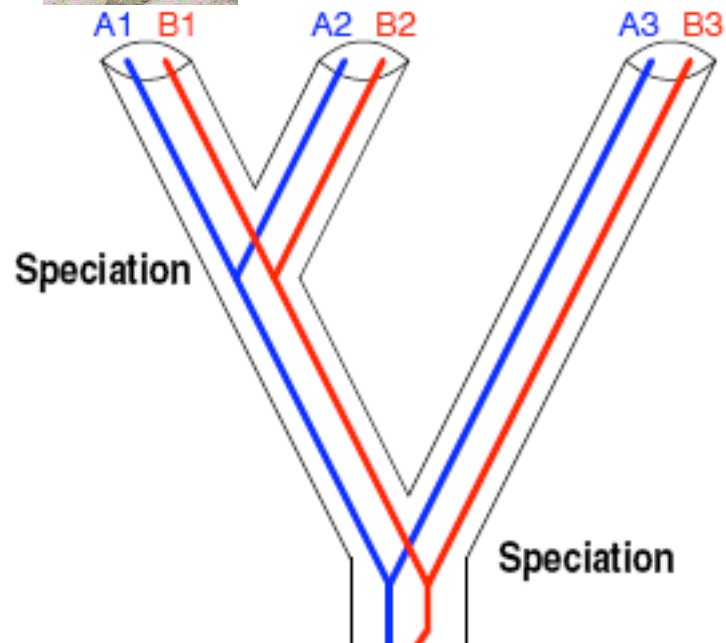
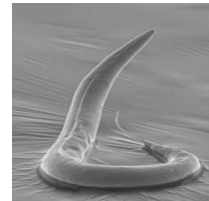
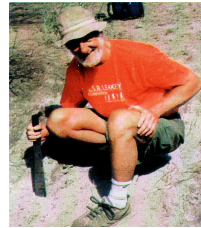
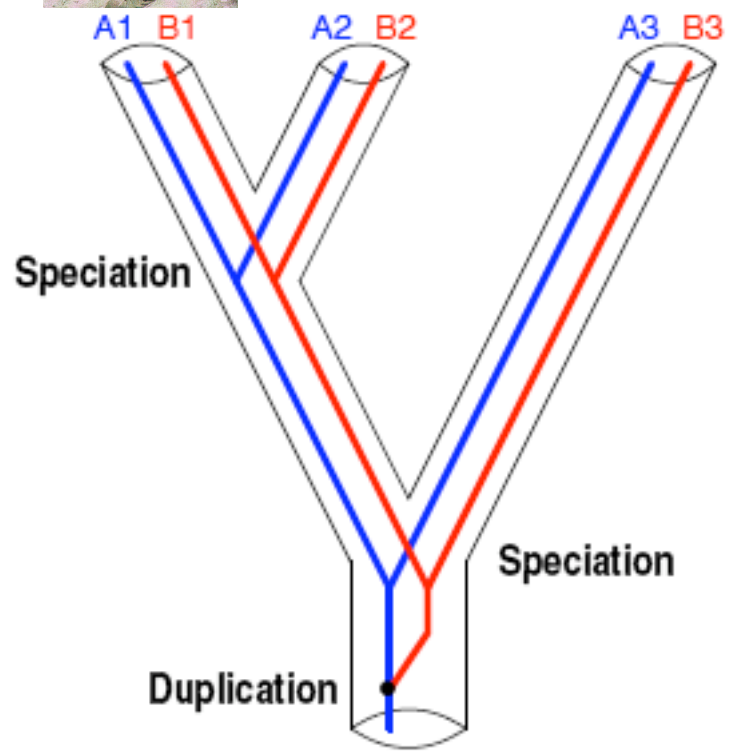
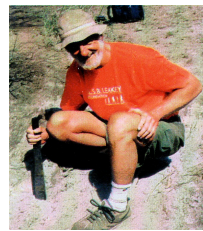


Inferring Orthology

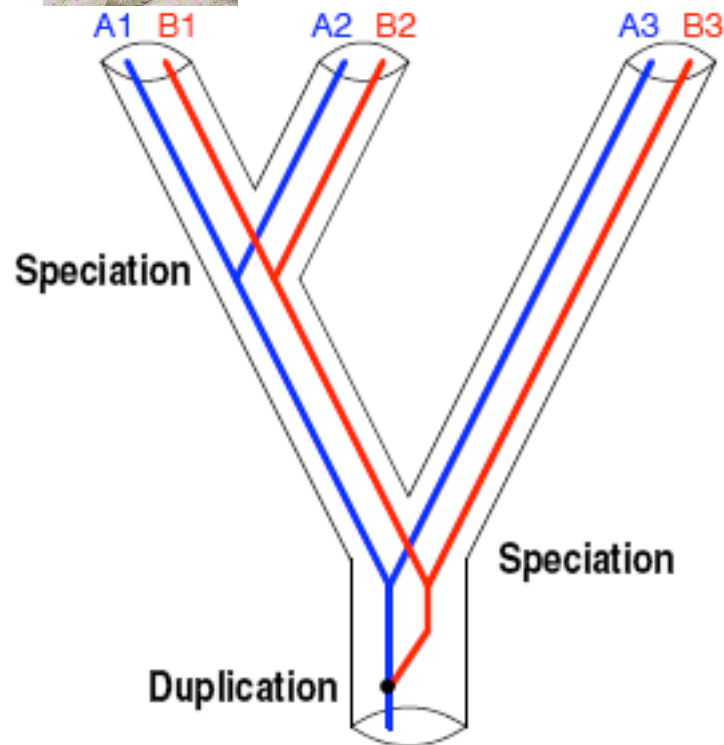
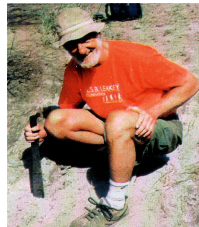
The evolutionary relationships of species...



The evolutionary relationships of species and their genes



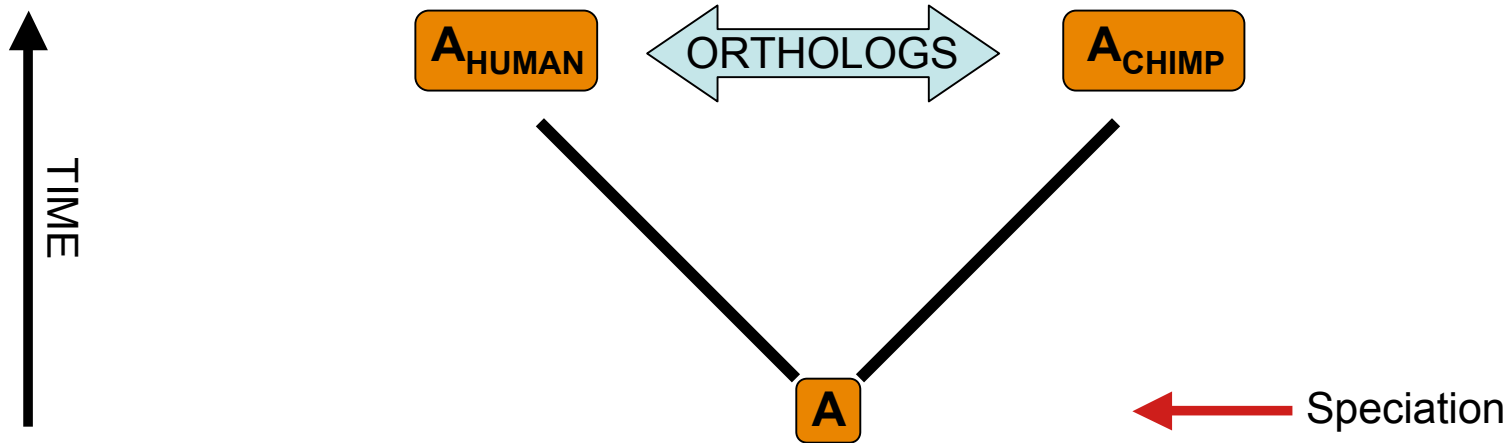
The evolutionary relationships of species and their genes



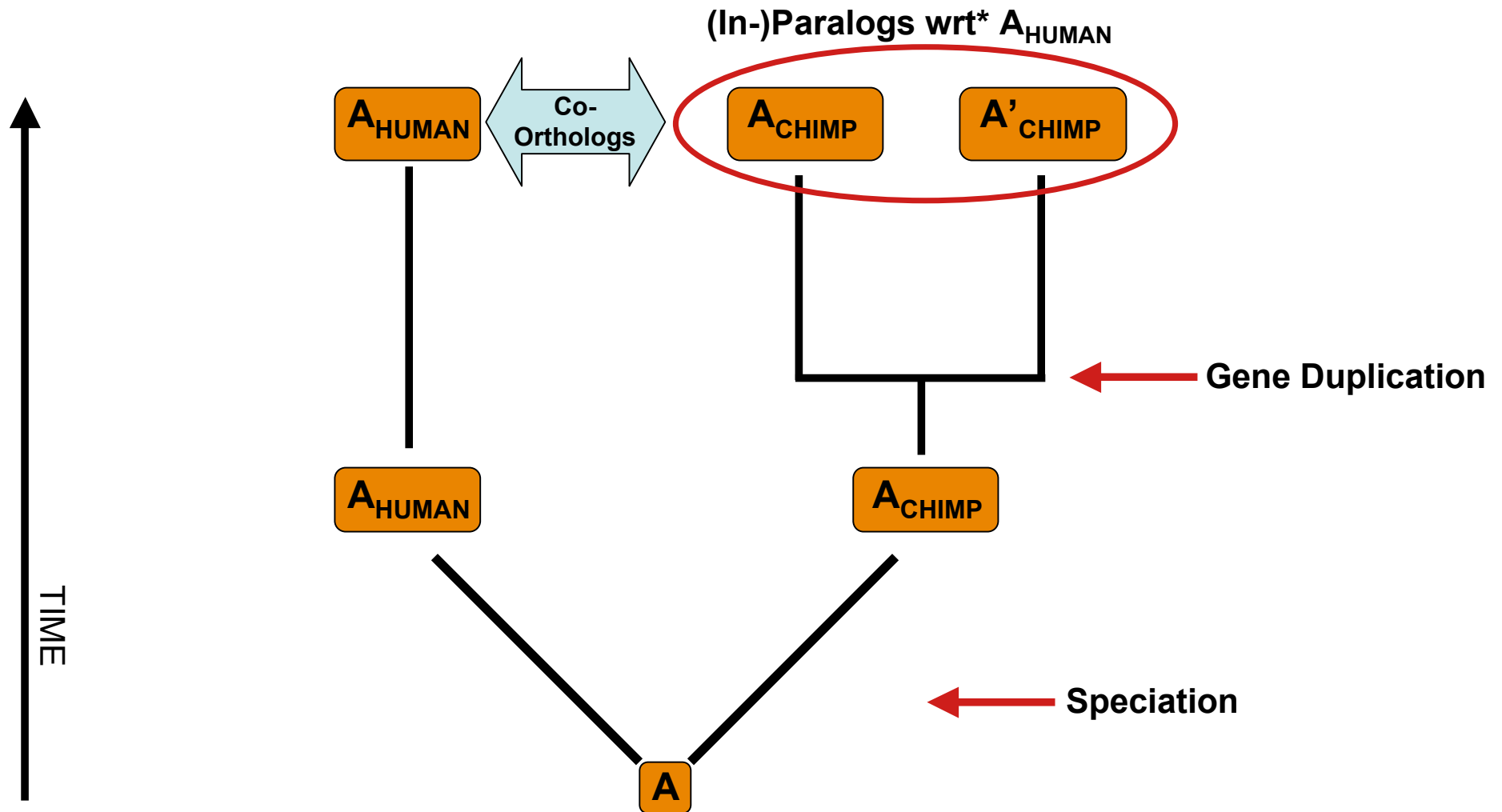
Arguments for orthology assumption:

- a sequence tree that is congruent to the species tree
- conservation of genomic position
- sequence similarity (typically, reciprocal best blast hit)
- similarity of function

Evolutionary Relationships between Genes and their Products

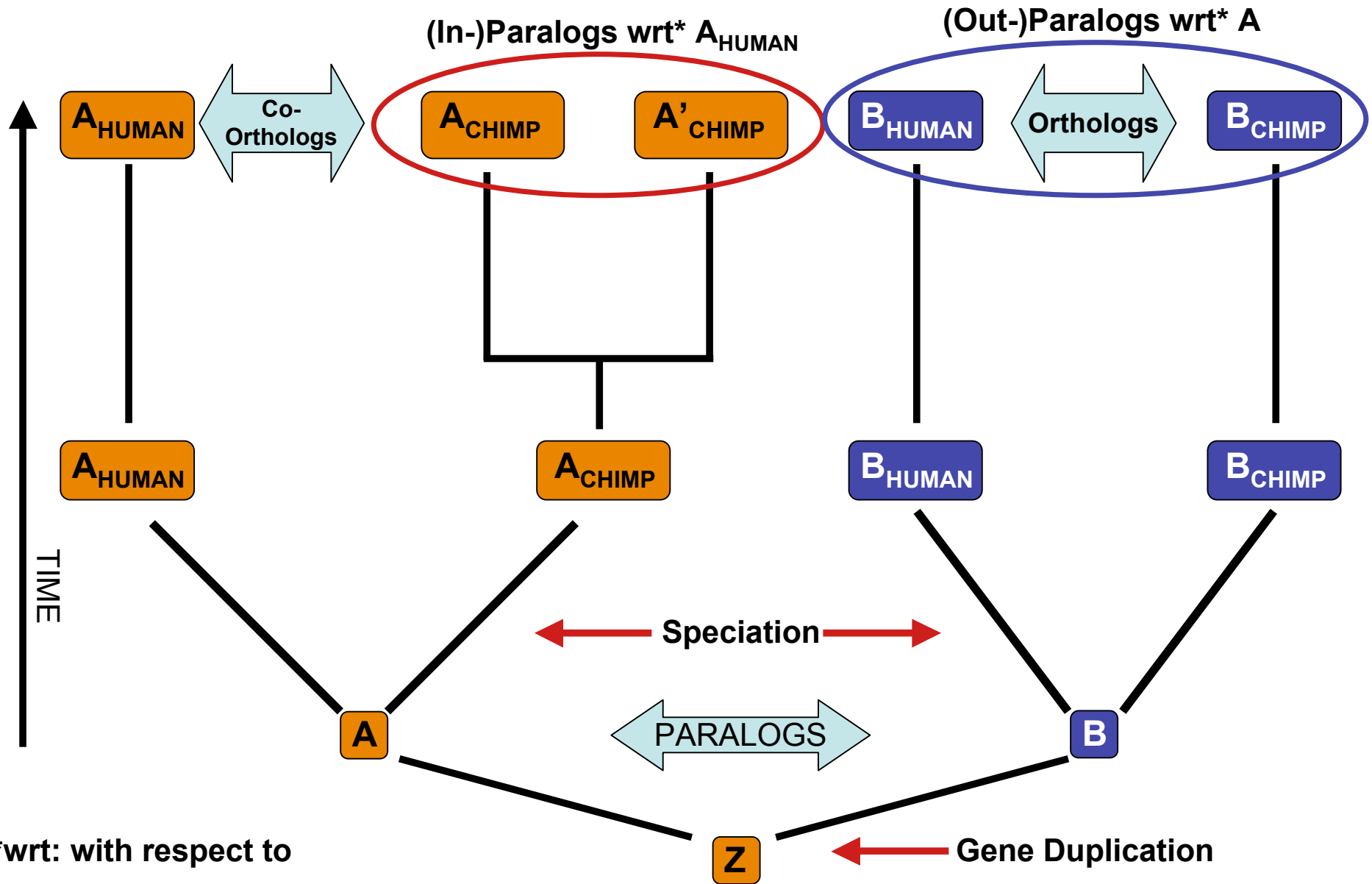


Evolutionary Relationships between Genes and their Products

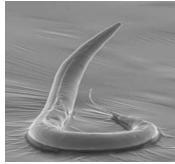


*wrt: with respect to

Evolutionary Relationships between Genes and their Products



Identification of Orthologous Genes: The KOGs (COGs) approach



C. elegans



Arabidopsis



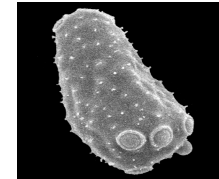
S. pombe



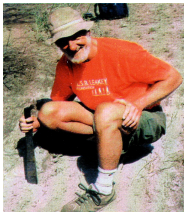
Drosophila

**Analysis of 7 completely
sequenced eukaryotic genomes**

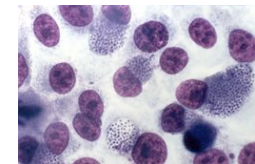
<http://www.ncbi.nlm.nih.gov/COG>



S. cerevisiae

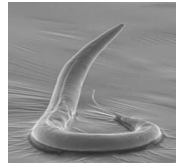


Human



E. cuniculi

Identification of Orthologous Genes: The KOGs approach



C. elegans



Arabidopsis

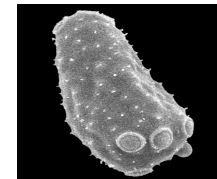


S. pombe

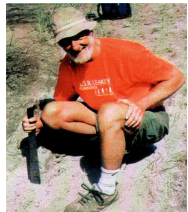


Drosophila

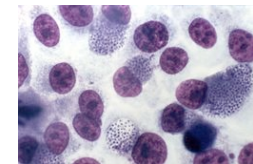
- 1) Detect and mask typical repetitive protein domains
- 2) All against all comparison of protein sequences (BLAST)
- 3) Detection of triangles of mutually consistent genome-specific best hits.



S. cerevisiae



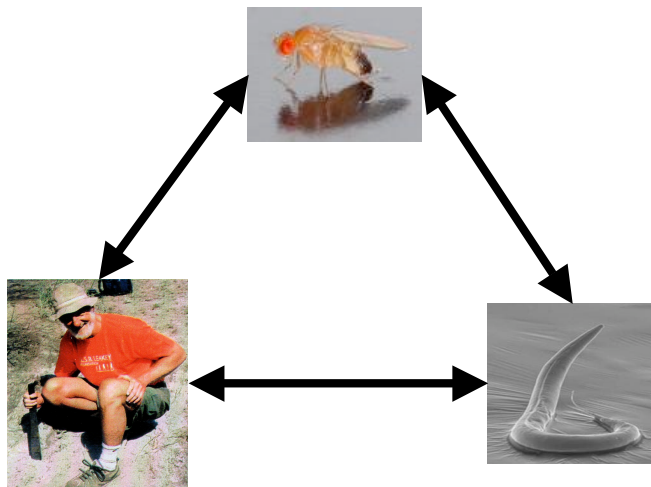
Human



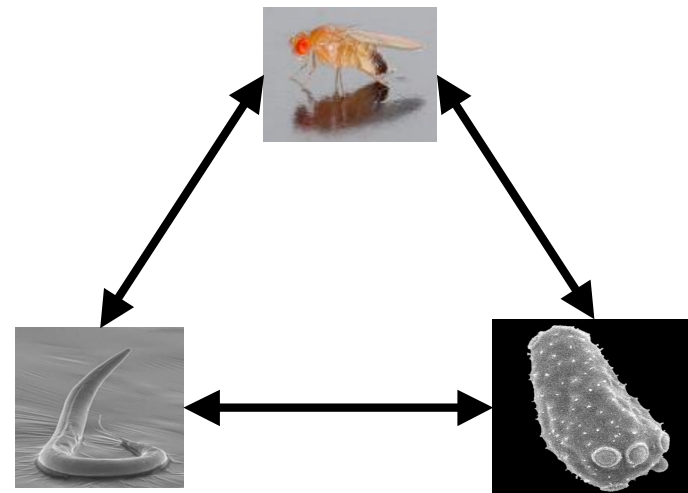
E. cuniculi

Identification of Orthologous Genes: Reciprocal Best Blast Hits

- 1) Detect and mask typical repetitive protein domains
- 2) All against all comparison of protein sequences (BLAST)
- 3) Detection of triangles of mutually consistent genome-specific best hits.



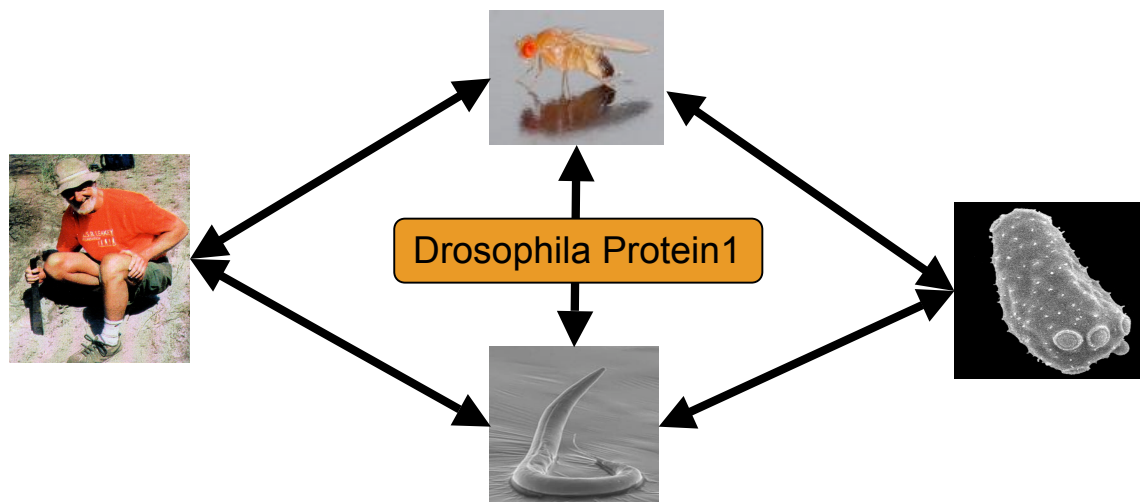
Drosophila Protein1



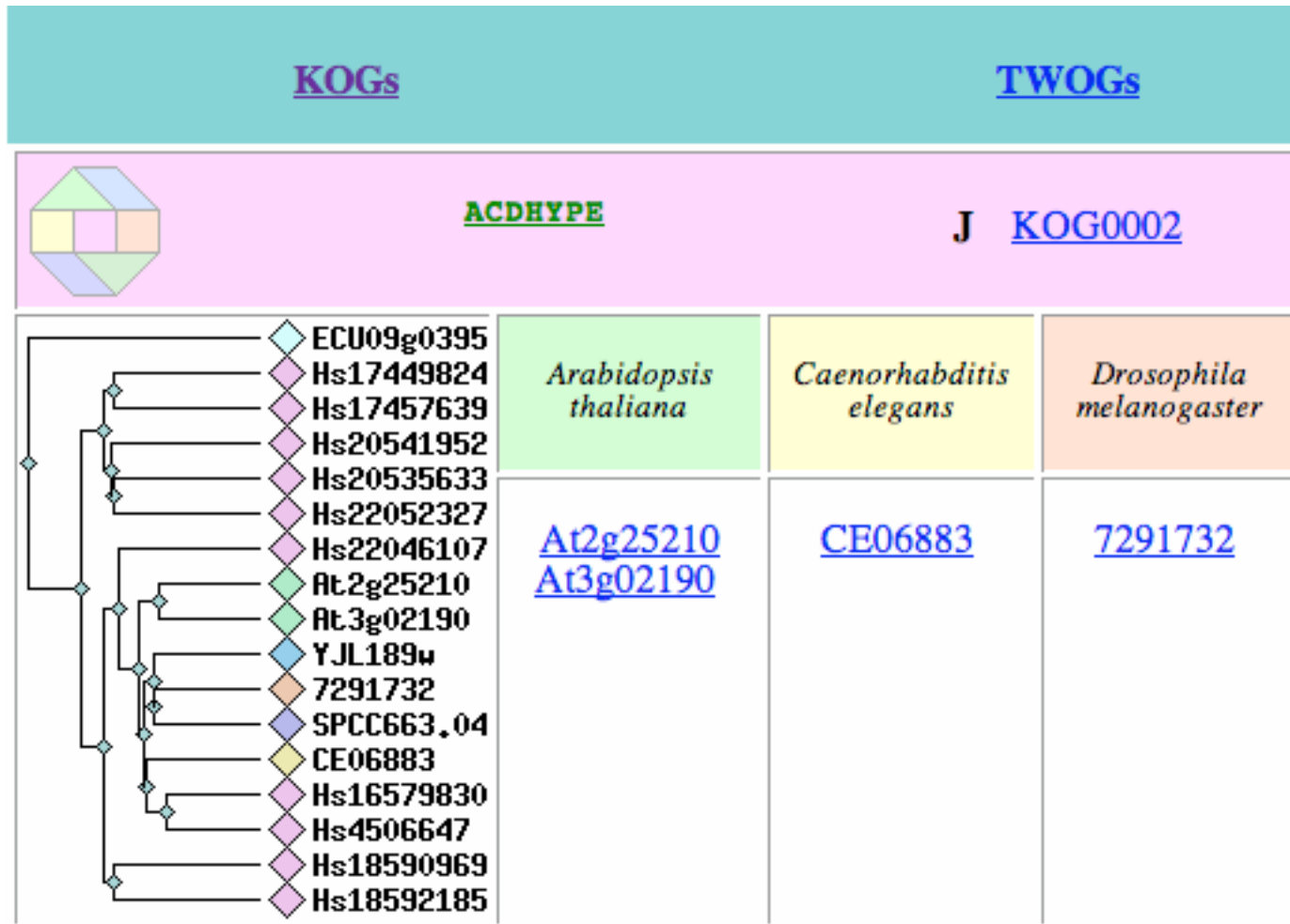
Drosophila Protein1

Identification of Orthologous Genes: Reciprocal Best Blast Hits

- 1) Detect and mask typical repetitive protein domains
- 2) All against all comparison of protein sequences (BLAST)
- 3) Detection of triangles of mutually consistent genome-specific best hits (BeTs).
- 4) Merging BeTs with a common side to form preliminary KOGs
- 5) Curate KOGs, e.g., split KOGs that are artificially bridged by multi-domain proteins.

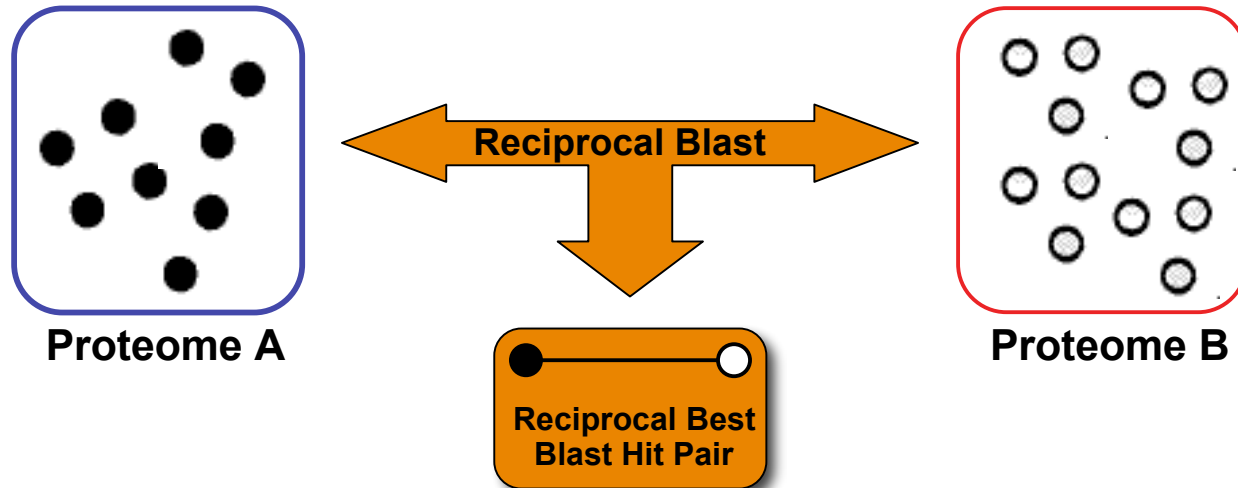


One example of a eukaryotic group of 'orthologous' genes

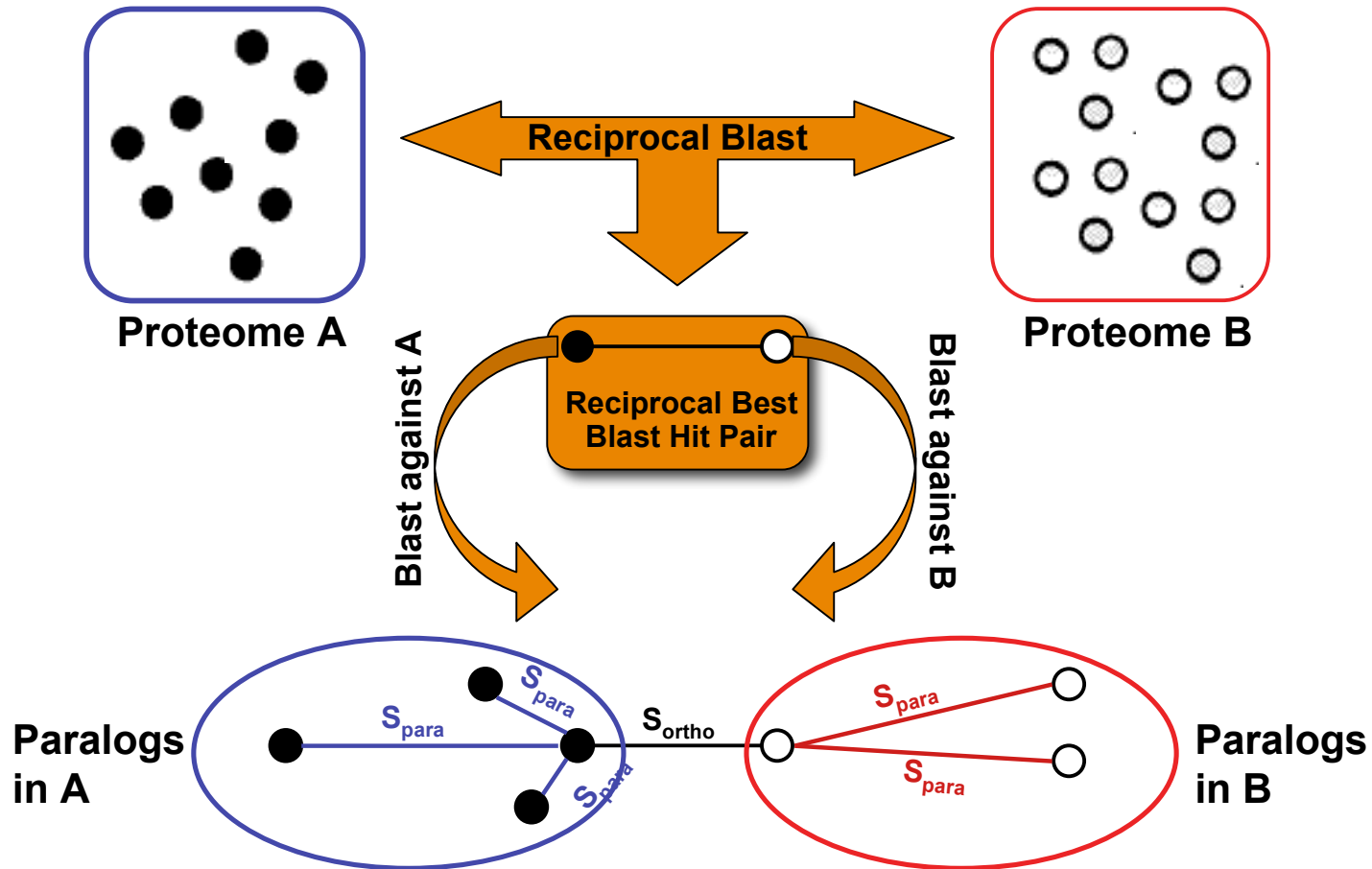


from: <http://www.ncbi.nlm.nih.gov/COG/grace/shokog.cgi?KOG0002>

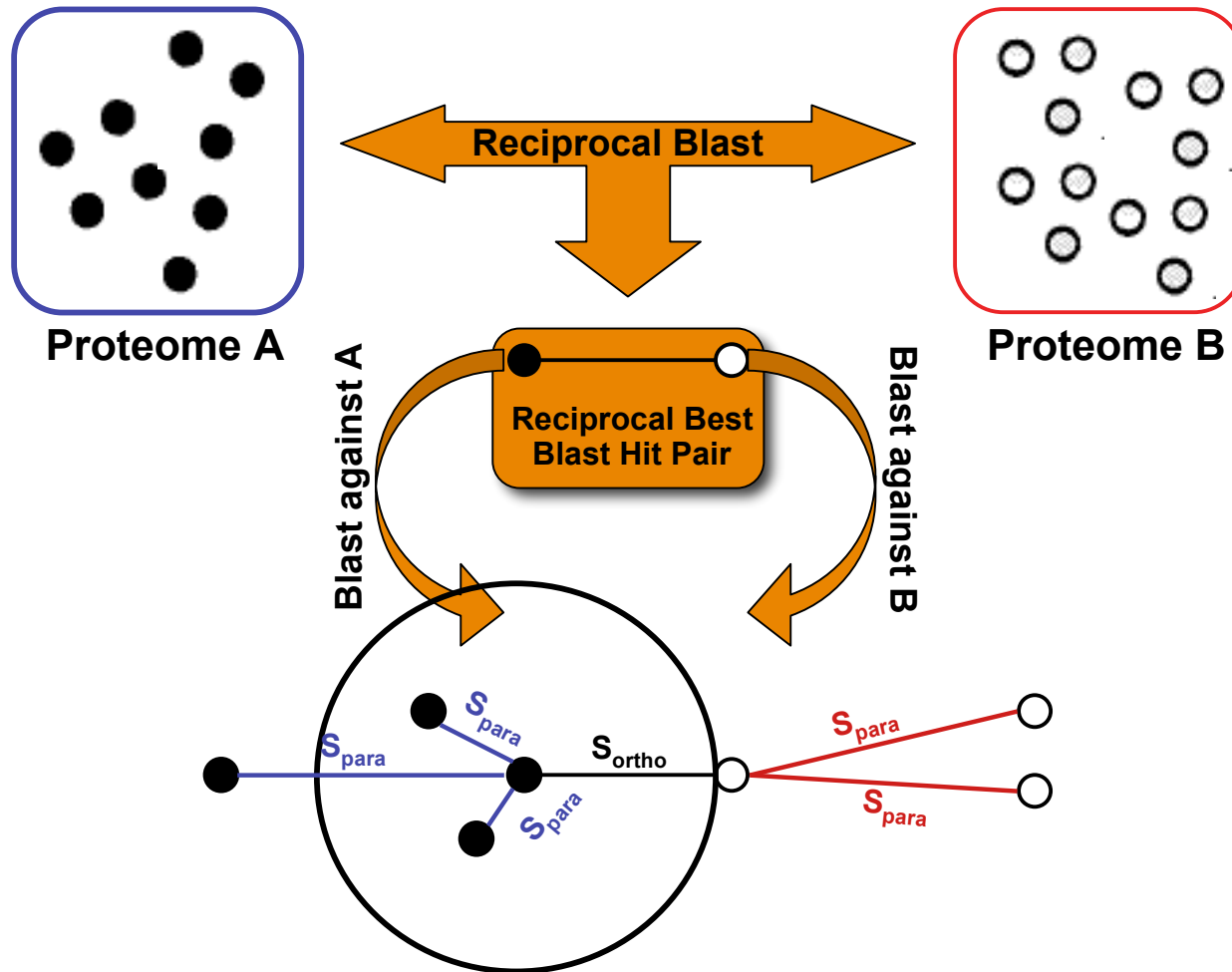
Identification of Orthologous Genes: The Inparanoid Approach



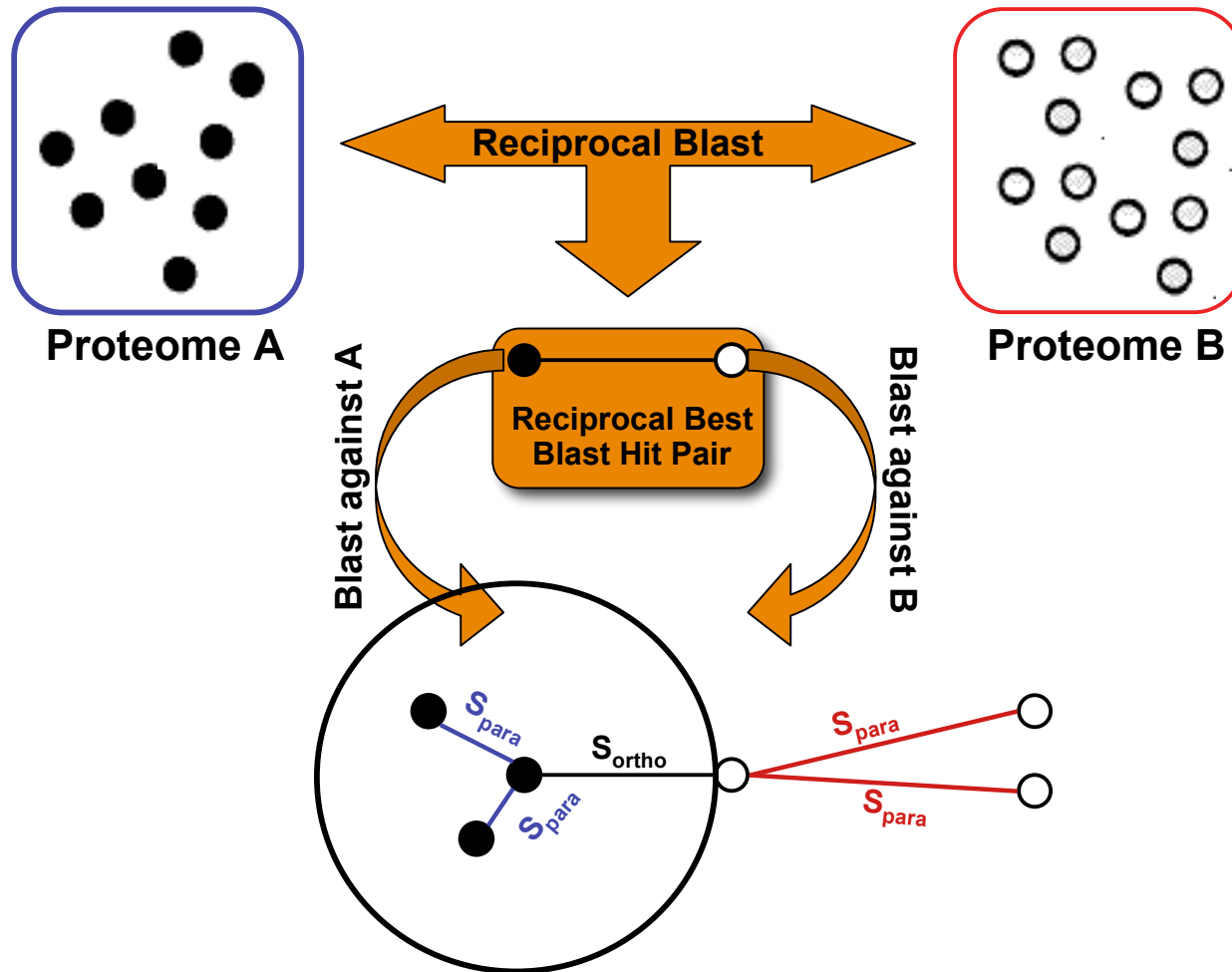
Identification of Orthologous Genes: The Inparanoid Approach



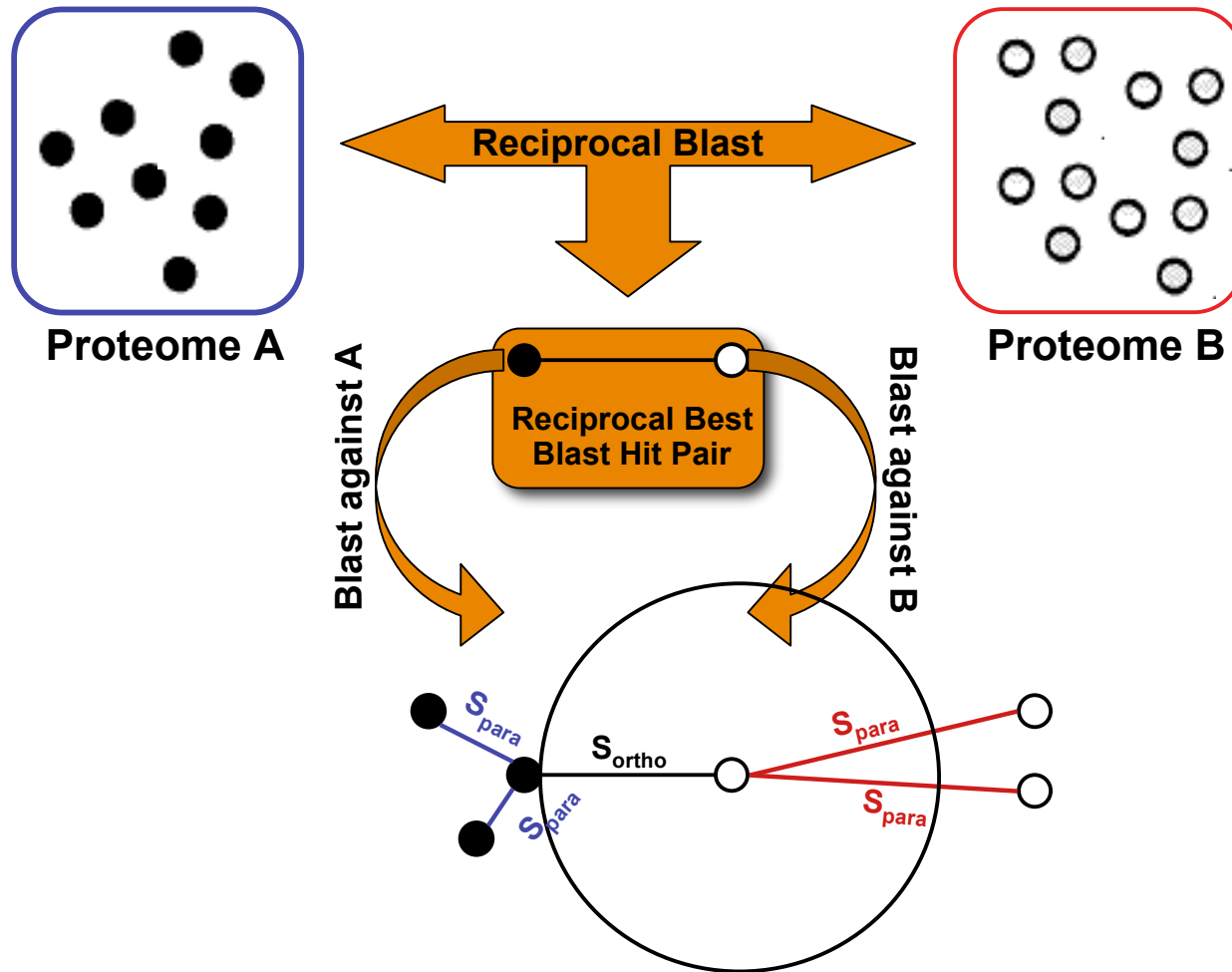
Identification of Orthologous Genes: The Inparanoid Approach (pruning)



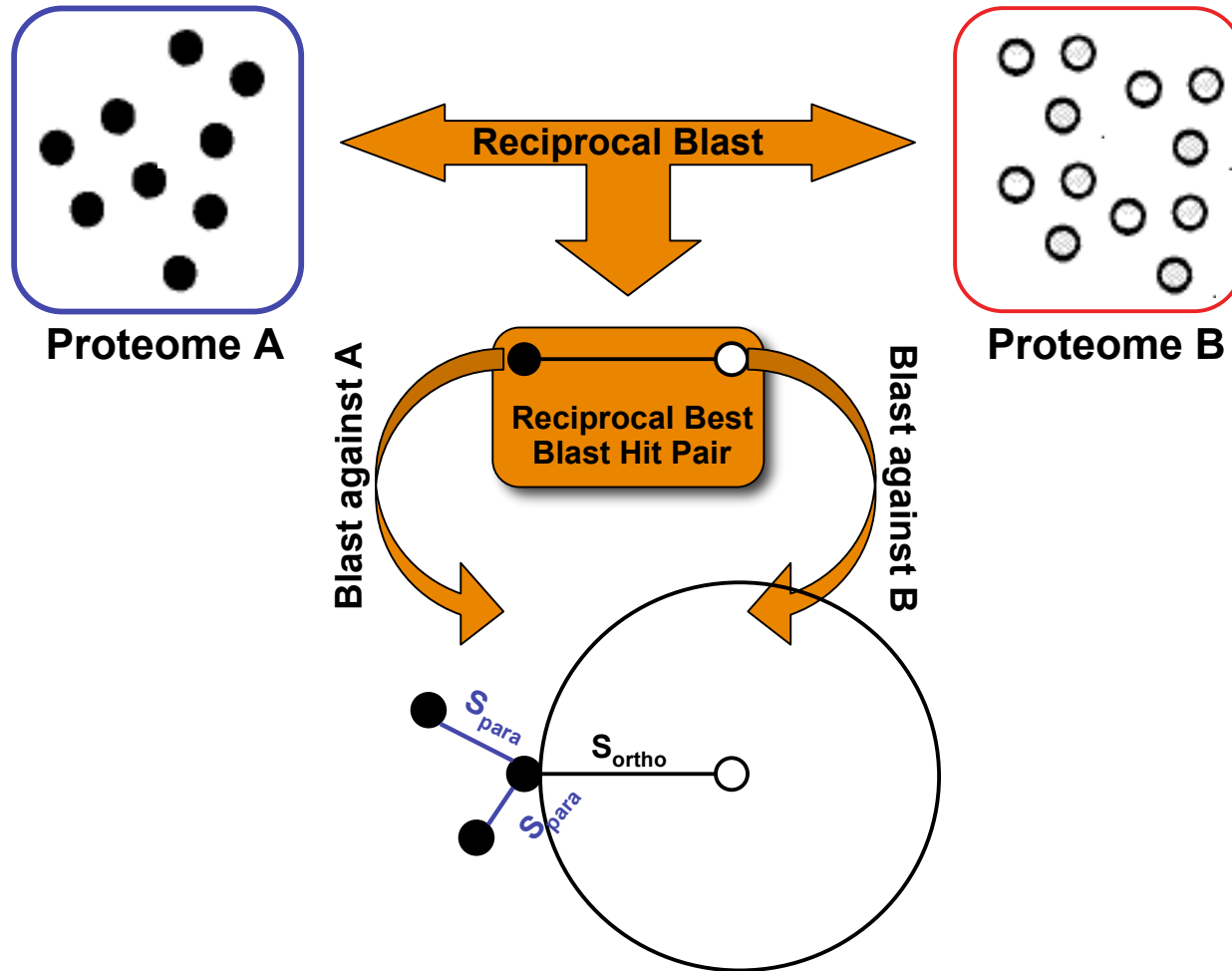
Identification of Orthologous Genes: The Inparanoid Approach (pruning)



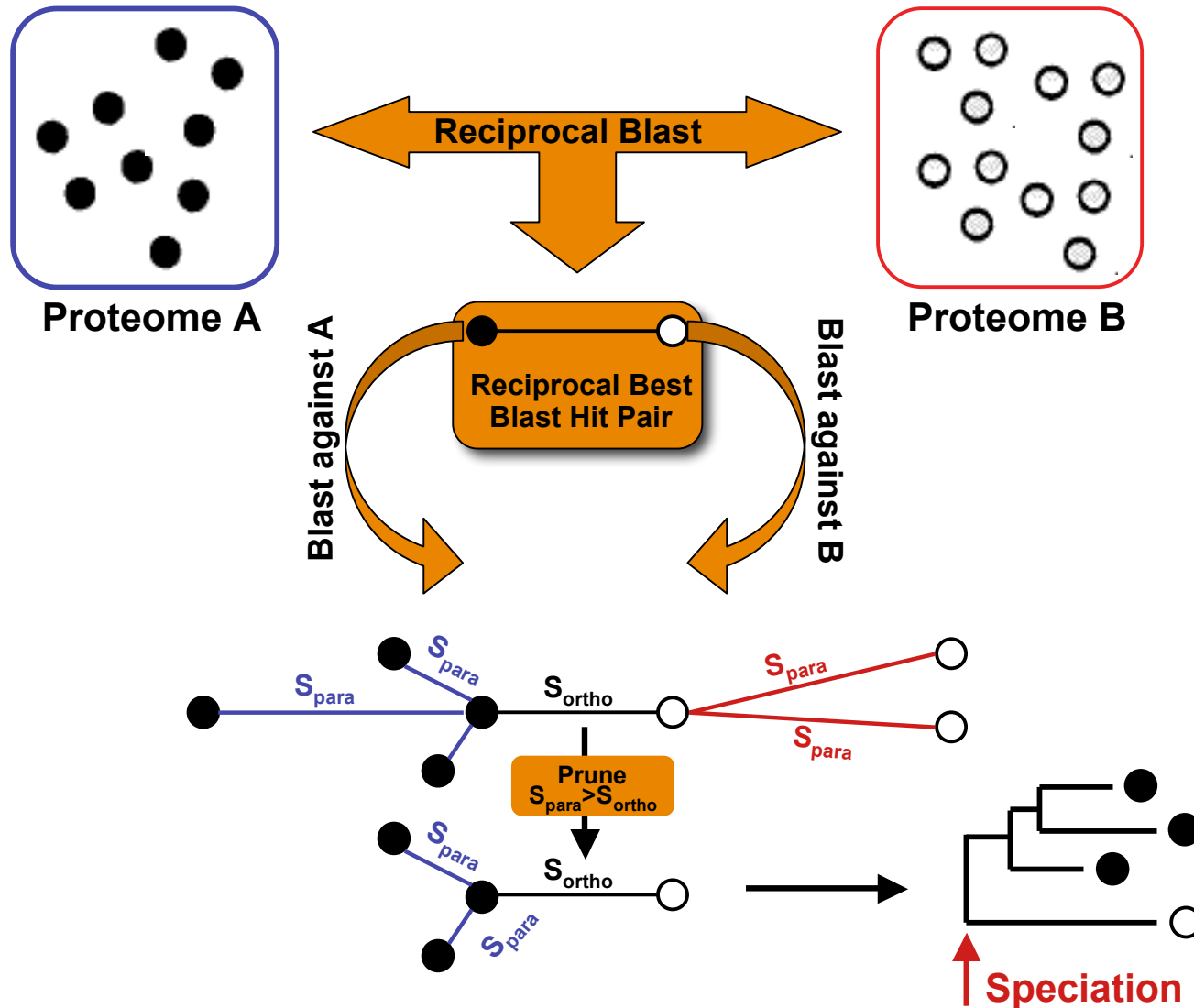
Identification of Orthologous Genes: The Inparanoid Approach (pruning)



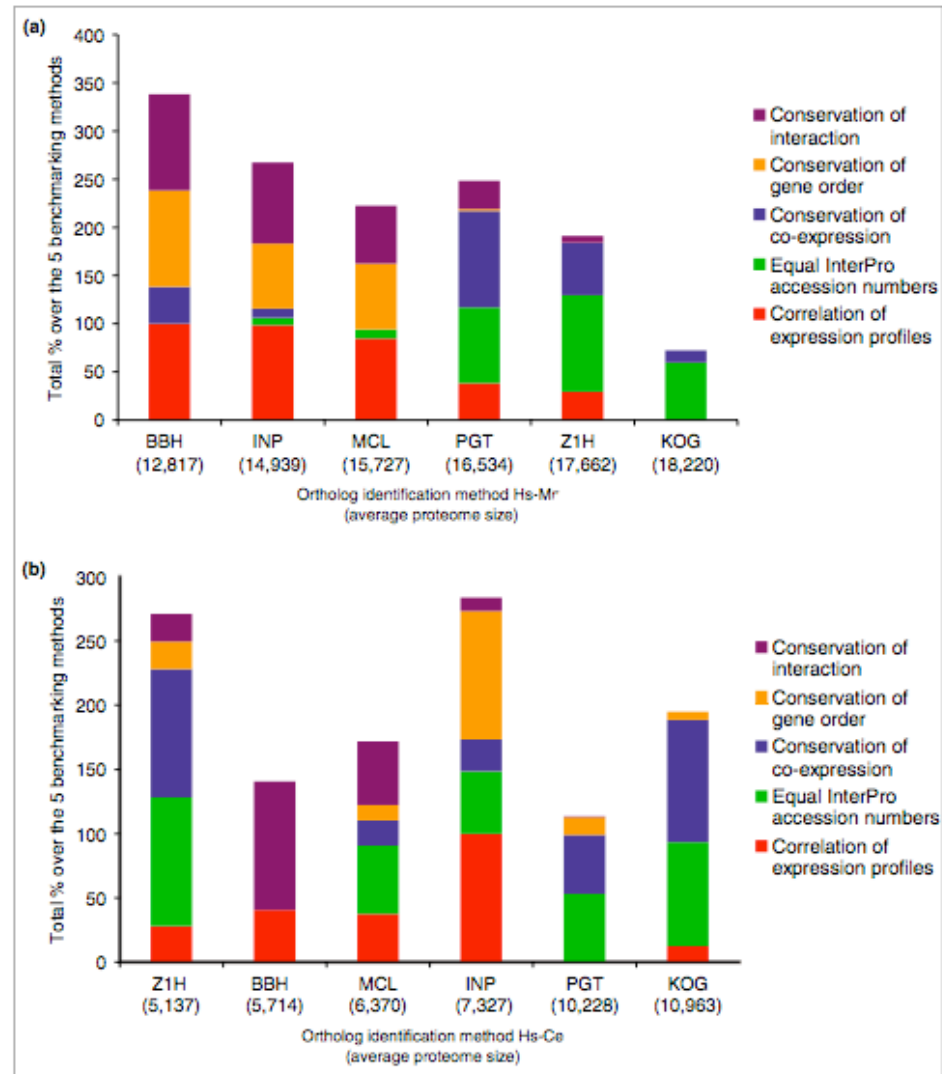
Identification of Orthologous Genes: The Inparanoid Approach (pruning)



Identification of Orthologous Genes: The Inparanoid Approach

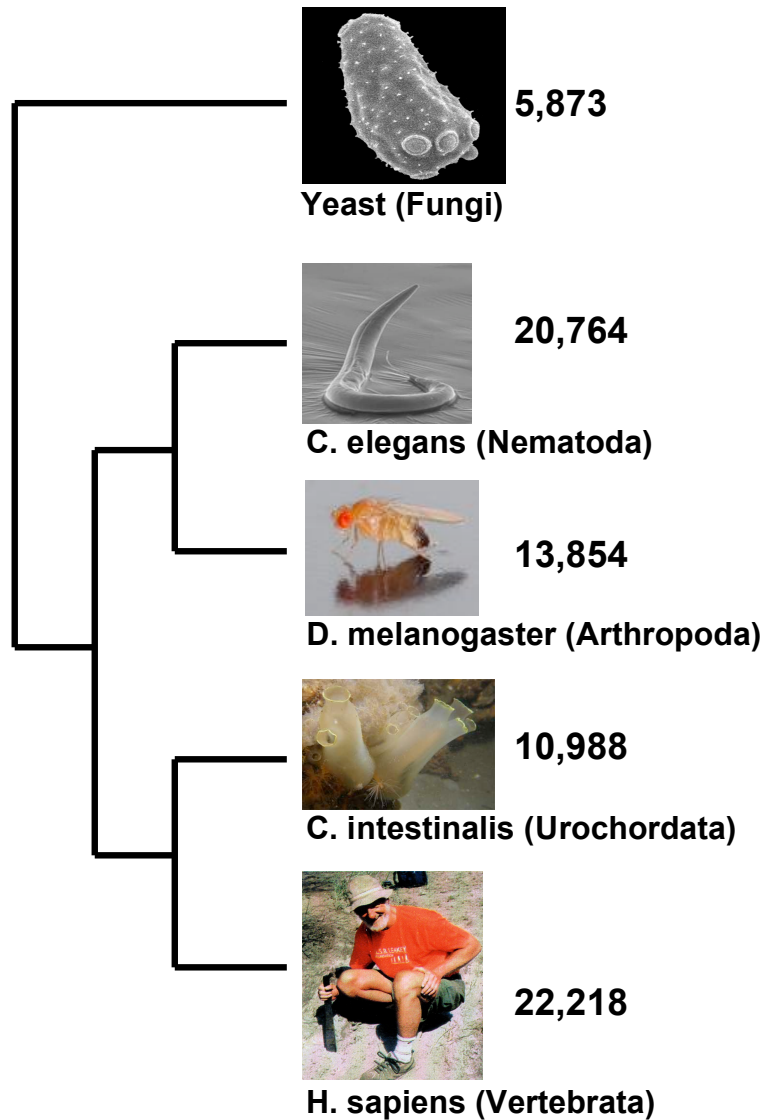


An evaluation of orthology prediction methods

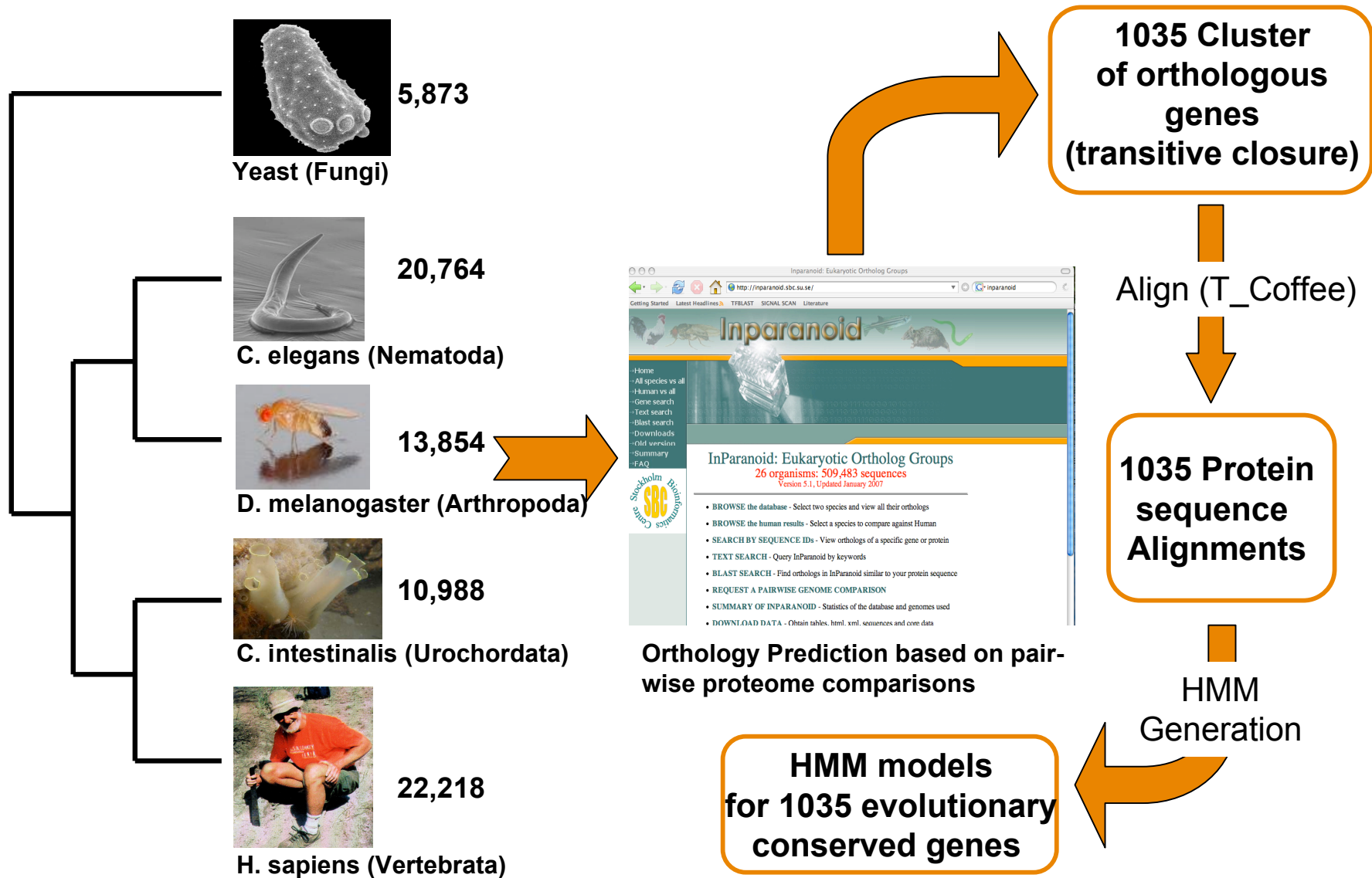


From: Hulsen et al. (2006) Benchmarking orthology prediction methods... Genome Biology 7:R31

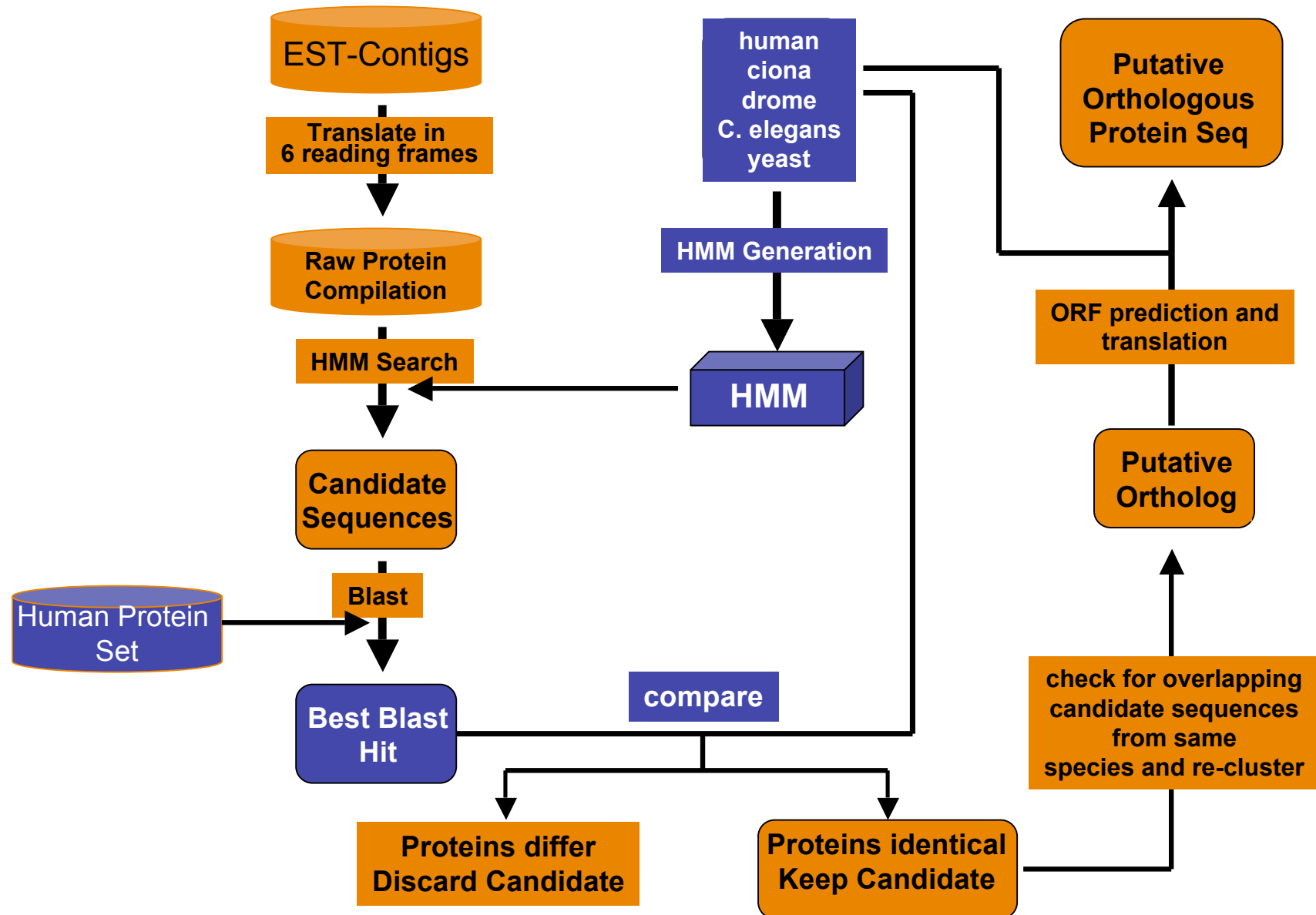
Determining a Core Gene Set for Phylogeny Reconstruction



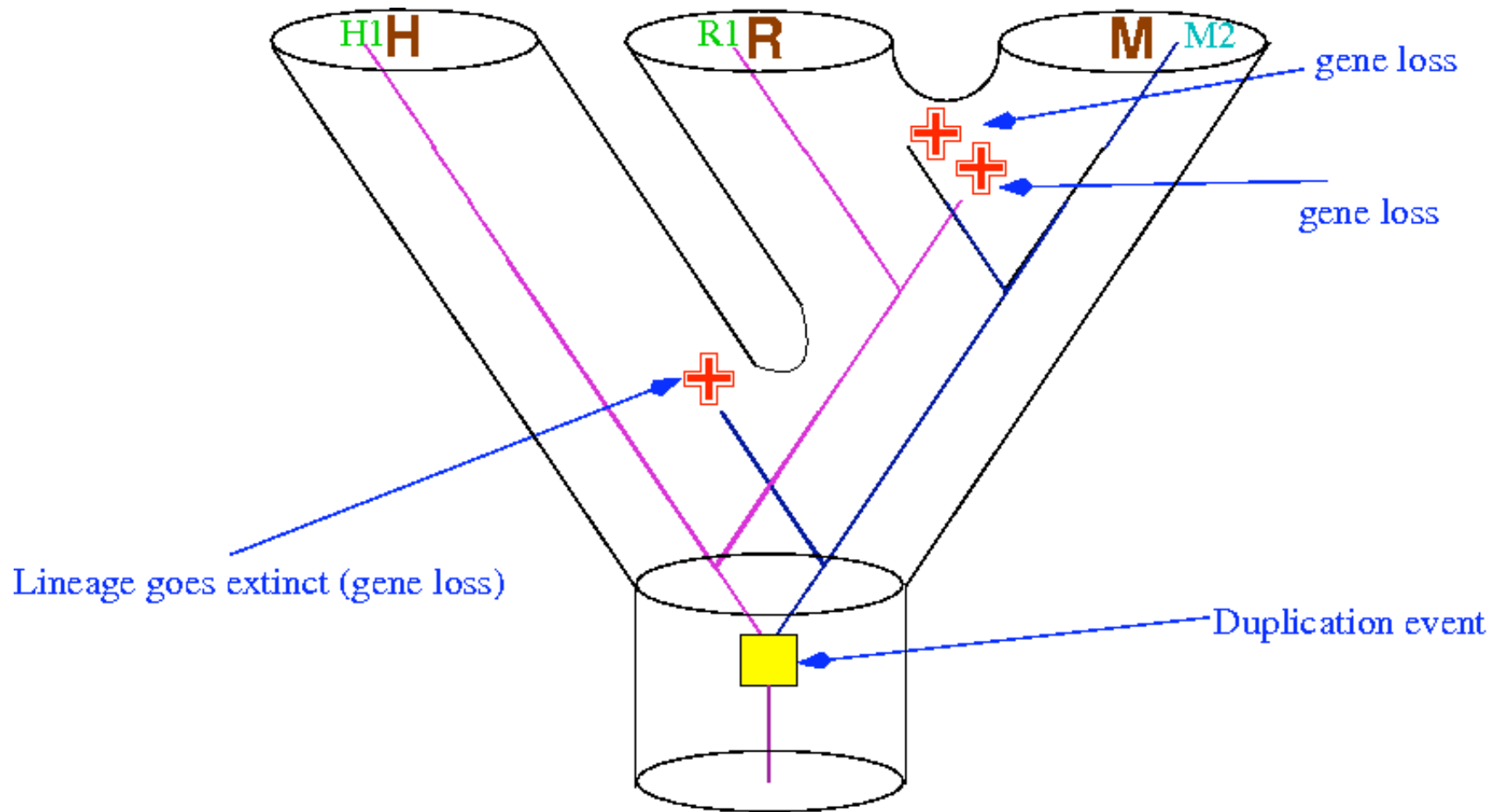
Determining a Core Gene Set for Phylogeny Reconstruction



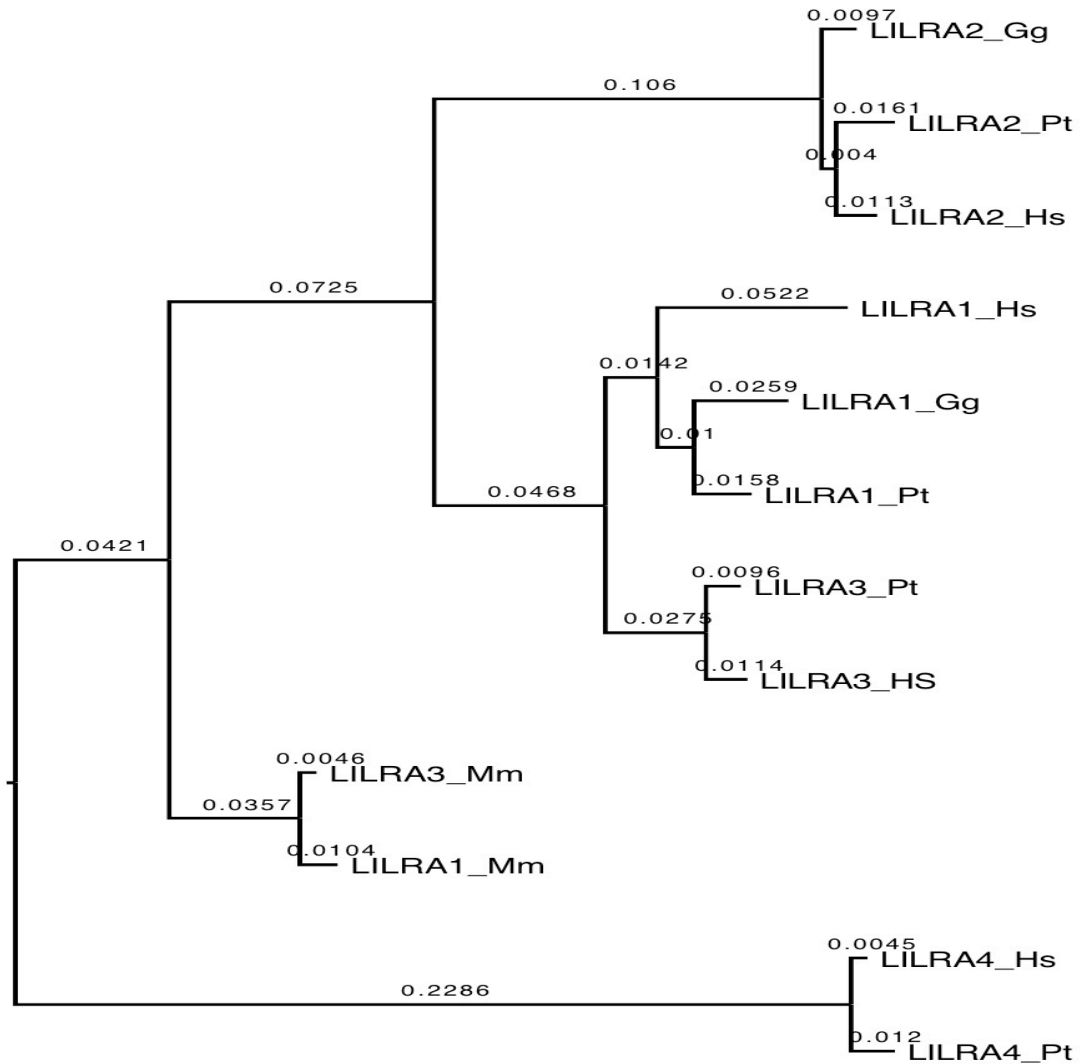
Screening EST-Data for the presence of orthologous genes



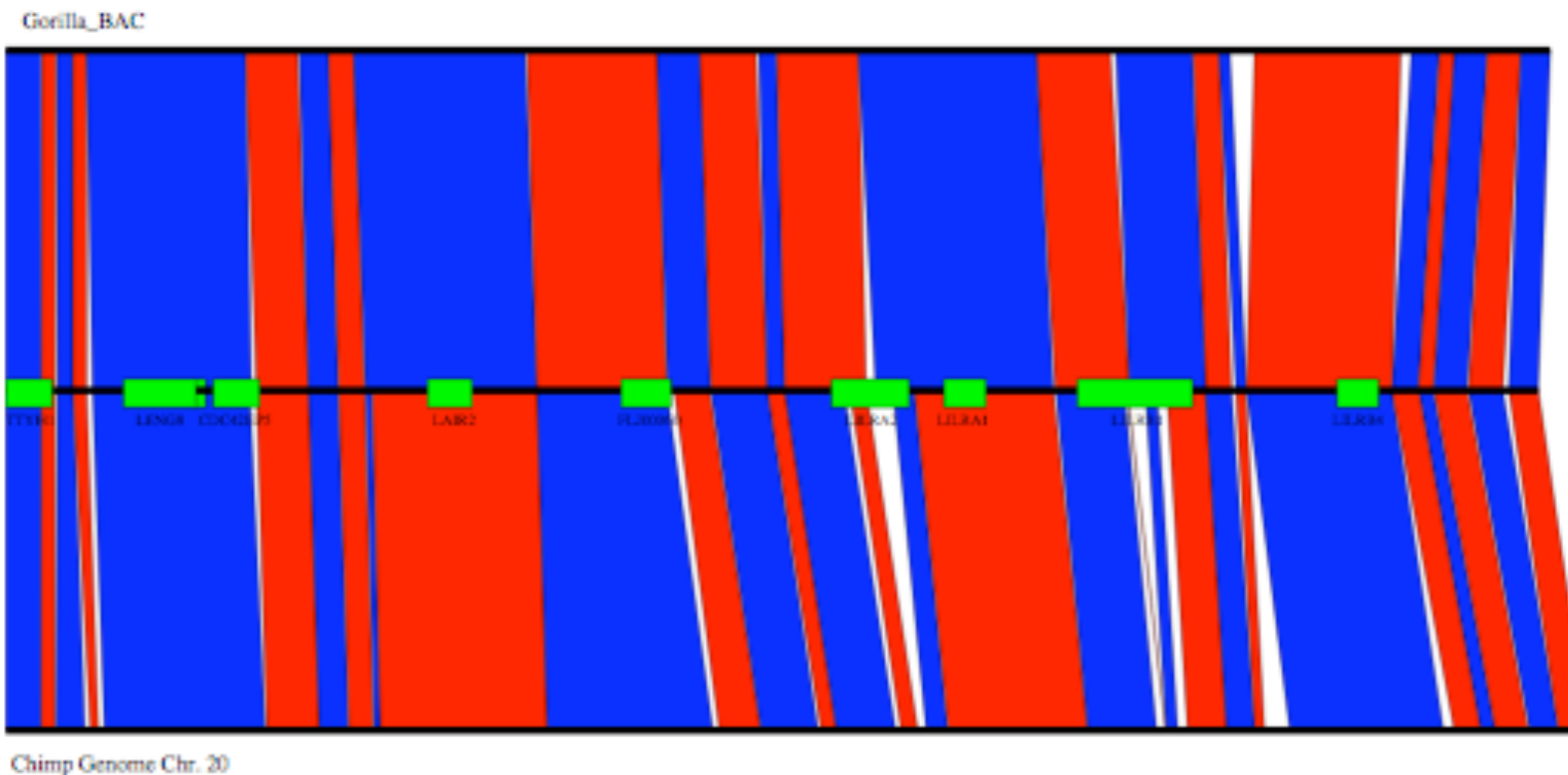
The Problem of Hidden paralogy



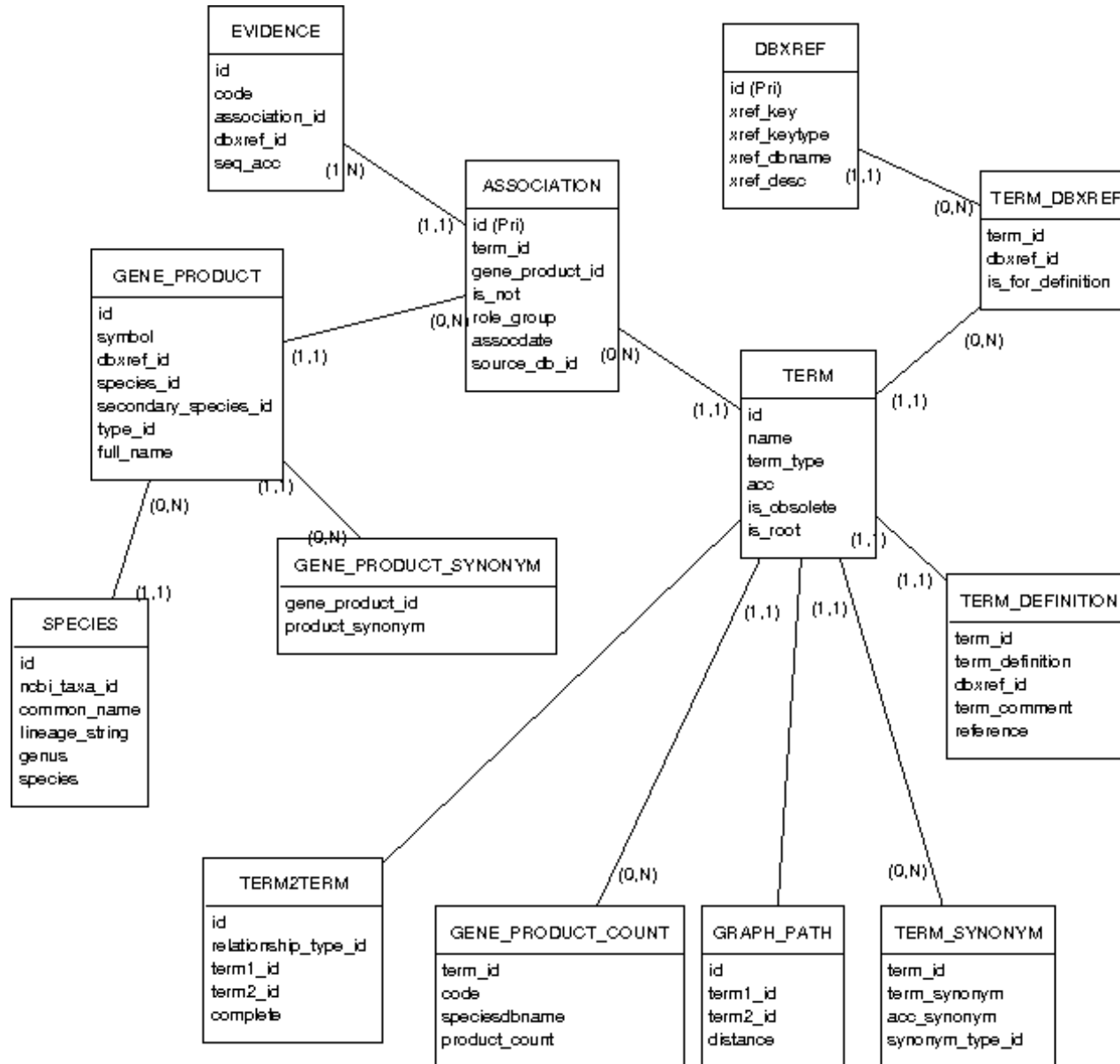
The ML-tree of the LILRA Gene Family (CDS)



Conservation of Gene order



Data management



An Example:

Cox1_human, a human protein that binds other proteins as well as metal ions (copper and iron), acts as an oxidoreductase and oxidizes cytochrome-c, hence we assign it a cytochrome-c oxidase activity

An Example:

Cox1_human, a human protein that binds other proteins as well as metal ions (copper and iron), acts as an oxidoreductase and oxidizes cytochrome-c, hence we assign it a cytochrome-c oxidase activity....

and, what I forgot to mention....,

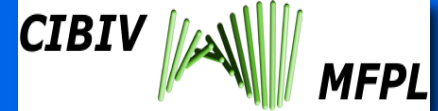
we think it belongs to a group of proteins with terminal oxidase activity, it has a hydrogen ion transporter activity -- which in fact it shares with at least 13 other proteins in humans (VPP1_HUMAN VATG1_HUMAN VATG3_HUMAN ATP5S_HUMAN VPP4_HUMAN AT1B4_HUMAN VPP2_HUMAN VATG2_HUMAN ATP6_HUMAN Q8WXQ4_HUMAN Q96F77_HUMAN ATPK_HUMAN VPP3_HUMAN), and it exerts its oxidoreductase activity on heme groups.....

Categorize Information



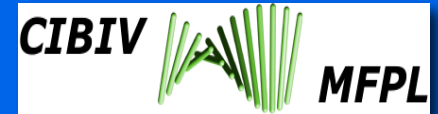
Protein_Name	Description
Cox1_human	<p>a human protein that binds other proteins as well as metal ions (copper and iron), acts as an oxidoreductase and oxidizes cytochrome-c, hence we assign it a cytochrome-c oxidase activity, it belongs to a group of proteins with terminal oxidase activity, it has a hydrogen ion transporter activity, which it shares with at least 13 other proteins in humans (VPP1_HUMAN VATG1_HUMAN VATG3_HUMAN ATP5S_HUMAN VPP4_HUMAN AT1B4_HUMAN VPP2_HUMAN VATG2_HUMAN ATP6_HUMAN Q8WXQ4_HUMAN Q96F77_HUMAN ATPK_HUMAN VPP3_HUMAN), and it exerts its oxidoreductase activity on heme groups.</p>

Organize Information



		Attribute 1 (data type/domain)	Attribute 2 (data type/domain)	
		Protein_Name (Char)	Description (Char)	
REALATION (TABLE)	Heading	Row (n-Tuple)	Cox1_human	a human protein that binds other proteins as well as metal ions (copper and iron), acts as an oxidoreductase and oxidizes cytochrome-c, hence we assign it a cytochrome-c oxidase activity, it belongs to a group of proteins with terminal oxidase activity, it has a hydrogen ion transporter activity, which it shares with at least 13 other proteins in humans (..), and it exerts its oxidoreductase activity on heme groups.

The concept of *keys*



Protein_Name (Char)	Description (Char)
Cox1_human	a human protein that binds other proteins as well as metal ions (copper and iron), acts as an oxidoreductase and oxidizes cytochrome-c, hence we assign it a cytochrome-c oxidase activity, it belongs to a group of proteins with terminal oxidase activity, it has a hydrogen ion transporter activity, which it shares with at least 13 other other proteins in humans (..), and it exerts its oxidoreductase activity on heme groups.

Key Attribute: Every row (tuple) in a table (**relation**) can be uniquely identified by the value of the key attribute. Hence, values of key attributes need to be unique and **NOT NULL** in the relation.

Protein_Name (Char, PRI)	Description (Char)
Cox1_human	a human protein that binds other proteins as well as metal ions (copper and iron), acts as an oxidoreductase and oxidizes cytochrome-c, hence we assign it a cytochrome-c oxidase activity, it belongs to a group of proteins with terminal oxidase activity, it has a hydrogen ion transporter activity, which it shares with at least 13 other other proteins in humans (..), and it exerts its oxidoreductase activity on heme groups.

Keys can consist of a **single attribute** or can be **composites** of several attributes.

Primary key: The key that is chosen among all available keys for a relation. It is used in transactions in preference over the **alternate keys**.

Protein_Name (Char, PRI, UNIQUE/NOT NULL)	Description (Char)
Cox1_human	a human protein that binds other proteins as well as metal ions (copper and iron), acts as an oxidoreductase and oxidizes cytochrome-c, hence we assign it a cytochrome-c oxidase activity, it belongs to a group of proteins with terminal oxidase activity, it has a hydrogen ion transporter activity, which it shares with at least 13 other other proteins in humans (..), and it exerts its oxidoreductase activity on heme groups.

Consistency is enforced not by rules built into the applications but rather by **constraints** declared as part of the logical schema.

Protein_Name (Char, PRI, UNIQUE/NOT NULL)	Description (Char)
Cox1_human	a human protein that binds other proteins as well as metal ions (copper and iron), acts as an oxidoreductase and oxidizes cytochrome-c, hence we assign it a cytochrome-c oxidase activity, it belongs to a group of proteins with terminal oxidase activity, it has a hydrogen ion transporter activity, which it shares with at least 13 other other proteins in humans (..), and it exerts its oxidoreductase activity on heme groups.

All information is represented by data values

Very nice...but



...how to manage this information a whee bit more efficiently??

PROTEIN

Protein_Name (Char, PRI, UNIQUE/NOT NULL)	Description (Char)	Species (Char)
Cox1	binds other proteins as well as metal ions (copper and iron), acts as an oxidoreductase and oxidizes cytochrome-c, hence we assign it a cytochrome-c oxidase activity, it belongs to a group of proteins with terminal oxidase activity, it has a hydrogen ion transporter activity, which it shares with at least 13 other proteins in humans (..), and it exerts its oxidoreductase activity on heme groups.	Human

Continue with the categorization: Split the description of functions



PROTEIN

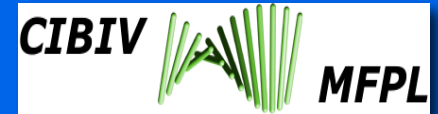
Protein_Name	Function (Char)	Description (Char)	Species (Char)
Cox1	binds other proteins as well as metal ions (copper and iron), acts as an oxidoreductase and oxidizes cytochrome-c, cytochrome-c oxidase activity, terminal oxidase activity, it has a hydrogen ion transporter activity,	shares hydrogen ion transporter activity, with at least 13 other proteins in humans (..), exerts its oxidoreductase activity on heme groups.	Human

PROTEIN

Protein_Name (Pri)	Function	Description	Species
Cox1_human	binds other proteins as well as metal ions (copper and iron), acts as an oxidoreductase and oxidizes cytochrome-c, cytochrome-c oxidase activity, terminal oxidase activity, it has a hydrogen ion transporter activity,	shares hydrogen ion transporter activity with at least 13 other proteins in humans (..), exerts its oxidoreductase activity on heme groups.	Human

- Attribute values must be atomic
- no attribute must occur a different number of times on different records (add rows rather than columns)

Normalization: 1st normal form



PROTEIN

Protein_Name (Pri)	Function	Description_id	Species
Cox1	protein binding	1	Human
Cox1	binds metal ions (copper and iron)	1	Human
Cox1	oxidoreductase	1	Human
Cox1	cytochrome-c oxidase	1	Human
Cox1	terminal oxidase	1	Human
Cox1	hydrogen ion transporter activity	1	Human

Normalization: 1st normal form



PROTEIN

Protein_Name (Pri)	Function	Description_id	Species
Cox1	protein binding	1	Human
Cox1	binds metal ions (copper and iron)	1	Human
Cox1	oxidoreductase	1	Human
Cox1	cytochrome-c oxidase	1	Human
Cox1	terminal oxidase	1	Human
Cox1	hydrogen ion transporter activity	1	Human

NO KEY

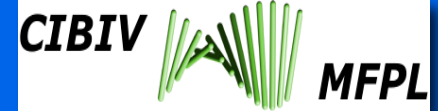
Normalization: 1st normal form



PROTEIN

Protein_Name (Pri)	Function (Pri)	Description_id	Species
Cox1	protein binding	1	Human
Cox1	binds metal ions (copper and iron)	1	Human
Cox1	oxidoreductase	1	Human
Cox1	cytochrome-c oxidase	1	Human
Cox1	terminal oxidase	1	Human
Cox1	hydrogen ion transporter activity	1	Human

Normalization: 1st normal form



PROTEIN

Protein_Name (Pri)	Function (Pri)	Description_id (Integer)	Species
Cox1	protein binding	1	Human
Cox1	binds metal ions (copper and iron)	1	Human
Cox1	oxidoreductase	1	Human
Cox1	cytochrome-c oxidase	1	Human
Cox1	terminal oxidase	1	Human
Cox1	hydrogen ion transporter activity	1	Human

ProteinDescription

id (Pri)	Description
1	shares hydrogen ion transporter activity with at least 13 other other proteins in humans (..), exerts its oxidoreductase activity on heme groups.

Foreign Key: integrity constraint. Value of the attribute is drawn from a key in another relation

PROTEIN

Protein_Name (Pri)	Function_Id (Pri)	Function	Description_id (Integer)	Species
Cox1	1	protein binding	1	Human
Cox1	2	binds metal ions (copper and iron)	1	Human
Cox1	3	oxidoreductase	1	Human
Cox1	4	cytochrome-c oxidase	1	Human
Cox1	5	terminal oxidase	1	Human
Cox1	6	hydrogen ion transporter activity	1	Human

PROTEIN

Protein_Name (Pri)	Function_Id (Pri)	Function	Description_id (Integer)	Species
Cox1	1	protein binding	1	Human
Cox1	2	binds metal ions (copper and iron)	1	Human
Cox1	3	oxidoreductase	1	Human
Cox1	4	cytochrome-c oxidase	1	Human
Cox1	5	terminal oxidase	1	Human
Cox1	6	hydrogen ion transporter activity	1	Human

- the table must be in 1NF
- none of the non-prime attributes are functionally dependent on a part of a composite key (automatically fulfilled if none of the keys are composite!)

PROTEIN

Protein_Name (Pri)	Function_Id (Pri)	Function	Desc_id (Integer)	Species
Cox1	1	protein binding	1	Human
Cox1	2	binds metal ions (copper and iron)	1	Human
Cox1	3	oxidoreductase	1	Human
Cox1	4	cytochrome-c oxidase	1	Human
Cox1	5	terminal oxidase	1	Human
Cox1	6	hydrogen ion transporter activity	1	Human

- the table must be in 1NF
- none of the non-prime attributes are functionally dependent on a part of a candidate (automatically fulfilled if none of the keys are composite!)

Normalization: 2nd normal form



PROTEIN

Protein_Name (Pri)	Function_Id	Desc_id (Integer)	Species
Cox1	1	1	Human

ProteinDescription

id (Pri)	Description
1	shares hydrogen ion transporter activity with at least 13 other other proteins in humans (..), exerts its oxidoreductase activity on heme groups.

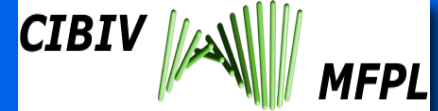
Functional_annotation

ID (Integer, Pri)	Prot_Func_Id (Integer,Pri)
1	1
1	2
1	3
1	4
1	5
1	6

PROTEIN_Function

ID (Integer, Pri)	Function
1	protein binding
2	binds metal ions (copper and iron)
3	oxidoreductase
4	cytochrome-c oxidase
5	terminal oxidase
6	hydrogen ion transporter activity

Cross referencing



PROTEIN

Protein_Name (Pri)	Function_Id (Pri)	Desc_id (Integer)	Species_id	Species
Cox1	1	1	92001	Human

ProteinDescription

id (Pri)	Description
1	shares hydrogen ion transporter activity with at least 13 other other proteins in humans (..), exerts its oxidoreductase activity on heme groups.

Functional_annotation

ID (Integer, Pri)	Prot_Func_Id (Integer, Pri)
1	1
1	2
1	3
1	4
1	5
1	6

PROTEIN_Function

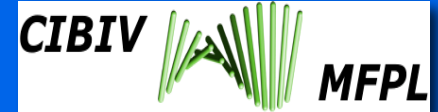
ID (Integer, Pri)	Function
1	protein binding
2	binds metal ions (copper and iron)
3	oxidoreductase
4	cytochrome-c oxidase
5	terminal oxidase
6	hydrogen ion transporter activity

PROTEIN

Protein_Name (Pri)	Function_Id	Desc_id (Integer)	Species_ id	Species
Cox1	1	1	92001	Human

- the table must be in 2NF
- no non-prime attribute is functionally dependent on any other non-prime attribute

Normalization: 3^d normal form



TAXON

Species_id (Pri)	Species
92001	Human

ProteinDescription

id (Pri)	Description
1	shares hydrogen ion transporter activity with at least 13 other other proteins in humans (..), exerts its oxidoreductase activity on heme groups.

PROTEIN

Protein_Name (Pri)	Function_Id	Species_id	Desc_id (Integer)	Species
Cox1	1	92001	1	Human

Functional_annotation

ID (Integer, Pri)	Prot_Func_Id (Integer, Pri)
1	1
1	2
1	3
1	4
1	5
1	6

PROTEIN_Function

ID (Integer, Pri)	Function
1	protein binding
2	binds metal ions (copper and iron)
3	oxidoreductase
4	cytochrome-c oxidase
5	terminal oxidase
6	hydrogen ion transporter activity

Schema of the GO Relational database

