**Interaction among different types of epigenetic mechanisms regulating oyster responses to environmental variation**

The rationale of these analyses is to 1, Annotate chromatin-associated proteins (CAP) involved in epigenetic metabolism in the Eastern oyster genome; and 2, examine the interaction between DNA methylation and the regulation of the function of CAP across different oyster populations. We hypothesize that the different environmental conditions will trigger specific epigenetic responses across populations, requiring changes in the expression of CAP that will involve modifications in inducible methylation marks. Consequently, we expect to see DNA methylation “footprints” (in the form of variation in CpGoe) in gene bodies of CAP, and that these footprints will differ across populations. These footprints will result from differential selective constraints operating on CpG sites relevant for regulation. This rationale is based on previous observations that a significant proportion of the variation between species is actually due to regulatory changes, rather than changes in coding regions. This is best illustrated by the case of human vs. chimpanzee (1% genomic sequence divergence, but that increases to 15% if we focus on CpG sites).

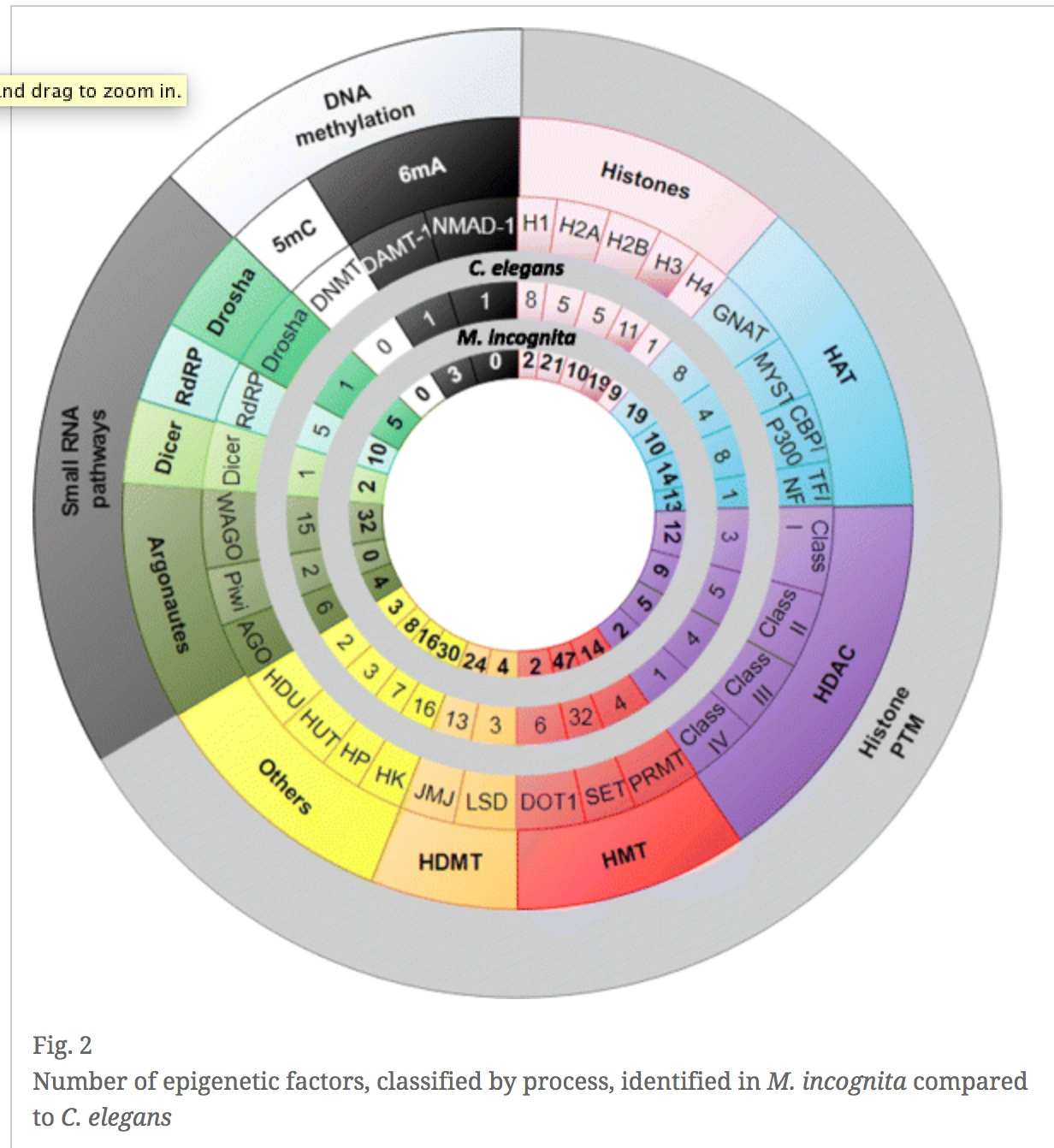
1. Annotation of chromatin-associated proteins in the oyster

Transcript sequences from the CHROMEVALOA database [(Suárez-Ulloa et al. 2013)](https://paperpile.com/c/gUoxaD/aFnb) were used as a reference to identify CAP in the Eastern oyster genome assembly. For that purpose, RNA sequences were aligned to the re-sequenced C. virginica genome assembly using tblastx. Only sequence hits showing e-values below 1e-6 and similarity over 80% were selected. The resulting sequences were further aligned against the transcriptomic set available in NCBI associated with the representative genome C\_virginica-3.0 (GCA\_002022765.4) using blastn to obtain the complete sequence of the transcripts. Subsequently, the transcripts identified as matches to CHROMEVALOA, and thus inferred as chromatin-associated, were aligned using blastn to the current re-sequenced assembly in order to identify the genomic coordinates (start and end positions) for the corresponding genes.

All blastn analyses were carried out setting a maximum threshold for e-value equal to 1e-6.

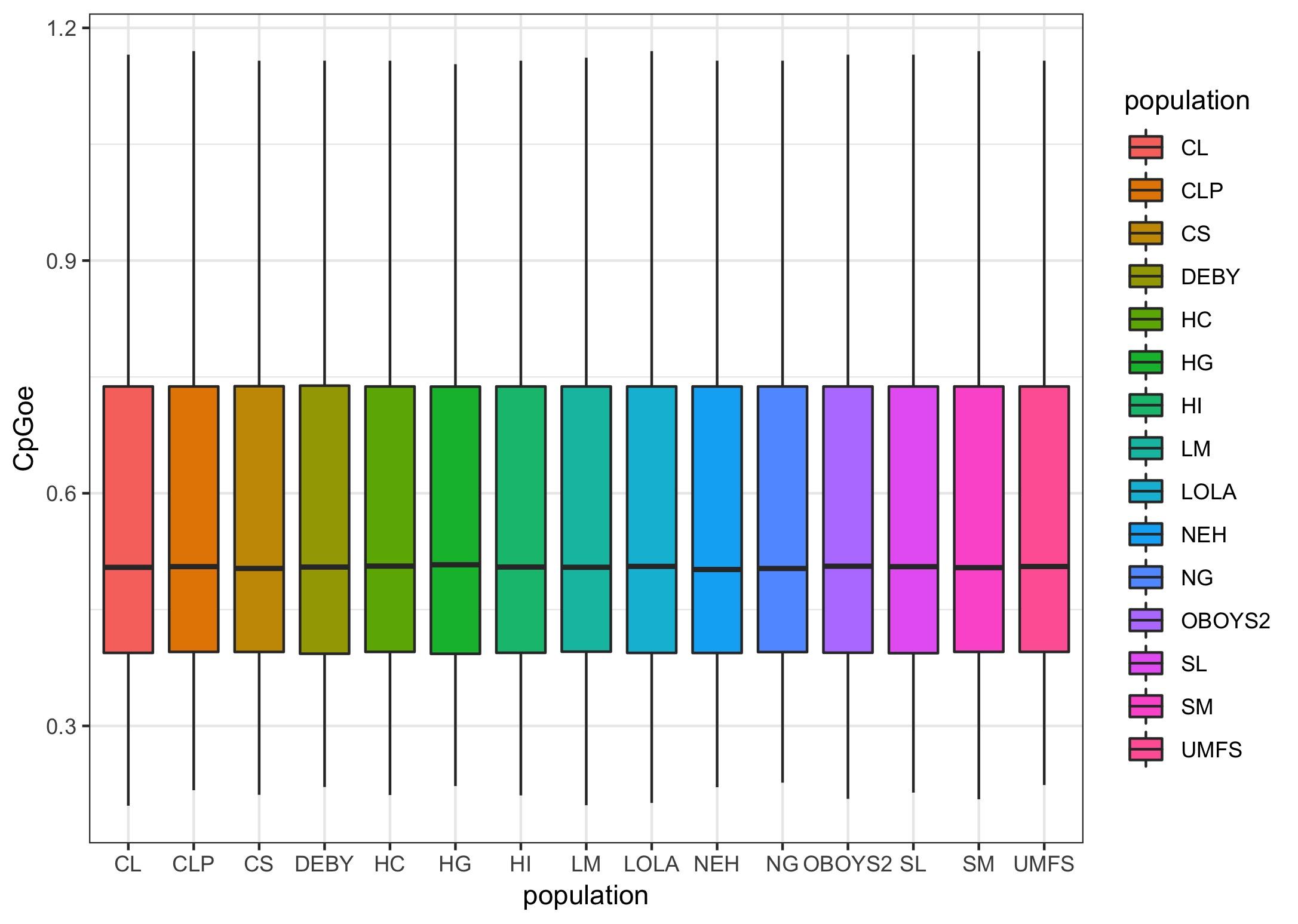
Redundant positions, as well as partial (incomplete) alignments, were manually discarded. The annotation process was further completed by manually identifying and curating genes encoding gene families (i.e., histones) and enzymatic components of the DNA methylation machinery and chromatin remodeling factors (DNMTs, TETs, HATs, HDACs, HMTs and HDMs). Overall, a total of approximately 470 genes were annotated.

ADDITIONAL analyses on this section may include the elaboration of a figure such as this one comparing different types of mollusc genomes:

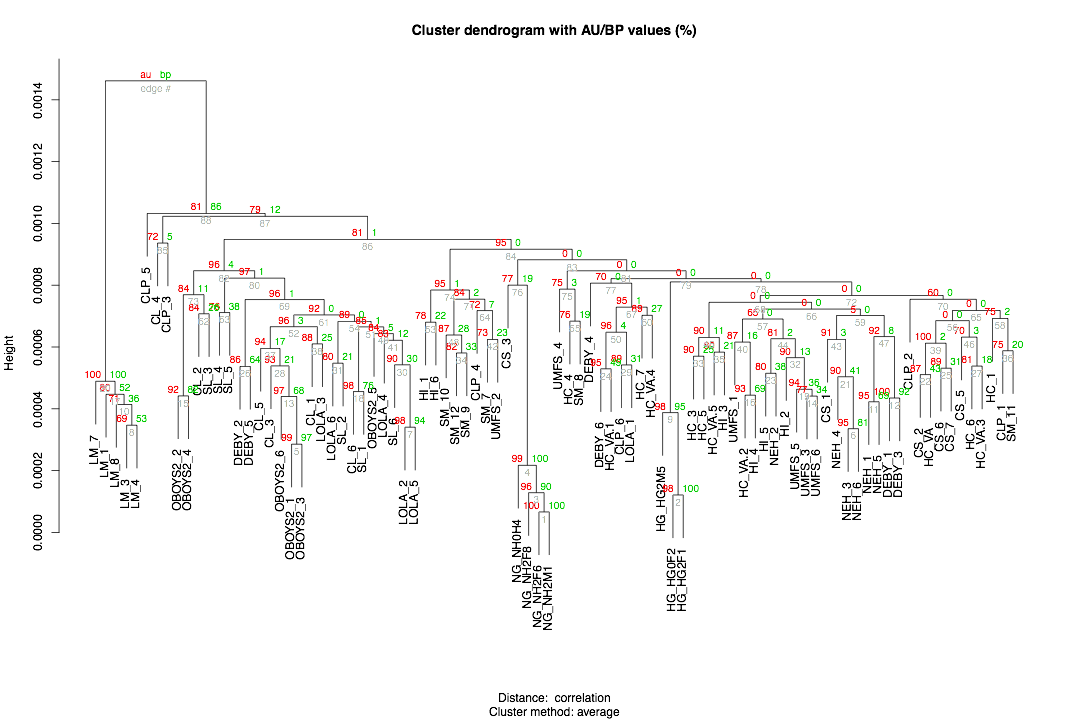


2. Analysis of CpGoe in chromatin-associated proteins across oyster populations

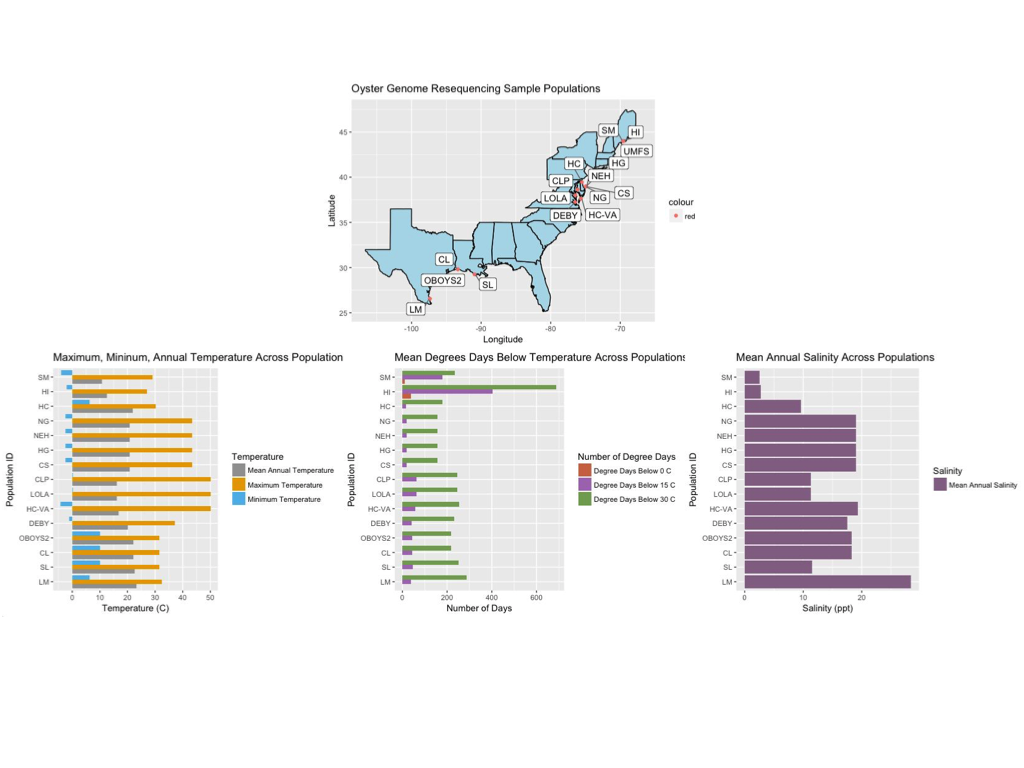
This analysis started using a CpGoe matrix generated for genes automatically annotated in the oyster genome. These gene intervals did not match those defined by our annotation, as the later was done mostly manually. Still, we were able to write the code to proceed with CpGoe analyses including all genome intervals containing the CAP genes, with the exception of 14 genes that were partial or not represented in these automatic intervals. The results do not show differences in CpGoe across populations, at least after analyzing all genes together.



Samples were analyzed across populations based on CpGoe ratio using unsupervised hierarchical clustering (agglomeration method = average and distance measure = Euclidean) was used to group sampling sites based on the similarities of their temporal patterns. A multiscale bootstrap resampling was performed to generate P values for each cluster [(Suzuki and Shimodaira 2006)](https://paperpile.com/c/gUoxaD/A60a), representing the uncertainty associated with the obtained cluster distribution. Approximately unbiased (au) P values represent the results of the multiscale bootstrap resampling, which is consid- ered more accurate than the bootstrap probability (bp). The clustering distribution was considered significant if au >95 [(Suzuki and Shimodaira 2006)](https://paperpile.com/c/gUoxaD/A60a).



Clustering of some populations may account for the environmental differences across oyster populations. For instance, LM population is at a hypersaline estuary in south Texas, displaying a potential significant effect of salinity on CpG sites.



In order to account for the genome intervals corresponding specifically to the genes identified in our manual annotation, the CpGoe matrix was recalculated only for those areas of the genome. Results, in this case, show COMPLETE ...

The analysis was further refined by discriminating among different types of chromatin-associated protein types, notably 3 categories: 1 histones (n=150 approx), 2 histone-modifying enzymes (n=27), 3 chromevaloa proteins (n=292). RESULTS SHOW THAT CpGoe display differences across types????

These analyses were performed by including complete genes for these CAPs. Whe discriminating between exons and introns for these genes, RESULTS SHOW THAT ...