NEW GENOMIC TOOLS FOR MOLECULAR STUDIES OF EVOLUTIONARY CHANGE IN THREESPINE STICKLEBACKS

by

DAVID M. KINGSLEY[1,2,3], **BAOLI ZHU**[4], **KAZUTOYO OSOEGAWA**[4], **PIETER J. DE JONG**[4], **JACQUELINE SCHEIN**[5], **MARCO MARRA**[5], **CATHERINE PEICHEL**[6], **CHRIS AMEMIYA**[7], **DOLPH SCHLUTER**[8], **SARITA BALABHADRA**[2], **BRIAN FRIEDLANDER**[2], **YEE MAN CHA**[9], **MARK DICKSON**[9], **JANE GRIMWOOD**[9], **JEREMY SCHMUTZ**[9], **WILLIAM S. TALBOT**[2] and **RICHARD MYERS**[9,10]

([1] HHMI; [2] Dept. of Developmental Biology, Stanford, CA USA 94305-5329; [4] BACPAC Resources, CHORI, Oakland, CA USA; [5] Genome Sciences Center, BC Cancer Agency, Vancouver BC V5Z4E6 Canada; [6] Fred Hutchinson Cancer Research Center, Seattle, WA USA; [7] Molecular Genetics Department, Benaroya Research Institute at Virginia Mason, Seattle WA; [8] Dept. of Zoology, University of British Columbia, Vancouver BC V6T1Z4; [9] Dept. of Genetics and Stanford Human Genome Center, Stanford, CA USA)

Summary

The dramatic radiation of sticklebacks in different post-glacial environments provides a unique opportunity to study the molecular mechanisms that underlie rapid evolutionary change in vertebrates. We have developed a number of genomic and genetic tools to facilitate further study of a wide range of morphological, physiological and behavioral traits in sticklebacks. A large collection of microsatellite markers has previously been developed for use in genome-wide linkage mapping of interesting traits in crosses between different stickleback forms. cDNA libraries have been generated and EST sequencing projects have begun to isolate stickleback homologs of developmental control genes. Large insert BAC libraries have been built to compare chromosome regions of interest from both anadromous and freshwater stickleback populations. Large scale fingerprinting of one of these libraries has been

used to assemble overlapping contigs of BAC clones for chromosome walking and positional cloning. Together with recent development of methods to make transgenic sticklebacks, these tools should make it possible to identify the molecular basis of many different evolutionary traits in stickleback, and to begin to answer longstanding questions about the numbers and types of mutations that control the appearance of new morphological, physiological, and behavioral traits during vertebrate evolution.

Introduction

The threespine stickleback *Gasterosteus aculeatus* offers a unique opportunity for genetic analysis of evolutionary differences in vertebrates. This small marine fish has undergone a very recent adaptive radiation in the freshwater streams and lakes created after the end of the last ice age (Bell & Foster, 1994). The widespread melting of glaciers led to dramatic changes in sea level and land elevation only 10-20 thousand years ago. As a result, populations of marine sticklebacks became isolated in thousands of newly created lakes in formerly ice-covered regions across North America, Northern Europe, and Northern Asia. These populations subsequently diverged into a large number of different forms that show marked variation in morphology, physiology, life history, and behavioral traits. Although many of the interesting differences between populations are as large as those normally seen between different species or genera, most stickleback populations can still be crossed by natural matings or artificial fertilization in the laboratory, making it possible to use genetic approaches to study the basis of evolutionary differences among populations.

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 the Nobel Prize in Physiology and Medicine in 1973 (Cronholm & Levi, 1973). A 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 books on the behavior, morphology, physiology, life-history, distribution, ecology, and paleontology of different stickleback populations (Wootton, 1976, 1984; Coad, 1981; Ziuganov, 1991; Bell & Foster, 1994; Paepke, 1996; Keivany, 2002).

Sicklebacks also have numerous experimental advantages for detailed molecular studies of vertebrate evolution. The extreme recency of the stickleback radiation in post-glacial lakes makes it possible to study the earliest steps of evolutionary divergence, and increases the chance that underlying genetic basis of many traits is relatively simple. The small size, relatively rapid generation time (six months in the laboratory), and large clutch size of sticklebacks makes it possible to raise large numbers of progeny for high-resolution genetic mapping. A set of random microsatellite markers has recently been developed that shows high levels of polymorphism in every population tested. These markers have been used to assemble an initial linkage map of the threespine stickleback (Peichel *et al.*, 2001). Finally, the genome size of sticklebacks is relatively compact (approximately 700 megabases, (Hinegardner, 1968)), less than a quarter the size of the mouse or human genome, and less than half the size of zebrafish.

Recent application of genome wide linkage mapping approaches to sticklebacks has begun to determine the number and location of chromosome regions responsible for evolutionary differences between populations (Peichel *et al.*, 2001). In order to identify the actual genes and mutations responsible for evolutionary change, it will still be necessary to clone the chromosome regions that control traits of interest, to identify candidate genes within these regions, to look for expression or coding differences between populations, and to verify that these differences are responsible for phenotypic differences using transgenic methods. Here we describe the development of a number of new genomic resources for threespine sticklebacks, including cDNA and genomic libraries, expressed sequence tag (EST) collections, and genome wide physical maps based on large scale fingerprinting of bacterial artificial chromosome (BAC) clones (see Table 1 for summary of abbreviations and terms). These tools will facilitate the isolation and study of many different stickleback genes and chromosome regions of interest, and should ultimately make it possible to determine the molecular mechanisms that have led to new behavioral, morphological, and physiological traits in this vertebrate model organism.

Materials and methods

cDNA library and EST sequencing

Total RNA was extracted using RNAzol from heads and mixed internal organs of two adult fish from the Salinas River in California. mRNA was prepared and used to construct an

TABLE 1. *Glossary of molecular terms and acronyms*

cDNA	Complementary DNA (a cloned DNA version of a single messenger RNA molecule).
EST	Expressed sequence tag (a DNA sequence determined from part or all of a cDNA clone).
BAC	Bacterial artificial chromosome (a DNA plasmid engineered to allow stable propogation of a large insert of foreign DNA in bacteria).
contig	A contiguous set of overlapping DNA clones used to derive a physical map or sequence of a given gene or chromosome region.
ICE	Internet Contig Explorer (a free JAVA-based program for searching and displaying sets of overlapping DNA clones from large-scale physical mapping projects).

amplified library in the lambda ZAP express vector (Stratagene Custom Library Services). First strand synthesis was primed with an oligodT primer containing a 5 prime synthetic XhoI site preceding the oligodT stretch. Five prime adaptors included a synthetic EcoR1 site (GAATTCGGCACGAGG). cDNAs were inserted into the ZAP express vector unidirectionally in the sense orientation with respect to the lacZ promoter of pBK-CMV. An amplified library was prepared from approximately 3 million primary clones. In vivo excision was used to generate individual pBK-CMV plasmid clones. Individual bacterial colonies were isolated and grown in the presence of kanamycin (50 micrograms/ml). Cultures of individual clones were supplemented with dimethylsulfoxide to a final concentration of 3.5% (volume/volume), arrayed in 96 well plates, and frozen at $-80°C$.

We sequenced the cDNA fragments using standard methods with Applied Biosystems (ABI) Big-Dye Terminator chemistry and performed the detection on an ABI 3730xl sequencer. 5′ end reads were generated with the universal T3 vector primer and 3′ end reads were generated with a poly-T primer with one variable end base. Two wells from each source plate were also resequenced and compared to the original sequences to verify sample tracking and correct for any mishandled plates. We counted successful reads as reads with greater than or equal to 200 bases of quality 20 as calculated by Phred (Ewing & Green, 1998; Ewing *et al.*, 1998). All sequences have been deposited in Genbank (accession numbers CD492707-CD51029) and dbEST (18525002-18542581).

BAC library construction

Two new stickleback BAC libraries were created, one from an anadromous population from the Salmon River in British Columbia (CHORI-213), and one from freshwater benthic fish collected from Paxton Lake on Texada Island (CHORI-215). For the Salmon River library, high molecular weight genomic DNA was prepared from a mixed blood sample derived from about 60 individuals, including a mixture of both males and females. The isolated blood cells (erythrocytes & lymphocytes) were mixed with Incert^{TM} agarose (FMC, 1.0% final concentration) and pipetted into disposable molds (BioRad) to form 80 μl agarose plugs. The solidified agarose plugs were treated with lysis buffer (1% lithium dodecyl sulfate, 10 mM

Tris, 100 mM EDTA, pH 9) to extract the high molecular weight DNA. The plugs contain 10 to 20 μg of high molecular weight DNA depending on the number of blood cells. For the Paxton Benthic population, the preparation of DNA plugs was similar, except that blood from each individual male fish was separately processed, and DNA from a single male was used to prepare segment 1 of the library, and DNA from a second male was used to prepare segment 2 of the library (see below).

Clones for each library were prepared essentially as described by Osoegawa *et al.* (1998) using two different strategies to partially digest and clone the high-molecular weight DNA in corresponding BAC vectors. Each library thus contains two different sub-segments prepared with different enzymes to maximize the diversity of clones. The first parts of each library ('segment 1') were prepared from DNA fragments partially digested with EcoRI restriction enzyme and EcoRI-methylase, cloned into the EcoRI site of the pTARBAC2.1 vector. The second parts of each library ('segment 2') were derived from genomic DNA fragments partially digested with MboI and cloned between the BamHI sites of the pTARBAC1.3 vector. Vector sequences and graphics can be found at: *http://bacpac.chori.org/vectorsdet.htm*. The backbone and utilities of the pTARBAC vectors have been described previously (Zeng *et al.*, 2001).

Prior to ligating the vectors and DNA fragments, small-size DNA fragment were removed through three consecutive sizing steps by pulsed-field electrophoresis. The size-purified DNA of the 150-200 kb fraction was isolated from the agarose slices by electro-elution. The fragments were dialyzed and then ligated to the appropriate vector. The ligation products were drop-dialyzed on floating membranes to remove salt and were then used for electroporation of commercially available DH10B electro-competent E. coli cells (T1-phage resistant, InVitrogen Inc). The cells from a large number of electro-transformations were incubated for 60 minutes with SOB media and glycerol to allow for expression of antibiotic (chloramphenicol) resistance. The pooled clones were stored at $-80°$C to allow for a sufficient number of cells to be obtained, to receive preliminary characterization for clone-insert size and to schedule clone arraying.

When the quality assessment indicated satisfactory results, then BAC clones were spread on large square Petri dishes (24×24 cm) with LB agarose, sucrose and chloramphenicol. The dishes contained sucrose to allow for preferential growth of recombinant clones, which include insert sequences disrupting the expression of the toxic SacBII gene. The dishes were incubated for about 16-18 hrs and the resulting colonies (1,000-3,000 per dish) were arrayed into 384-well microtiter dishes (Genetix) using a colony-picking robot (Genetix QpixII). The 384-well dishes contained LB media supplemented with chloramphenicol and glycerol. The newly picked clones were incubated overnight at 37°C. The glycerol is not metabolized and permits cryo-preservation of the arrayed clones at -80 following overnight growth.

For the purpose of identifying and isolating genes of interest by probe hybridization, copies of the arrayed clone collections were prepared. One of the copies of each library was dedicated for use in a clone-gridding robot (BioRobotics) to create high-density colony filters to be used for DNA probe screening. The colony filters (prepared on 22×22 cm nylon membranes) contain 18,000 independent clones, each gridded in duplicate from 48 microtiter dishes to make 36,000 total colonies. After overnight growth of the colonies on the filters (on square LB agarose dishes), the filters were processed by pronase & SDS treatment and the extracted DNA was denatured and fixed to the membranes by UV cross-linking. Filters were made for a total of 107,087 independent clones of the Salmon River library (four screening filters for 71,087 clones in segment 1 and two screening filters for 36,000 clones in

segment 2). Filters were made for a total of 72,000 independent clones of the Paxton Benthic library (two screening filters for 36,000 clones in segment 1 and two screening filters for 36,000 clones in segment 2).

BAC physical mapping

BAC clones from the Salmon River library were fingerprinted in 96-well format using an agarose gel methodology (Marra *et al.*, 1997, 1999; McPherson *et al.*, 2001; Schein *et al.*, 2003). The bacterial clones were cultured overnight and bacterial pellets collected by centrifugation. BAC DNA was isolated by alkaline lysis purification, digested with *Hin*dIII and the resulting restriction fragments resolved by electrophoresis on 1.2% agarose gels. Gels were stained post-electrophoresis with SYBR Green I (Molecular Probes, Inc.) and digital images acquired using a Molecular Dynamics Fluorimager 595. Fingerprint lanes were tracked on the gel images by technical staff using the program Image (*www.sanger,.uk/Software/Image*) (Sulston *et al.*, 1988, 1989). The gel images were analyzed using BandLeader software (Fuhrmann *et al.*, 2003) for identification and size determination of the *Hin*dIII restriction fragments for each clone. Automated assembly tools in the software program FPC (Soderlund *et al.*, 1997, 2000; Ness *et al.*, 2002) were employed to assemble contigs comprised of overlapping BAC clones based on fingerprint similarity. The automated FPC assembly is publicly available for download (*www.bcgsc.ca/lab/mapping/data*). The data can also be viewed via the Internet using iCE (Fjell *et al.*, 2003), a Java-based application for viewing FPC data (*http://www.bcgsc.ca/about/news/ice*).

Results

To facilitate isolation of stickleback homologs of genes of interest, cDNA clones were picked individually from a mixed organ cDNA library, grown up and arrayed in 96 well plates, and used for sequencing reactions with both 5 prime and 3 prime primers. From 9,600 attempts on each end we produced 8,956 successful (93.3% of attempts) 5′ end reads and 8,501 (88.5% of attempts) successful 3′ end reads.

Cluster analysis of the 8,501 successful 3 prime reads was used to estimate locus coverage of the EST sequencing to date. We ran a complete pairwise comparison of all reads using NCBI's megablast program with filtering disabled. We then clustered each read with every other read that had at least a 200 bp or longer alignment at greater than 96% identity within the first 300 bp of the reads. This clustered all reads that contained a similar 3'UTR and end coding region together, but prevented chimeric subclones from merging two distinct loci. This analysis identified a total of 4,780 distinct stickleback gene clusters. Most clusters are still represented by only one or a small number of clones, suggesting that further sequencing of this library will continue to identify new stickleback genes (Table 2).

TABLE 2. *Distribution of EST numbers in stickleback EST clusters*

Sequence reads per cluster	Number of clusters found
1	3775
2 to 10	928
11 to 20	46
21 to 30	17
>30	14
Total:	**4780**

BAC libraries from different stickleback populations

To facilitate detailed genomic studies of genes and chromosome regions in threespine sticklebacks, two large insert bacterial artificial chromosome libraries were prepared. The first library was made from pooled blood samples of approximately 60 male and female fish from an anadromous population in the Salmon River, British Columbia. Anadromous fish are thought to be the progenitor of many of the freshwater populations established after the end of the last ice age. The library from Salmon River fish thus provides a sample of the type of genetic information present in a fully armored migratory fish population, and is likely to represent the type of ancestoral genetic variation present in the founders of recently derived freshwater populations.

The second library was made from blood cell DNA isolated from male fish from the Paxton Lake Benthic population. This well studied population is a member of a species pair found in a small lake on Texada Island in the Strait of Georgia (McPhail, 1992). The Paxton Benthic library samples the type of genetic information found in a highly derived freshwater population with extensive armor loss and trophic modifications for bottom feeding. The large size of Paxton fish, and improvements in blood recovery and DNA isolation, made it possible to build the entire Paxton benthic BAC library from only two individuals, thus reducing the likely DNA sequence variation among separate but overlapping clones isolated from the same genetic region.

Tables 3 and 4 summarize the basic characteristics of each library. Based on the number of clones and average sizes of inserts in these libraries, we estimate that the two libraries represent approximately 30-fold coverage of the Salmon River fish genome, and 20× coverage of the Paxton benthic genome.

TABLE 3. *CHORI-213 BAC library from anadromous sticklebacks*

Segment	1	2	Total
Cloning vector	pTARBAC2.1	pTARBAC1.3	
DNA source	Pooled blood sample	Pooled blood sample	
	from 60 Salmon River fish	from 60 Salmon River fish	
Restriction enzyme	EcoRI	MboI/BamHI	
Total 384-well dishes	192	96	288
Recombinant clones	Approx. 71,087	Approx. 36,000	107,087
Insert size (average)	190 Kbp	~160 Kbp	
Library redundancy	20 fold	10 fold	30 fold
(0.7 Gbp genome)			

TABLE 4. *CHORI-215 BAC library from freshwater sticklebacks*

Segment	1	2	Total
Cloning vector	pTARBAC2.1	pTARBAC1.3	
DNA source	Blood from single	Blood from a different	
	Paxton Lake Benthic male	Paxton Lake Benthic male	
Restriction enzyme	EcoRI	MboI/BamHI	
Total 384-well dishes	192	96	288
Recombinant clones	Approx. 36,000	Approx. 36,000	72,000
Insert size (average)	165 Kbp	~155 Kbp	
Library redundancy	10 fold	10 fold	20 fold
(0.7 Gbp genome)			

Physical mapping

Fingerprints were attempted for all BAC clones in the two segments of the Salmon River library (CHORI 213). A total of 102,178 (92%) clones was successfully fingerprinted and exported to FPC, representing approximately 26-fold coverage of the *Gasterosteus aculeatus* genome. The remaining clones either failed to grow during overnight culturing, were nonrecombinant, were refractive to the fingerprinting process, did not have a sufficiently informative fingerprint, or failed for other technical reasons. The average insert size of the clones in the FPC database, derived from the sum of the fingerprint fragments, is 191,973 bp for clones from segment 1 of the library and 166,892 bp for clones from segment 2 of the library. Following automated assembly, 71,211 clones were assembled into 8,397 contigs containing two or more clones. The mean contig size was 8.4 clones. The

TABLE 5. *Distribution of contig sizes in initial stickleback physical map assembly*

Clones per contig	Number of contigs	Total number of clones	% of all BAC clones in contigs
2	4,524	9,048	12.7%
3-9	2,902	12,309	17.3%
11-50	660	16,338	22.9%
>50	311	33,516	47.1%
Total:	**8,397**	**71,211**	**100%**

remaining 30,967 clones were not assigned to contigs with the assembly parameters used ('singletons'). The singletons will be further analyzed to determine whether they can be used to create new contigs or be placed into existing contigs. The distribution of contig sizes is shown in Table 5. More than 87% of the assembled clones are in contigs of 3 or more clones. Given the average relationship between physical and genetic distance in sticklebacks (approximately 0.5 Mb per centiMorgan), and the large average clone size in the BAC library, even the smallest contigs should be useful for isolating genomic regions surrounding genes or traits of interest.

Discussion

The number and type of genes and mutations that contribute to the evolution of new traits in vertebrates are still unknown. Previous development of a genome-wide linkage map for the threespine sticklebacks (Peichel *et al.*, 2001) has made it possible to map the chromosome regions responsible for interesting evolutionary differences between stickleback populations that have adapted to different environments (Peichel *et al.*, 2001, 2004; Colosimo *et al.*, 2004; Cresko *et al.*, 2004; Shapiro *et al.*, 2004). The new resources described here provide both library and physical mapping resources for detailed molecular studies of stickleback genes and chromosome regions of interest.

All of the EST sequences have been deposited in Genbank, and can be searched through standard BLAST tools either from the NCBI website (EST division database), or a dedicated BLAST searcher that we have developed specifically for stickleback ESTs (*http://cegs.stanford.edu/blast/blast_stickleback.html*). Such searches represent one of the simplest ways to

test whether a stickleback homolog of a gene of interest is already available. Aliquots of the amplified cDNA library, and any individual clones from the EST sequencing project, are being made available from the Stanford Genome Evolution Center (see: *http://cegs.stanford.edu/stickleback_EST_sequences. jsp*). We plan to expand the EST sequencing project to additional tissues and developmental stages over the next two years. This should further expand the number of distinct genes represented in the cDNA clone collections, and facilitate the isolation, mapping and expression analysis of candidate genes for many different developmental, morphological, physiological, and behavioral traits in sticklebacks.

The availability of highly redundant, large-insert BAC libraries from anadromous and freshwater fish provides a general mechanism for recovering clones from a particular chromosome region of interest, and comparing the genomic structure between regions in two different populations of sticklebacks. Nylon filters are available for screening either stickleback library by hybridization with a probe of interest and detailed protocols for screening the libraries have been posted on: *http://bacpac.chori.org/protocols.htm*. Once a hybridization-positive clone is identified, the gridding pattern permits the identification of the clone address with respect to the precise microtiter dish and row/column intersection. Clones can then be obtained by requesting the corresponding address location from BACPAC Resources (see: *http://bacpac.chori.org/ordering_information.htm*). As an example of this approach, BAC library screening has recently been used to isolate stickleback homologs of several different genes known to be involved in hindlimb specification in other vertebrates. The stickleback BAC clones made it possible to develop microsatellite markers closely linked to each candidate gene, and to show that the stickleback *Pitx1* gene maps to a major chromosome region that controls hindlimb reduction in freshwater populatons (Shapiro *et al.*, 2004).

The fingerprinting of BAC clones represents an efficient strategy for detecting overlaps between different clones, and assembling sets of genomic cones that cover different regions of the genome (Marra *et al.*, 1997). For researchers currently interested in studying any particular gene or chromosome region in sticklebacks, we recommend developing a probe that can be used to screen the Salmon River library, determining the library coordinates of one or more clones that map to the region, and then using the contig information from the physical mapping project to look for additional BAC clones that

extend further from that region. All of the fingerprint and contig information from the physical mapping project can be searched through web-based tools, providing immediate access to the most up-to-date version of the current clone assemblies (*http://www.bcgsc.ca/lab/mapping/*). The BAC screening and contig building approach has recently been used to isolate a series of overlapping DNA clones from the major chromosomal region controlling sex determination in sticklebacks. Sequencing of the BAC clones from male and female fish provides the first molecular evidence for an evolving X/Y chromosome system in *Gasterosteus aculeatus* (Peichel *et al.*, 2004).

The availability of a deep, large-insert library from an anadromous population, and an initial physical map of clones from this library, now makes it possible to physically isolate the genetic information from fish representing the presumed ancestral state of many different derived freshwater populations. The corresponding regions from a freshwater lake population can also be recovered from the Paxton benthic BAC library. The large-scale sequence, gene content, and sequence variation can then be compared in anadromous and freshwater populations.

Final demonstration that a particular gene or sequence variant is responsible for a trait of interest will require methods to transfer genes between populations, and test for rescue or phenocopy of the corresponding trait. We have recently developed methods that allow successful introduction, expression, and germ line transmission of exogenous DNA in sticklebacks (see Hosemann *et al.*, this volume). A major goal of future research will be to expand these methods to the study of cDNA clones and BAC clones that map to the same chromosome regions that are already known to control evolutionary differences between stickleback forms. The Salmon River BAC clones are ideally suited for testing whether transfer of particular genetic information from an anadromous fish can modify the morphological, physiological, or behavioral differences seen in derived freshwater forms.

The recency of the stickleback radiation in post-glacial lakes greatly increases the chances that the underlying genetic basis of many traits are relatively simple. Ten to twenty 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 groups of mammals. The time for stickleback divergence 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 (Ingman *et al.*, 2000). Further study of the adaptive radiation of sticklebacks may reveal genetic mechanisms that are broadly relevant to the evolution of complex traits and physiological differences between other groups, including humans.

References

Bell, M.A. & Foster, S.A., eds (1994). The evolutionary biology of the threespine stickleback. — Oxford University Press, Oxford.

Coad, B.W. (1981). A bibliography of sticklebacks. — Syllogeus 35, p. 1-42.

Colosimo, P.F., Peichel, C.L., Nereng, K.S., Blackman, B.K., Shapiro, M.D., Schluter, D. & Kingsley, D.M. (2004). The genetic architecture of parallel armor plate reduction in threespine sticklebacks. — PLOS Biology 2, p. 635-641.

Cresko, W.A., Amores, A., Wilson, C., Murphy, J., Currey, M., Phillips, P., Bell, M.A., Kimmel, C.B. & Postlethwait, J.H. (2004). Parallel genetic basis for repeated evolution of armor loss in Alaskan threespine stickleback populations. — Proc. Natl. Acad. Sci. USA 101, p. 6050-6055.

Cronholm, B. & Levi, L. (1973). Genetically programmed and learned behaviors. A presentation of this year's Nobel Prize winners in physiology and medicine. — Lakartidningen 70, p. 3807-3811.

Ewing, B. & Green, P. (1998). Base-calling of automated sequencer traces using phred. II. Error probabilities. — Genome Res. 8, p. 186-194.

— —, Hillier, L., Wendl, M.C. & Green, P. (1998). Base-calling of automated sequencer traces using phred. I. Accuracy assessment. — Genome Res. 8, p. 175-185.

Fjell, C.D., Bosdet, I., Schein, J.E., Jones, S.J. & Marra, M.A. (2003). Internet Contig Explorer (iCE) — a tool for visualizing clone fingerprint maps. — Genome Res. 13, p. 1244-1249.

Fuhrmann, D.R., Krzywinski, M.I., Chiu, R., Saeedi, P., Schein, J.E., Bosdet, I.E., Chinwalla, A., Hillier, L.W., Waterston, R.H., McPherson, J.D., Jones, S.J. & Marra, M.A. (2003). Software for automated analysis of DNA fingerprinting gels. — Genome Res. 13, p. 940-953.

Hinegardner, R. (1968). Evolution of cellular DNA content in teleost fishes. — Am. Nat. 102, p. 517-523.

Ingman, M., Kaessmann, H., Paabo, S. & Gyllensten, U. (2000). Mitochondrial genome variation and the origin of modern humans. — Nature 408, p. 708-713.

Keivany, Y. (2002). Bibliography of sticklebacks website. — *http://www.geocities.com/ CapeCanaveral/Hall/1345/stickbibl.html*

Marra, M.A., Kucaba, T.A., Dietrich, N.L., Green, E.D., Brownstein, B., Wilson, R.K., McDonald, K.M., Hillier, L.W., McPherson, J.D. & Waterston, R.H. (1997). High throughput fingerprint analysis of large-insert clones. — Genome Res. 7, p. 1072-1084.

— —, — —, Sekhon, M., Hillier, L., Martienssen, R., Chinwalla, A., Crockett, J., Fedele, J., Grover, H., Gund, C., McCombie, W.R., McDonald, K., McPherson, J., Mudd, N., Parnell, L., Schein, J., Seim, R., Shelby, P., Waterston, R. & Wilson, R. (1999). zA map for sequence analysis of the Arabidopsis thaliana genome. — Nat. Genet. 22, p. 265-270.

McPhail, J.D. (1992). Ecology and evolution of sympatric sticklebacks (Gasterosteus): evidence for a species pair in Paxton Lake, Texada Island British Columbia. — Can. J. Zool. 70, p. 361-369.

McPherson, J.D., Marra, M., Hillier, L., Waterston, R.H., Chinwalla, A., Wallis, J., Sekhon, M., Wylie, K., Mardis, E.R., Wilson, R.K., Fulton, R., Kucaba, T.A., Wagner-McPherson, C., Barbazuk, W.B., Gregory, S.G., Humphray, S.J., French, L., Evans, R.S., Bethel, G., Whittaker, A., Holden, J.L., McCann, O.T., Dunham, A., Soderlund, C., Scott, C.E., Bentley, D.R., Schuler, G., Chen, H.C., Jang, W., Green, E.D., Idol, J.R., Maduro, V.V., Montgomery, K.T., Lee, E., Miller, A., Emerling, S., Kucherlapati, Gibbs, R., Scherer, S., Gorrell, J.H., Sodergren, E., Clerc-Blankenburg, K., Tabor, P., Naylor, S., Garcia, D., de Jong, P.J., Catanese, J.J., Nowak, N., Osoegawa, K., Qin, S., Rowen, L., Madan, A., Dors, M., Hood, L., Trask, B., Friedman, C., Massa, H., Cheung, V.G., Kirsch, I.R., Reid, T., Yonescu, R., Weissenbach, J., Bruls, T., Heilig, R., Branscomb, E., Olsen, A., Doggett, N., Cheng, J.F., Hawkins, T., Myers, R.M., Shang, J., Ramirez, L., Schmutz, J., Velasquez, O., Dixon, K., Stone, N.E., Cox, D.R., Haussler, D., Kent, W.J., Furey, T., Rogic, S., Kennedy, S., Jones, S., Rosenthal, A., Wen, G., Schilhabel, M., Gloeckner, G., Nyakatura, G., Siebert, R., Schlegelberger, B., Korenberg, J., Chen, X. N., Fujiyama, A., Hattori, M., Toyoda, A., Yada, T., Park, H.S., Sakaki, Y., Shimizu, N., Asakawa, S., *et al.* (2001). A physical map of the human genome. — Nature 409, p. 934-941.

Ness, S.R., Terpstra, W., Krzywinski, M., Marra, M.A. & Jones, S.J. (2002). Assembly of fingerprint contigs: parallelized FPC. — Bioinformatics 18, p. 484-485.

Osoegawa, K., Woon, P.Y., Zhao, B., Frengen, E., Tateno, M., Catanese, J.J. & de Jong, P.J. (1998). An improved approach for construction of bacterial artificial chromosome libraries. — Genomics 52, p. 1-8.

Paepke, H.-J. (1996). Die Stichlinge. — Westarp Wissenschaften, Germany.

Peichel, C.L., Nereng, K.S., Ohgi, K.A., Cole, B.L., Colosimo, P. F., Buerkle, C.A., Schluter, D. & Kingsley, D.M. (2001). The genetic architecture of divergence between threespine stickleback species. — Nature 414, p. 901-905.

— —, Ross, J.A., Matson, C.K., Dickson, M., Grimwood, J., Schmutz, J., Myers, R.M., Mori, S., Schluter, D. & Kingsley, D.M. (2004). The master sex-determination locus in threespine sticklebacks is on a nascent Y chromosome. — Curr. Biol. 14, p. 1416-1424.

Schein, J., Kucaba, T., Sekhon, M., Smailus, D., Waterston, R. & Marra, M. (2003). High throughput BAC fingerprinting. — In: Bacterial artificial chromosomes: methods adn protocols (S. Zhao & M. Stodolsky, eds). Humana Press Inc.

Shapiro, M.D., Marks, M.E., Peichel, C.L., Blackman, B.K., Nereng, K.S., Schluter, D. & Kingsley, D.M. (2004). Genetic and developmental basis of evolutionary pelvic reduction in threespine sticklebacks. — Nature 428, p. 679-782.

Soderlund, C., Humphray, S., Dunham, A. & French, L. (2000). Contigs built with fingerprints, markers, and FPC V4.7. — Genome Res. 10, p. 1772-1787.

— —, Longden, I. & Mott, R. (1997). FPC: a system for building contigs from restriction fingerprinted clones. — Comput. Appl. Biosci. 13, p. 523-535.

Sulston, J., Mallett, F., Durbin, R. & Horsnell, T. (1989). Image analysis of restriction enzyme fingerprint autoradiograms. — Comput. Appl. Biosci. 5, p. 101-106.

— —, — —, Staden, R., Durbin, R., Horsnell, T. & Coulson, A. (1988). Software for genome mapping by fingerprinting techniques. — Comput. Appl. Biosci. 4, p. 125-132.

Wootton, R.J. (1976). The biology of the sticklebacks. — Academic Press, London.

— — (1984). A functional biology of sticklebacks. — University of California Press, Berkeley and Los Angeles.

Zeng, C., Kouprina, N., Zhu, B., Cairo, A., Hoek, M., Cross, G., Osoegawa, K., Larionov, V. & de Jong, P. (2001). Large-insert BAC/YAC libraries for selective re-isolation of genomic regions by homologous recombination in yeast. — Genomics 77, p. 27-34.

Ziuganov, V.V. (1991). The family Gasterosteidae of the world fauna. Fauna of the USSR. Fishes, Vol. 5. — Nauka, Leningrad (in Russian).