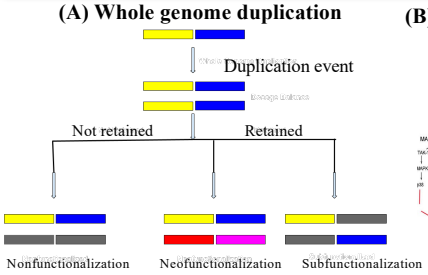
A Computational Pipeline for Investigating Protein Pathway Diversification Post-Genome Duplication

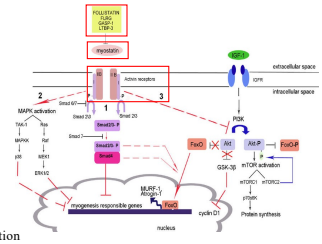
Laura McDonnell (Andrew ID: lmcdonne)
Ray and Stephanie Lane Computational Biology Department, Carnegie Mellon University

Background

(A) Whole genome duplication



(B) TGF-Beta Signaling Pathway



(C) Species Tree

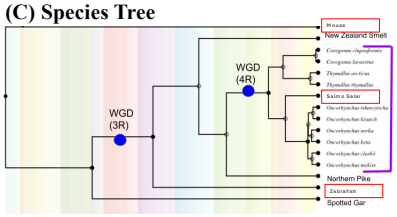


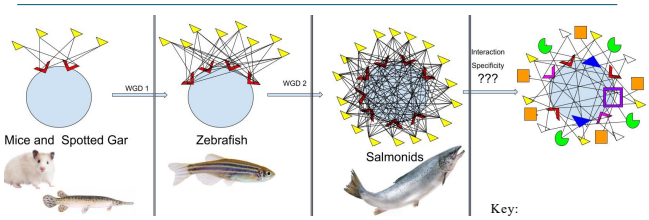
Figure 1: (A) Whole Genome Duplication (WGD) event followed by loss function of one of the copies (nonfunctionalization), OR followed by the retention of these genes by changing the function of one of the copies (neofunctionalization) or the two copies sharing parts of the original function (subfunctionalization).

(B) Myostatin is a protein that functions as a skeletal muscle growth inhibitor and is a part of the TGF-Beta Signaling Pathway. When Myostatin is knocked out, it leads to excess growth of skeletal muscles.

(C) The species tree of interest, with two recent WGD's marked on the branches and salmonids marked by the purple bracket

Hypothesis

Did the mechanisms involved in maintaining duplicated genes after consecutive whole genome duplication events in the Salmonid lineage lead to pathway diversification and increased specificity of the interactions between molecules in the myostatin pathway?



1 copy of each gene:	2 copies of each gene:	4 copies of each gene:
Activin A Receptor A	Activin A Receptor IA	8 Types of Activin A Receptors
Activin A Receptor B	Activin A Receptor IIA	Myostatin 1
Myostatin	Activin A Receptor IB	Myostatin 2
	Activin A Receptor IIB	Myostatin 3
	Myostatin I	Myostatin 4
	Myostatin II	

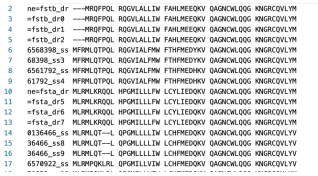
Figure 2: A representation of the effect consecutive WGDs had on myostatin and its interacting partners and the species that represent each pathway.

Phylogenetic Analysis Pipeline and Results

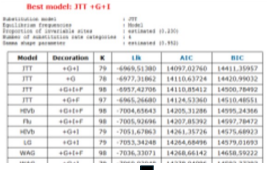
1. Find Homologs (BLASTP)
- Query myostatin genes (and interacting partners) in mouse, zebrafish, and salmon genomes

Myostatin a
Follistatin b
Activin A receptor IB
Activin A receptor IIB

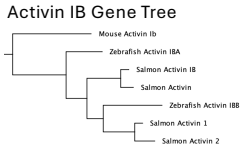
2. Multiple Sequence Alignment (MAFFT)
- Align matches for each gene

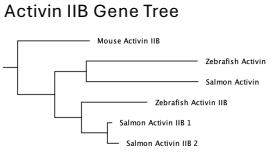


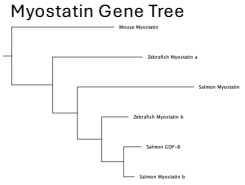
3. Model Selection (PhyML)
- Estimate maximum likelihood phylogenies

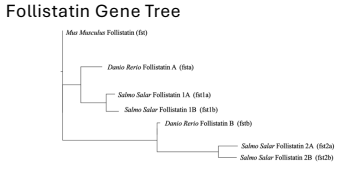


4. Gene Tree Reconstruction
- Construct and reconcile gene trees

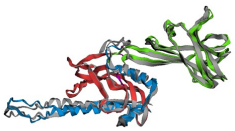






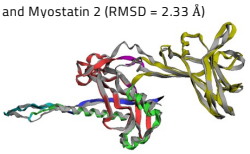


5. Structural Alignment (FATCAT)
- Visualize superposition of structures
- Determine degree of structural divergence



Superposition of salmonid Myostatin 1a and Myostatin 2 (RMSD = 2.33 Å)

6. Species Specific Docking (Replica Dock)
- Investigate interactions between interacting partners



Superposition of salmonid Myostatin 1a and Myostatin 1b (RMSD = 1.79 Å)

Conclusion & Next Steps

- Using the high throughput method of building models with FATCAT, the preliminary results show that there is significant structural similarity between Atlantic salmon myostatin 2 with the myostatin 1a. Additionally, myostatin 1a and 1b do not show structural divergence.
- Future work involves more detailed structural modeling as well as modeling of all interacting partners, coupled to docking to look at potential interaction changes and a gain of specificity.
- Further future work involves detailed (dN/dS-based) selection analysis on myostatin and all interacting partners, which when coupled with the protein structural analysis, will illuminate the patters of pathway and protein regulatory diversification.

References

Altschul, S.F., Gish, W., Miller, W.., Myers, E.W. and Lipman, D.J. (1990). "Basic local alignment search tool." *Journal of Molecular Biology,* **215**: 403-410.

Elkina, Y., von Haehling, S., Anker, S. D. and Springer, J. (2011). The role of Myostatin in muscle wasting: an overview. *J Cachexia Sarcopenia Muscle,* **2**(3): 143-151.

Guindon, S., Dufayard, J.F., Lefort, V., Anisimova, M., Hordijk, W. and Gascuel, O. (2010). New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0. *Systematic Biology,* **59**(3): 307-321.

Katoh, Kuma, Toh, Miyata. (2005). MAFFT version 5: improvement in accuracy of multiple sequence alignment. *Nucleic Acid Res.* **33:** 511- 518.

Lawson CB, Niino T, Hermansen RA, Brok-Volchanskaya V, Jackson MF, Garikipati DK, Liberles DA, Rodgers BD. (2012). The salmonid myostatin gene family: a novel model for investigating mechanisms that influence duplicate gene fate. *BMC Evol Biol.* **12:**202.

Ye, Y., Godzik A. (2004). FATCAT: A web server for flexible structure comparison and structure similarity searching. *Nucleic Acids Research*, **32:** W582-W585.