GENOMIC DIFFERENCES BETWEEN DIAPAUSING AND NON-DIAPAUSING SPECIES

Master's thesis by:
Ane Naur Jensen
15th of June 2026

AARHUS UNIVERSITY
Faculty of Natural Science
Section for Bioinformatics and Computational Biology

Supervisors:
Thomas Bataillon - Section for Bioinformatics and Computational Biology
Vera Anna van der Weijden - Department of Molecular Biology and Genetics

INSTITUT FOR MOLEKYLÆRBIOLOGI OG GENETIK
AARHUS UNIVERSITET

Abstract:

Some mammalian species have evolved the capacity to undergo diapause, allowing them to time the birth of their offspring independent of the time of mating. Because some of the species employing diapause are not closely related, questions have arisen regarding how the underlying mechanisms of diapause have evolved. This thesis aims to answer part of this question by comparing the difference in non-synonymous/synonymous rate ratios (dN/dS) between diapausing and non-diapausing species. Analyses were performed using the program CODEML, part of the PAML package, on a total of 44 genes, including 21 genes suspected to be relevant for diapause and 23 control genes with no known relevance for diapause. The results have shown different evolutionary tendencies for the two different types of diapause: obligate and facultative. Facultative diapausing species generally exhibited stronger purifying selection, whereas obligate diapausing species showed weaker purifying selection relative to non-diapausing species. This pattern was observed for both diapause-related and control genes, although it was more pronounced among the diapause-related genes. No evidence was found for positive selection in any of the genes analysed. Phylogenetic patterns suggest that facultative and obligate diapause may have arisen independently, with facultative diapause potentially evolving once and subsequently being maintained by strong purifying selection.

Acknowledgement:

I would like to express my sincere gratitude to my external supervisor Vera Anna van der Weijden at Department of Molecular Biology and Genetics for welcoming me into her group, and to my internal supervisor Thomas Bataillon at Section for Bioinformatics and Computational Biology for matching me up with Vera and her group. A big thanks to both of them for providing guidance and perspective throughout the whole process. And thank you to the rest of the van der Weijden group for support and feedback during group meetings, and a special thanks to Krista Agathe Bossow for helping me get started with the project and providing me with relevant literature, as well as continuous input and advise.

Declaration of use of Generative Artificial Intelligence:

☒ **I/we used generative artificial intelligence (GAI) to complete this project** *(tick the box).* List the GAI tool(s) you used (remember to specify version): - OpenAI. *ChatGPT (GPT-5.5).* https://chatgpt.com
I/we used GAI tools in the following way *(See accompanying list of possible uses for inspiration)* For alternative ways of formulating text: - Finding synonyms for words. - Rephrasing of sentences sounding clumsy or weird. To understand a topic better: - Clarifying the meaning of paragraphs in literature. - Asking if my understandings of a topic are correct. For programming tasks: - Specifying error messages and possible ways to fix it. - Help tweak my code to do exactly what I want. - Restructuring of code to be more “reader-friendly”.

Abbreviations:

OrthoMaM: Orthologous Mammalian Markers.

CDS: Coding sequence.

dN: Non-synonymous substitution rate.

dS: Synonymous substitution rate.

ω: Non-synonymous/synonymous rate ratio.

IGF: Insulin-like growth factor.

ESC: Endometrial stromal cell.

IGF-1R: Insulin-like growth factor 1 receptor.

GH: Growth hormone.

IGFBP: Insulin-like growth factor binding protein.

PAPP-A: Pregnancy-associated plasma protein-A, pappalysin-1.

STC: Stanniocalcin.

PAML: Phylogenetic Analysis by Maximum Likelihood.

LRT: Likelihood ratio test.

ML: Maximum Likelihood

CTMC: Continuous-Time Markov Chain.

π: Equilibrium frequency.

κ: Transition/transversion rate ratio.

FMutSel: Frequencies of Codons based on Mutation and Selection.

L: Likelihood.

ℓ or lnL: log-likelihood.

MLE: Maximum Likelihood Estimated.

$2\Delta l$: Twice the log-likelihood or the LRT statistic

BEB: Bayes empirical Bayes.

NEB: Naïve empirical Bayes.

Table of content