Genomic analysis of the tryptome reveals molecular mechanisms of gland cell evolution

Principal Investigator: Leslie S. Babonis

Draft or Version Number: v.0.1

08 Jan 2019

LIST OF ABBREVIATIONS

<Insert text>	<Insert text>
Nvec	Nematostella vectensis
Elin	Edwarsiella lineata
Adig	Acropora digitifera
Rren	Renilla renilla
Hmag	Hydra magnipapillata
Ccrux	Calvadosia cruximellitensis
Bflo	Branchiostoma floridae
Mlei	Mnemiopsis leidyi
Avan	Atolla vanhoeffeni
Aala	Alatina alata

1 INTRODUCTION: BACKGROUND INFORMATION AND SCIENTIFIC RATIONALE

1.1 BACKGROUND INFORMATION

Nematostella has nearly 100 trypsin-domain containing genes. Is this a case of Nematostella-specific duplication/divergence or was this trypsin diversity present in the last common ancestor of all animals? If it is Nematostella-specific, do we see duplication and divergence of these genes in other animal lineages as well? How many trypsin genes were in the common ancestor of animals? How many were in the common ancestor of cnidarians/bilaterians? How many were in the common ancestor of Anthozoans and Edwarsidae?

1.2 RATIONALE

To understand how new cell funcitons arise, we need to understand how new genes arise and how they get parsed into different cell types/lineages. We know the expression patterns of several of the trypsin genes in Nematostella but we don't know whether the ancestor of all trypsins had multiple functions/expression patterns that have been partitioned as this gene family duplicated. I.e., do we see evidence of subfunctionalization in the trypsin genes in Nematostella?

1.3 HYPOTHESES

1. The common ancestor of animals had only few trypsins and this gene family has undergone massive expansion in multiple lineages of cnidarians (including Nematostella).

2. Sister trypsins in Nematostella are more likely to be expressed in the same cell/tissue type than more distantly related trypsins.

1.4 OBJECTIVES

1. To characterize the evolutionary history of trypsin genes in animals as a means to understand how gene duplication and divergence gives rise to specialized cell function.

2. To combine expression data and evolutionary history to make inferences about ancestral functions of individual trypsin genes and determine whether subfunctionalization and/or neofunctionalization have influenced the evolution of cell type specialization in Nematostella.

2 STUDY DESIGN AND ENDPOINTS

We will build a phylogeny of all the trypsin-domain containing genes from representative animal taxa to reconstruct the origin/diversification of this gene family. We will then plot onto this phylogeny the known expression patterns of trypsins from Nematostella (and other taxa, as available). From these data we will make inferences about the relationship between duplication events and changes in gene expression. We will generate a tree of trypsins from ctenophores, sponges, Trichoplax, cnidarians, and bilaterians to determine how many trypsins were present in the common ancestor of all animals and determine if there have been multiple lineage-specific expansions of this gene family throughout the evolution of animals. We will then build a cnidarian-only phlogeny with greater taxon sampling to more thoroughly reconstruct the cnidarian ancestor and determine whether the expansion of trypsin genes in Nematostella occurred before the diversification of cndiarians. Taxa for the cnidarian-only phylogeny: N. vectensis (corrected trypsin sequences only), Edwardisiella lineata, Acropora digitifera, Renilla renilla, Hydra magnipapillata, Atolla vanhoeffeni, Alatina alata, and Calvadosia cruximellitensis.

##################################Commands#####################################

#run hmmscan in: /bwdata1/babonis/04-Gland_cells/06-HMM_SCAN

hmmscan --domtblout Rren_scan_domtab.tab /usr/local/pfam/Pfam-A.hmm /bwdata1/babonis/04-Gland_cells/Reni_reni.v2.aa &

perl hmmscan_parse.pl ../Rren_scan_domtab.tab > Rren_parse.out

#run HMM2ALN and trees in: /bwdata1/babonis/04-Gland_cells/19-CNIDARIAN_ONLY_TREE

hmm2aln.pl --hmm=../01-DATA/Trypsin.hmm --name=flarff --fasta_dir=01-FASTA_DIR --threads=1 > zookie

mv zookie cnidarian_tryp.aln.fa

Gblockswrapper cnidarian_tryp.aln.fa

perl -pi -e 's/ //g' cnidarian_tryp.aln.fa-gb

perl -pi -e 's/\|/_/g' cnidarian_tryp.aln.fa-gb

raxmlHPC-PTHREADS-SSE3 -d -T 44 -p 12345 -m PROTGAMMAVT -s cnidarian_tryp.aln.fa-gb -n cnid_tryp_boots -# 1000 -x 54321 &raxmlHPC-PTHREADS-SSE3 -T 12 -p 12345 -# 25 -m PROTGAMMAVT -s cnidarian_tryp.aln.fa-gb -n cnid_tryp_mp &

raxmlHPC-PTHREADS-SSE3 -d -T 12 -p 12345 -# 25 -m PROTGAMMAVT -s cnidarian_tryp.aln.fa-gb -n cnid_tryp_rt &

iqtree-omp -nt AUTO -s cnidarian_tryp.aln.fa-gb -m VT -pre cnid_tryp_iqVT &

raxmlHPC -f e -m PROTGAMMAVT -t cnid_tryp_iqVT.treefile -s cnidarian_tryp.aln.fa-gb -n iq.compare

raxmlHPC -m PROTGAMMAVT -p 12345 -f b -t RAxML_bestTree.cnid_tryp_XX -z RAxML_bootstrap.cnid_tryp_boots -n RAxML_bestTree.cnid_tryp_XX_bootstraps_applied

3 WORK COMPLETED SO FAR W DATES

Trypsin genes identified by HMMSCAN – April 2018

Original trypsin alignment made with HMM2ALN – May 2018

RaxML_mp, RaxML_rt, iqTree original run – May 2018 (same for Trypsin_2 tree)

RaxML_mp, RaxML_rt, iqTree for ShK genes – Jun 2018

Trypsin x trypsin blast (Nvec only) – Sept 2018

Manual curation/removal of duplicate gene predictions – Nov 2018

Removed Mlei (long branches), removed Bflo (too many sequences), re-ran tree with bootstraps – Dec 2018

performed HMMSCAN in Renilla reniformis (Rren) and Calvadosia cruximellitensis (Ccrux) – Jan 2019

added Mlei back to animal alignment, re-ran trypsin and trypsin_2 trees – Feb 2019

added Crux and Rren to cnidarian alignment, re-ran trypsin and trypsin_2 trees – Mar 2019

made Nvec only tree by pruning all other taxa from cnidarian trypsin tree – Mar 2019

inference criteria (p-values, bayes factors, model fit indices

Only count/analyze trypsin domains with ≤ 0.05 Eval (per Koch et al 2012)

link to repo with custom scripts

00-README files with commands and custom scripts found in: /bwdata1/babonis/04-Gland_cells

List of custom scripts:

hmmscan_parse.pl, hmm2aln.pl

4 LITERATURE REFERENCES

Koch BJ, Ryan JF, Baxevanis AD (2012) The Diversification of the LIM Superclass at the Base of the Metazoa Increased Subcellular Complexity and

Promoted Multicellular Specialization. PLoS ONE 7(3): e33261. doi:10.1371/journal.pone.0033261

5 PHYLOTOCOL AMENDMENT HISTORY

Version	Date	Significant Revisions