Adaptation in genes, duplicates, families, functional modules and genomes

François Serra

October 2011

Contents

1 Introduction

Contents

2 Random-like structure of DNA

2.1 Introduction

2.2 Results and Discussion

2.3 Material and methods

3 Adaptation in genes, duplicated (and families)

3.1 BRANCHED1

3.2 Protamines Rodents and Primates

3.3 Selective pressure on duplicated genes in Drosophila

4 Searching for evolutionary patterns in funcionally linked group of genes

4.1 Introduction

4.2 Material and Methods

4.2.1 Dataset

Five mammals

Complete genomes of 5 mammals species (*Homo sapiens*, *Pan troglodytes*, *Mus musculus*, *Rattus norvegicus* and *Canis familiaris*) where retrieved from *Ensembl* [Flicek *et al.*2011]. Also orthology prediction between each pair of species possibly done between human and the others was retrieved from *Ensembl Compara* [Vilella *et al.*2009] using biomart [Kinsella *et al.*2011]. Only groups of orthologs *one-to-one* with one representative of each species where kept in the final dataset. 4.1 NUMBERS

6 Drosophila

4.2.2 Alignments

Each of the group of orthologous sequences were aligned with Muscle [Edgar2004], and, once aligned sequences were cleaned with trimal [Capella-Gutiérrez *et al.*2009] keeping all sequences but trimming aligment columns with the euristic1 method.

4.3 open on colocalization $\rightarrow notrandom$

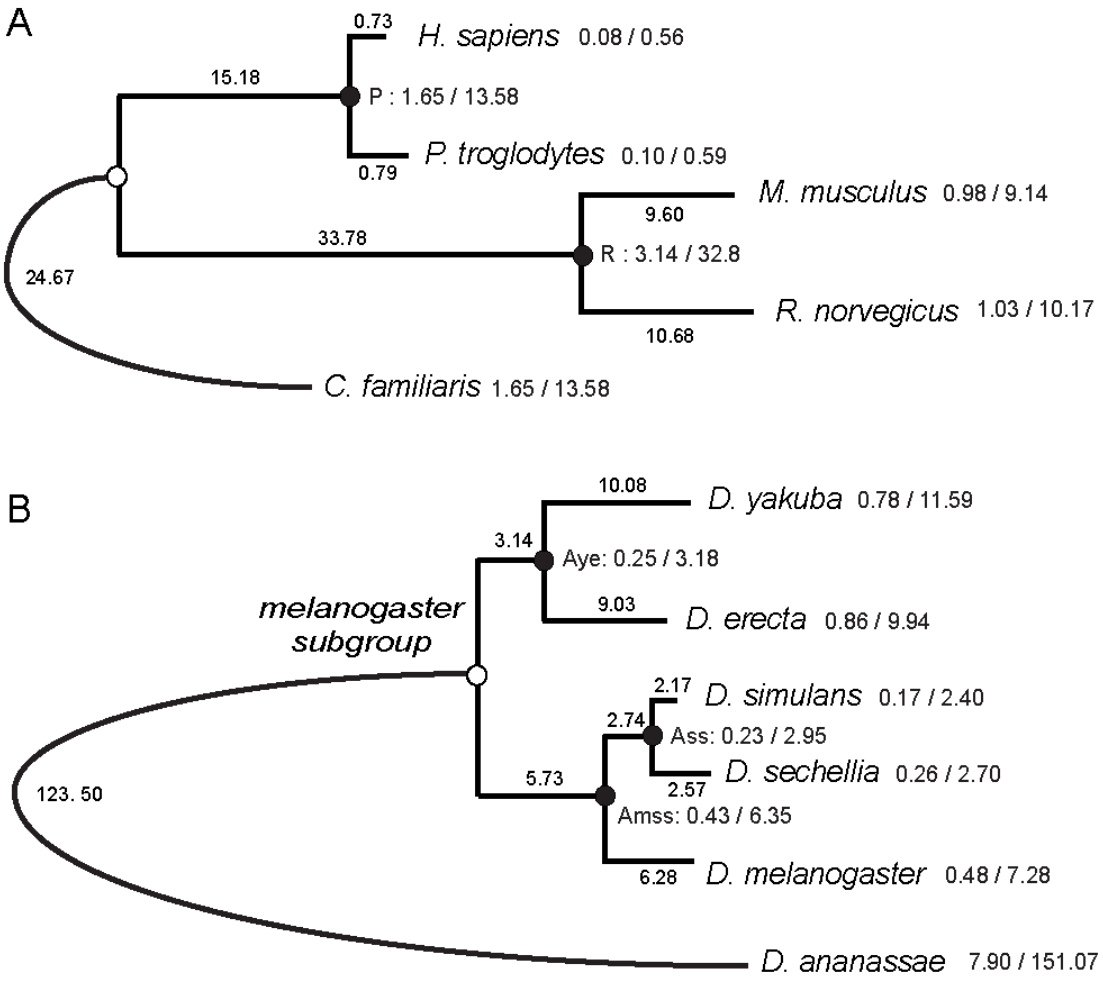

Figure 4.1: Mammals and *Drosophila* phylogeny. blabli blob lu dkfnlskjdf

5 Life inside genomes, dynamics and predictions

5.1 Genomic elements, dispersion and abundance

5.2 Species Abundance Diversity in genomes

5.3 Neutrality of SAD

6 Tools, programs, methods

6.1 ETE-evol plugin

6.2 Phylemon

6.3 Ecolopy

6.4 Isoform selection?

7 Conclusions

Bibliography

[Capella-Gutiérrez *et al.*2009] Salvador Capella-Gutiérrez, José M Silla-Martínez, and Toni Gabaldón, trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics (Oxford, England)* **25**(15) (2009), 1972–3.

[Edgar2004] Robert C Edgar, MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic acids research* **32**(5) (2004), 1792–7.

[Flicek *et al.*2011] Paul Flicek, M Ridwan Amode, Daniel Barrell, Kathryn Beal, Simon Brent, Yuan Chen, Peter Clapham, Guy Coates, Susan Fairley, Stephen Fitzgerald, Leo Gordon, Maurice Hendrix, Thibaut Hourlier, Nathan Johnson, Andreas Kähäri, Damian Keefe, Stephen Keenan, Rhoda Kinsella, Felix Kokocinski, Eugene Kulesha, Pontus Larsson, Ian Longden, William McLaren, Bert Overduin, Bethan Pritchard, Harpreet Singh Riat, Daniel Rios, Graham R S Ritchie, Magali Ruffier, Michael Schuster, Daniel Sobral, Giulietta Spudich, Y Amy Tang, Stephen Trevanion, Jana Vandrovcova, Albert J Vilella, Simon White, Steven P Wilder, Amonida Zadissa, Jorge Zamora, Bronwen L Aken, Ewan Birney, Fiona Cunningham, Ian Dunham, Richard Durbin, Xosé M Fernández-Suarez, Javier Herrero, Tim J P Hubbard, Anne Parker, Glenn Proctor, Jan Vogel, and Stephen M J Searle, Ensembl 2011. *Nucleic acids research* **39**(Database issue) (2011), D800–6.

[Kinsella *et al.*2011] Rhoda J. Kinsella, Andreas Kähäri, Syed Haider, Jorge Zamora, Glenn Proctor, Giulietta Spudich, Jeff Almeida-King, Daniel Staines, Paul Derwent, Arnaud Kerhornou, Paul Kersey, and Paul Flicek, Ensembl BioMarts: a hub for data retrieval across taxonomic space. *Database : the journal of biological databases and curation* **2011** (2011), p. bar030.

[Vilella *et al.*2009] Albert J Vilella, Jessica Severin, Abel Ureta-Vidal, Li Heng, Richard Durbin, and Ewan Birney, EnsemblCompara GeneTrees: Complete, duplication-aware phylogenetic trees in vertebrates. *Genome research* **19**(2) (2009), 327–35.

List of Figures

List of Tables