**GIMAP Gene Family Descriptive Statistics**

Erin Roberts

PhD Student

Gomez-Chiarri Lab

May 15, 2018

\*\*Data was gathered by first searching (using grep command in bash) for all “GTP IMAP” genes in the GCA\*rna.gz reference genome. Exons for each gene were found by searching the LOC GeneID in the .gff annotated reference genome for the eastern oyster.

|  |  |
| --- | --- |
| **Table 1: Total Genes** |  |
| *GIMAP GENES* | 53 |
| *GIMAP4* | 42 |
| *GIMAP7* | 9 |
| *GIMAP8* | 2 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Table 2: GIMAP Genes per chromosome** |  |  |  |  |
| *Chromosome* | *Number of GIMAP Genes on CHR* | *Number of GIMAP 4* | *Number of GIMAP 7* | *Number of GIMAP 8* |
| CHR2 | 3 | 1 | 0 | 2 |
| CHR4 | 5 | 5 | 0 | 0 |
| CHR5 | 1 | 1 | 0 | 0 |
| CHR6 | 1 | 1 | 0 | 0 |
| CHR7 | 10 | 10 | 0 | 0 |
| CHR8 | 25 | 18 | 7 | 0 |
| CHR9 | 8 | 6 | 2 | 0 |
| TOTAL | 53 | 42 | 9 | 2 |

|  |  |  |
| --- | --- | --- |
| **Table 3. Exons Per GIMAP gene Per Chromosome (cont. on next page)** | | |
| *Gene* | *Number* | *Chromosome* |
| LOC111120314 | 3 | CHR2 |
| LOC111119582 | 48 | CHR2 |
| LOC111119581 | 3 | CHR2 |
| LOC111129932 | 9 | CHR4 |
| LOC111129930 | 9 | CHR4 |
| LOC111130155 | 9 | CHR4 |
| LOC111130153 | 11 | CHR4 |
| LOC111132212 | 2 | CHR5 |
| LOC111100020 | 9 | CHR6 |
| LOC111102552 | 4 | CHR7 |
| LOC111103040 | 8 | CHR7 |
| LOC111105336 | 4 | CHR7 |
| LOC111105335 | 4 | CHR7 |
| LOC111105339 | 8 | CHR7 |
| LOC111105333 | 43 | CHR7 |
| LOC111103458 | 4 | CHR7 |
| LOC111103460 | 3 | CHR7 |
| LOC111104840 | 12 | CHR7 |
| LOC111103161 | 3 | CHR7 |
| LOC111108760 | 6 | CHR8 |
| LOC111106989 | 29 | CHR8 |
| LOC111110635 | 3 | CHR8 |
| LOC111106328 | 4 | CHR8 |
| LOC111106343 | 4 | CHR8 |
| LOC111109878 | 5 | CHR8 |
| LOC111105744 | 4 | CHR8 |
| LOC111108559 | 2 | CHR8 |
| LOC111109668 | 3 | CHR8 |
| LOC111109667 | 7 | CHR8 |
| LOC111108253 | 4 | CHR8 |
| LOC111110321 | 6 | CHR8 |
| LOC111109557 | 8 | CHR8 |
| LOC111108220 | 2 | CHR8 |
| LOC111105930 | 1 | CHR8 |
| LOC111109343 | 1 | CHR8 |
| LOC111109344 | 1 | CHR8 |
| LOC111106079 | 1 | CHR8 |
| LOC111106081 | 3 | CHR8 |
| LOC111109357 | 3 | CHR8 |
| LOC111109358 | 3 | CHR8 |
| LOC111109853 | 2 | CHR8 |
| LOC111109737 | 2 | CHR8 |
| LOC111108121 | 80 | CHR8 |
| LOC111107002 | 4 | CHR8 |
| LOC111110115 | 3 | CHR8 |
| LOC111111775 | 12 | CHR9 |
| LOC111113702 | 9 | CHR9 |
| LOC111115664 | 5 | CHR9 |
| LOC111112005 | 3 | CHR9 |
| LOC111111454 | 3 | CHR9 |
| LOC111115902 | 2 | CHR9 |
| LOC111110875 | 5 | CHR9 |
| LOC111111241 | 62 | CHR9 |

|  |  |  |
| --- | --- | --- |
| **Table 4: GIMAP Gene List** |  |  |
| Gene ID | Gene Name | Chromosome |
| LOC111120314 | GTPase IMAP family member 8-like | CHR2 |
| LOC111119582 | GTPase IMAP family member 4-like | CHR2 |
| LOC111119581 | GTPase IMAP family member 8-like | CHR2 |
| LOC111129932 | GTPase IMAP family member 4-like | CHR4 |
| LOC111129930 | GTPase IMAP family member 4-like | CHR4 |
| LOC111130155 | GTPase IMAP family member 4-like | CHR4 |
| LOC111130153 | GTPase IMAP family member 4-like | CHR4 |
| LOC111108760 | GTPase IMAP family member 4-like | CHR4 |
| LOC111132212 | GTPase IMAP family member 4-like | CHR5 |
| LOC111100020 | GTPase IMAP family member 4-like | CHR6 |
| LOC111102552 | GTPase IMAP family member 4-like | CHR7 |
| LOC111103040 | GTPase IMAP family member 4-like | CHR7 |
| LOC111105336 | GTPase IMAP family member 4-like | CHR7 |
| LOC111105335 | GTPase IMAP family member 4-like | CHR7 |
| LOC111105339 | GTPase IMAP family member 4-like | CHR7 |
| LOC111105333 | GTPase IMAP family member 4-like | CHR7 |
| LOC111103458 | GTPase IMAP family member 4-like | CHR7 |
| LOC111103460 | GTPase IMAP family member 4-like | CHR7 |
| LOC111104840 | GTPase IMAP family member 4-like | CHR7 |
| LOC111103161 | GTPase IMAP family member 4-like | CHR7 |
| LOC111106989 | GTPase IMAP family member 7-like | CHR8 |
| LOC111110635 | GTPase IMAP family member 4-like | CHR8 |
| LOC111106328 | GTPase IMAP family member 4-like | CHR8 |
| LOC111106343 | GTPase IMAP family member 4-like | CHR8 |
| LOC111109878 | GTPase IMAP family member 4-like | CHR8 |
| LOC111105744 | GTPase IMAP family member 4-like | CHR8 |
| LOC111108559 | GTPase IMAP family member 7-like | CHR8 |
| LOC111109668 | GTPase IMAP family member 7-like | CHR8 |
| LOC111109667 | GTPase IMAP family member 4-like | CHR8 |
| LOC111108253 | GTPase IMAP family member 4-like | CHR8 |
| LOC111110321 | GTPase IMAP family member 7-like | CHR8 |
| LOC111109557 | GTPase IMAP family member 7-like | CHR8 |
| LOC111108220 | GTPase IMAP family member 7-like | CHR8 |
| LOC111105930 | GTPase IMAP family member 4-like | CHR8 |
| LOC111109343 | GTPase IMAP family member 4-like | CHR8 |
| LOC111109344 | GTPase IMAP family member 4-like | CHR8 |
| LOC111106079 | GTPase IMAP family member 4-like | CHR8 |
| LOC111106081 | GTPase IMAP family member 4-like | CHR8 |
| LOC111109357 | GTPase IMAP family member 4-like | CHR8 |
| LOC111109358 | GTPase IMAP family member 4-like | CHR8 |
| LOC111109853 | GTPase IMAP family member 4-like | CHR8 |
| LOC111109737 | GTPase IMAP family member 4-like | CHR8 |
| LOC111108121 | GTPase IMAP family member 7-like | CHR8 |
| LOC111107002 | GTPase IMAP family member 4-like | CHR8 |
| LOC111110115 | GTPase IMAP family member 4-like | CHR8 |
| LOC111111775 | GTPase IMAP family member 4-like | CHR9 |
| LOC111113702 | GTPase IMAP family member 4-like | CHR9 |
| LOC111115664 | GTPase IMAP family member 4-like | CHR9 |
| LOC111112005 | GTPase IMAP family member 7-like | CHR9 |
| LOC111111454 | GTPase IMAP family member 7-like | CHR9 |
| LOC111115902 | GTPase IMAP family member 4-like | CHR9 |
| LOC111110875 | GTPase IMAP family member 4-like | CHR9 |
| LOC111111241 | GTPase IMAP family member 4-like | CHR9 |

Figure 1: Phylogenetic Tree of GIMAP Amino Acid Sequences. Amino acid sequences were gathered from the GCA\*rna.gz file and were then aligned using ClustalW with default settings. Amino acid sequences from individual chromosomes, particularly 4,7 and 8 all tend to cluster together.

Figure 2: GIMAP Phylogenetic Tree of Amino Acid Sequences. This tree was generated in using using the program MSA. Default ClustalW parameters were used to generate the tree and a neighbor joining algorithm in MSA was employed to generate the distance matrix. No bootstrapping was performed. Currently RAxML is being run on the URI cluster to perform bootstrap replicates.

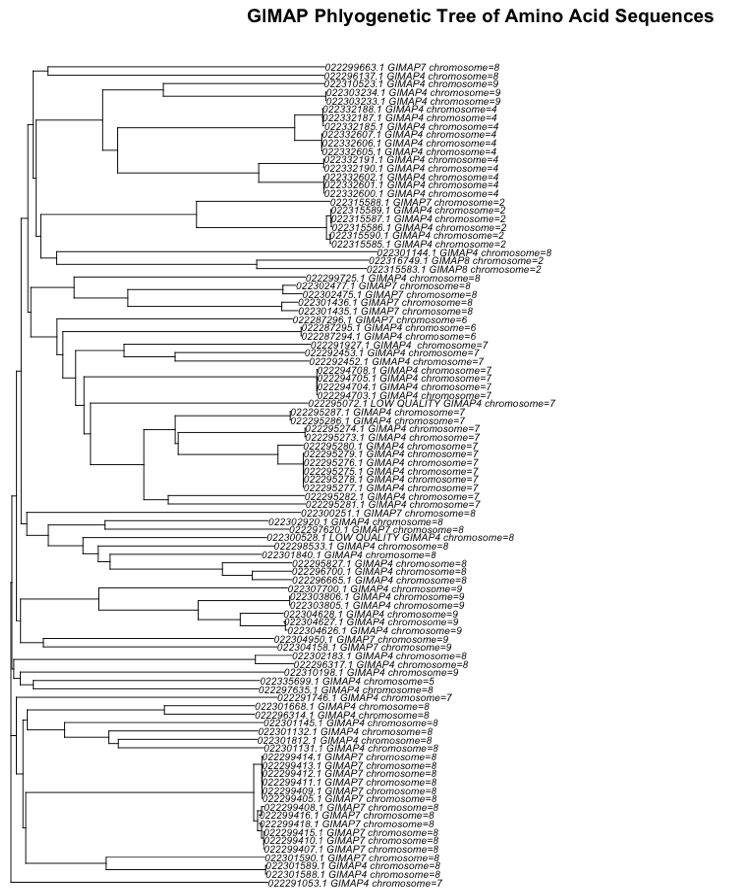


Figure 3: IGV comparison of two adjacent GIMAP 4 genes on Chromosome 7. Identical DNA sequence and amino acid sequence.

