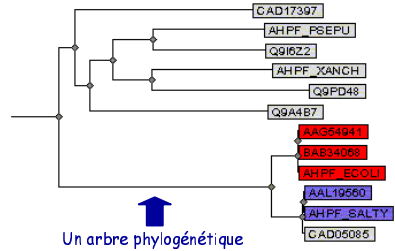


Recueils d'arbres phylogénétiques

HOVERGEN :
9800 arbres de gènes issus de vertébrés.

HOBACGEN :
11500 arbres de gènes issus de procaryotes.



Un arbre phylogénétique

HOVERGEN et HOBACGEN sont des bases de données de familles de gènes homologues développées au Pôle Bio-Informatique Lyonnais. Un arbre phylogénétique est calculé pour chaque famille.

Problématique

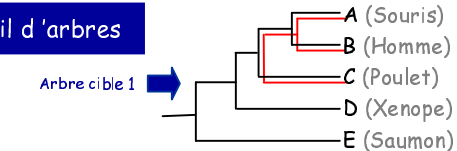
Question d'un biologiste:
- « Je souhaite retrouver 3 gènes homologues: un chez la souris, un chez le poulet, et un chez l'homme. Je désire que le gène de la souris et celui de l'homme soient les deux plus proches. »

Recherche d'information dans un recueil d'arbres

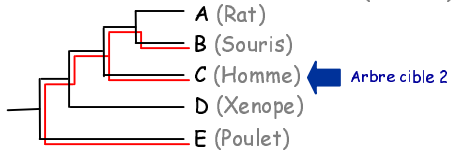
Première solution

Recherche « à la main », avec présélection par les espèces recherchées:
Long et sensible à l'erreur humaine.

Arbre cible 1



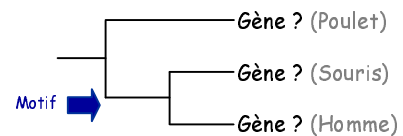
Arbre cible 2



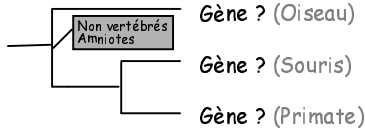
Seconde solution

Recherche de toutes les occurrences d'un sous arbre, appelé **motif d'arbre**, dans un recueil donné. Cela revient à résoudre le problème de la recherche de motifs non ordonnés dans les arbres, où les **arbres cibles** sont les arbres phylogénétiques.

Motif

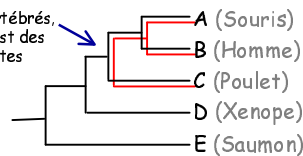


Utilisation de différents niveaux taxonomiques



Améliorations et enrichissement des motifs d'arbres

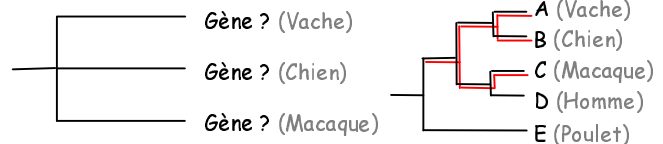
Pas de vertébrés, si ce n'est des amniotes



Possibilité d'utiliser **n'importe quel niveau taxonomique** (mammifères, bactéries, rongeurs, hominidés). Possibilité de **contraindre des parties de l'arbre** sur les taxons représentés.

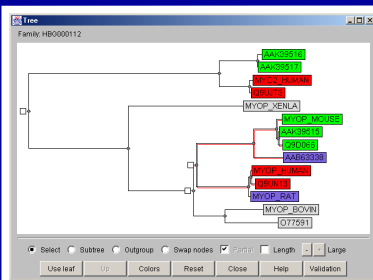
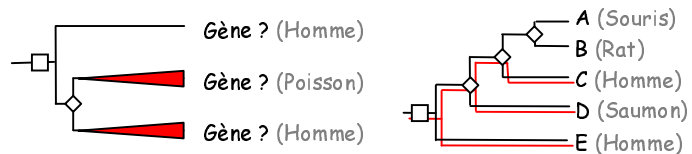
Topologies mal résolues

Possibilité de rechercher des **motifs non binaires**. N'importe quelle topologie binaire correspondante sur l'arbre cible sera acceptée lors de la recherche.



Nœuds de duplication / spéciation

Nœuds de duplication placés dans HOVERGEN, par une méthode de réconciliation d'arbres. Possibilité d'utiliser cet information au cours de la recherche de motifs.



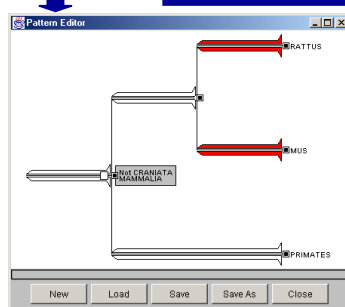
Visualisation d'arbre phylogénétique

Outils d'édition



Fenêtre d'édition

Intégration dans Famtech



Fenêtre principale

Nœud de duplication

Nœud de spéciation

Pas de nœuds de duplication sur le chemin correspondant

Families

Total number of families in Hovergen: 18167

Family ID	Count	Family Name
HB000000	1922	UNCLASSIFIED PARTIAL SEQUENCES
HB000001	4	100 RBA PROTEIN; 130011000RBA PROTEIN; 25.0 KDA PROTEIN; KIAA
HB000002	59	14-3-3 FAMILY
HB000003	3	PROTEIN 1-4
HB000004	3	15E1-1 FAMILY; 181001500RBA PROTEIN; 8.8 KDA PROTEIN; PROTEIN
HB000005	2	GATC FAMILY; 501003018RBA PROTEIN; PROTEIN 15E1.2
HB000006	5	2-19 PROTEIN; ONCOPROTEIN INHIBIT PROTEIN 1; PROTEIN G3705
HB000007	3	CETASTIN RELATED PROTEIN 1; CETASTIN RELATED PROTEIN 2
HB000008	6	ADULT BRAIN PROTEIN 239; PETAL BRAIN PROTEIN 239
HB000009	18	PHOSPHATASE 2A REGULATORY SUBUNIT B FAMILY
HB000010	4	PHOSPHATASE 2A REGULATORY SUBUNIT B FAMILY
HB000011	10	PHOSPHATASE 2A REGULATORY SUBUNIT A FAMILY
HB000012	21	PHOSPHATASE 2A REGULATORY SUBUNIT B FAMILY
HB000013	3	SERINE/TRYPHANYLASE PROTEIN PHOSPHATASE 2A; 72/130 KDA REGULATORY
HB000014	33	BETA-200 FAMILY
HB000015	13	31.9 KDA PROTEIN; C8H FL31219 PFS; CLONE NT3P4001849; HEARL
HB000016	2	SH3 DOMAIN-LINKING PROTEIN SH3-2
HB000017	1	P52-700 CELL ASSOCIATED PROTEIN SCL; SCL_BHMAN
HB000018	3	061001307RBA PROTEIN; 3-HYDROXYMETHYLMALATE 3:4-DIOXYGENASE
HB000019	5	3-METHYL-ADENINE DNA GLUCOSYLASE; DNA-3-METHYLADENINE GLUCOSYL
HB000020	78	ALDO/KETO REDUCTASE 1 FAMILY; ALDO/KETO REDUCTASE FAMILY
HB000021	4	4-189 LIGAND RECEPTOR; D3692P13.1; TUMOR NECROSIS FACTOR RECEPT