

## Title

Erives AJ, Fassler JS (2015) Metabolic and Chaperone Gene Loss Marks the Origin of Animals: Evidence for Hsp104 and Hsp78 Chaperones Sharing Mitochondrial Enzymes as Clients. PLoS ONE 10(2): e0117192. <https://doi.org/10.1371/journal.pone.0117192>

## Reference

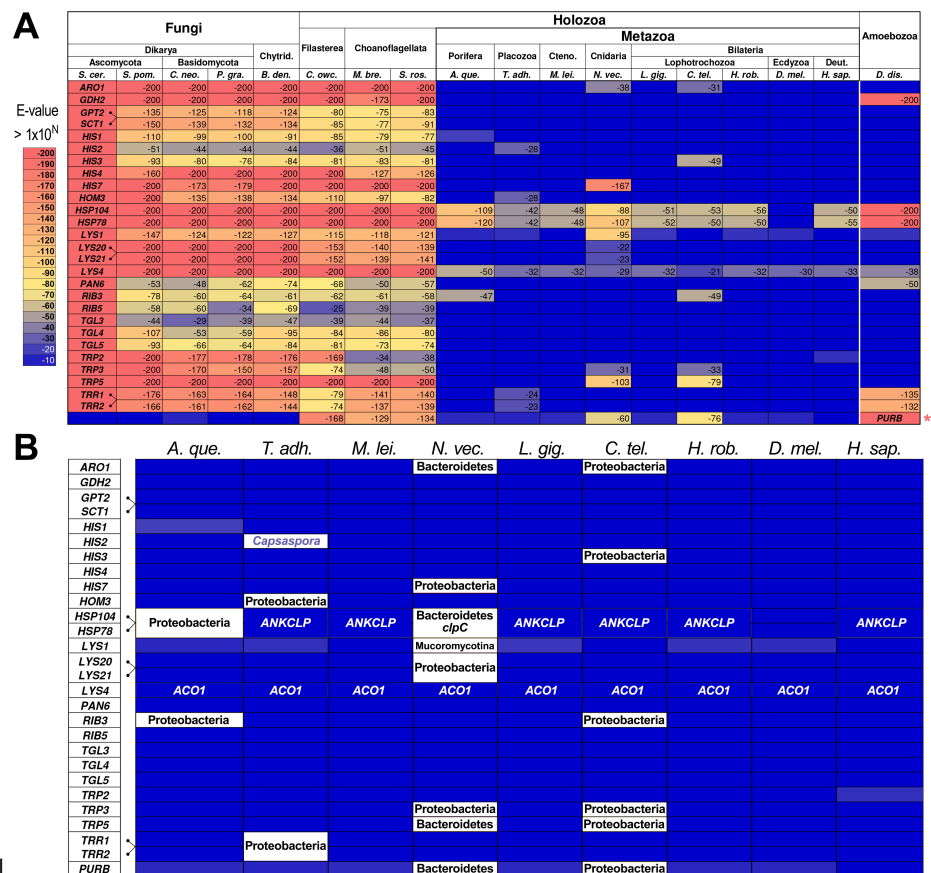
Link Link 2

## Introduction and Abstract

The authors of this paper were interested in seeing which genes were lost in the evolution of animals. While the expansion of genes, particularly transcription factors, in animals has been well-studied, the loss of genes from single-celled eukaryotes to animals has been poorly studied. Understanding which genes were lost is important because it may lend insight into the context, origin, and constraints associated with new transcription programs in animals.

They predicted 27 genes were maintained in non-animal eukaryotes but lost in animals. To test this hypothesis, they used BLASTP to compare the fungal (*S. cerevisiae*) homolog of those genes with 17 other organisms from fungi and holozoa. Their results confirmed their hypothesis. However, they only looked at 7 non-animal genomes and 10 animal genomes. Since that time, many more genomes have been sequenced. This project expands on their finding by using BLASTP to compare the *S. cerevisiae* genes to the 64 genomes in Paps et al., 2018.

## Original Figure



~Figure 1

adhaerens; M. lei., Mnemiopsis leidyi; N. vec., Nematostella vectensis; L. gig., Lottia gigantea; C. tel., Capitella teleta; H. rob., Helobdella robusta; D. mel., Drosophila melanogaster; and H. sap., Homo sapiens. Other abbreviations: Chytridio., Chytridiomycota; Placo., Placozoa; Ctenoph., Ctenophora; Ecdyso., Ecdysozoa; and Chor., Chordata. The last row is from an alternative screen for lost animal genes (see text) in which Amoebozoa PFAM domains that are absent in animals were identified (five genes already identified plus one new gene, PURB). (B) Close-up of the lost animal gene columns from panel A. This table shows either the taxonomic origin of a weak animal match, or the name of the gene if it is not directly related (i.e., not orthologous) to the named yeast gene (first column) for all matches with BLASTP expectation values  $< 1e-20$ . Most of the animal matches to the candidate lost genes correspond to lineage-specific horizontal gene transfers and/or environmental contaminants from unrelated (non-opisthokont) clades. Some matches correspond to non-orthologous genes such as ANKCLP (S2 Fig.) in the case of the eukaryotic clpB genes HSP78 and HSP104, or ACO1 in the case of LYS4 (Fig. 3). Only HIS2 might have been lost soon after early animal radiation based on the presence of a weak match in Trichoplax adhaerens. This gene is most similar to the version from Capsaspora and not to either of the two choanoflagellates, so it is of uncertain origin.

## My description of the figure

Figure 1 of this paper uses computational methods to identify genes and genetic functions retained in fungi and choanoflagellates but lost in animals. The left column is the list of genes predicted by the Ensembl pipeline to be conserved in fungi and choanoflagellates, but absent in animals. The boxes are the log of the e-values. The red boxes indicate a close BLASTP match with the yeast gene, and the blue boxes represent a weak to no match. This figure clearly shows that these genes are conserved in fungi and choanoflagellates and absent in animals (metazoans). There is a weak match in some metazoans with Hsp104 and Hsp78, but this is a nonfunctional variant of the gene.

## My Project

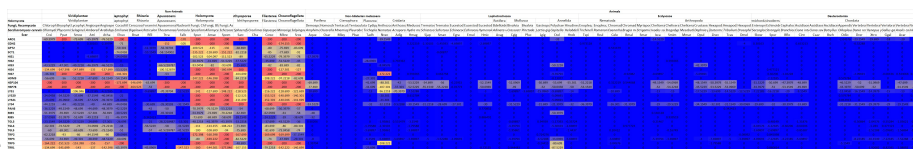
This is a powerful figure, but lots of other genomes have been sequenced since this figure was made by Erives and Fassler. These genomes can be found in Paps et al., 2018. My objective is to redo their BLASTP analysis with more genomes, including more non-animal eukaryotes (besides fungi and choanoflagellates), and more animal genomes.

## Materials and Methods

1. Obtain the protein sequences for the 27 yeast genes in FASTA format from the supplemental figures of Erives and Fassler, 2015.
2. Obtain the list of sequenced genomes from Paps et al., 2018
3. Use the RefSeq\_protein database in NCBI BLASTP to check for alignments between the 27 candidate genes and the sequenced genomes. For genomes not available in the RefSeq database, use the non-redundant database
4. Put the E-values in Microsoft excel, and obtain the log of the e-values. Enter an e-value of 1 for no matches (log of 1 is 0), and a manually enter a value of -200 for matches with an e-value of 0
5. To make the heat map use the conditional formatting feature in Excel. Set -200 as red, -100 as yellow, and the highest value as blue.
6. For the trimmed figure, remove genomes (columns) from the heat map with no BLASTP matches.
7. Copy and paste the heat map cells into Microsoft paint
8. Save the file as a JPEG

## My Figure

~My Figure



~Trimmed Version

