



HAL
open science

Predicting isoform transcripts: What does the comparison of known transcripts in human, mouse and dog tell us?

Nicolas Guillaudeau, Catherine Belleannée, Samuel Blanquart, Jean-Stéphane Varré

► To cite this version:

Nicolas Guillaudeau, Catherine Belleannée, Samuel Blanquart, Jean-Stéphane Varré. Predicting isoform transcripts: What does the comparison of known transcripts in human, mouse and dog tell us?. JOBIM 2019 - Journées Ouvertes Biologie, Informatique et Mathématiques, Jul 2019, Nantes, France. 8, pp.1, 2019, 10.7490/f1000research.1117311.1 . hal-02267357

HAL Id: hal-02267357

<https://inria.hal.science/hal-02267357>

Submitted on 19 Aug 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

To fill in the knowledge about **transcript isoforms** expressed from a gene, we have proposed a **comparative genomics** method allowing to identify orthologous exons shared by a pair of genes [1]. We predict transcript isoforms in human, mouse and in a non-model organism, dog, and we identify **135 conserved genes** having common gene structures and common potential transcriptomes.

Modelling gene structures using comparative genomics to predict isoforms [1]

Our **structure of a target gene** (fig.1a)

- based on **functional sites** of known transcripts:
 - [: start codon] : stop codon < > : splice sites **a letter**: (part of) a coding exon
- using functional sites of known transcripts in an orthologous source gene (fig.1b):
 - to reveal **new functional sites** or **coding exons** on the target gene (fig.1c)
 - to predict **new transcripts** of the target gene, using predicted functional sites (fig.1d)

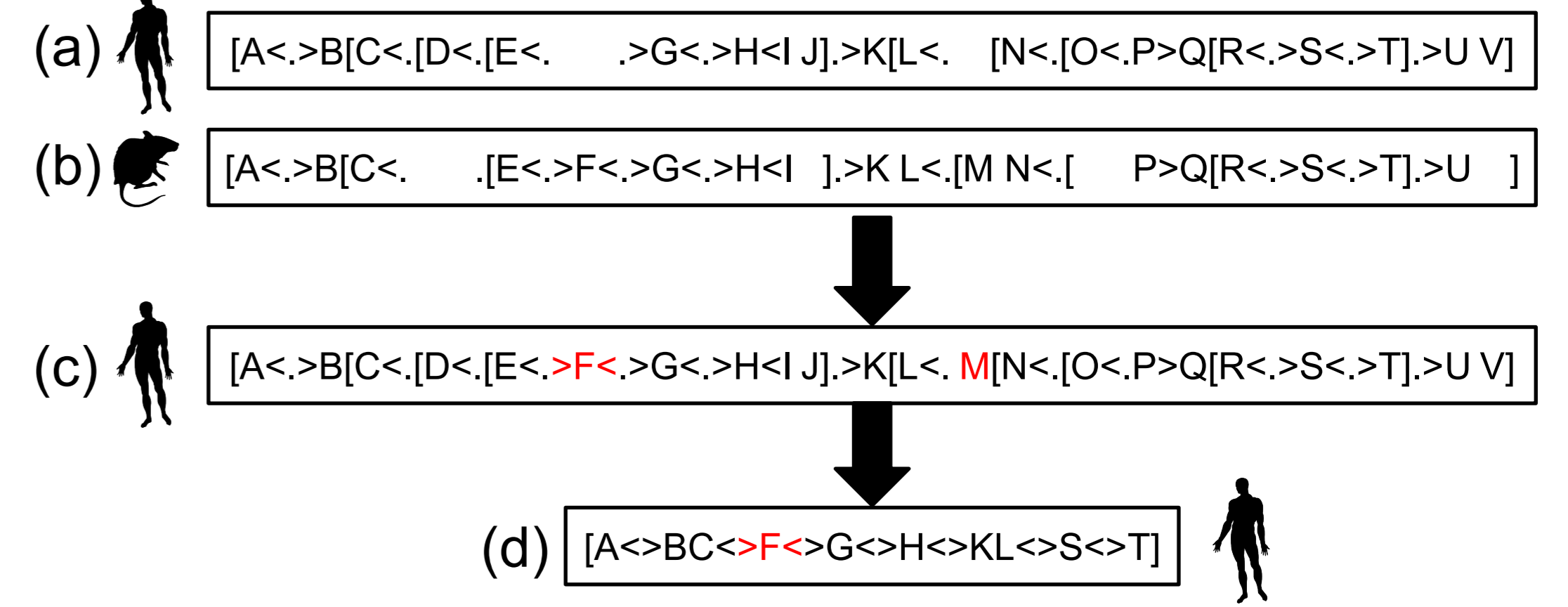
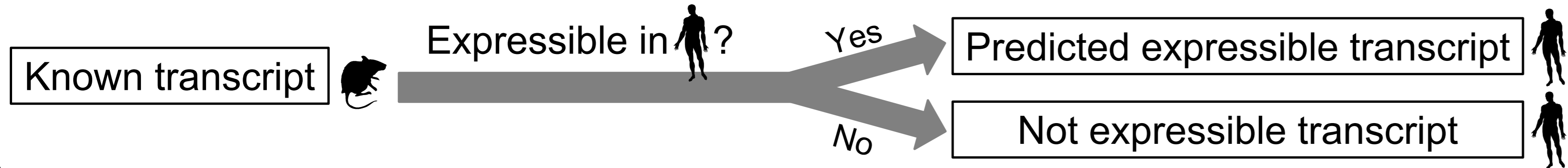


Figure 1. CREM gene: pairwise comparison between human (a) and mouse (b) reveals new functional sites or exons (red) on human gene (c). This lead to predict new expressible transcripts (d) in human.

Application: Predicting transcripts in human, mouse and dog

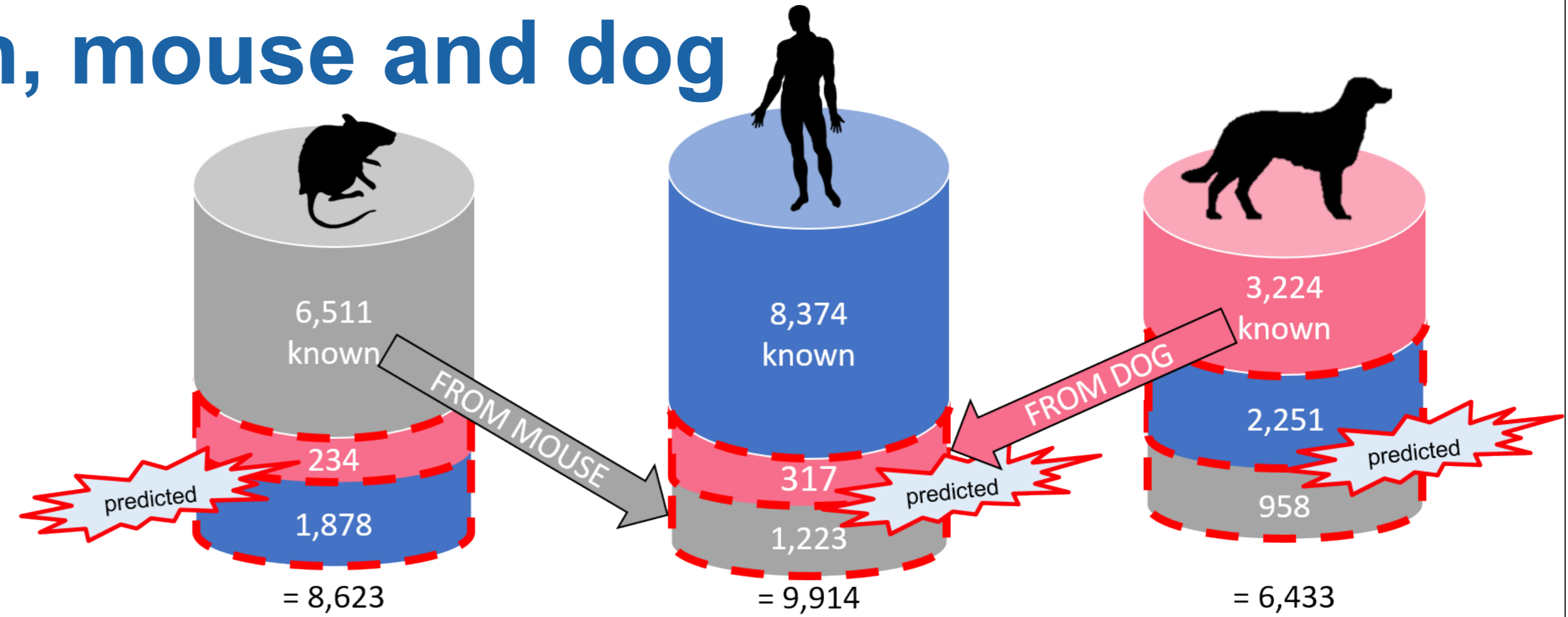
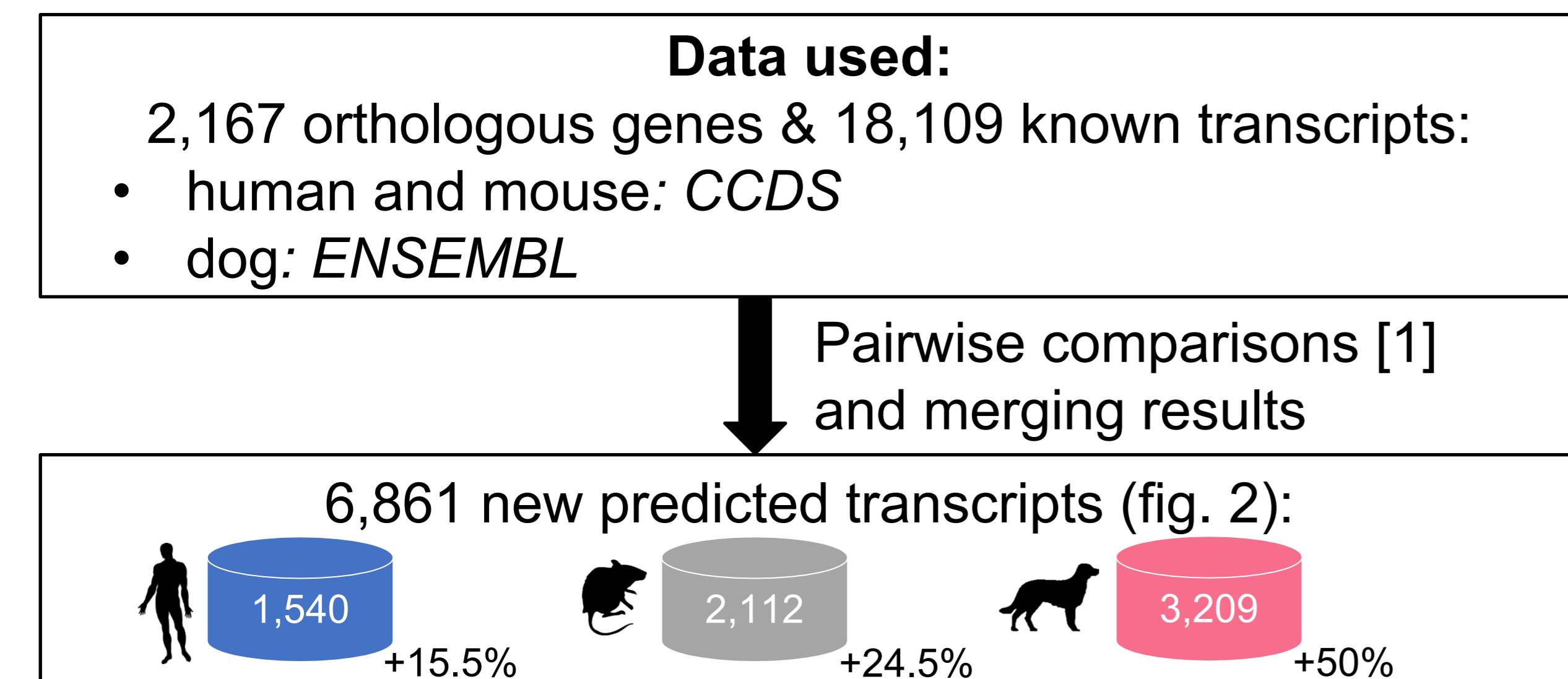


Figure 2. Estimated transcriptomes: known and predicted transcripts obtained using pairwise comparisons.

Analyzing conservation over functional sites and transcripts

We build two conservation graphs for each gene:

- conserved **functional site** graph (fig.4) & conserved **transcript** graph (fig.5)
- a graph component shows orthology relationships between species
- classification of graph components (3-species case):
 - shared in the 3 species (fig.3a), in 2 species (fig.3b), specific to a species (fig.3c)
- only graphs without “ambiguous” components (fig.3d) are **considered for analysis**

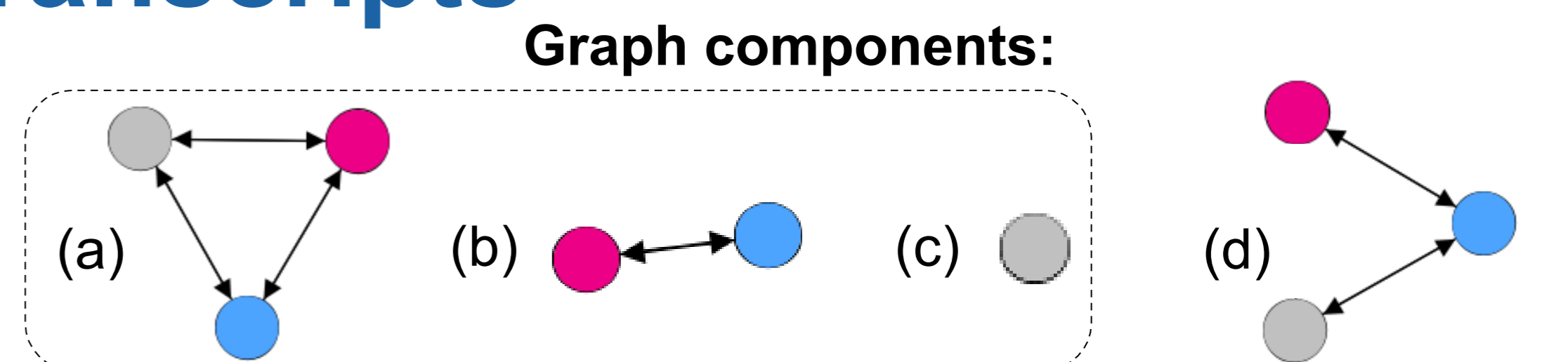


Figure 3. Considered gene components: (a) triplet, (b) duplicate, (c) singleton. “Ambiguous” graph components (d) are not considered.

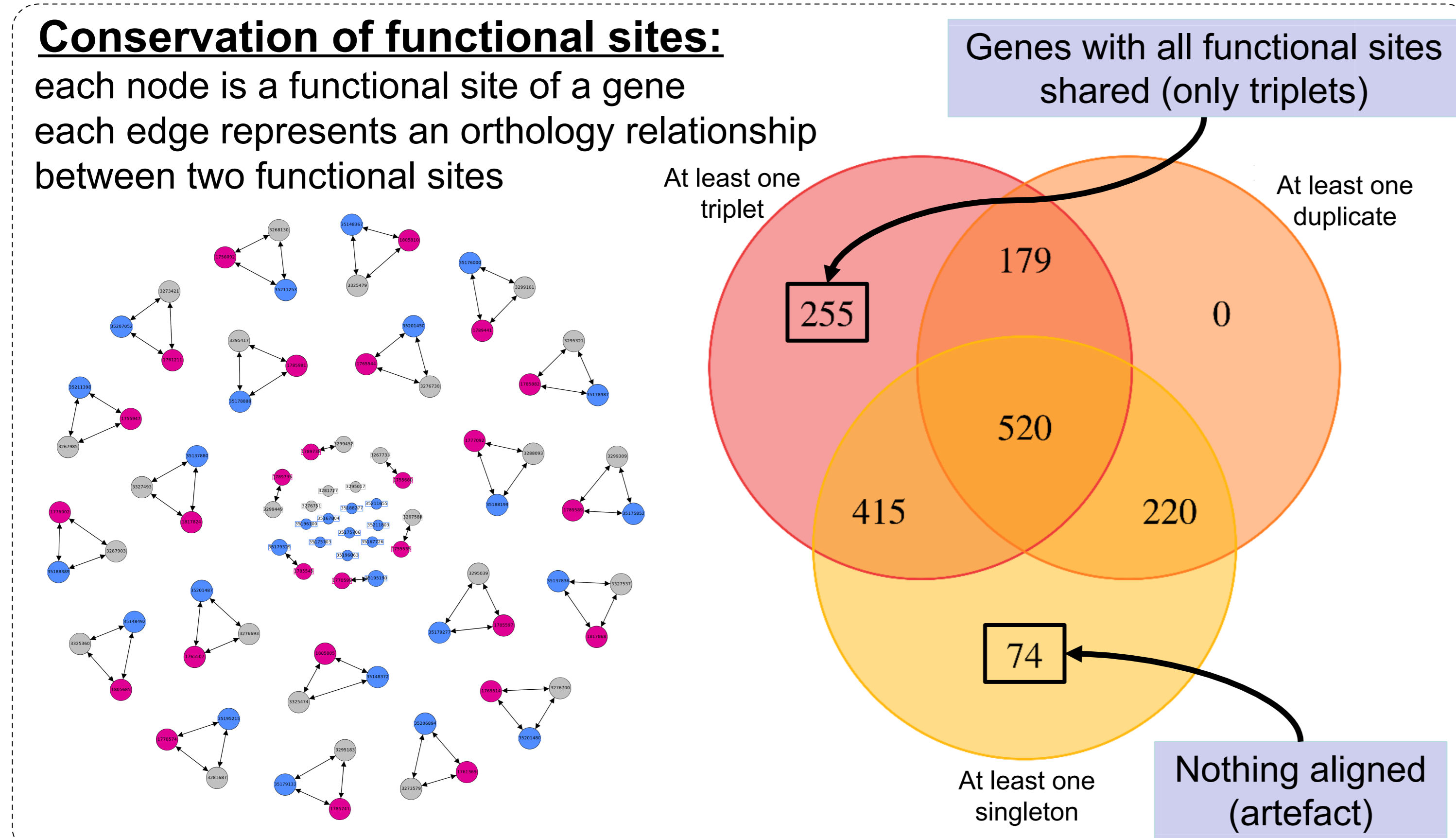


Figure 4. Functional site graph of CREM gene (left). Analysis of 1,663 considered functional site graphs. 255 graphs reveal genes having common gene structures (right).

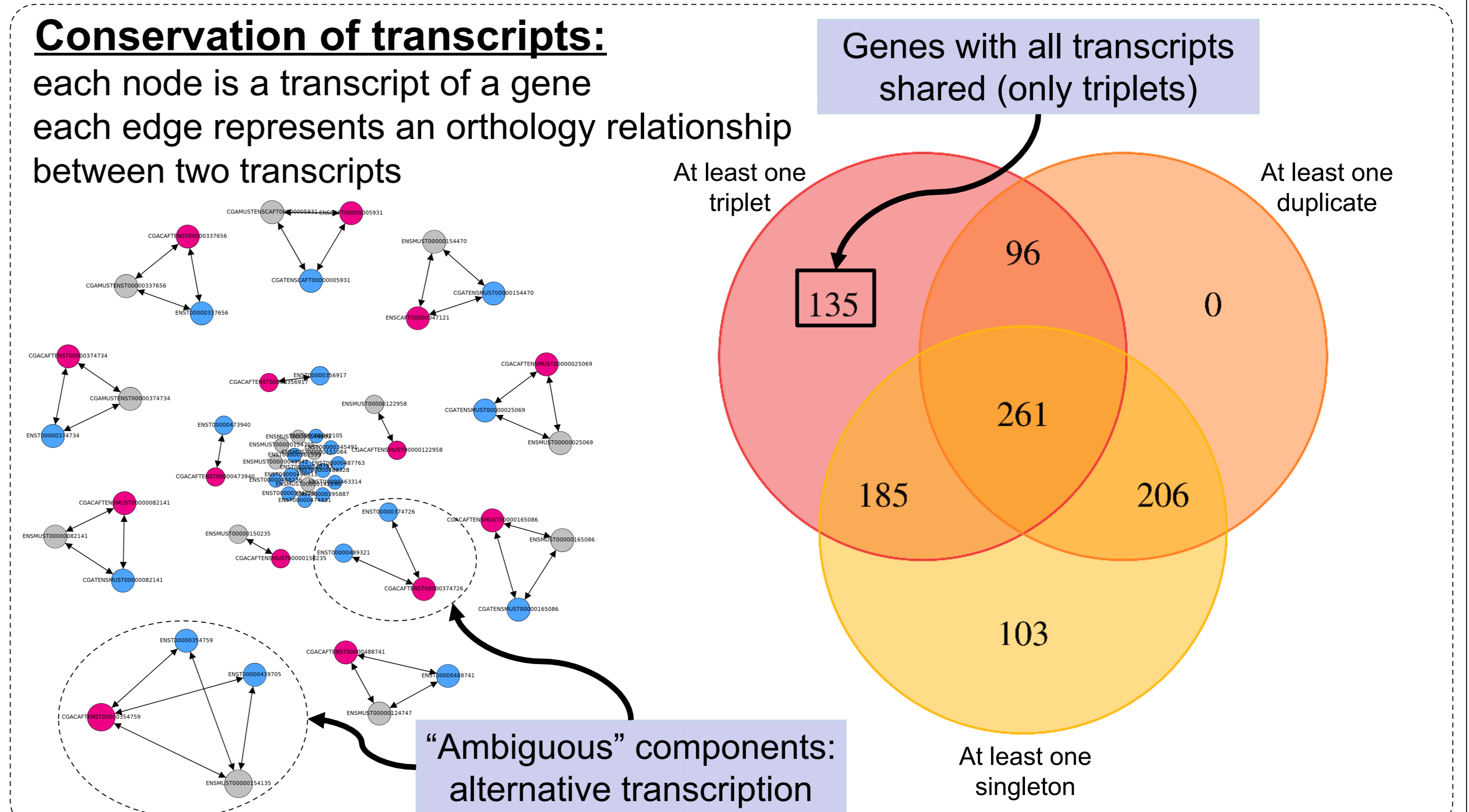


Figure 5. Transcript graph of CREM gene (ambiguous) (left). Analysis of 986 considered transcript graphs. 135 graphs reveal genes having common potential transcriptomes (right).

Phylogenetic interpretation of gene structures

Most analyzed genes encountered structure divergence

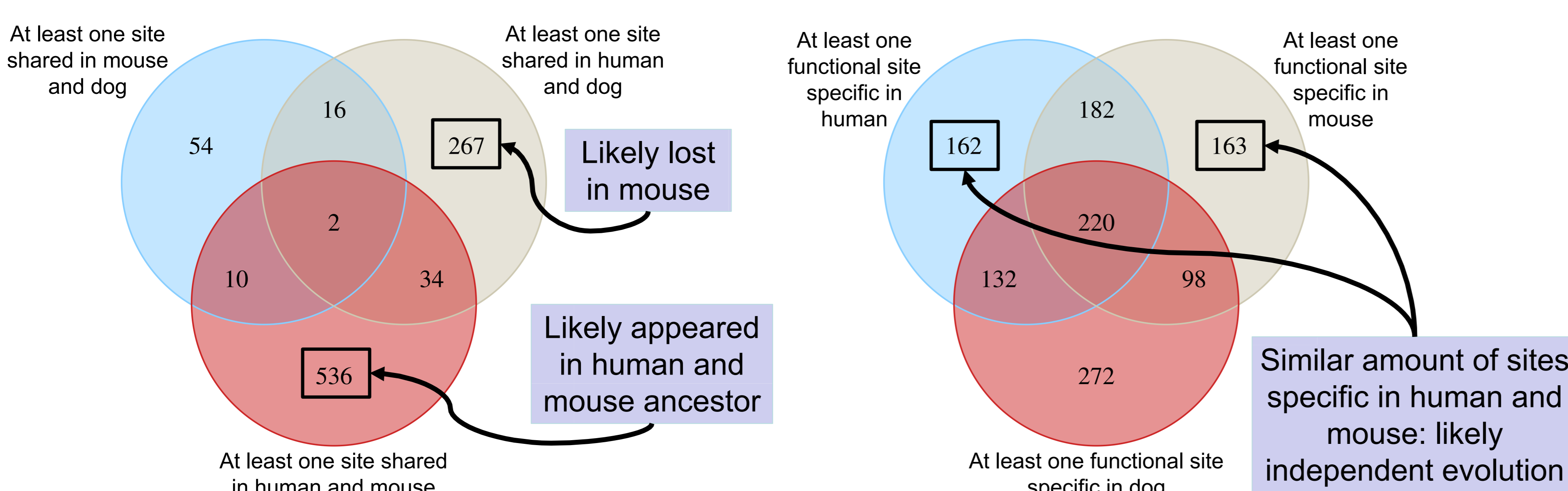


Figure 6. Distribution of duplicates (left) and singletons (right) components in functional site graphs.

Some genes have the same gene structure but different transcriptomes

