Regarding NSF panel review of Proposal 1120512.

Criticism 1: No funds for students and post-docs: "Perhaps the biggest concern raised, particularly with the ambitiousness of the proposal, is why there are no funds budgeted for graduate students or post-docs."

The funds that we requested are sufficient to complete the research objectives, and in fact favors funding intensive training of dozens of students and post-docs (through the workshop) rather than extended training of only a couple. Funds are requested by Wes Warren's group for sequencing and assembly – their personnel costs are a portion of their operating costs included in their budget. Their group has extensive experience sequencing and assembling genomes, and has budgeted for this project with full knowledge of what it will take to complete this portion of the proposed research. Funds are requested by Shaw/Colbourne's group for organizing the bioinformatics pipeline, including support for the bioinformaticians that are experts at doing this. Their group has used the same people and pipeline for two daphnia genomes and the pea aphid genome, and has budgeted appropriately given their extensive experience. Therefore, sufficient funds are budgeted for sequencing, assembly, and bioinformatics. What remains is funding for hypothesis testing. Funds are requested for a total of 18 months salary among the PIs. Funds are requested for the training of dozens of post-docs and students during the annotation workshop. These funds are sufficient for successful completion of the project. PIs may access additional sources of funds (fellowships, TA-ships, etc…) for additional students and post-docs who will work on the many projects, beyond those proposed here, and for which this study will provide the foundation.

Criticism 2: The project is too ambitious: "The panel still felt that the proposed research was extremely ambitious."

It appears that this criticism is leveled at our sequencing and assembly strategy. Wes Warren's group has direct experience in every aspect of our proposed sequencing and assembly strategy. His group has participated in the sequencing and assembly of many published genomes (zebra finch, frog, platypus, leaf-cutter ant, nematode, etc.). They have successfully created and paired-end sequenced small and large insert libraries, including the proposed 40-kb insert libraries, which is stated in section G.1.1 (a criticism of reviewer #3). They have sequenced and assembled de novo the chicken and pigeon genomes using the proposed strategy; this is summarized in section G.1.1 (Reviewer #3 seemed to question whether we had experience with de novo genome assemblies). Our timeline estimate was based on the Warren group's direct experience, and was conservative. In Warren's group's experience, library construction takes two weeks; we allow for six weeks. In their experience, the paired-end sequencing of libraries, which will total only 7 Illumina lanes, will require two weeks; we allow for six weeks. In their experience, assembly of a 1.5 Gb genome will require one month; we allow for three months. This is specifically outlined in section G.1.1. Because the technology is advancing so rapidly, we understand how our proposal appears ambitious since it takes several years for state-of-the-art

workflows to become common knowledge through the literature, by which time the state-of-the-art has progressed. That is why we emphasized our group's current and hands-on experience with these techniques in section G.1.

Criticism 3: The sequencing strategy may not produce a quality assembly: "There was some concern that the genome sequencing would fail to produce a quality assembly"

Wes Warren's group has direct experience with this sequencing and assembly strategy, and has demonstrated that it is appropriate and feasible. We described his unpublished data showing this in section G.1.1. Other un-published research from Dr. Warren's colleagues at other world-class genome sequencing centers are already using this strategy successfully for sequencing and de novo assembly of large complex eukaryote genomes, including primate genomes. We cited the Li et al (2010 Nature) paper that assembled a high-quality panda genome using all Illumina sequence, even without our proposed pair-end sequencing of 40 kb inserts. We also cited Gnerre et al (PNAS early edition) which developed and applied the ALLPATHs-LG assembler that produces high-quality de novo assemblies using all second-generation sequence for large and complex mammalian genomes. These techniques are cutting-edge, where results are just starting to be widely published (2010, 2011), so it is perhaps not surprising that reviewers are still cautious about the accessibility of genome sequencing for non-traditional models. However, such projects appear to be routinely funded by the NIH, and we hope that these techniques and strategies will soon become more widely recognized by NSF reviewers so that basic science in non-traditional models is brought up to speed.

Criticism 4: Community-based annotation would not be reliable: "There was also some concern over whether the community-based annotation of the genome would produce reliable annotations, and justification for the quality of work produced from this approach would be useful."

The source of this criticism is unclear. The NSF does not allow letters of support from the community (a criticism of Reviewer #3). However, the large community of researchers that has coalesced around the development of genomic resources for Fundulus was outlined in the first submission of this grant. Since that first review panel clearly recognized the large community, we removed this list from subsequent submissions (from the August 2009 panel summary: "The acquisition of a robust annotated genome benefits a large community of researchers across a diverse array of scientific disciplines"). We state in the proposal that our annotation strategy is proven (citing communities that similarly annotated genomes of fruitfly, mouse, sea urchin, skate, etc.) and that it has been successful in our hands (citing the daphnia, pea aphid, and wasp genome projects). It is unclear what additional justification could be more convincing or helpful to reviewers.