

BIO-INFORMATICA

COURSE 3B VERGELIJKENDE GENOOMANALYSE

WEEK 6: FYLOGENIE

OVERZICHT ONDERZOEKSPLAN 3B

Week	Onderwerp	Activiteiten	M&M
1	Oriëntatie op onderwerp Sequenties verzamelen	Inlezen Stap 1	NCBI database
2	Genen en GC percentage	Stap 2	Python script
3	Codon gebruik	Stap 3	Python script
4	Eiwit karakterisering / aminozuren	Stap 4	Python script
5	Kenmerken oppervlakte proteïnen	Stap 5	Python script
6	Fylogenetisch onderzoek	Stap 6	Bioinf tools
7	Onderzoeksverslag		

DEZE WEEK

Onderwerpen

- Fylogenie
 - Wat is het?
 - Hoe kunnen we fylogenie bekijken/fylogenetische bomen maken?
 - ClustalO tool

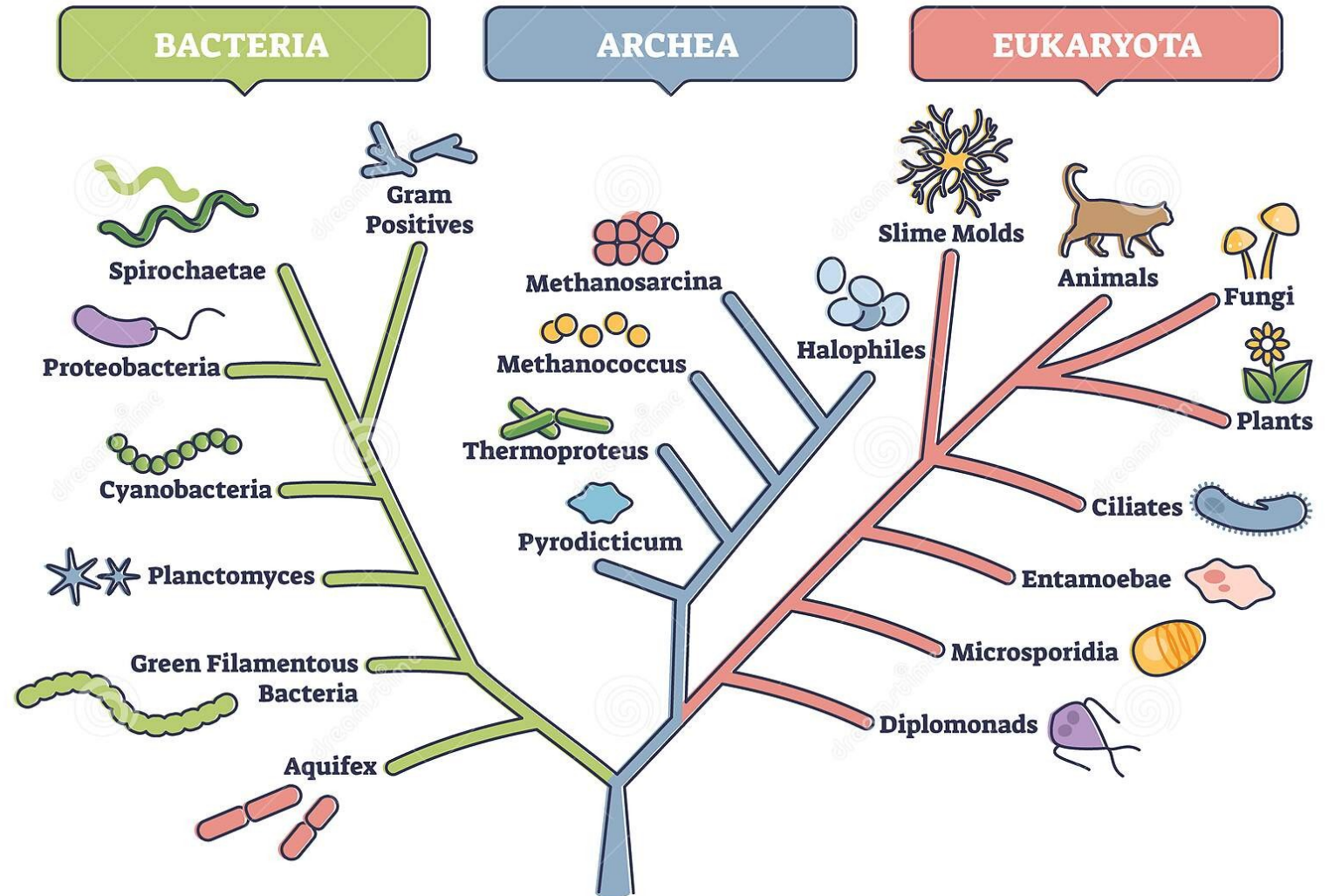
Materialen

- Campbell 22.3 blz 527-532
 - Phylogenetic trees
- Campbell Hfst 22.4
- Campbell Hfst 26 Evolution of viruses, blz 618 -620
- Campbell Hfst 27.2 blz 632 Evolution of prokaryotes
- Pevsner Hfst 15: 705-710
- Pevsner Hfst 16: 762-764, 766-768

PHYLOGENETIC TREE

FYLOGENIE

- De studie naar de evolutionaire geschiedenis en relatie tussen individuen of groepen organismen.
- Aan de hand van genetische en/of anatomische eigenschappen.

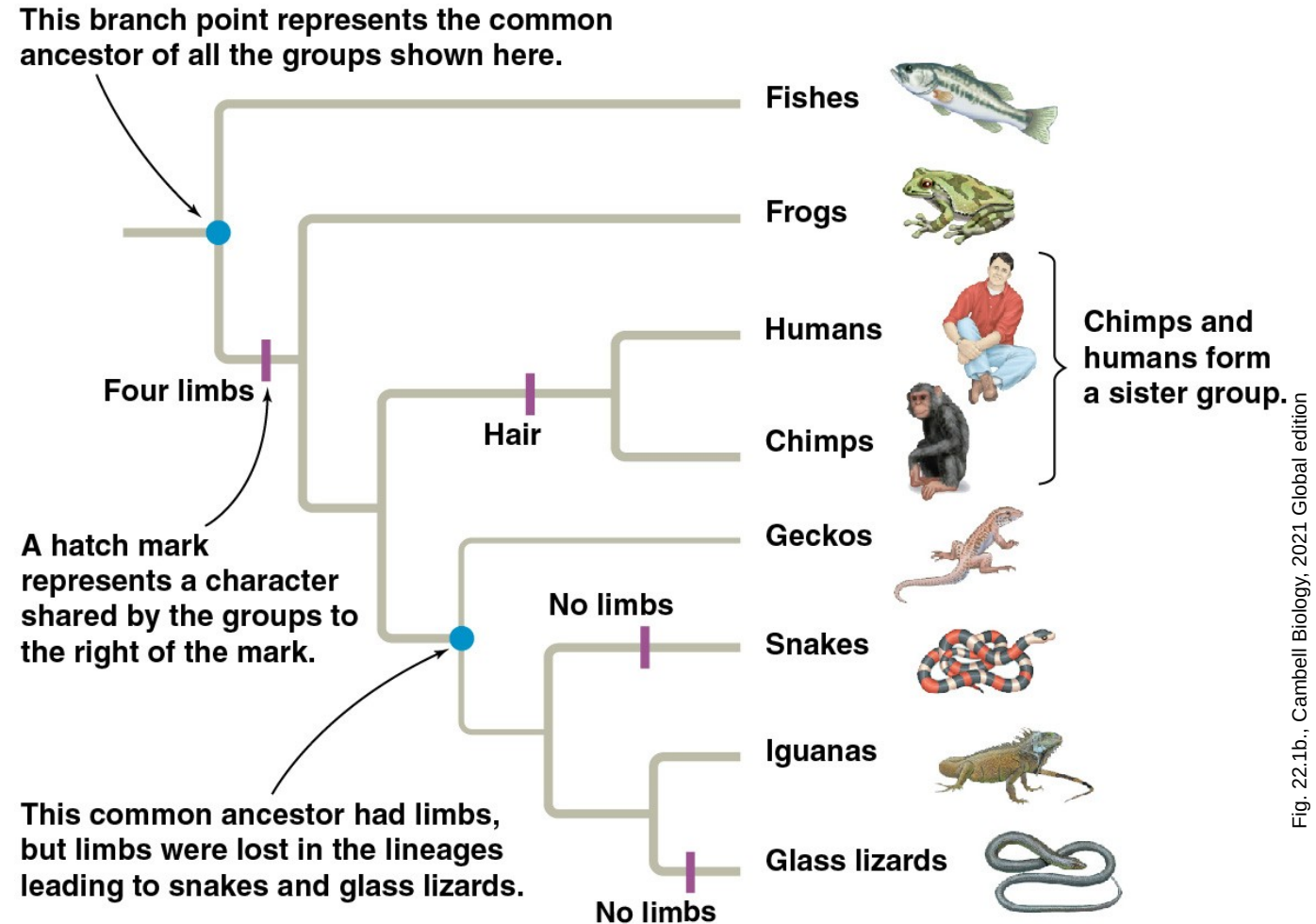


FYLOGENIE

- Aan de hand van anatomische eigenschappen.

How do biologists distinguish and categorize the millions of species on Earth?

Traits shared due to **common ancestry** are used to classify organisms into groups that reflect their **evolutionary history**:



FYLOGENIE

- Aan de hand van genetische eigenschappen (gen sequenties).
Verwantschap kan worden bepaald aan de hand van een multiple sequence alignment. Sequentie similariteit duidt op homologie

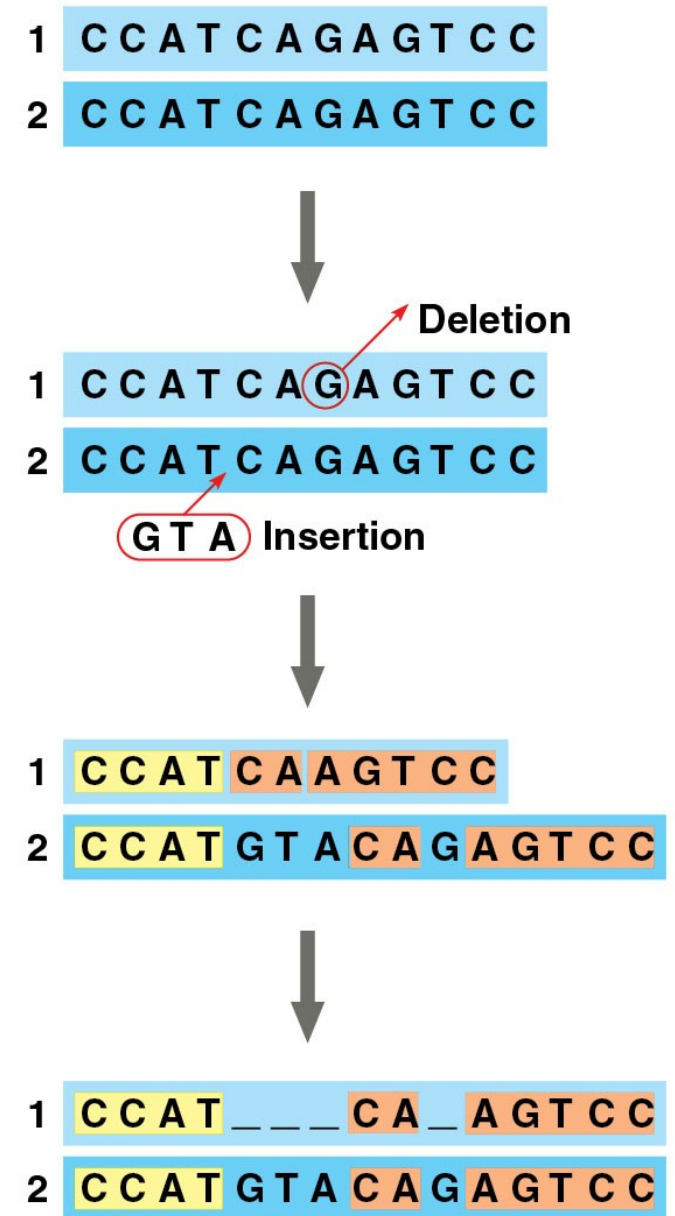


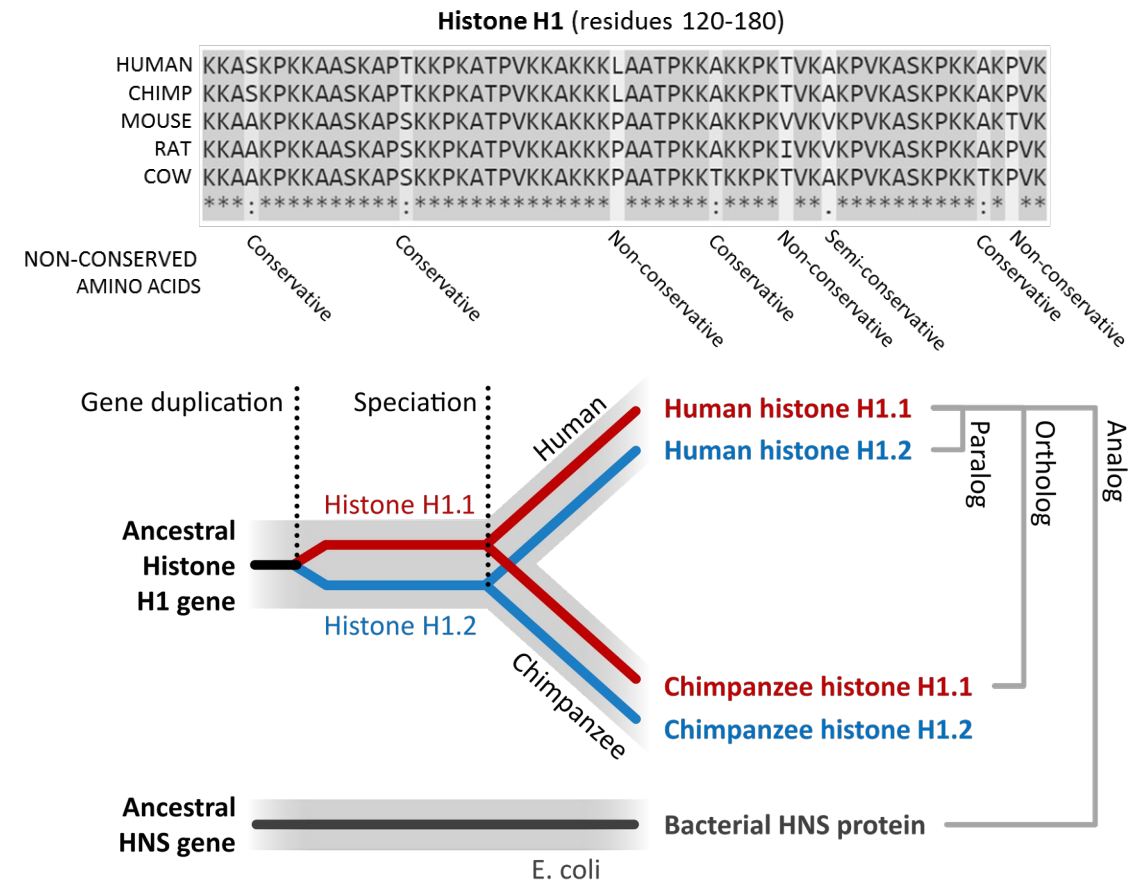
Fig. 22.8., Cambell Biology, 2021 Global edition

HOMOLOGE SEQUENTIES (HERHALING)

- Sequenties kunnen op verschillende manieren een gedeeld voorouderlijke sequentie hebben:

Door een speciation (soortvorming): een gen in een voorouderlijk organisme bestaat nu in twee aparte organismen. We noemen dit **orthologen**.

Door een duplicatie in: een gen in een (voorouderlijk) organisme is gedupliceerd en beide kopieën kunnen zich bevinden in nieuwe organismen. We noemen dit **paralogen**.



FYLOGENETISCHE BOOM

- Multiple sequence alignments kunnen een fylogenetische boom creëren.
 1. Een set aan (homologe) sequenties verkrijgen.
 2. Afstanden bepalen tussen de sequenties d.m.v. pairwise alignments.
 3. Vormen van een similariteit boom met deze afstanden.
 4. Combineren van de alignments uit stap 1 in de volgorde van stap 2.

Tool die dit voor je kunnen doen zijn bijvoorbeeld ClustalO, PhyloTree, MEGA...

[Creating a Phylogenetic Tree - YouTube](#)

```
1 CCAT___CA_AGTCC
2 CCATGTA CAGAGTCC
```

Fig. 22.8, Cambell Biology, 2021 Global edition

FYLOGENETISCHE BOOM BOUWEN STAP 1

- Afstanden bepalen tussen sequenties geeft een getal die we in een matrix kunnen zetten.

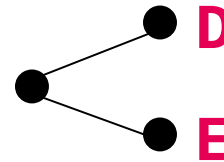
	A	B	C	D	E
A	0	6	9	11	9
B	6	0	7	9	7
C	9	7	0	8	6
D	11	9	8	0	4
E	9	7	6	4	0

A ATGCGATACCGACGAA
B ATCCGGAACAGAGAAA
...
...

FYLOGENETISCHE BOOM BOUWEN STAP 2

- Boom maken met deze afstanden

	A	B	C	D	E
A	0	6	9	11	9
B	6	0	7	9	7
C	9	7	0	8	6
D	11	9	8	0	4
E	9	7	6	4	0



Kleinste afstand eerst, die zit in de boom het dichtstbij

FYLOGENETISCHE BOOM BOUWEN STAP 2

- Boom maken met deze afstanden

	A	B	C	D	E
A	0	6	9	11	9
B	6	0	7	9	7
C	9	7	0	8	6
D	11	9	8	0	4
E	9	7	6	4	0

	A	B	C	DE
A	0	6	9	10
B	6	0	7	8
C	9	7	0	7
DE	10	8	7	0

Gemiddelde nemen maakt een nieuwe afstand

FYLOGENETISCHE BOOM BOUWEN STAP 2

- Boom maken met deze afstanden

	A	B	C	D	E
A	0	6	9	11	9
B	6	0	7	9	7
C	9	7	0	8	6
D	11	9	8	0	4
E	9	7	6	4	0

	A	B	C	DE
A	0	6	9	10
B	6	0	7	8
C	9	7	0	7
DE	10	8	7	0

Gemiddelde nemen maakt een nieuwe afstand

FYLOGENETISCHE BOOM BOUWEN STAP 2

- Boom maken met deze afstanden

	A	B	C	D	E
A	0	6	9	11	9
B	6	0	7	9	7
C	9	7	0	8	6
D	11	9	8	0	4
E	9	7	6	4	0

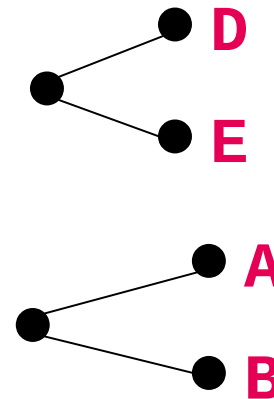
	A	B	C	DE
A	0	6	9	10
B	6	0	7	8
C	9	7	0	7
DE	10	8	7	0

Gemiddelde nemen maakt een nieuwe afstand

FYLOGENETISCHE BOOM BOUWEN STAP 2

- Boom maken met deze afstanden

	A	B	C	DE
A	0	6	9	10
B	6	0	7	8
C	9	7	0	7
DE	10	8	7	0



Nieuwe relatie met minste afstand

FYLOGENETISCHE BOOM BOUWEN STAP 2

- Boom maken met deze afstanden

	A	B	C	DE
A	0	6	9	10
B	6	0	7	8
C	9	7	0	7
DE	10	8	7	0

	AB	C	DE
AB	0	8	9
C	8	0	7
DE	9	7	0

Gemiddelde nemen maakt een nieuwe afstand

FYLOGENETISCHE BOOM BOUWEN STAP 2

- Boom maken met deze afstanden

	A	B	C	DE
A	0	6	9	10
B	6	0	7	8
C	9	7	0	7
DE	10	8	7	0

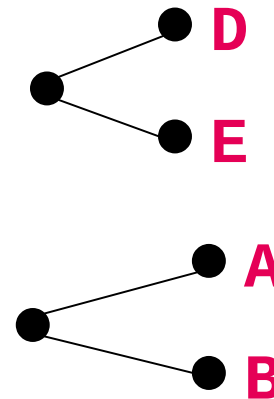
	AB	C	DE
AB	0	8	9
C	8	0	7
DE	9	7	0

Gemiddelde nemen maakt een nieuwe afstand

FYLOGENETISCHE BOOM BOUWEN STAP 2

- Boom maken met deze afstanden

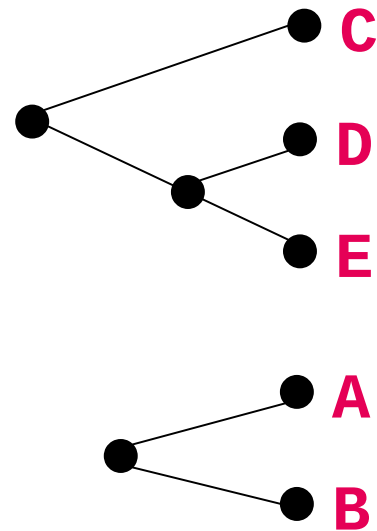
	AB	C	DE
AB	0	8	9
C	8	0	7
DE	9	7	0



FYLOGENETISCHE BOOM BOUWEN STAP 2

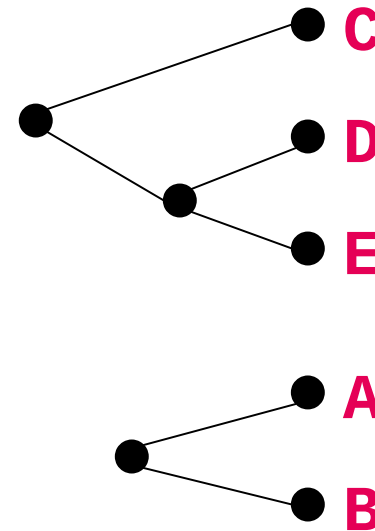
- Boom maken met deze afstanden

	AB	C	DE
AB	0	8	9
C	8	0	7
DE	9	7	0



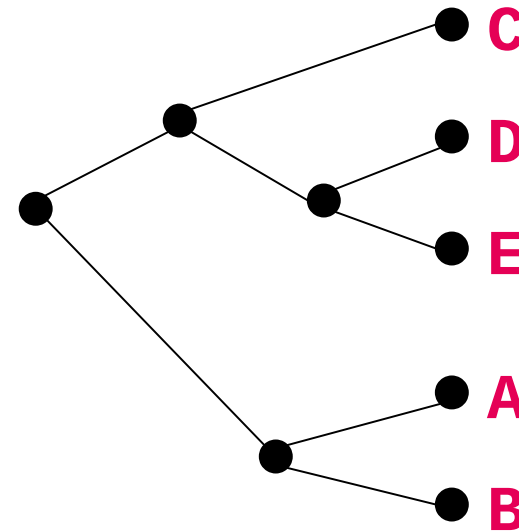
FYLOGENETISCHE BOOM BOUWEN STAP 2

	AB	CDE
AB	0	8.5
CDE	8.5	0



FYLOGENETISCHE BOOM BOUWEN STAP 2

	AB	CDE
AB	0	8.5
CDE	8.5	0



FYLOGENETISCHE BOMEN

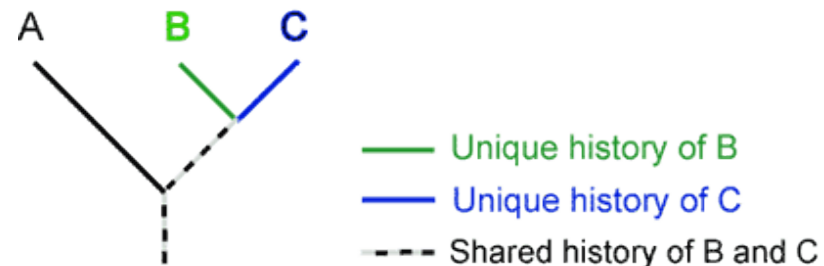
- Vroeger voornamelijk op basis van morfologie/anatomische eigenschappen.
- Nu moleculair (genen, eiwitten, mRNA, 3D structuren van eiwitten, etc.).
- Wat we wel en niet kunnen leren van bomen:

Een tak punt (node) representeert (een vertakking/afsplitsing van) een gemeenschappelijk voorouder → soortvorming.

Een voorouderlijke lijn is een volgorde van organismen die leid tot de afstammelingen.

Zustergroepen delen een gemeenschappelijk voorouder

Zustergroepen hebben deels hun eigen geschiedenis, deels een gedeelde geschiedenis.



FYLOGENETISCHE BOMEN

Parts of a Phylogenetic Tree

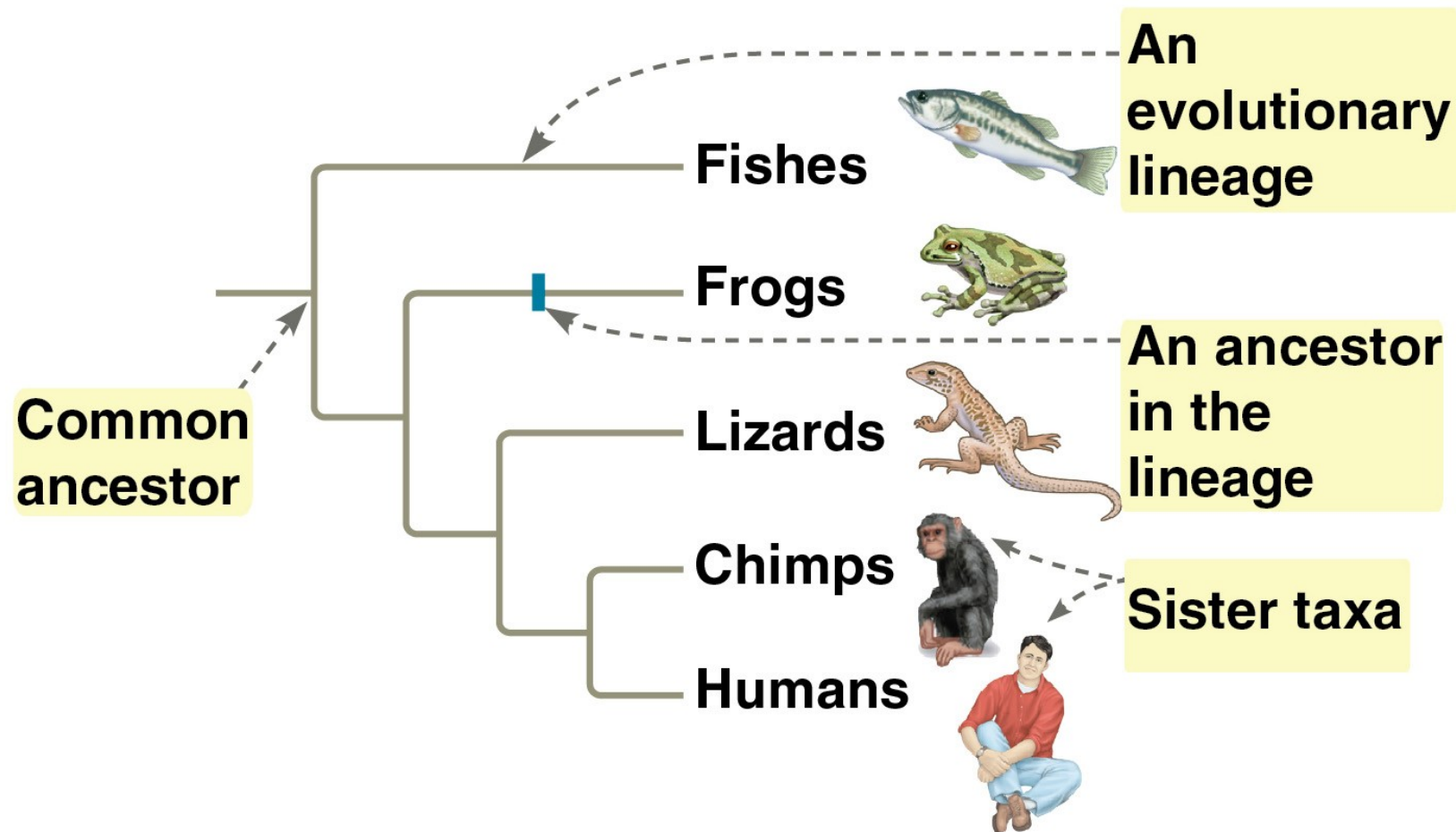
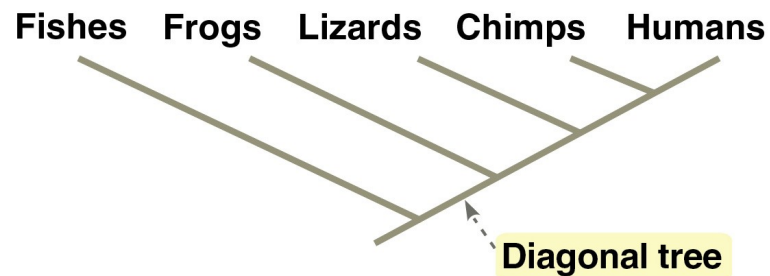
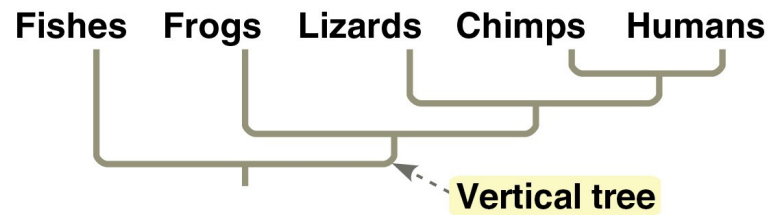


Fig. 22.5., Cambell Biology, 2021 Global edition

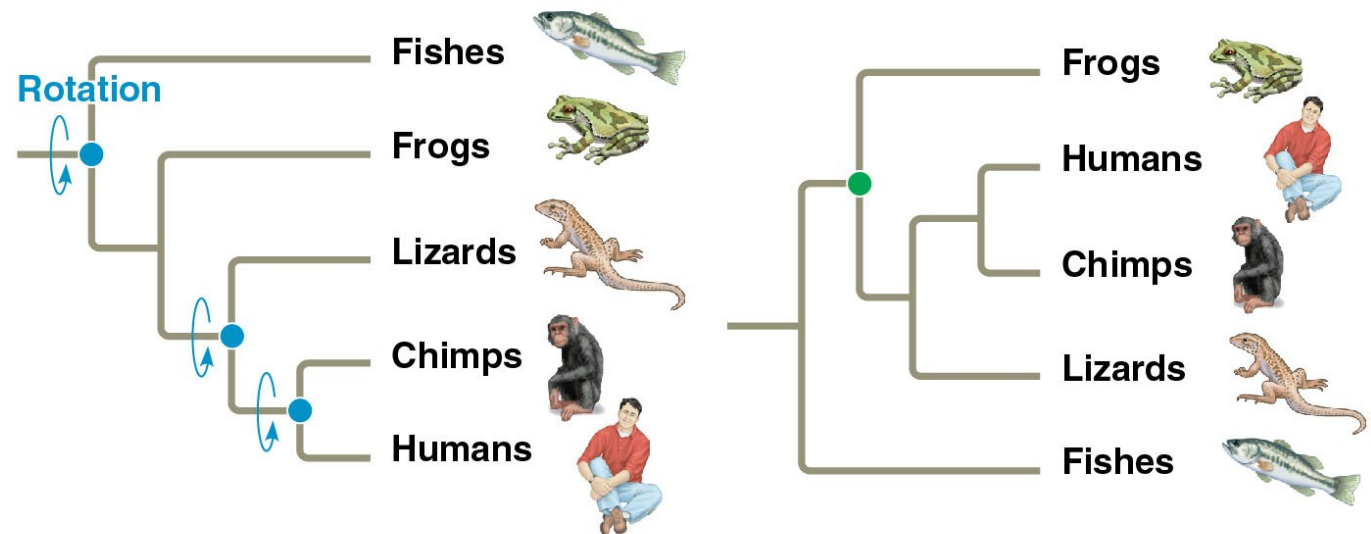
VERSCHILLENDE VORMEN VAN FYLOGENETISCHE BOMEN

- Bomen kunnen diagonaal of verticaal zijn.
- Ze kunnen om hun tak punt (node) draaien.

Alternative Forms of Tree Diagrams



Rotating Around Branch Points



CLADOGRAM VS. FYLOGRAM

Cladogram

- Laat alleen een relatie zien tussen verschillende organismen in relatie tot een gemeenschappelijke voorouder.
- Hypothese van een relatie tussen organismen.
- Takken zijn allemaal dezelfde lengte (geen informatie dus).
- Kan geen evolutionaire tijd weergeven wanneer organismen uit elkaar zijn geëvolueerd.

Fylogram (fylogenetische boom)

- Laat een relatie zien tussen verschillende organismen in relatie tot de evolutionaire tijd en de hoeveelheid verschillen die zijn ontstaan in die tijd.
- Een representatie van evolutionaire historie tussen organismen.
- De takken hebben een andere lengte die een evolutionaire afstand weergeeft.
- Geeft een indicatie van evolutionaire tijd van wanneer organismen uit elkaar zijn geëvolueerd.

VERSCHILLENDE VORMEN VAN FYLOGENETISCHE BOMEN

- De tak lengtes kunnen de genetische verandering weergeven.
Kortere takken betekend minder genetische verandering.
- Soms wordt tijd gebruikt m.b.v. fossielen.
- We kunnen nooit 100% weten wat de juiste (vorm van een) boom is. We kunnen gebruik maken van:

Maximum parsimony is een methode die aanneemt dat de meest waarschijnlijke boom de boom is die het minst evolutionaire gebeurtenissen heeft.

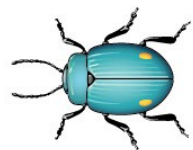
- Wanneer we DNA/eiwit sequenties gebruiken, betekend dit de minste base veranderingen.

Maximum likelihood identificeert welke boom het meest waarschijnlijke een bepaalde set van DNA/eiwit sequenties produceert, d.m.v. regels/kennis over hoe DNA/eiwitten over tijd veranderen. Bijvoorbeeld, alle nucleotide substituties hebben een gelijke kans voor te komen.

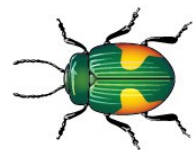
VERSCHILLENDE VORMEN VAN FYLOGENETISCHE BOMEN

- Veel tools gemaakt om bomen te vinden die het meest parsimony (en likelihood) hebben.

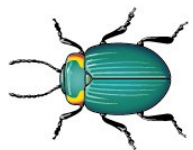
Technique



Species I



Species II



Species III

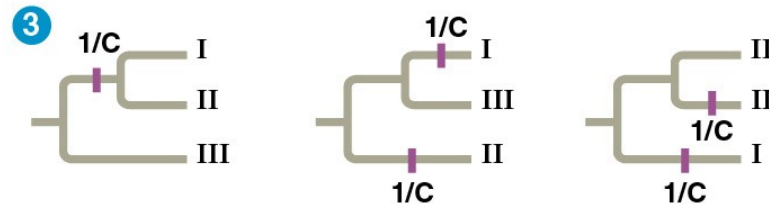
1 Three phylogenetic hypotheses:



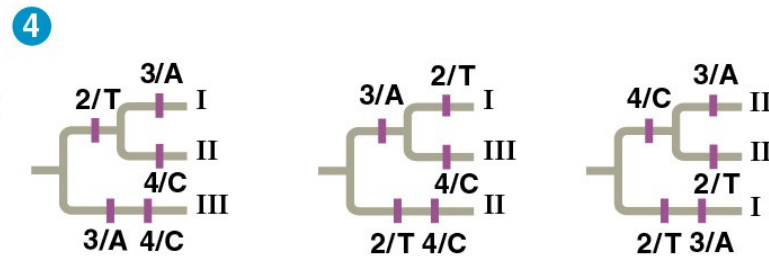
2

	Site			
	1	2	3	4
Species I	C	T	A	T
Species II	C	T	T	C
Species III	A	G	A	C
Ancestral sequence	A	G	T	T

3



4



Results

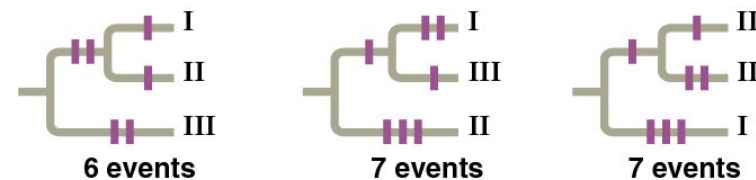


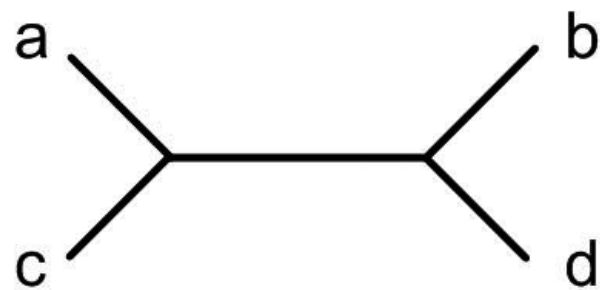
Fig. 22.15., Cambell Biology, 2021 Global edition

FYLOGENETISCHE BOMEN “ROOTEN”

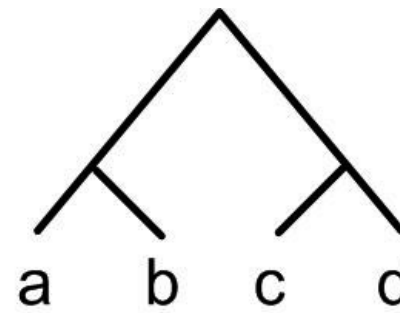
- Om een juist idee te krijgen van hoe organismen (of genen) gerelateerd zijn aan elkaar en in welke richting evolutie gaat, moeten we de juiste “root” hebben.

Een root is het oudste punt in een boom die de theoretische voorouder van alle organismen in die boom weergeeft.

- Hoe krijgen we een root in een boom, of in andere woorden: hoe kunnen we een idee krijgen van de voorouderlijke sequentie?



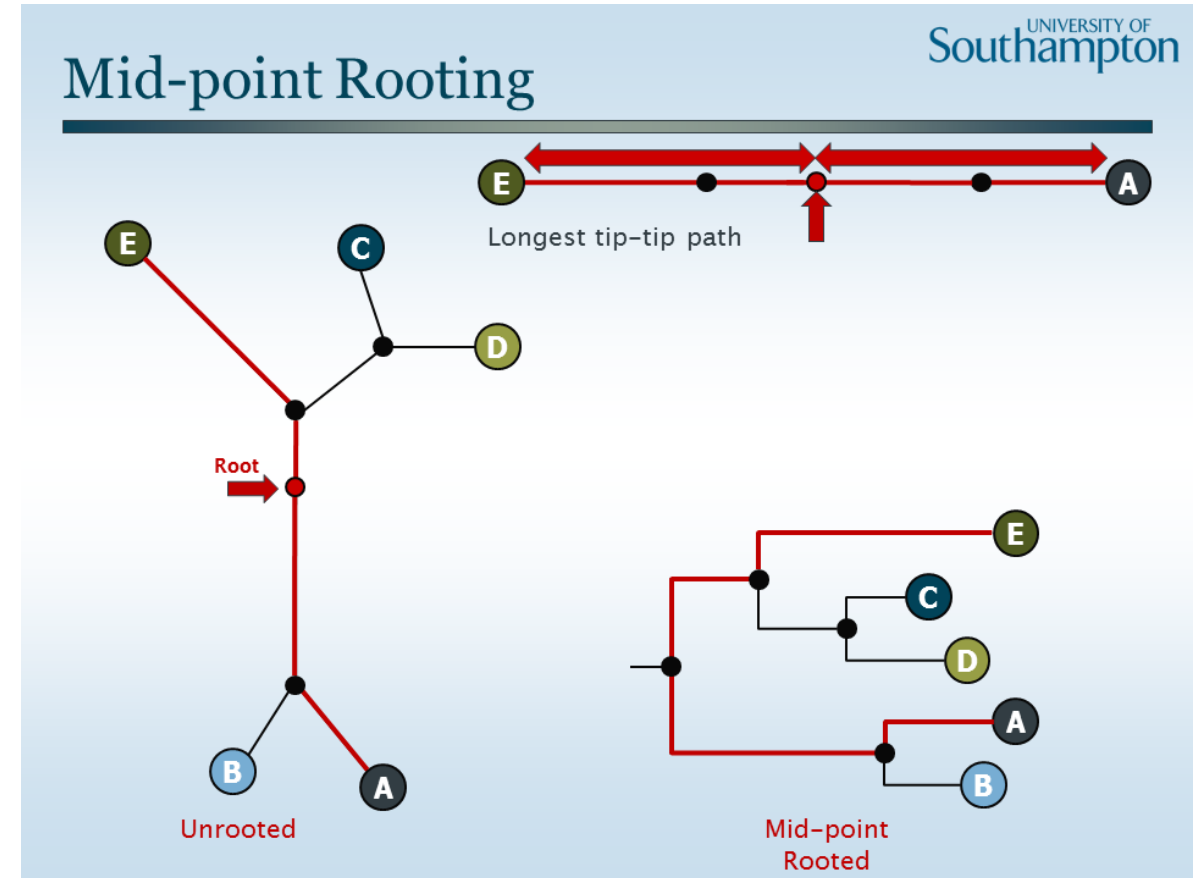
(a)



(b)

FYLOGENETISCHE BOMEN “ROOTEN”

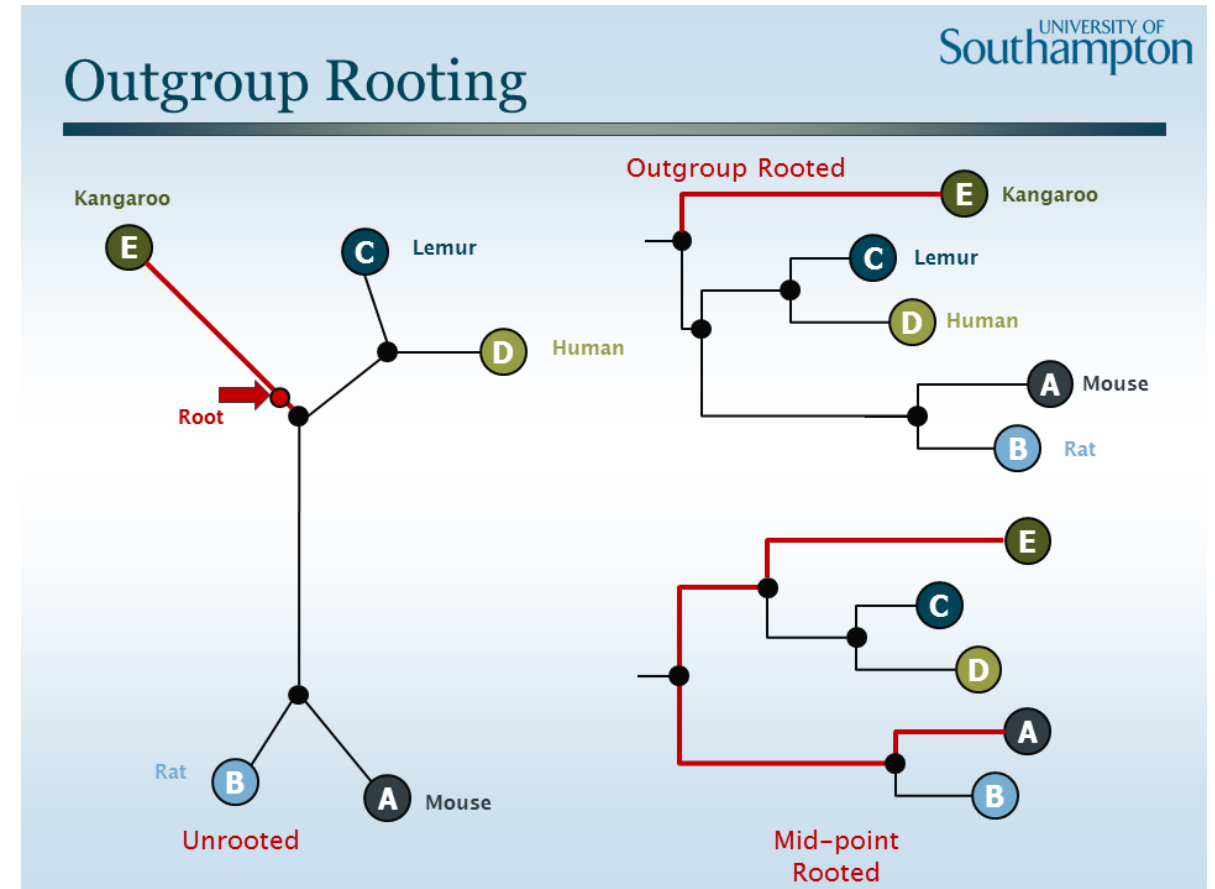
- **Midpoint rooting:** *Langste* lengte vinden (A-E) in de boom van topje tot topje. De root wordt dan in het midden van die tak geplaatst.
- Dit werkt als de evolutie snelheid tussen de sequenties redelijk gelijk is of als de organismen nauw verwant zijn.



FYLOGENETISCHE BOMEN “ROOTEN”

- **Outgroup rooting:** gebruikt een sequentie van een organisme waarvan we weten dat die ver van de andere organismen in de boom staan (ver verwant) → outgroup. Bijvoorbeeld: een vis kan een outgroup zijn van zoogdieren (zoals mens en aap). Bijvoorbeeld: een buideldier (kangoeroe) kan een outgroup zijn van de placentadieren (mens, muis, ...).

[The Cabbages of Doom: How to root a phylogenetic tree](#)



EVOLUTIE SNELHEDEN EN RELATIES

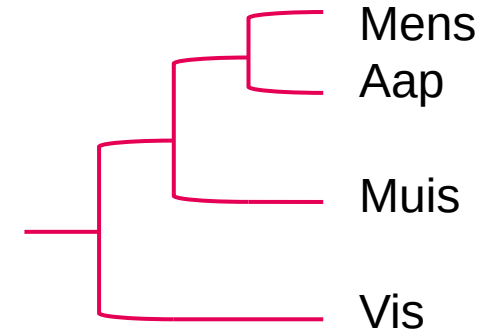
- We kunnen genen gebruiken om te kijken wat de relatie is tussen organismen.
Genen kunnen in verschillende snelheden evolueren in (dezelfde lijn) van organismen (nucleotiden muteren sneller).
 - Ribosomaal RNA evolueert langzaam, dus die kunnen we gebruiken om hele oude relaties tussen organismen op te sporen.
 - Mitochondriaal DNA evolueert heel snel, dus kunnen we die (alleen) gebruiken voor nieuwe relaties (nauw verwante) organismen op te sporen.

GEN VS. ORGANISME BOMEN

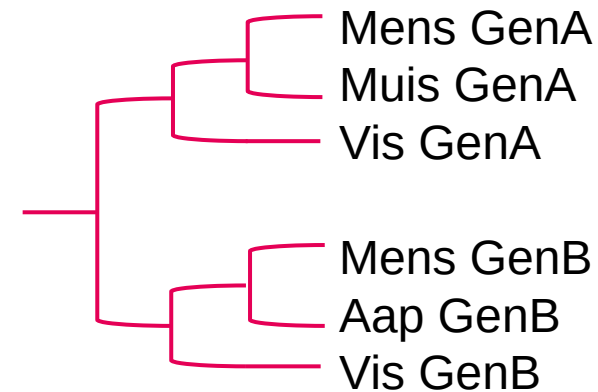
- Ook al gebruiken we genen om organisme bomen te maken, moeten we onthouden dat: **gen bomen en organisme bomen zijn twee verschillende dingen zijn!**
- Een gen boom reflecteert een uniek verhaal over de evolutie van een gen, die gelinkt is aan de evolutie van een organisme, maar vaak anders loopt dat de evolutie van een organisme (Szöllősi and Daubin 2012).
Bijvoorbeeld: in sommige gevallen heeft een bepaald organisme (mens) een bepaald gen, maar een zuster organisme (aap) niet. Is het gen verloren in aap, of een nieuw gen in mens?

GEN VS. ORGANISME BOMEN

- Als we een organisme boom hebben, kunnen we een gen boom “reconcileren”. Het proces van een twee bomen met elkaar vergelijken, die van een gen en die van de organismen met dat gen. De gen boom laat de evolutionaire historie zien van het gen, de organisme boom laat de evolutionaire historie zien van de organismen.
- Stel: gen A en gen B in vier organismen



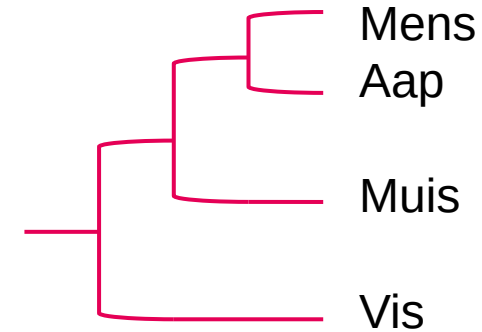
Organisme boom



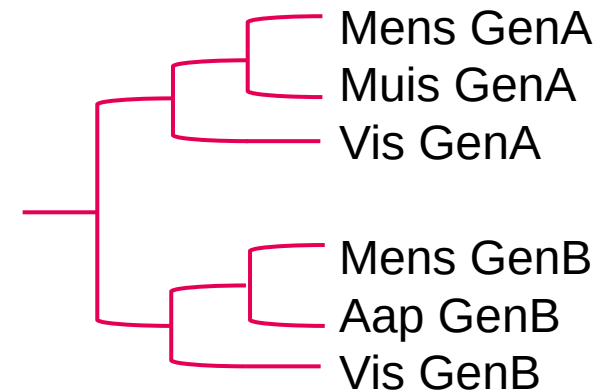
Gen boom

GEN VS. ORGANISME BOMEN

- Gen A mist in Aap, maar we weten dat de voorouder met deze organismen het gen had.
Aap is gen A verloren.
- Gen B mist in Muis, maar ook hier weten we dat het gen in de voorouder zat.
Muis is gen B verloren.

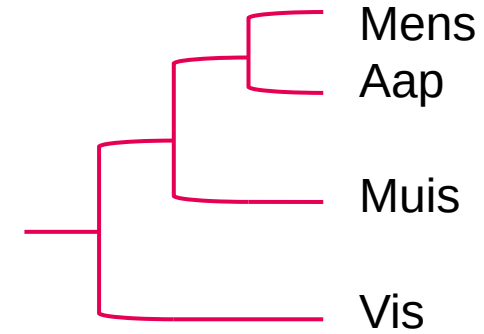
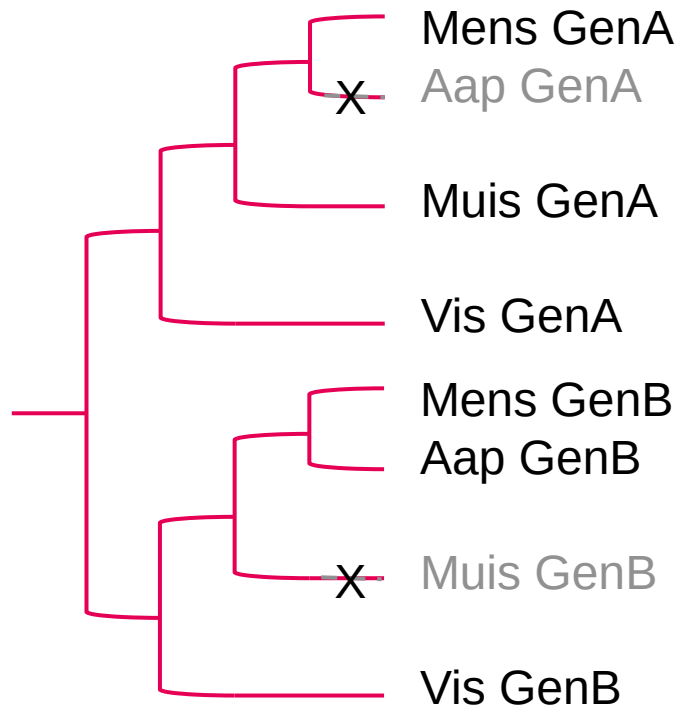


Organisme boom

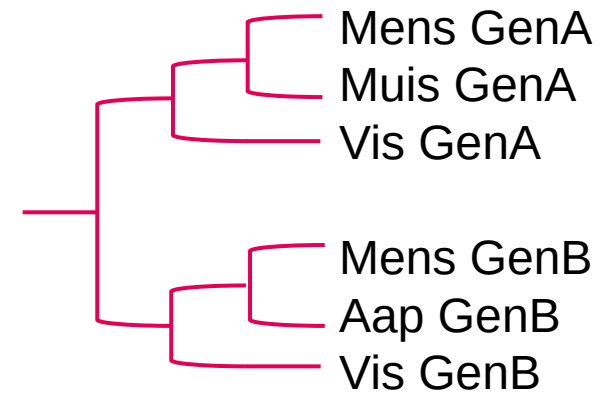


Gen boom

GEN VS. ORGANISME BOMEN



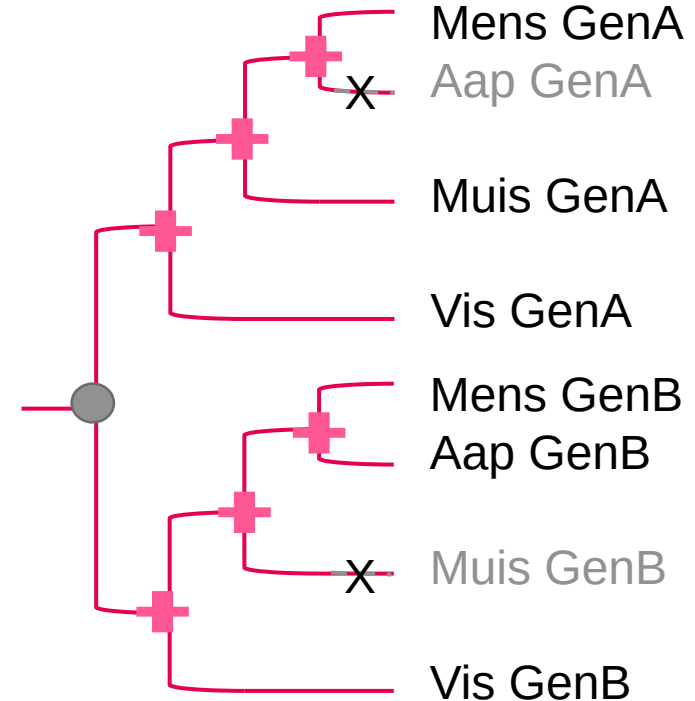
Organisme boom



Gen boom

GEN VS. ORGANISME BOMEN

- Gen A en B lijken ook op elkaar (maar minder dan Gen A op de GenA in andere organismen).
- We kunnen concluderen dat een gen is gedupliceerd in de voorouder van deze organismen en daarna de organismen zijn gevormd (soortvorming).



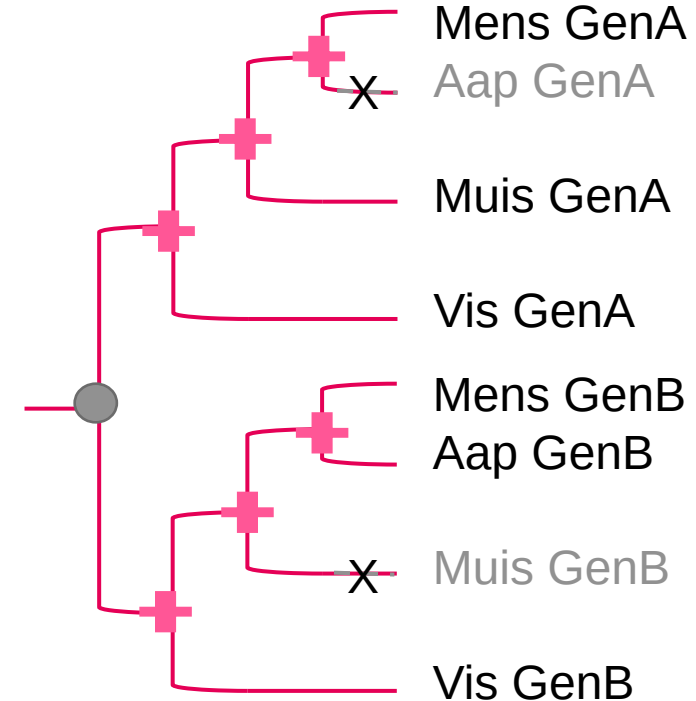
● Duplicatie
+ Soortvorming

HOMOLOGE SEQUENTIES (HERHALING)

- Sequenties kunnen op verschillende manieren een gedeeld voorouderlijke sequentie hebben:
 - Door een speciation (soortvorming): een gen in een voorouderlijk organisme bestaat nu in twee aparte organismen. We noemen dit **orthologen**.
 - Door een duplicatie in: een gen in een (voorouderlijk) organisme is gedupliceerd en beide kopieën kunnen zich bevinden in nieuwe organismen. We noemen dit **paralogen**.

GEN VS. ORGANISME BOMEN

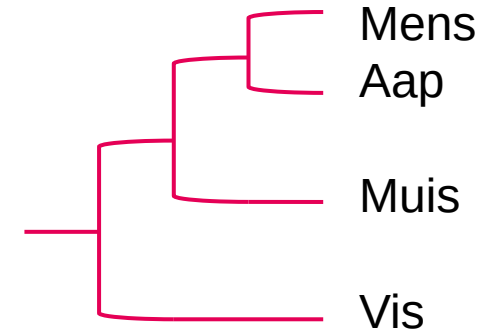
- Gen A en B lijken ook op elkaar (maar minder dan Gen A op de genen A in andere organismen).
- We kunnen concluderen dat een gen is gedupliceerd in de voorouder van deze organismen en daarna de organismen zijn gevormd (soortvorming).
 - Gen A en gen B zijn paralogen
 - Gen A in mens, muis en vis zijn orthologen.
 - Gen B in mens, aap en vis zijn orthologen.



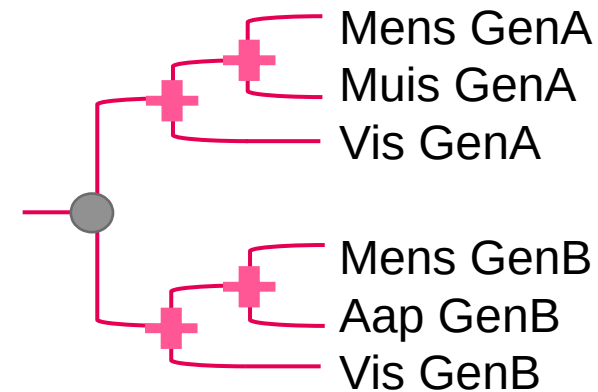
● Duplicatie
+ Soortvorming

GEN VS. ORGANISME BOMEN

- Gen A en B lijken ook op elkaar (maar minder dan Gen A op de genen A in andere organismen).
- We kunnen concluderen dat een gen is gedupliceerd in de voorouder van deze organismen en daarna de organismen zijn gevormd (soortvorming).
- Gen A en gen B zijn paralogen
- Gen A in mens, muis en vis zijn paralogen.



Organisme boom

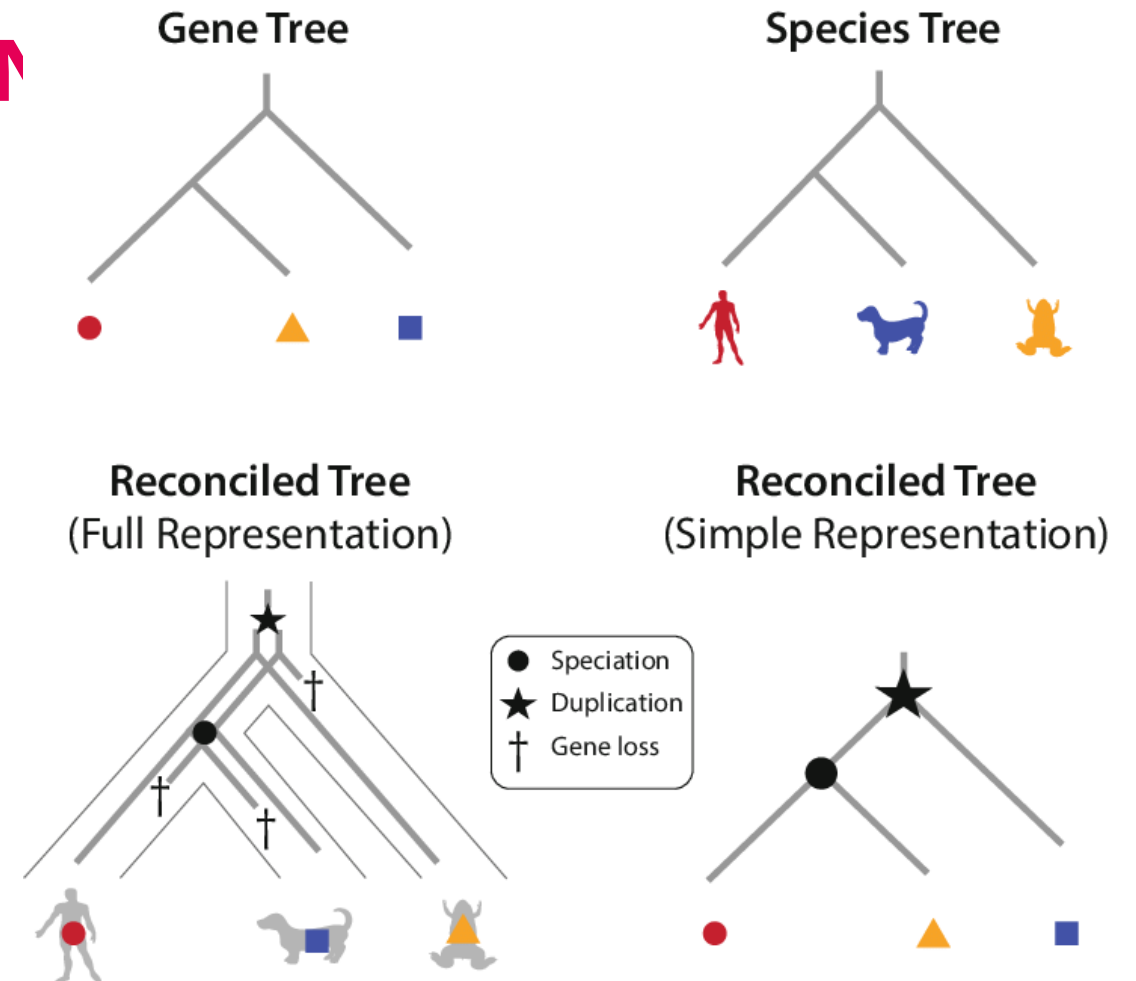


Gen boom



GEN VS. ORGANISME BOMEN

- Voorbeeld: genen lijken niet de evolutie te volgen van die van de organismen (mens, hond en kikker).
- Wat we kunnen afleiden is dat de genen van de mens en die van de kikker het meest op elkaar lijken
 - Rood = rond = mens
 - Geel = driehoek = kikker
 - Blauw = vierkant = hond



Schematic example of the gene/species tree reconciliation. The gene... | Download Scientific Diagram (researchgate.net)

GEN BOMEN

- Gen bomen laten ons zien welke genen gedupliceerd, verloren, of gevormd zijn.
- Hoe genfamilies ontstaan, hoe de evolutie van genen is etc.

GEN DUPLICATIES EN GEN FAMILIES

- Gen duplicaties (kopieën) zorgen voor een verhoging in gen aantallen in een genoom en zorgt voor evolutionaire veranderingen (nieuw functies).
- Paralogen ontstaan na een duplicatie en nemen een andere functie.
- Orthologen ontstaan na een soortvorming en hebben dezelfde functie in verschillende organismen. In andere woorden, is hetzelfde gen in verschillende organismen.
- Meerdere gen duplicaties achter elkaar resulteert in gen families, groepen gerelateerde genen in een genoom.

HOMOLOGIE EN FYLOGENIE

- Als we een organisme boom willen maken, willen we dan orthologen or paralogen genen om te vergelijken?
We gebruiken orthologen want die weergeven de historie van een soortvorming.

GENOOM EVOLUTIE

- Soorten die heel erg lang geleden uit elkaar zijn gegaan, kunnen veel orthologe genen nog delen.
 - Mens en muis delen 99% orthologe genen.
 - Mens en muis zijn ~65 miljoen jaar geleden uit elkaar gegaan.
- Onthoud: Gen aantal (die door duplicatie verhoogt bijvoorbeeld) en complexiteit van organismen zijn niet gecorreleerd. Mens heeft bijvoorbeeld maar 4x meer genen dan gist.
 - Komt o.a. door alternative splicing.

VOLGENDE WEEK

- Aan het onderzoeksverslag