

## 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.

<b>Table 1: Total Genes</b>	
<i>GIMAP GENES</i>	53
<i>GIMAP4</i>	42
<i>GIMAP7</i>	9
<i>GIMAP8</i>	2

<b>Table 2: Genes per chromosome</b>				
<i>Chromosome</i>	<i>Number of Total Genes</i>	<i>Number of GIMAP 4</i>	<i>Number of GIMAP 7</i>	<i>Number of GIMAP 8</i>
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

<b>Table 3. Table 3. Exons Per GIMAP gene Per Chromosome (cont. on next page)</b>		
<i>Gene</i>	<i>Number</i>	<i>Chromosome</i>
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

<b>Table 4: GIMAP Gene List</b>		
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.

GIMAP Chr7	
NC_035786.1	NC_035786.1
NC_035786.1:38086481-38093250	NC_035786.1:38056270-38063543
<p>CGTACTTCATTGTTGTTTCACATCTGgaatattatacaaaattgaaaaaaac</p> <p>V L H C S H L E Y L Y K L K K T</p> <p>S Y F I V V H I W N I F I Y T N I E K N</p> <p>R T S L L F T S G I F I Q I E K N</p>	<p>TTCGTTGTTGTTCCACATCTGgaatattatacaaaattgaaaaaaac</p> <p>S L L S T S G I F I Q I E K K L</p> <p>L R C C P H L E Y L Y K L K K N</p> <p>L F V V V H I W N I Y T N K K T</p>
7	7
LOC111105335	LOC111105336
XM_022439573.1	XM_022439574.1