

Panel Summary #1

Proposal Number: 1051176

Panel Summary:

Panel Summary

Evolutionary Genetics Fall 2010 Panel

Results Of Prior NSF Support (If Applicable):

Results of prior support Criterion I (Intellectual Merit):

The PIs report on three prior NSF grants, and have produced a substantial number of papers as a result of previous funding.

Results of prior support Criterion II (Broader Impacts):

Broader impacts from the grant to AW are not reported. The other PIs have reported significant broader impacts, including training of undergraduates, graduate students, and postdocs. The Daphnia resources produced by the award to JS are also very impressive.

If this is a resubmission, how have previous criticisms been addressed?

The previous criticisms involved a general lack of cohesion regarding the plans to integrate additional data into the genome sequence. The proposal has been revised to more explicitly describe the nature of the additional data and its relevance.

CRITERION I: [Intellectual Merit]:

Intellectual Merit Strengths:

With regard to Intellectual Merit, the strengths of the proposal remain the same, and they remain compelling. This is an excellent group of scientists, working on an important question and using an important model system. The sequencing of the *Fundulus* genome will be very valuable. The enthusiasm for this work among the panelists was very high.

Intellectual Merit Weaknesses:

Although enthusiasm was high, there were some concerns raised in the panel discussion. Whether the proposed work will result in identification of causal genes is uncertain. There was some concern regarding whether the work could be completed in 3 years.

CRITERION II: [Broader Impacts]:

The broader impacts are centered around annotation and curation of the newly sequenced genome. This will be accomplished using an annotation workshop, which will involve training of postdocs and graduate students. There is also some inclusion of minority undergraduates in this process.

Broader Impact Strengths:

The training associated with the workshop will be valuable, and will facilitate the ultimate usage of the sequenced genome to its fullest potential. The production of a quality genome sequence for *Fundulus* will be an important addition to research infrastructure in a variety of disciplines.

Broader Impact Weaknesses:

N/A

POST-DOC MENTORING PLAN:

N/A

SYNTHESIS AND RECOMMENDATION

The PIs propose to sequence and annotate the genome of *Fundulus heteroclitus*. They will then integrate a vast array of additional experimental data and resequence targeted populations. Ultimately, they hope to identify the underlying genetic mechanisms that explain rapid evolution of pollution tolerance. The PIs also plan to investigate the genetic basis of convergence and whether this convergence results from standing variation or new mutations. This is an important and timely topic.

There is a great deal of enthusiasm for the investigators and the system.

The panel recommendation is Superior

This summary was read by the assigned panelists and they concurred that the summary accurately reflects the panel discussion.

Panel Recommendation: Superior

Review #1

Proposal Number:	1051176
Performing Organization:	La St U
NSF Program:	Evolutionary Genetics
Principal Investigator:	Whitehead, Andrew
Proposal Title:	Collaborative Research: The genomic basis of dramatic, rapid, convergent evolution in the killifish <i>Fundulus heteroclitus</i>
Rating:	Excellent

REVIEW:

What is the intellectual merit of the proposed activity?

This proposal is a collaborative effort among some of the world's best groups in environmental genomics, to develop a reference genome sequence for killifish, *Fundulus heteroclitus*, and to use that genome as a basis for the discovery of patterns of evolutionary adaptation to contaminated environments. The project has both fundamental and applied merits, and is of additional value as a community building effort.

On the fundamental side, the project will shed light on the synchrony between phenotypic and molecular convergent evolution. Convergent evolution of morphological and physiological traits being a relatively common phenomenon in nature, evolutionary biologists have assumed that such convergences must involve similar genetic mechanisms. However, the idea that similar adaptive phenotypes can be accounted for by the same genetic changes does not seem to hold in general, as shown by work on lake whitefish and beach mice. The present proposal is designed to test the hypothesis of synchrony between phenotypic and molecular convergent evolution in a highly promising model system, eminently suited to generate detailed insights to test the hypothesis.

On the applied side the proposal will shed new light on the rapid but still enigmatic development of tolerance to organochlorine contaminants seen in killifish. There is a suggestion that suppression of the AhR pathway, or mutations in one or more of the downstream AhR targets (e.g. DREs) might be among of the mechanism involved. In this light it is very appropriate to aim for a complete genome sequence of the killifish, because the AhR-induced CYP genes are usually members of a large family. Knowing all paralogs of CYPs in the genome is crucial to understand xenobiotic metabolism. If some aspects of the AhR-pathway are blocked, many other aspects of cellular biochemistry may be affected and a genome-wide expression analysis will be in place.

The proposal aims for a three-fold strategy to identify genes involved with pollutant tolerance: (1) QTL markers segregating with tolerance, (2) population genetics outlier analysis using *Fst*, (3) detecting genes with expression diverging between populations. This three-fold strategy is sure to find the genes involved, even if they are few. The combination of these three approaches is a unique and very strong aspect of this proposal.

The proposal logically builds on a very strong experimental basis developed by the groups involved over the years. A lot of pilot and preliminary work has been done to make the work planned successful. A proof of concept for the strategy to assemble full genome sequences from next generation sequencing technology has been given before (CGB, WUGC).

The groups involved have a very strong laboratory and computational infrastructure. A number of up-to-date bioinformatics techniques are planned.

What are the broader impacts of the proposed activity?

The proposal will act as a crystallization point for a community of *Fundulus* researchers. The proposal aims to involve young researchers in an annotation workshop and other activities aimed at increasing understanding of the killifish genome. Thus, there is a strong community-building aspect in the proposal. As a side-effect the project will contribute to raising awareness among young scientists of the exciting avenues taken by modern biological sciences and it will hopefully increase enthusiasm among selected students to pursue a career in genome biology.

Another aspect of the impact will be the issue of risk assessment of contaminated sediments and aquatic ecosystems. Ever since the 1992 Rio Convention on Biological Diversity, concerns have been raised on the possible loss of genetic variation caused by population bottlenecks and directional selection on wild populations exposed to pollution (genetic erosion). This issue received, however, little attention in research. Several studies have demonstrated a loss of genetic diversity under pollution, but other studies have not been able to demonstrate such effects. Given a full array of modern genomics technology as proposed in the present proposal on *Fundulus heteroclitus* it will be a prime case for analysing the issue of genetic erosion.

Summary Statement

This is a surely outstanding proposal that deserves to be funded.

Review #2

Proposal Number: 1051176
Performing Organization: La St U
NSF Program: Evolutionary Genetics
Principal Investigator: Whitehead, Andrew
Proposal Title: Collaborative Research: The genomic basis of dramatic, rapid, convergent evolution in the killifish *Fundulus heteroclitus*
Rating: Multiple Rating: (Excellent/Very Good)

REVIEW:

What is the intellectual merit of the proposed activity?

The focus of this proposal is to identify the genetic basis of convergent evolution of pollution tolerance in *Fundulus*.

The PIs propose to produce a reference genome for *Fundulus*, identify candidate genes that may be associated with adaptation to polluted sites, and perform population re-sequencing to support studies of molecular evolution.

Strengths

Production of a *Fundulus* genome sequence would be very useful to the broader research community. There is no question it will greatly enhance the utility of an already important model.

In addition, understanding the genetic basis of evolved pollution tolerance is important.

The hypothesis that the observed pattern of convergent evolution is a result of standing genetic variation is intriguing from a conservation standpoint and worth testing.

The sequencing requirements, and assembly and annotation strategies are well thought out, clearly described, and well supported.

This is a quality group of researchers. However, I still wonder whether they can pull off such an ambitious project on this aggressive timeline.

Weaknesses

I agree this is an important organism. I'm not sure a genome sequence for *Fundulus* would 'transform the research landscape for environmental, ecological, and evolutionary biology'.

This proposal is very ambitious. Are Aims 2 and 3 really feasible? Integration of these data with a genome sequence would be a great accomplishment. On the other hand, the data are largely generated using other support, and the specific details are difficult to evaluate. The metaanalyses and integration of all the different data types requires detailed descriptions of methodology, assumptions, and preliminary data.

The timeline indicates the sequencing will be completed and assembled within 8 months, and annotation completed about 1 year from the start date. Again, this is very aggressive.

What are the broader impacts of the proposed activity?

The proposed broader impacts are associated with training, primarily through joint annotation of the genome in a 'jamboree' style conference. This will provide excellent training to diverse members of the Fundulus research community.

An additional broader impact will be the production of the Fundulus genome database using GMOD tools. This database will be very broadly used in diverse fields beyond evolutionary biology.

Summary Statement

Review #3

Proposal Number: 1051176
Performing Organization: La St U
NSF Program: Evolutionary Genetics
Principal Investigator: Whitehead, Andrew
Proposal Title: Collaborative Research: The genomic basis of dramatic, rapid, convergent evolution in the killifish *Fundulus heteroclitus*
Rating: Very Good

REVIEW:

What is the intellectual merit of the proposed activity?

By studying the killifish *Fundulus heteroclitus*, this grant tries to determine the genetic basis of evolutionary convergence. The hypothesis framing the proposed research activity is that ancestral polymorphism in key loci are responsible for subsequent parallel adaptive processes in different populations via selective sweeps. The nature of these polymorphisms (protein-coding versus cis-regulatory) and the relative role of selection on standing genetic variation versus selection on de novo mutations will also be evaluated. In the first place, this collaborative effort proposes to generate a genome assembly for *F. heteroclitus* and to perform its annotation. The approaches planned to be used seem appropriate for the proposed goals but there are a few aspects in relation to the use of complementary information (e.g. genome sequence from other species) that are not explained in enough detail. It is also unclear if the PIs have in mind other molecular approaches (e.g. FISH) to orient and order some poorly represented regions in the final genome assembly. The variety of additional information to be anchored to the genome assembly is impressive and, along with the population-level resequencing, it seems reasonable that they use it for the identification of the loci responsible for the convergent adaptive response in different populations. The possibility that this adaptive response can have a polygenic nature, genotype x environment interactions, and that similar adaptive responses might result from different combination of mutations in different populations are not discussed.

What are the broader impacts of the proposed activity?

An annotation training workshop, web-based conferences, and a wiki page are major elements of the broader impact of this grant proposal. Up to 40 students (including 3 undergraduates) and post-docs will benefit from the training associated with this research proposal. In addition, the genome sequence of *F. heteroclitus* will be deposited in public repositories and the impact of this project for the research community working on *F. heteroclitus* is expected to be remarkable.

Summary Statement

Ambitious collaborative effort using an excellent model for Ecological Genetics. There is a clear unbalance in the amount of preliminary data generated with respect to the assembly and annotation of the reference genome as compared to the various types of information to be anchored to that reference genome.

Review #4

Proposal Number:	1051176
Performing Organization:	La St U
NSF Program:	Evolutionary Genetics
Principal Investigator:	Whitehead, Andrew
Proposal Title:	Collaborative Research: The genomic basis of dramatic, rapid, convergent evolution in the killifish <i>Fundulus heteroclitus</i>
Rating:	Excellent

REVIEW:

What is the intellectual merit of the proposed activity?

Strengths:

Excellent assembly of well established and well accomplished researches in a broad array of research disciplines utilizing *Fundulus heteroclitus* as a model teleosts. An extensive amount of genomic research effort has been accomplished with this species despite the obvious lack of a detailed genome sequencing effort. The current proposal culls these efforts and establishes a solid foundation in which to build a significant genomic resource. *Fundulus heteroclitus* is a highly relevant model species in evolutionary biology and marine environmental genomics. There is high potential for success and significant impact within multiple research communities interested in effort. Research aims are well established and follow a logical and progressive order for defining *Fundulus heteroclitus* genome, mapping defined traits and establishing genome/gene-environmental associations. Aim 2 incorporates data from several laboratories and sets a sound foundation for establishing trait loci with defined phenotypes through integrating OTL, RAD, and EST-linked microsatellite data with sequenced genome. Aim 3 is well designed and provides a mechanism to establish polymorphic nature of tolerant and resistant populations.

Weakness:

Individual aims of the proposal are linked in such that aim 2 and 3 are highly dependent upon success of aim one i.e. sequencing of *Fundulus heteroclitus* genome. If aim one does not proceed as planned with significant (60X) coverage, productivity in aims two and three may be compromised. No alternate approaches are established to ensure success of other parts of the proposal in lieu of progress of aim 1 (this may not be possible). This is inherently dangerous for any proposal. With this said, the strengths of the collaborative research team are extremely strong with extensive experiences in this model, sequencing technologies and assembly of novel genomes. Thus there is a high likely hood of success for all aims.

Regarding chromosomal sequence organization a note of caution that significant interchromosomal recombination and gene rearrangement has occurred between teleosts genomes- syntenic relationships and comparisons made to zebrafish, tetraodon and medaka or other teleosts, may not prove to be fruitful in this instance.

Little mention is given to how gene duplicates with a high degree of sequence similarity/identity will be handled in the assembly, gene finding and annotation- specifically regarding paired end Illumina sequences between (200-350) bp

Proposal focuses solely on mutations in structural and regulatory regions and convergent evolution as the basis for adaptation/tolerance between native and exposed *Fundulus heteroclitus* populations. No mention is given to alternate mechanisms regulating gene expression and genomic structure such as epigenetic modifications.

In aim 3 it is assumed that 'loci that have experienced recent, independent selective sweeps do to similar ecological pressures will share signatures of reduced variation'. Is it not possible that different loci are independently responsible for tolerant phenotypes? Why would it be assumed that all populations have

converged upon a similar/exact mechanism?

Investigators:

All investigators are well qualified to direct the proposed research with training in evolutionary genetics, toxicology, molecular biology, and genomics. Cumulatively investigators have extensive experience with the *Fundulus heteroclitus* model sequencing technologies and genome annotation bioinformatics. All investigators have strong records of publication in the field and will significantly contribute the over goal of the project..

What are the broader impacts of the proposed activity?

Strength:

Excellent BI model, with opportunity to establish community based annotation methods and practices. Out reach to broad research community possibly impacting large numbers of students interested in cutting edge techniques in genomics and bioinformatics.

Weakness:

None

Summary Statement

Excellent assembly of well established and well accomplished researches.

Fundulus heteroclitus is a highly relevant model species in evolutionary biology and marine environmental genomics.

There is high potential for success and significant impact within multiple research communities interested in effort.

Individual aims of the proposal are linked in such that aim 2 and 3 are highly dependent upon success of aim one i.e. sequencing of *Fundulus heteroclitus* genome. If aim one does not proceed as planned with significant (60X) coverage, productivity in aims two and three may be compromised

Review #5

Proposal Number: 1051176
Performing Organization: La St U
NSF Program: Evolutionary Genetics
Principal Investigator: Whitehead, Andrew
Proposal Title: Collaborative Research: The genomic basis of dramatic, rapid, convergent evolution in the killifish *Fundulus heteroclitus*
Rating: Excellent

REVIEW:

What is the intellectual merit of the proposed activity?

This proposal will examine the genetic basis of adaptive evolution of pollution tolerance in the killifish. Multiple independent populations of the killifish have evolved tolerance of high pollution levels, and other populations are highly susceptible to this pollution. This has been a very rapid evolutionary change. The proposed work will use a genome of the killifish combined with QTL mapping of RAD SNPs to identify the potential target genes of selection. Crosses of susceptible and tolerant lines will be used in this analysis. The project will also be able to explore the role of convergence in generating this tolerance and the relative role of standing genetic variation versus new mutations in this adaptation.

The proposed work does a good job of examining both regulatory and protein variation elements, and the PIs appreciate the difficulty of this. Production of the killifish genome seems achievable and will open up many new avenues of research. The logistics of the genome assembly and annotation are well outlined. The proposed work integrates across many labs and approaches and this is a positive aspect. The PIs also acknowledge and have a plan for dealing with sequencing errors. This is a very solid and well thought out proposal likely to produce exciting and important results.

What are the broader impacts of the proposed activity?

Broader impacts include a training network for students and post-docs in analytical bioinformatics tools. The killifish genome itself will be a broader impact as it will allow further scientific development of this system.

Summary Statement

Review #6

Proposal Number:	1051176
Performing Organization:	La St U
NSF Program:	Evolutionary Genetics
Principal Investigator:	Whitehead, Andrew
Proposal Title:	Collaborative Research: The genomic basis of dramatic, rapid, convergent evolution in the killifish <i>Fundulus heteroclitus</i>
Rating:	Excellent

REVIEW:

What is the intellectual merit of the proposed activity?

This is the third time I have reviewed this proposal. It is very worthy of support.

The lead-PI has assembled a strong team of first-rate collaborators. Together they show that they have command of the various steps involved in sequencing, assembling and annotating a novel eukaryotic genome and to make this resource available to the larger community.

The PIs have responded to most of the criticisms raised, but have ignored others. I have only a few minor comments:

- There is still the occasional sloppiness in writing (maybe they ran out of time for proof-reading?), but overall the proposal is now more readable and easier to follow.
- It would have been helpful to this reviewer, however, if the PIs had made some effort to present the gene annotation and bioinformatics pipeline in a graphical manner. Similarly, a chart would have made it easier to follow Specific Aim 2.
- BLAST results with an e-value of $e < 10 \times 10^{-10}$ will be considered "significant hits". What is the justification for this threshold, which could be a bit too liberal.
- It would have been useful to learn a bit more about the transcriptome data set: What kind of physiological function is CYP1A involved in? Which tissues were assayed in this experiment?
- The management plan is still a bit weak on specifics.

Despite these little annoyances, all the proposed aims make eminent sense, are very feasible, and will make important contributions to a variety of fields.

What are the broader impacts of the proposed activity?

The Broader Impacts have been improved in that they now are more clearly designed to bring an entirely new generation of young researchers into a growing field. There is some emphasis on underrepresented groups in this regard. Outreach activities are not described in any detail.

Summary Statement

The proposal by Whitehead and collaborators, the second revision of an earlier submission, proposes the whole genome sequencing of the small teleost fish *Fundulus heteroclitus*, which has become a useful model system in studies of natural variation, adaptation and environmental toxicology. Overall, this is a strong proposal by a competent team of first-rate collaborators.