10**, 575-579 (2000).
58. Cole, N. J., Tanaka, M., Prescott, A. & Tickle, C. A. Expression of limb initiation genes and clues to the morphological diversification of threespine stickleback. *Current Biology*, (in press) (2003).
59. Chen, T. R. & Reisman, H. M. A comparative chromosome study of the North American species of sticklebacks (Teleostei: Gasterosteidae). *Cytogenetics* **9**, 321-32 (1970).
60. Grimwood, J., Schmutz, J., Dickson, M., Myers, R. M. & Kingsley, D. M. Initial survey of stickleback genomic sequences. (unpublished data).
61. McPhail, J. D. Ecology and evolution of sympatric sticklebacks (Gasterosteus): evidence for a species pair in Paxton Lake, Texada Island British Columbia. *Can. J. Zool.* **70**, 361-369 (1992).
62. Larson, G. L. & Mcintire, C. D. Food-habits of different phenotypes of threespine stickleback in Paxton Lake, British Columbia. *Transactions of the American Fisheries Society* **122**, 543-549 (1993).
63. Law, T. C. & Blake, R. W. Comparison of the fast-start performances of closely related, morphologically distinct threespine sticklebacks (Gasterosteus spp). *Journal of Experimental Biology* **199**, 2595-2604 (1996).
64. Ahn, D. G. & Gibson, G. Axial variation in the threespine stickleback: genetic and environmental factors. *Evolution & Development* **1**, 100-112 (1999).
65. Kingsley, D. M., Zhu, B., Osoegawa, K., de Jong, P. J., Schein, J., Marra, M., Peichel, C. L., Amemiya, C., Schluter, D., Balabhadra, S., Friedlander, B., Cha, Y. M., Dickson, M., Grimwood, J., Schmutz, J., Talbot, W. S. & Myers, R. M. New genomic tools for molecular studies of evolutionary change in sticklebacks. *Behavior*, (in press) (2004).