

# Uncovering the phylogeny of Hox proteins

Tomasz Korbak

June 22, 2018

## 1 Uncovering the phylogeny of Hox proteins

The goal of this notebook was to analyze the evolution of [Hox genes](#). These genes play a role in morphological development in a wide range of animals (including human), i.e. they control the body plan of an embryo along the head-tail axis.

### 1.1 Inferring the gene tree

Our dataset consists of 101 sequences coding for genes from Hox family. The first step of the analysis is to reconstruct the gene tree of these genes based on the similarity of our sequences.

For estimating similarity scores between each pair of sequences we shall use BLAST hit scores against whole set of 839 sequences.

Then a tree will be constructed using neighbor joining algorithm.

### 1.2 Reconciling the gene tree with a species tree

The next step of our analysis is to assume a particular topology of a species tree and reconcile it with inferred gene tree. Reconciliation here means mapping each gene tree node  $g$  to a species tree node  $s$  such that all  $g$  non-terminal are a subset of  $s$  non-terminals. Non-terminals here correspond to particular species (human, fruitfly, beetle or mouse).

Then, let us compute deep coalescence score (a measure of how well gene tree and species tree align) and the number of duplications predicted by the inferred reconciliation.

```
In [13]: compute_deep_coalescence(species_tree, gene_tree)
```

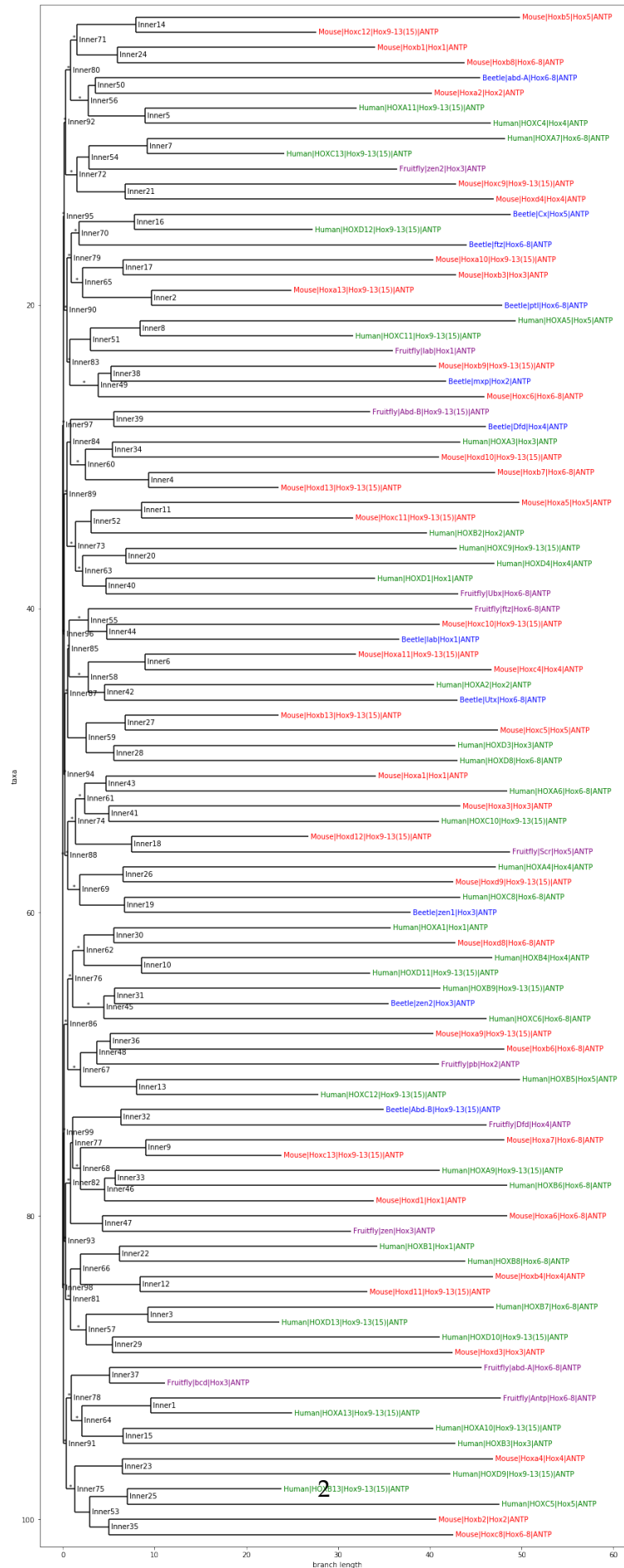
```
Out[13]: 151
```

```
In [14]: duplications = find_duplications(gene_tree)
```

```
In [16]: len(duplications)
```

```
Out[16]: 44
```

Let us plot the gene tree with marked duplications. Duplication events are marked with stars. We can observe that all occur near the root of the tree.



## 2 Protein domain analysis

The last task we turn to is protein domain analysis in gene classes in our subset. We will use HMMer to compute the number the score between each pair of gene classes.

We have obtain a 128x128 matrix of scores for each family pair. Let us visualize a heatmap of this matrix.

