

生物信息学：导论与方法

Bioinformatics: Introduction and Methods



<https://www.coursera.org/course/pkubioinfo>



生物信息学：导论与方法

Bioinformatics: Introduction and Methods

北京大学生物信息学中心 高歌、魏丽萍

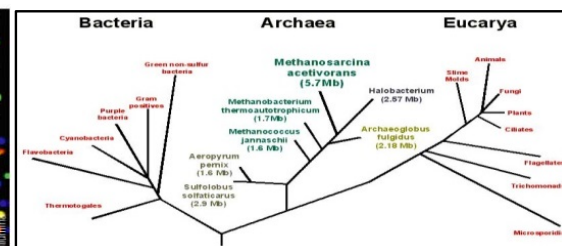
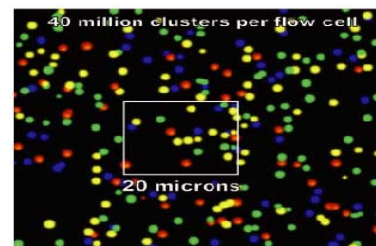
Ge Gao & Liping Wei

Center for Bioinformatics, Peking University





TAACCCTAACCCTAACCCTAACCCTAACCCTA
CCTAACCCTAACCCTAACCCTAACCCTAACC
CCCTAACCCTAACCCTAACCCTAACCCTAAC
AACCCTAACCCTAACCCTAACCCTAACCCTA
ACCCTAACCCTAACCCTAACCCTAACCCTAAC
CTACCCTAACCCTAACCCTAACCCTAACCCTA
ACCCTAACCCTAACCCTAACCCTAACCCTAA

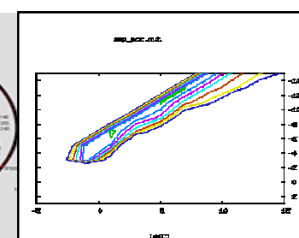
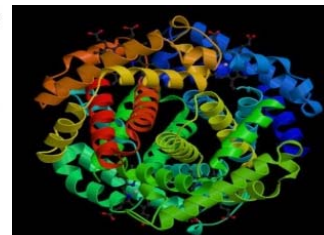
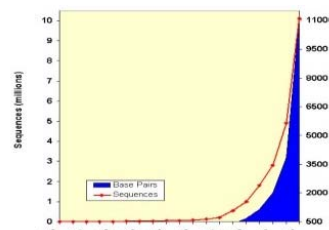
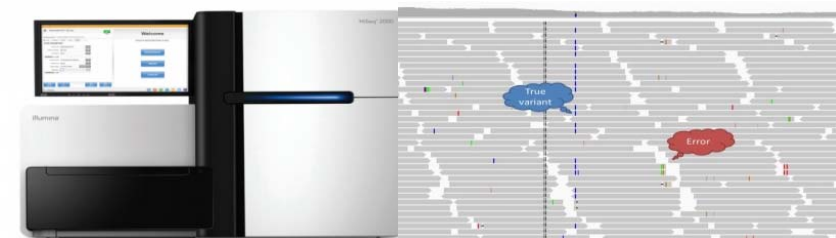


Case Study 2: From Dry to Wet, an Evolutionary Story

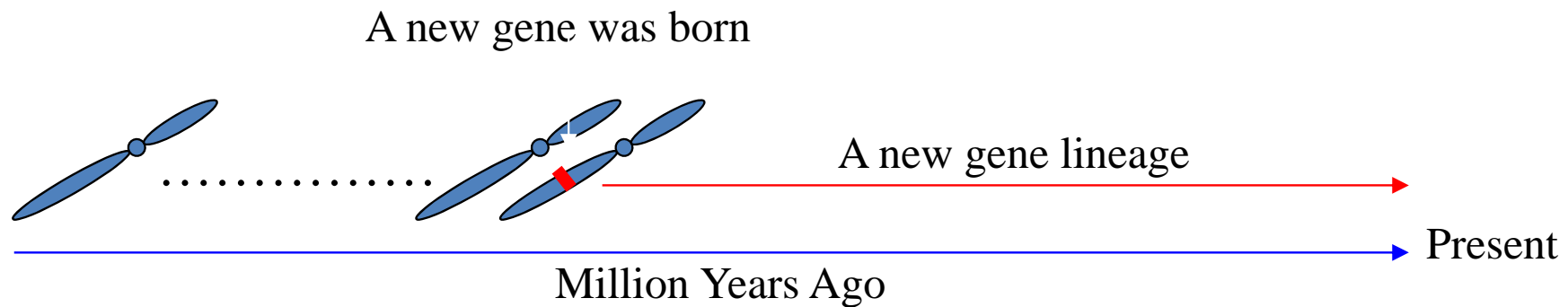
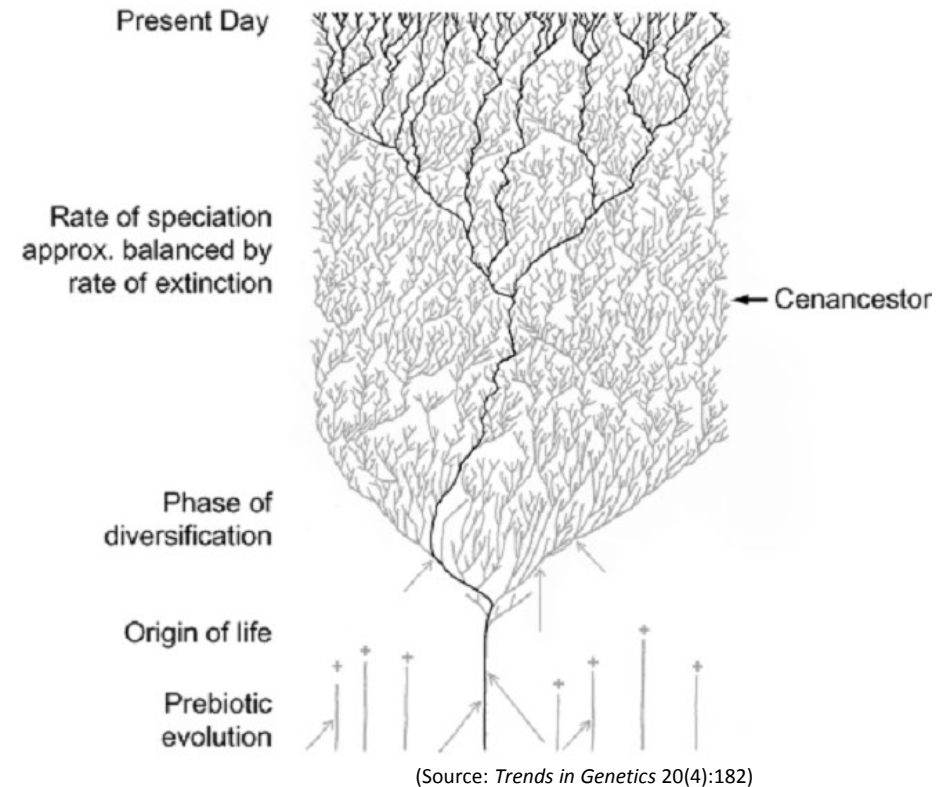
北京大学生物信息学中心 高歌

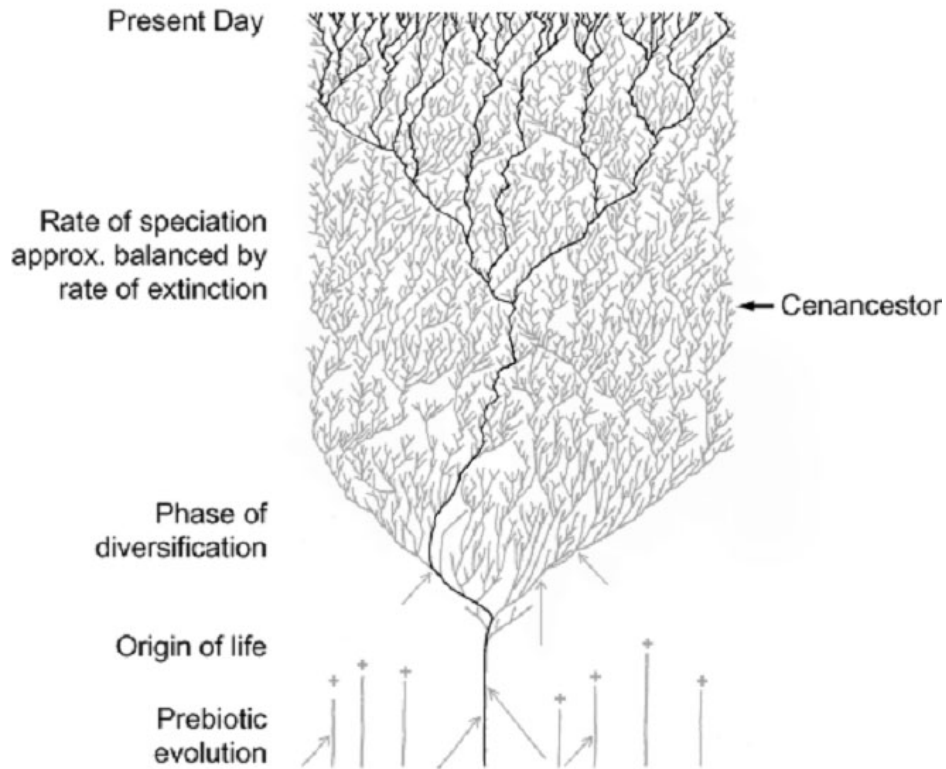
Ge Gao, Ph.D.

Center for Bioinformatics, Peking University

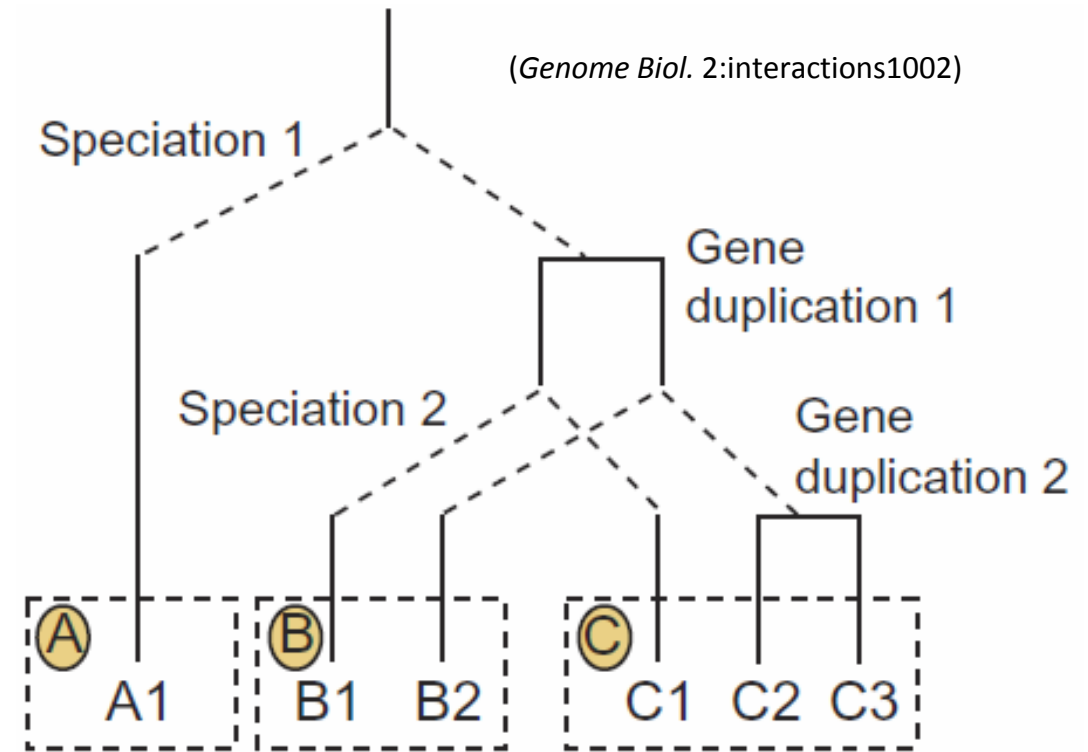


Newly emerging genes are key sources of evolutionary novelty

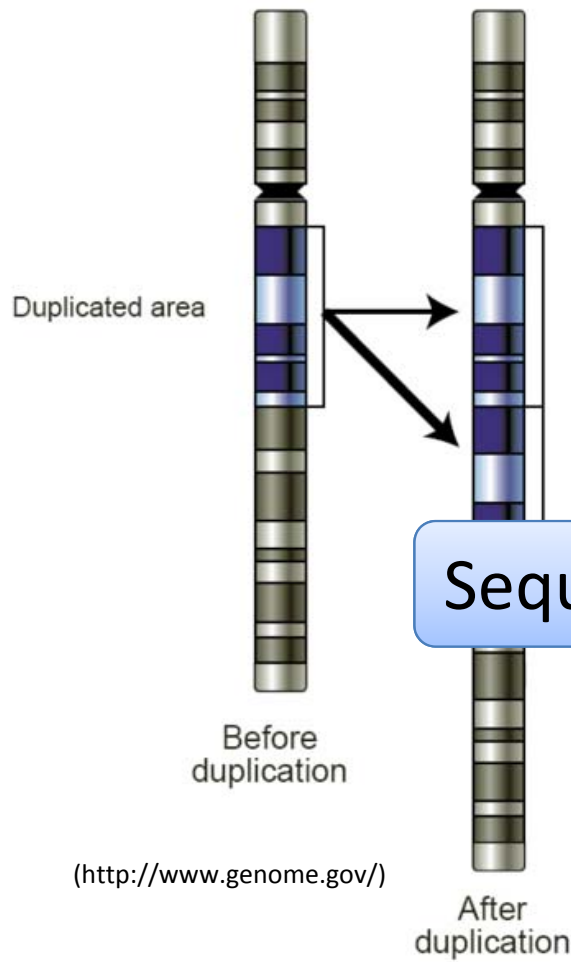




(Source: *Trends in Genetics* 20(4):182)

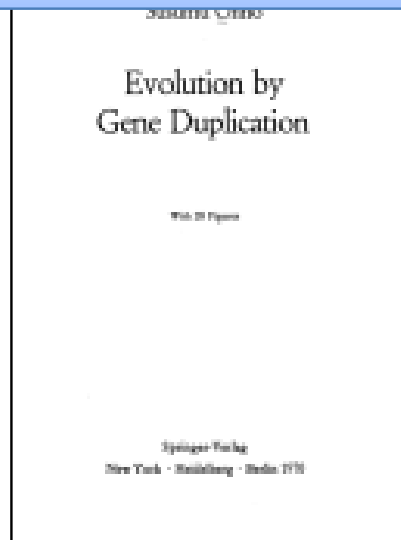


- **Orthologs**: homologous genes result from **speciation event**
- **Paralogs**: homologous genes result from **duplication event**



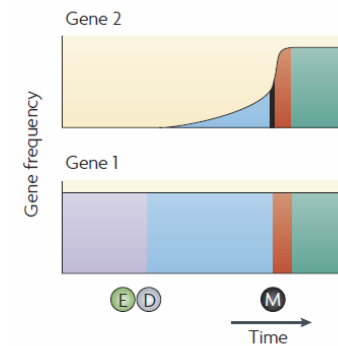
(<http://www.genome.gov/>)

Susumu Ohno
 Susumu Ohno
 February 1, 1928–January 13,
 2000

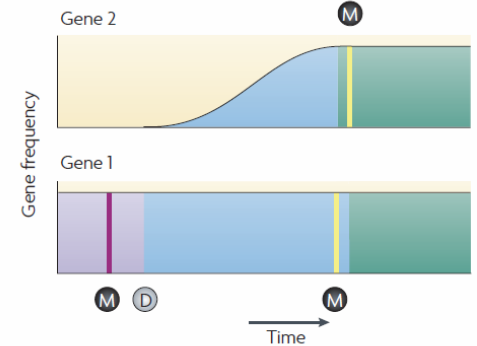


Copyright © Peking University

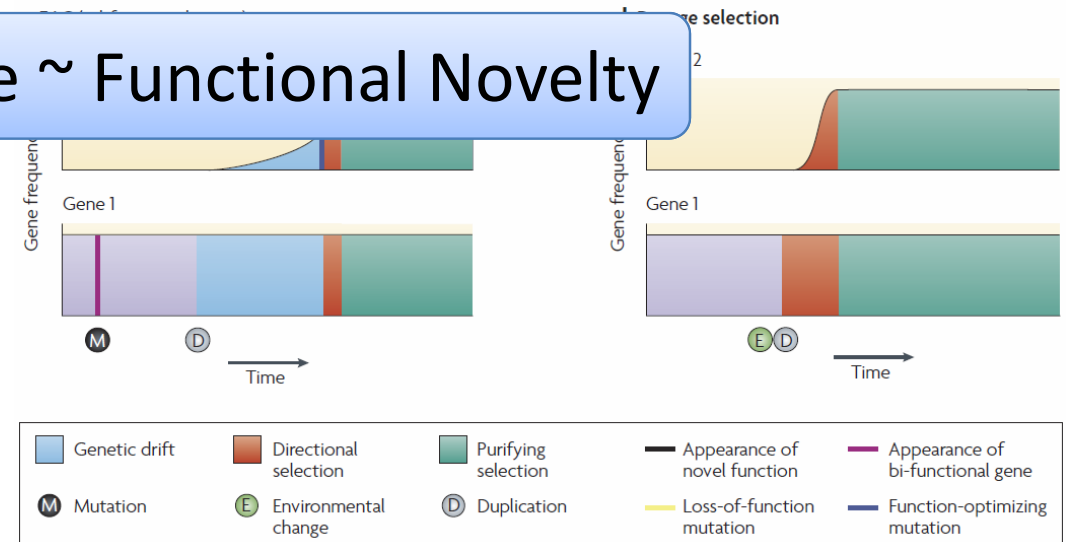
a MDN (neofunctionalization)



b DDC (subfunctionalization)



Sequence Divergence ~ Functional Novelty



(Conant *et al.* 2008)

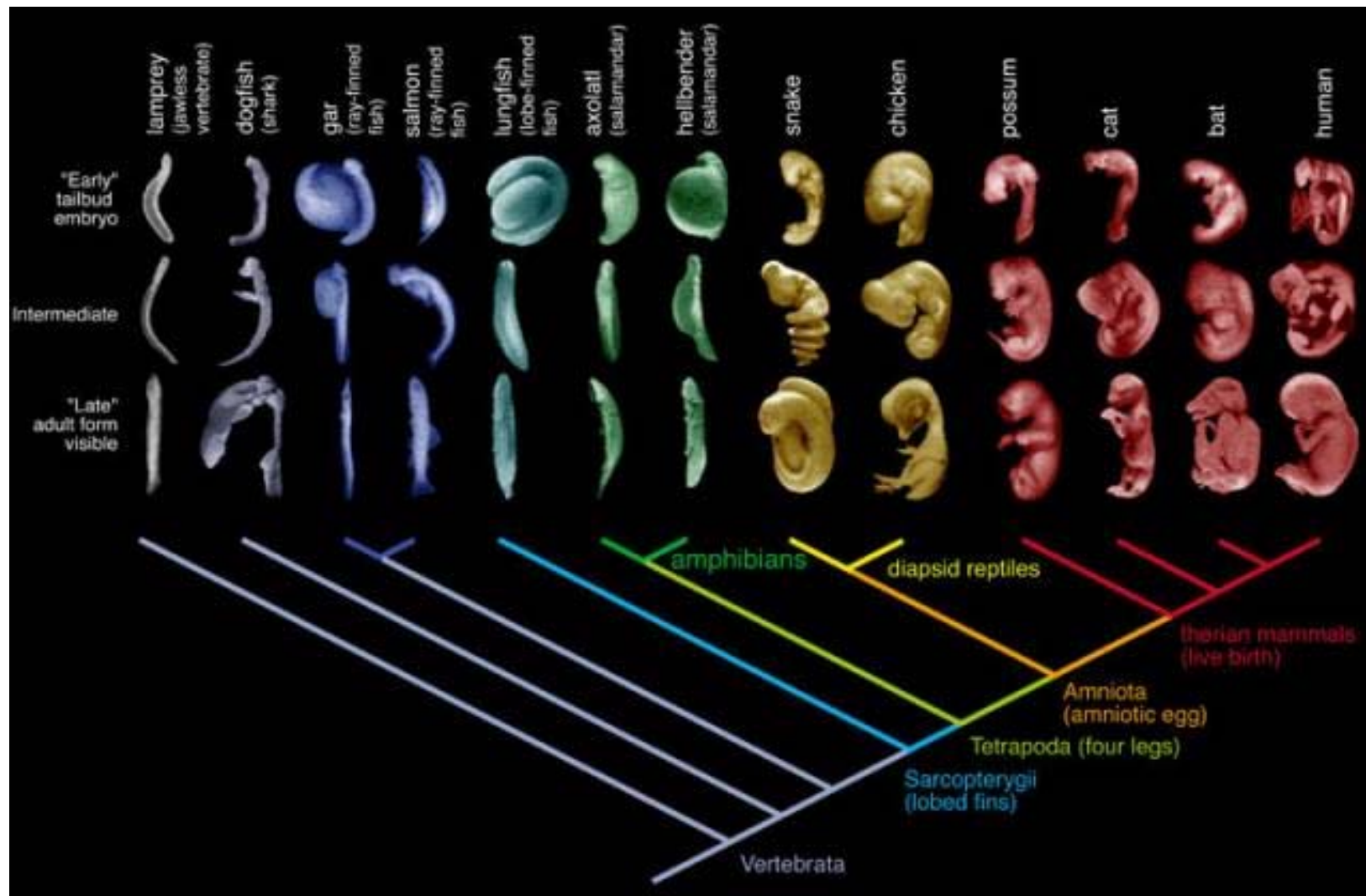
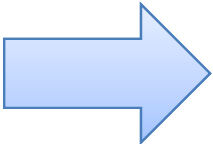


Figure 8. Developmental sequences of various vertebrates shown in phylogenetic context. Note the shared similarities of some closely related taxa, particularly the amniotes (modified from Richardson et al. 1998.) (Figure Source: ncseprojects.org/image/icons-evolution-figure-8)

Computationally screening for function divergent genes involved in early development regulation



Sequences from 14 species



BLAST

Similarity Matrix

Identify similar sequences among inter-species as well as intra-species

Insert an in-video survey here.

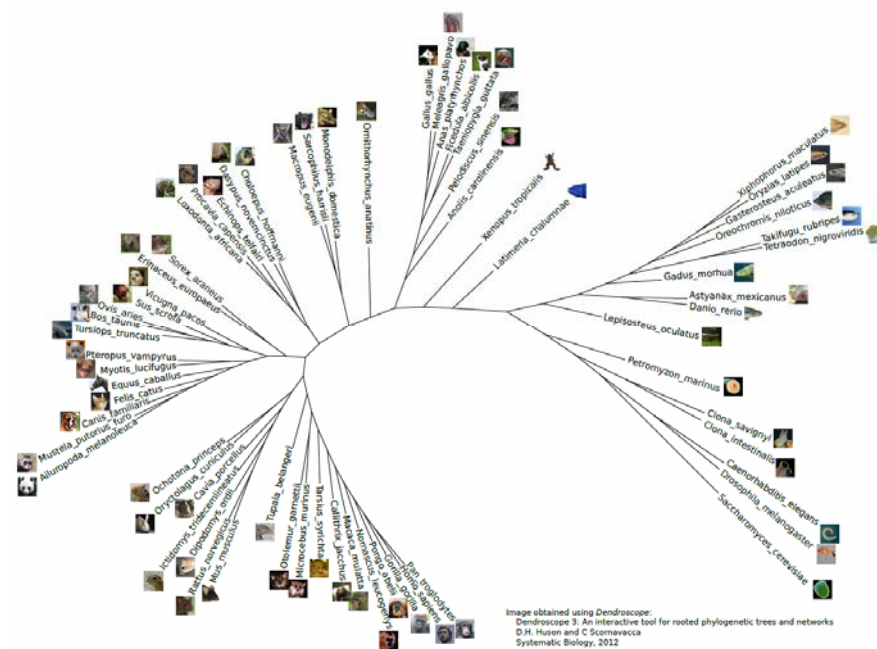
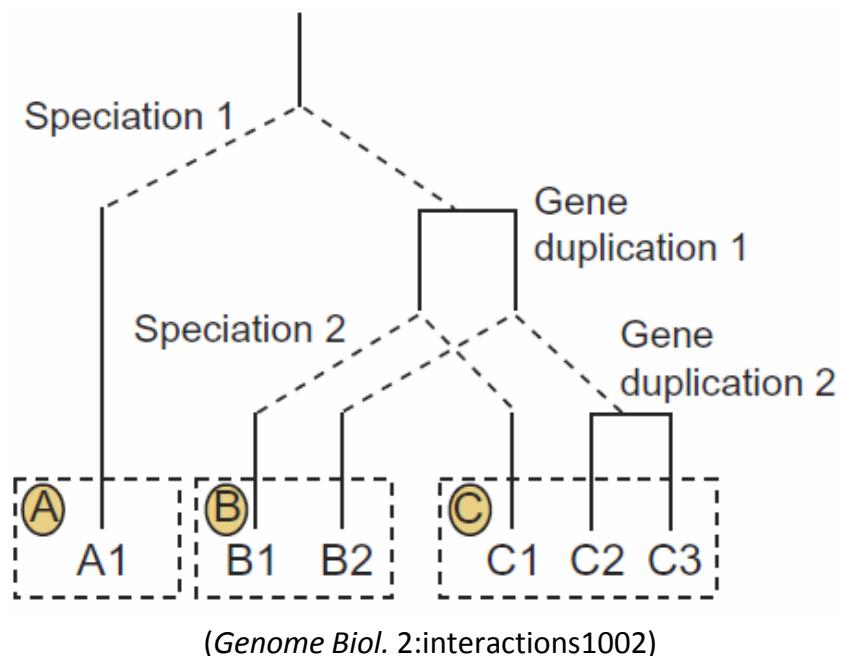
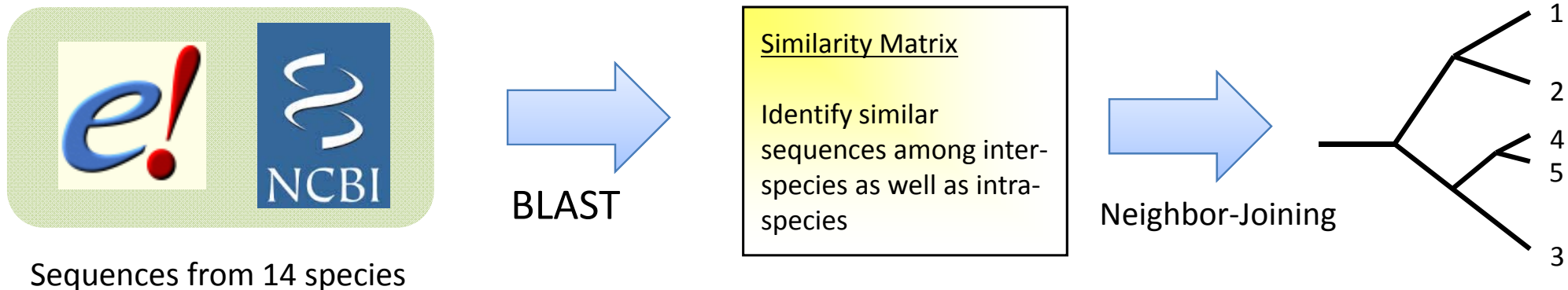
Question: Which Scoring Matrix would you like to use here?

- A) PAM1
- B) BLOSUM80
- C) BLOSUM62
- D) PAM2

C is the right answer, but B is also okay.

Either the answer is right or wrong, just go on.

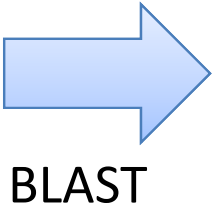
Computationally screening for function divergent genes involved in early development regulation



Computationally screening for function divergent genes involved in early development regulation

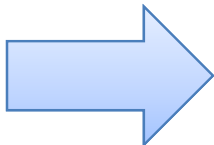


Sequences from 14 species

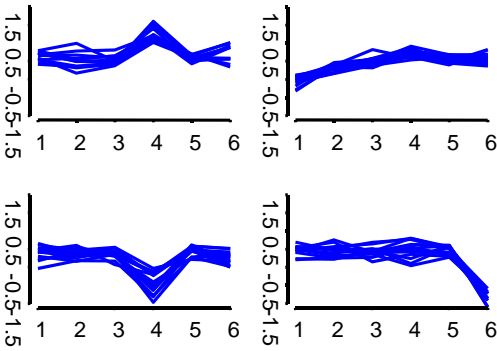
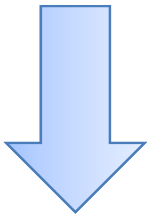
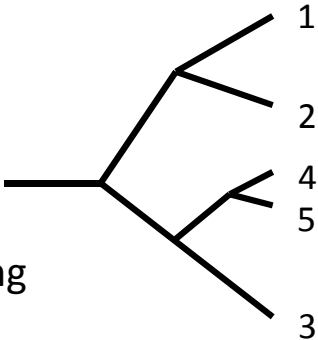


Similarity Matrix

Identify similar sequences among inter-species as well as intra-species



Neighbor-Joining



Expression Profiles

Insert an in-video survey here.

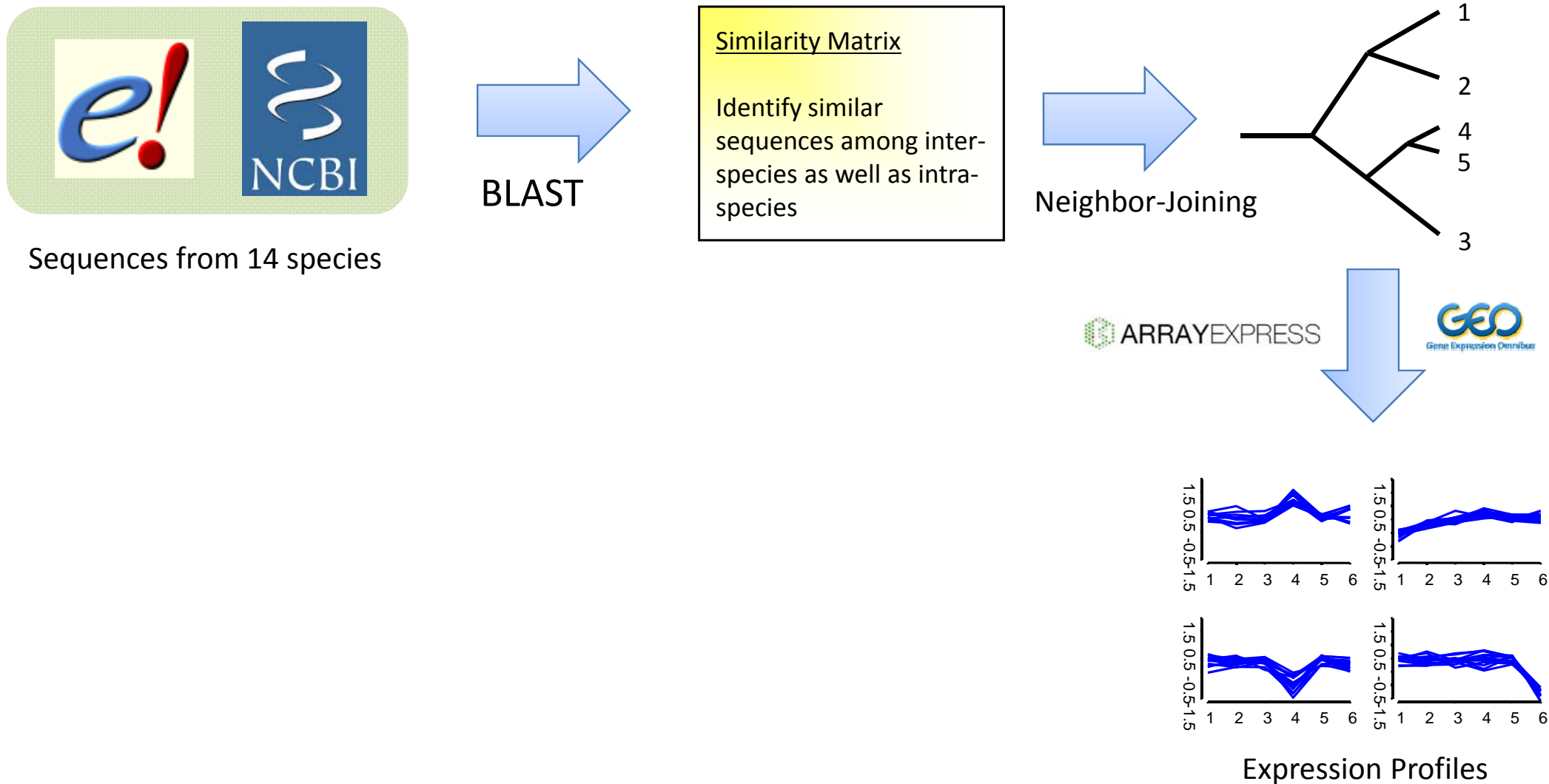
Question: Which database(s) you would NOT use here?

- A) PDB
- B) NCBI GEO
- C) NCBI SRA
- D) EBI ArrayExpress

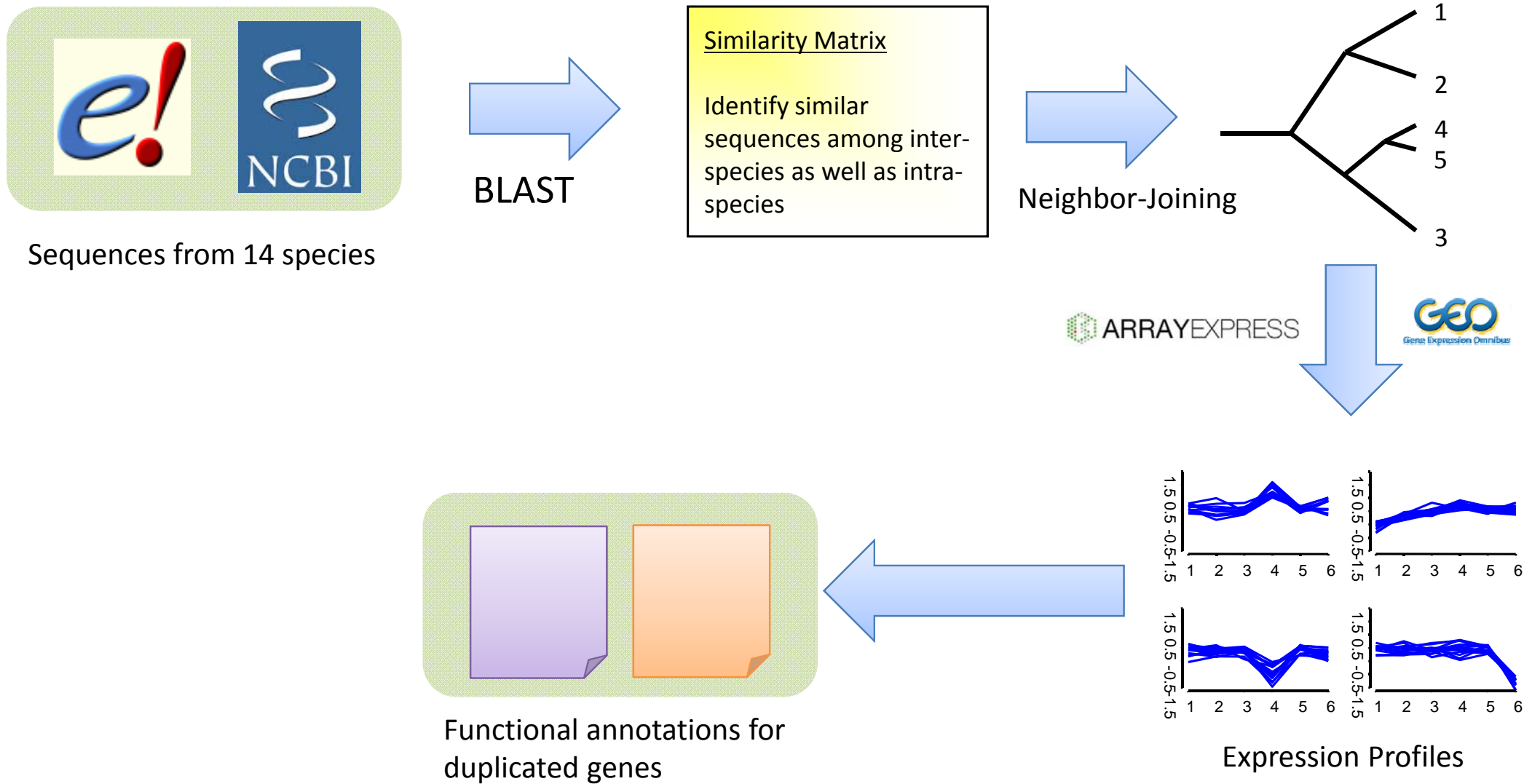
A is the right answer.

Either the answer is right or wrong, just go on.

Computationally screening for function divergent genes involved in early development regulation



Computationally screening for function divergent genes involved in early development regulation



Insert an in-video survey here.

Question: Which database(s) would you like use here?

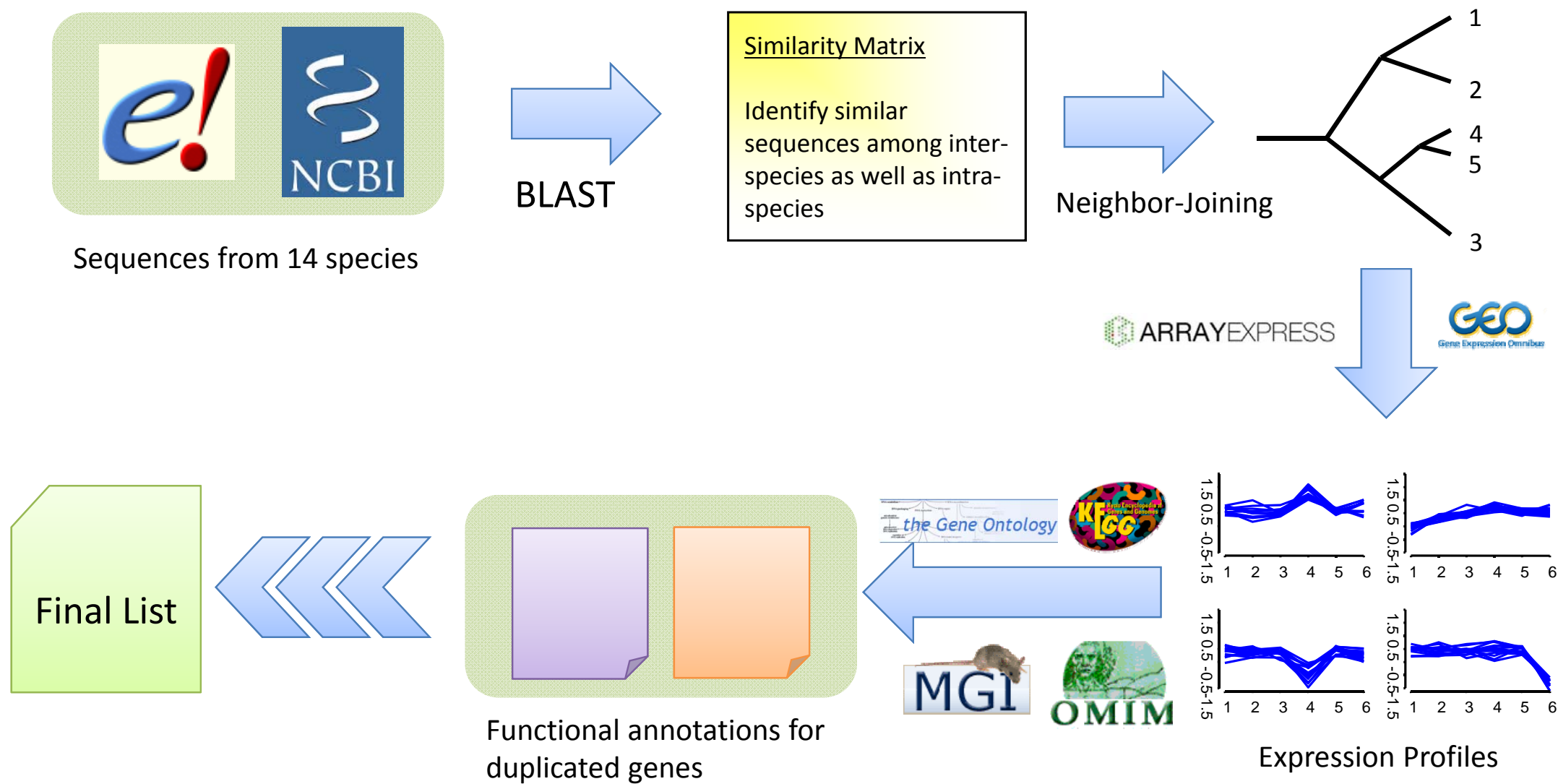
A) KEGG

B) Gene Ontology Annotation

Both A and B are correct

Either the answer is right or wrong, just go on.

Computationally screening for function divergent genes involved in early development regulation



Computationally screening for function divergenced genes involved in early development regulation

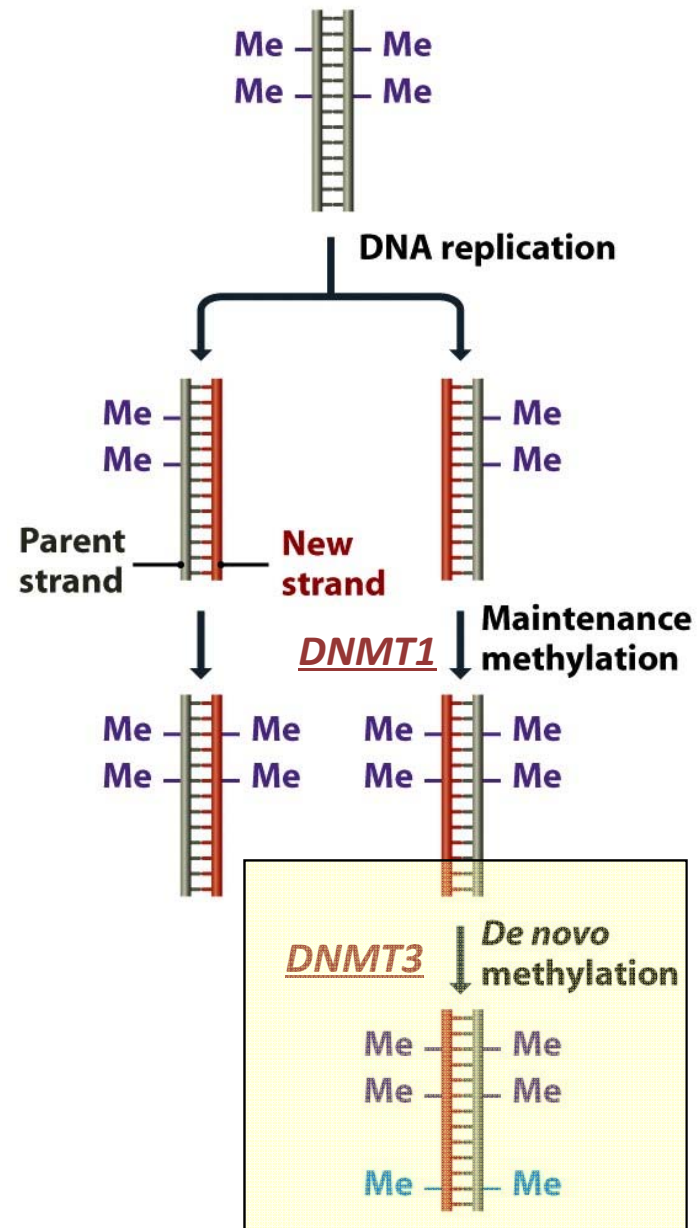
Computational Genomic Analysis and Bioinformatics through MAPK:

1. Sequence databases were constructed directly from the Ensembl website;
2. Each peptide sequence in the database was used to search the database using BLAST package.
3. Phylogenetic trees were constructed and paralogous pairs are identified from the resulting alignments based on a minimal amino acid identity (e.g. 50% and 70%) and an overlap of ≥ 35 amino acids in the region of local alignment.
4. Coding regions of pairs that meet these criteria will be aligned with the corresponding region and inspected for putative function divergence hallmarks.
5. Local warehouse were searched for further indicators derived from high-throughput data (esp. genetic, genomic, transcriptomic, proteomic and pathway data).

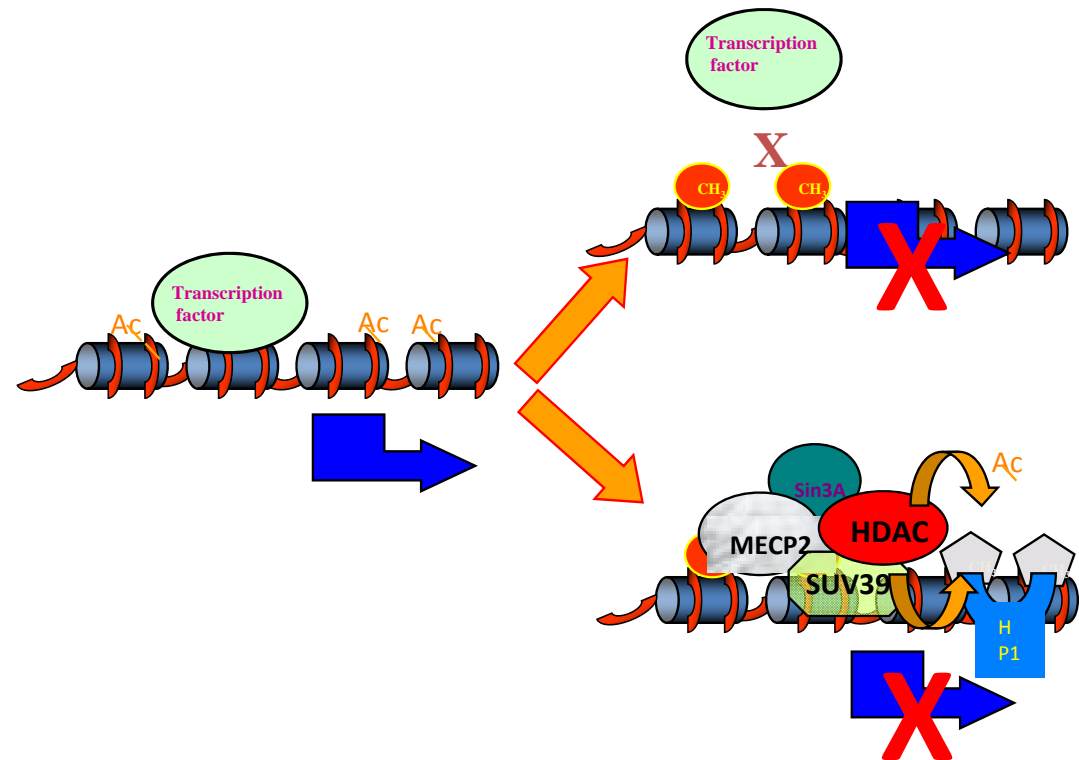
7 out of 50000+ new paralogous pairs showed clear functional divergence features involved in early development regulation.

XXX: should be the only one

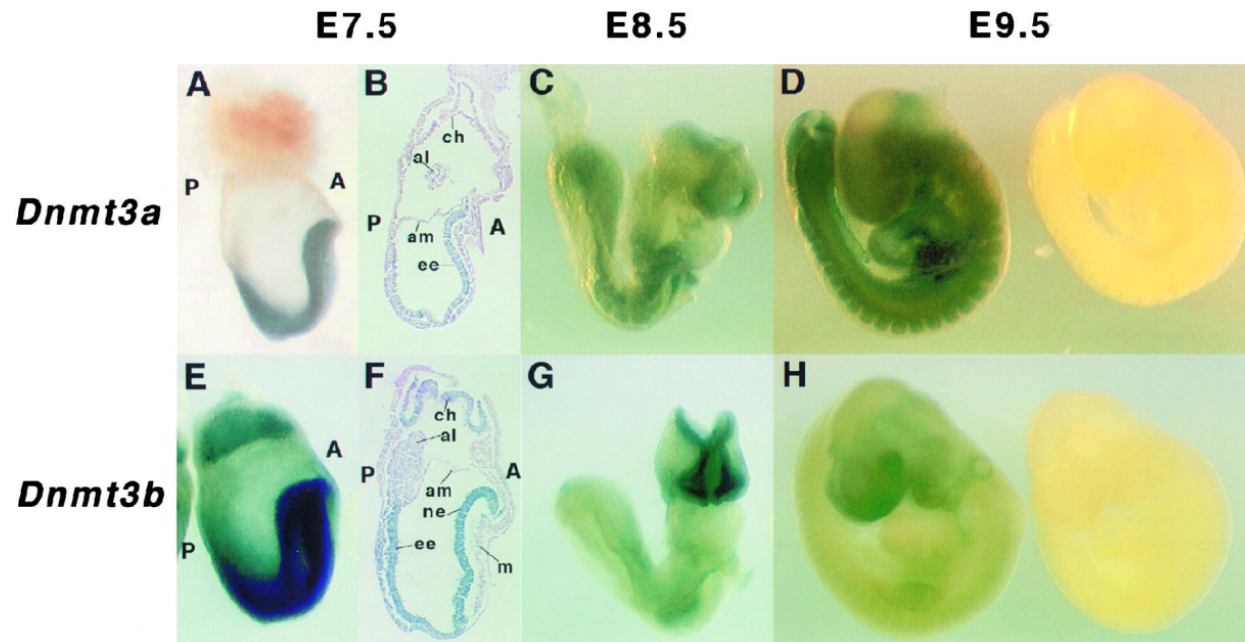
extract_blast_result3.pl 4% (69,0) (CPerl)



DNA methylation silences gene expression by two mechanisms



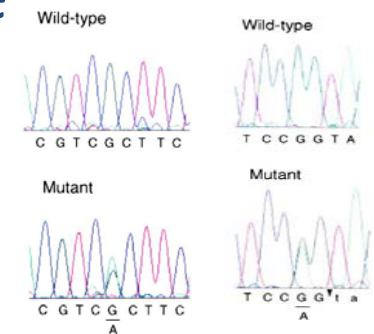
(Modified from Moshe Szyf's slide)



DNMT3-induced methylation is critical for early mouse embryo development



...and also for human ICF Syndrome





Pei Lab at Shanghai

生物信息学：导论与方法

Bioinformatics: Introduction and Methods



<https://www.coursera.org/course/pkubioinfo>