Introduction

• Atlantic Wood Industries is a superfund site on the Elizabeth River (VA), that was once used for wood processing and waste disposal and is highly contaminated with heavy metals and polycyclic aromatic hydrocarbons (PAH).

• One species that has been able to adapt to this highly polluted environment is the Atlantic killifish (*Fundulus heteroclitus*), a small brackish water fish.

• This study investigated the genomes of *F. heteroclitus* populations collected from the Atlantic Wood Industries site and nearby sites with different levels of PAH pollution to identify potentially adaptively important genes.

Purpose and Methods

• Fish were collected, and PAH concentrations were recorded from various sites in the Elizabeth River (Fig 1) by Dr. DiGiulio’s lab (Duke University).

• The fish were sequenced with RAD-seq (Fig 2), and the sequences were deposited in the National Center for Biotechnology Information.

• Computational approaches were used to call Single Nucleotide Polymorphisms (SNPs), determine pairwise *F*<sub>ST</sub> values, and find genes of interest.

• Significant SNPs were those that had high *F*<sub>ST</sub> values (P<0.01), in the two most highly polluted sites (Republic and Atlantic Wood) and the least polluted site (Mains).

Results

• RAD-seq identified ~100,000 SNPs among 160 individuals in 5 populations.

• After filtering, 10,242 SNPs in 148 individuals were analyzed.

• 66 loci showed significant *F*<sub>ST</sub> values (P<0.01) in the two most highly polluted sites (Republic and Atlantic Wood) compared to the least polluted site (Mains, Fig 4).

• 39 SNPs remained significant after a multiple test correction.

Discussion and Conclusions

• From the 39 loci, 21 genes are associated with those sites. Of the genes found to be under selection because of pollution, two of them (ahr<sup>2</sup> and lrrc49) are involved in detoxification of pollutants.

• The genes kmt2cb<sup>7</sup> and eya1<sup>8</sup> have higher allele frequencies in response to pollution.

• *F. heteroclitus* spend their embryonic stage on the beds of rivers and estuaries. Because the pollutants found at Atlantic Wood Industries are found in the sediments, embryonic fish are the most exposed to them, and two developmentally important genes, tshz1<sup>6</sup> and rfnr<sup>2</sup> were identified as potentially adaptively important.

• In polluted sites, some genes may be fixed which can reduce genetic diversity in these populations.

Future Research

• Further investigation into genes associated with pollution at different superfund sites would help understand the genetic components of adaptation to environmental degradation.

• Future studies analyzing the phenotypic response to selection associated with these genes would create a better understanding of the importance of these genes.

References


Acknowledgements

I would like to thank my advisors Dr. Douglas Crawford and Dr. Marjorie Oleksiak for giving me advice at every step of this research project. I would also like to thank the graduate students Melissa Drown, Amanda DiBerard and Samantha Sierra-Marín for their encouragement and insights into bioinformatics. Next, I want to thank the online bioinformatics community for making computational biology more accessible to scientists such as myself.

Fig 1. Population Genetics of Fundulus heteroclitus in Variably Polluted Sites

Fig 2. The RAD-seq method, in this process the DNA is cut at restriction sites and put into a database of genomes.

Fig 3. Visualization of which SNPs were used to find genes under selection. The loci in the green part of the Venn Diagram are expected to have high *F*<sub>ST</sub> because of pollution.

Fig 4. Fat per Locus between Atlantic Wood and Mains. Loci are considered significant if they have an FDR corrected P-value less than 0.01. The red points are loci with large F<sub>ST</sub> values that wouldn’t be due to random drift between populations.

Fig 5. Principal Component Analysis of Non-significant Loci. This principal component analysis used the 10,203 loci that were not under selection and found that populations tend to cluster based on location.

Fig 6. Principal Component Analysis of Significant Loci. This principal component analysis used the 39 loci that were positively adaptively important showing a clustering based on pollution level.

Fig 7. Minor allele frequency over log(PAH Concentration) for the six sites that were found to be different between populations. In these genes we can see a correlation between allele frequency and pollution level.


