Preliminary GIMAP Gene Family Analysis May 6th, 2018

QUESTION 1: Is it likely that the assembler was correct in classifying all of the GIMAP genes differently?

METHODS:

1. Gather exon information for all 55 GIMAP genes identified in the genome
   1. Compare intron exon boundaries of each - are some across different genes exactly the same, suggesting they should have been called as the same gene?

QUESTION 2: What are the mechanisms of GIMAP gene family diversification

METHODS:

1. Create multiple sequence alignments of all GIMAP RNA sequences that code for GIMAP proteins and all amino acid sequences using ClustalW and graph using MSA in R
   1. May 5, 2018: Gather all XP for GTPase IMAP in the reference GCA..protein.faa and use these XPs to pull out all of the GIMAP sequences. Then use the gene ID
   2. Graph the multiple alignments in R using MSA (DONE)
2. Create phylogenetic trees using RAxML with 500 bootstrap replicates and other default parameters (DONE)

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

GIMAP Family Analysis Next Steps: From Committee Meeting May 16th, 2018

TO-DO LIST

Random:

* Tell Marta on Tuesday MAY 22nd, 2018 to tell Kate to set up my account with Cayoos so that I can download the packet
* Remember to give my committee the more detailed criteria for the grant
* Only need to write the **program abstract and the Narrative**

1. **TO DO AFTER PROPOSAL IS WRITTEN:** Analyze the quality of the genome assembly for GIMAP. Were different alleles called as different genes on accident?
   1. TASK 1: Map genome resequencing raw read files to the reference genome. Compare where the reads for “tandem duplications” map back to.
      1. If half map to one and half map to the other that would be suspicious and indicate that they only mapped to one
   2. TASK 2: Map genome reads from genome SRA back to itself on the genome assembly to check for the same things
   3. TASK 3: Map GIMAP reads from eastern oyster to the long reads for the eastern oyster and see if they map in the same place, with the same intron exon boundaries and sequences in between the introns and exons
2. **For Paper 1 Population Level Paper/Preliminary Data for Proposal**
   1. TASK 1 (more for paper): Need to use a sequence based approach to find all of the apoptosis family genes across all of the genome resequencing data across the population range
      1. This sequence based approach will mean using HMMER to find the proteins using their conserved sequence and confirming those hits with BLAST
   2. **TOP PRIORITY** TASK 2 (for preliminary data): Get the full number of all the GIMAP genes from across the re-sequenced genomes and compare
      1. Compare whether there are specific patterns of GIMAP across all of the resequenced genomes or if there are population specific patterns. These would both be interesting, however if there are no patterns at all that would be concerning
   3. TASK 3: Compare the full number of GIMAP genes in C. virginica and closely related species
3. For Proposal: PROPOSED TOPIC 1 PERFORMING CHALLENGE IN INDIVIDUALS
   1. TASK 1: Present Pilot Study Preliminary Data from Dermo challenge in oysters and their apoptotic responses ? (not going to happen in time)
   2. TASK 2: Present preliminary GIMAP data from Above
   3. TASK 3: Add low coverage genome sequencing to the project so that we can tell which of the overall gene repertoire in these oysters is actually being expressed in the qRT-PCR assay
   4. TASK 4: **TIP:** During the course of your literature review, create a list of key words that describe the field, as well as thennames of investigators who have been active in the area during the last ten years.
   5. TASK 5: Decide the number of individuals I would like to sample and how many populations they are going to be from. Are they going to be from different populations as well? (No they won’t be from different populations”
   6. TASK 6: Include preliminary data from the transcriptomic study I did. To do this I need to compare the amino acid sequences from the different challenges and characterize which are alike and which are different
   7. TASK 7: Include preliminary data with a comparison of the GIMAP gene families between different closely related species (not critical for the preliminary data)
   8. TASK 8: Include apoptosis diagram
   9. TASK 9: Finish Apoptosis gene confirmation