

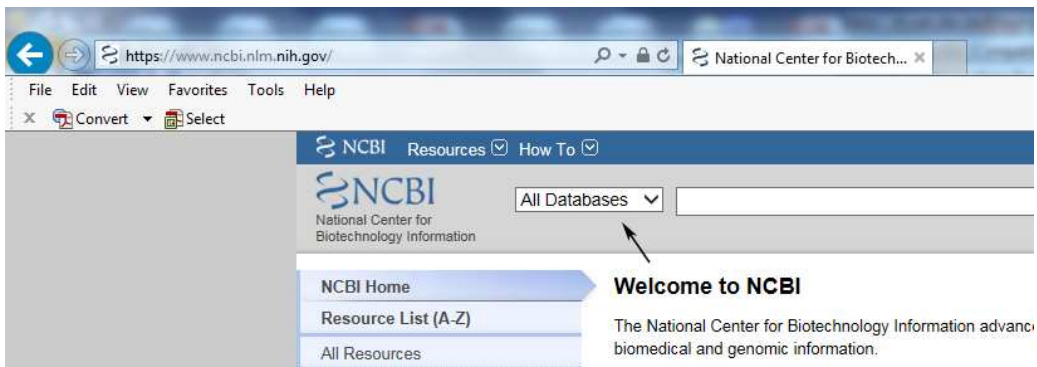
0. Bekijk vooraf bijvoorbeeld de kennisclip over evo-devo:

[https://www.youtube.com/watch?v=2RN1\\_sF5VKA&index=12&list=PLzugOrS2Z8ooyVJNuUxp5xelqATlh3i](https://www.youtube.com/watch?v=2RN1_sF5VKA&index=12&list=PLzugOrS2Z8ooyVJNuUxp5xelqATlh3i)

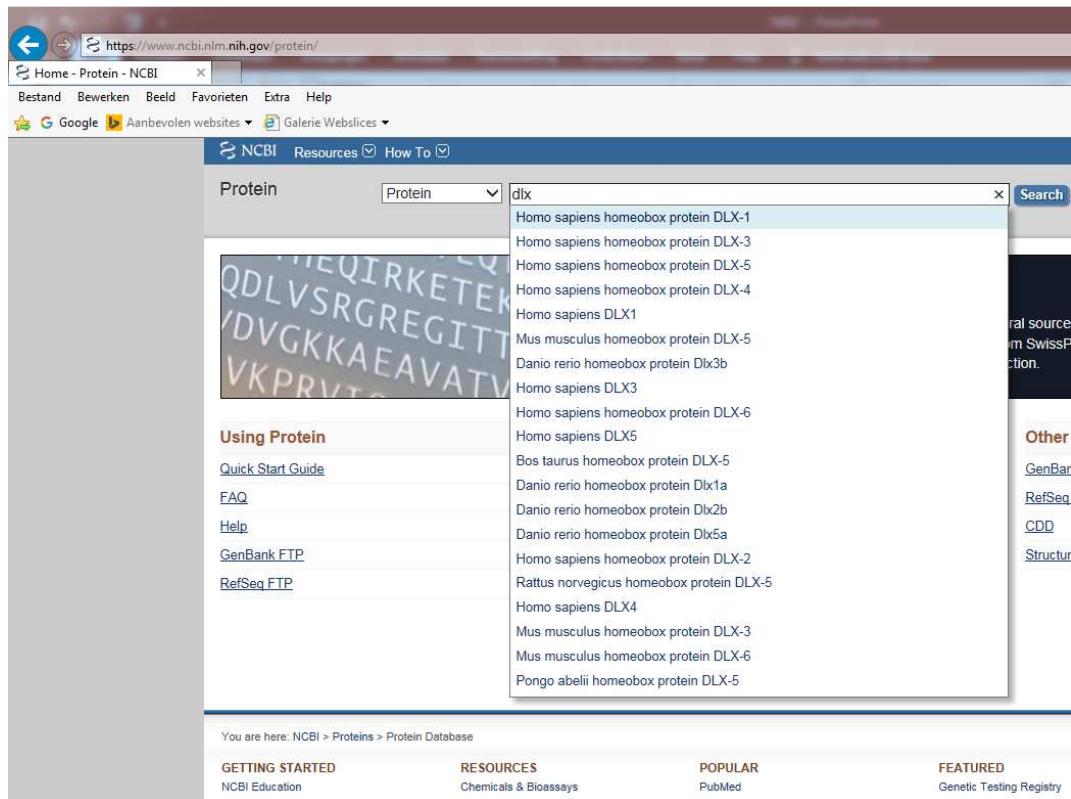
1. Ga naar NCBI <http://www.ncbi.nlm.nih.gov>

2. Geef aan dat je een eiwit “Protein” zoekt (bij pijltje in screenshot). We doen deze opdracht met eiwitsequenties, omdat de DNA sequenties teveel van elkaar verschillen. Druk op enter (zonder dat je tekst ingevoerd hebt, maar met je cursor in het zoekveld); je gaat dan meteen naar:

<https://www.ncbi.nlm.nih.gov/protein/?term=>

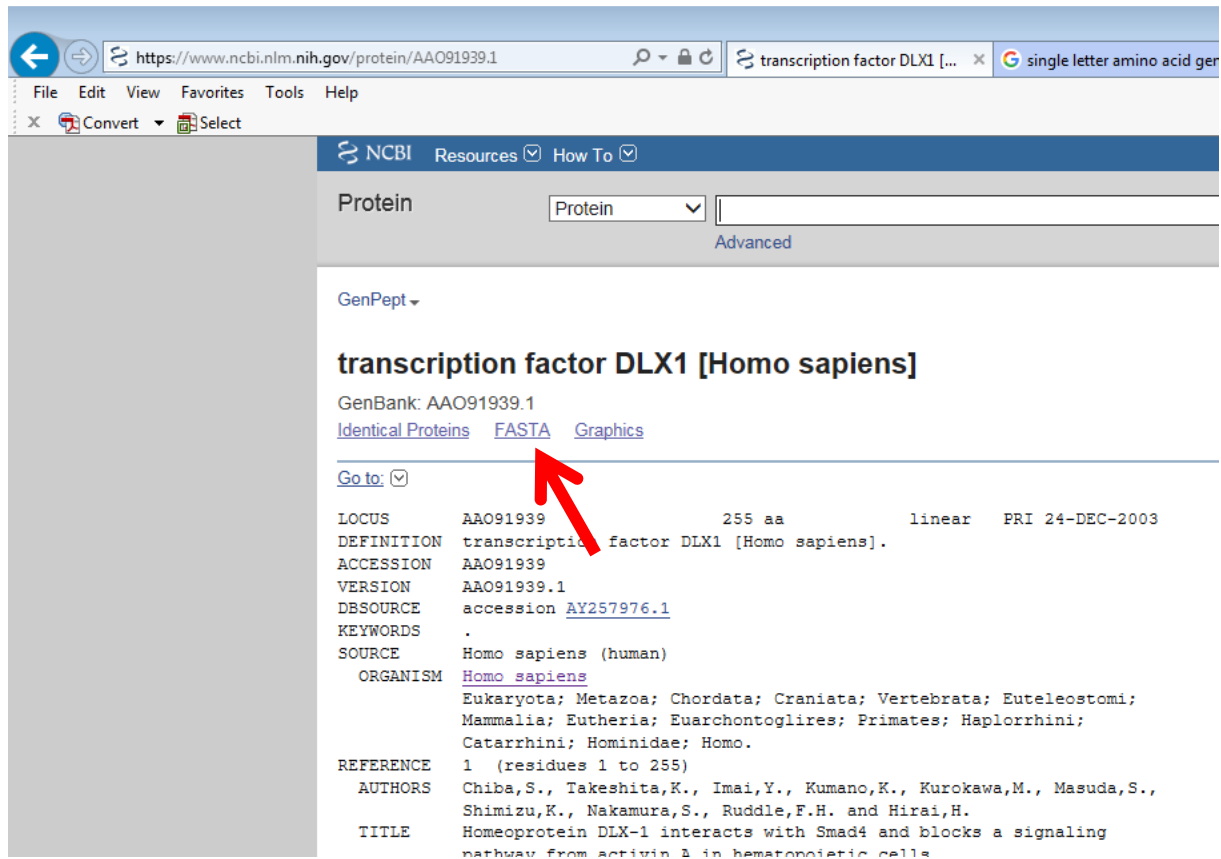


3. Typ in “dlx” (de afkorting van distalless in de mens), meteen verschijnt al een juiste optie. “Isoforms” zijn trouwens verschillende versies van het eiwit, afkomstig van hetzelfde gen (bijvoorbeeld door alternative splicing). Voor onze doeleinden maakt het niet zoveel uit welke splice vorm je neemt



4. De eerste link is meteen een link naar een eiwitsequentie (Transcription factor DLX1 Homo sapiens). Klik daarop. Als je helemaal naar onder scrolt, zie je de hele eiwitsequentie.

5. Een handige manier om met de sequenties te werken, is in FASTA-format. Klik op de link (rode pijl in screenshot).



The screenshot shows the NCBI protein page for transcription factor DLX1 [Homo sapiens]. The page includes a search bar, navigation links (Protein, GenPept, Identical Proteins, FASTA, Graphics), and a detailed description of the protein. A red arrow points to the 'FASTA' link in the top navigation bar.

transcription factor DLX1 [Homo sapiens]

GenBank: AAO91939.1

[Identical Proteins](#) [FASTA](#) [Graphics](#)

Go to: ☐

LOCUS AAO91939 255 aa linear PRI 24-DEC-2003

DEFINITION transcription factor DLX1 [Homo sapiens].

ACCESSION AAO91939

VERSION AAO91939.1

DBSOURCE accession [AY257976.1](#)

KEYWORDS .

SOURCE Homo sapiens (human)

ORGANISM [Homo sapiens](#)

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini; Catarrhini; Hominidae; Homo.

REFERENCE 1 (residues 1 to 255)

AUTHORS Chiba, S., Takeshita, K., Imai, Y., Kumano, K., Kurokawa, M., Masuda, S., Shimizu, K., Nakamura, S., Ruddle, F.H. and Hirai, H.

TITLE Homeoprotein DLX-1 interacts with Smad4 and blocks a signaling pathway from activin A in hematopoietic cells

6. Copy-paste die sequentie in een leeg, simpel tekst-bestandje (bijvoorbeeld van Notepad. Word is onhandig). Hou het fasta-format aan. Dat betekent: eerst een haak, dan de naam van je sequentie. Dan een enter, gevolgd door de sequentie (kan DNA zijn, maar in dit geval dus een enkele-letter aminozuursequentie). Geef de sequentie een handige naam. Als we uiteindelijk een stamboom gaan maken, wordt vaak alleen het eerste gedeelte van de naam meegenomen. Als daar alleen maar getalletjes staan, weet je niet meer wat wat was. Zet meteen in het eerste gedeelte dus “Mens” ofzo. Bijvoorbeeld:

```
>NP_835221.2 homeobox protein DLX-1 isoform 1 [Homo sapiens]
MTMTTTPESLNSPVSGKAVFMEFGPPNQMQSPSPMSHGHYSMHCLHSAGHSQPDGAYSSASSFSRPLGYP
YVNSVSSHASSPYISSVQSYPGSASLAQSRLEDPGADSEKSTVVEGGEVRFNGKGKKIRKPRTIYSSLQL
QALNRRFQQTQYLALPERAELAASLGLTQTQVKIWFQNKRSKFKKLMKQGGALEGSALANGRALSAGSP
PVPPGWNPNSSSGKSGGNAGSYIPSYTSWYPSAHQEAMQQPQLM
```

wordt:

```
>MensDistalless
MTMTTTPESLNSPVSGKAVFMEFGPPNQMQSPSPMSHGHYSMHCLHSAGHSQPDGAYSSASSFSRPLGYP
YVNSVSSHASSPYISSVQSYPGSASLAQSRLEDPGADSEKSTVVEGGEVRFNGKGKKIRKPRTIYSSLQL
QALNRRFQQTQYLALPERAELAASLGLTQTQVKIWFQNKRSKFKKLMKQGGALEGSALANGRALSAGSP
PVPPGWNPNSSSGKSGGNAGSYIPSYTSWYPSAHQEAMQQPQLM
```

7. Nu gaan we op zoek naar soortgelijke sequenties in andere dieren met BLAST. Ga naar : <http://blast.ncbi.nlm.nih.gov/Blast.cgi>. Copy-paste de menselijke sequentie in het query-veld. Selecteer bij "Organism"(onder choose search set) "Chimpanzee" (je mag ook meteen "Pan troglodytes" intypen). Anyway, taxon ID 9598. Verander verder niets en klik op "Blast" (links onderaan).

The screenshot shows the NCBI BLAST Standard Protein BLAST interface. The browser address bar displays [https://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastp&PAGE\\_TY](https://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastp&PAGE_TY). The page header includes the NIH logo and the text "U.S. National Library of Medicine" and "NCBI National Center for Biotechnology Information". The main heading is "BLAST® >> blastp suite".

The "Standard Protein BLAST" section is active. Under "Enter Query Sequence", the text input field contains the following FASTA sequence:

```

YVNSVSSHASSPYISSVQSYFGSASLAQSRLEDPGADSEKSTVVEGGEVRFNGKGGKIRKPRTIYSS
LQL
QALNRRFQQTOYLALPERAELAASLGLTQTQVKIWFQNKRSKFKKLMKQGGAALEGSALANGRALSA
GSP
FVPPGWNPNSSSGKSGGNAGSYIPSYTSWYPSAHQEAMQQPQLM
  
```

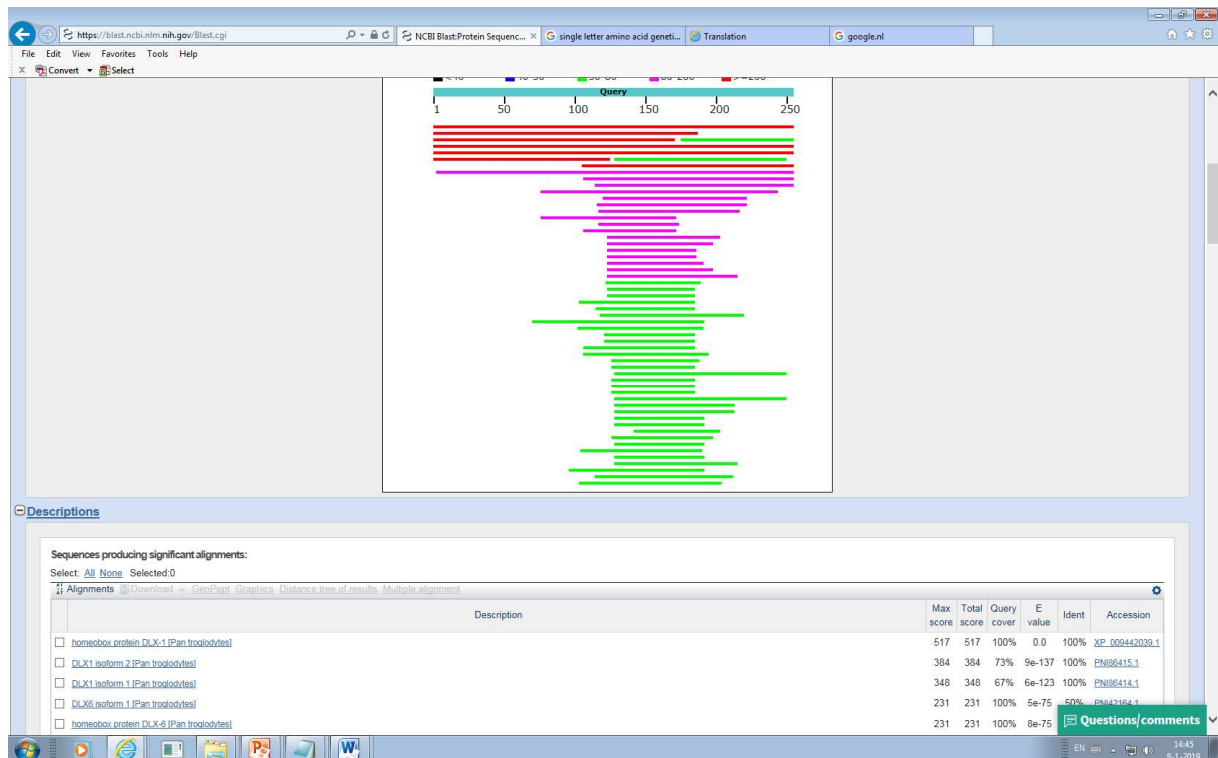
Below the sequence input, there are fields for "Or, upload file" (with a "Browse..." button), "Job Title" (with a text input field), and a checkbox for "Align two or more sequences".

The "Choose Search Set" section shows the "Database" dropdown set to "Non-redundant protein sequences (nr)". The "Organism" dropdown is open, displaying a list of search sets. The top entry is "chimpanzee (taxid:9598)". Other entries include "chimpanzees (taxid:9596)", "Chimpanzee immunodeficiency virus (SIV(CPZ)) (taxi...", "Chimpanzee immunodeficiency virus (taxid:11723)", "pygmy chimpanzee (taxid:9597)", "Chimpanzee stool avian-like circovirus Chimp17 (taxid...", "Pygmy chimpanzee papillomavirus (taxid:10576)", "Pygmy chimpanzee papillomavirus type 1 (taxid:10576)", "pygmy chimpanzee papillomavirus type 1 PCPV-1 (tax...", "Chimpanzee cytomegalovirus (taxid:188763)", "Plasmodium sp. chimpanzee clade C2 (taxid:879268)", "Chimpanzee alpha-1 herpesvirus (taxid:332937)", "Chimpanzee herpesvirus strain 105640 (taxid:332937)", "Plasmodium sp. chimpanzee clade C3 (taxid:880533)", "Plasmodium sp. chimpanzee clade C1 (taxid:880532)", and "Chimpanzee polyomavirus (taxid:305677)".

At the bottom left, there is a blue "BLAST" button. Below it, there is a link for "Algorithm parameter".

8. Wacht even.

9. In het resultaat zie je bovenaan een schematische weergave van de BLAST-resultaten. De kleurtjes geven de gelijkenis van de gevonden sequenties aan.



Als je helemaal naar beneden scrolt op de pagina, zie je de eigenlijke alignments. Je ziet: chimpansee Distalless is identiek aan menselijk distalless!

☐ ventral anterior homeobox 1 isoform X1 [Pan troglodytes]

☐ diencephalon/mesencephalon homeobox protein 1 isoform X1 [Pan troglodytes]

**Alignments**

Download [v] GenPept Graphics

homeobox protein DLX-1 [Pan troglodytes]  
Sequence ID: [XP\\_009442039.1](#) Length: 255 Number of Matches: 1  
[See 1 more title\(s\)](#)

Range 1: 1 to 255 [GenPept] [Graphics] [Next Match] [Previous Match]

| Score          | Expect   | Method                       | Identities    | Positives     | Gaps      |
|----------------|--|------------------------------|---------------|---------------|-----------|
| 517 bits(1332) | 0.0  | Compositional matrix adjust. | 255/255(100%) | 255/255(100%) | 0/255(0%) |
| Query 1        | MTMTMPESLNSPVSGKAVFMEFGPPNQMSFSPMSHGHYSMHCLHSAGHSQPDGAYSSA   | 60                           |               |               |           |
| Sbjct 1        | MTMTMPESLNSPVSGKAVFMEFGPPNQMSFSPMSHGHYSMHCLHSAGHSQPDGAYSSA   | 60                           |               |               |           |
| Query 61       | SSFSRPLGYPYVNSVSSHASSPYISSVQSYPGSASLAQSRLEDPGADSEKSTVVEGGEVR | 120                          |               |               |           |
| Sbjct 61       | SSFSRPLGYPYVNSVSSHASSPYISSVQSYPGSASLAQSRLEDPGADSEKSTVVEGGEVR | 120                          |               |               |           |
| Query 121      | FNGKGGKIRKPRTIYSSLQLQALNRRFQQTQYLALPERAELAASLGLTQTQVKIWFQNK  | 180                          |               |               |           |
| Sbjct 121      | FNGKGGKIRKPRTIYSSLQLQALNRRFQQTQYLALPERAELAASLGLTQTQVKIWFQNK  | 180                          |               |               |           |
| Query 181      | SKFKKLMKQGGAALEGSAANGRALSGSPVPVPGWNPNSGKSGGNAGSYIPSYTSW      | 240                          |               |               |           |
| Sbjct 181      | SKFKKLMKQGGAALEGSAANGRALSGSPVPVPGWNPNSGKSGGNAGSYIPSYTSW      | 240                          |               |               |           |
| Query 241      | YPSAHQEAMQPPQLM  | 255                          |               |               |           |
| Sbjct 241      | YPSAHQEAMQPPQLM  | 255                          |               |               |           |

**10.** Klik op de link naar de eiwitsequentie van de chimpansee (rode pijl in screenshot). Nu ga je hetzelfde doen als voor de menselijke sequentie: ga naar fasta en copy de chimpansee-sequentie. Plak die sequentie nu gewoon in dezelfde file als de menselijke sequentie, maar een regeltje eronder. Dat ziet er dan zo uit:

```
>NP_835221.2 homeobox protein DLX-1 isoform 1 [Homo sapiens]
MTMTTMPESLNSPVSGKAVFMEFGPPNQQMSPSPMSHGHYSMHCLHSAGHSQPDGAYSSASSFSRPLGYP
YVNSVSSHASSPYISSVQSYPGSASLAQSRLEDPGADSEKSTVVEGGEVRFNGKGKKIRKPRTIYSSLQL
QALNRRFQQTQYLALPERAELAASLGLTQTQVKIWFQNKRSKFKKLMKQGGAALEGSALANGRALSAAGSP
PVPPGWNPNSSSGKSGGNAGSYIPSYTSWYPSAHQEAMQQPQLM
```

```
>XP_009442039.1 homeobox protein DLX-1 [Pan troglodytes]
MTMTTMPESLNSPVSGKAVFMEFGPPNQQMSPSPMSHGHYSMHCLHSAGHSQPDGAYSSASSFSRPLGYP
YVNSVSSHASSPYISSVQSYPGSASLAQSRLEDPGADSEKSTVVEGGEVRFNGKGKKIRKPRTIYSSLQL
QALNRRFQQTQYLALPERAELAASLGLTQTQVKIWFQNKRSKFKKLMKQGGAALEGSALANGRALSAAGSP
PVPPGWNPNSSSGKSGGNAGSYIPSYTSWYPSAHQEAMQQPQLM
```

Of, met handige namen:

```
>MensDistalless
MTMTTMPESLNSPVSGKAVFMEFGPPNQQMSPSPMSHGHYSMHCLHSAGHSQPDGAYSSASSFSRPLGYP
YVNSVSSHASSPYISSVQSYPGSASLAQSRLEDPGADSEKSTVVEGGEVRFNGKGKKIRKPRTIYSSLQL
QALNRRFQQTQYLALPERAELAASLGLTQTQVKIWFQNKRSKFKKLMKQGGAALEGSALANGRALSAAGSP
PVPPGWNPNSSSGKSGGNAGSYIPSYTSWYPSAHQEAMQQPQLM
```

```
>ChimpanseeDistalless
MTMTTMPESLNSPVSGKAVFMEFGPPNQQMSPSPMSHGHYSMHCLHSAGHSQPDGAYSSASSFSRPLGYP
YVNSVSSHASSPYISSVQSYPGSASLAQSRLEDPGADSEKSTVVEGGEVRFNGKGKKIRKPRTIYSSLQL
QALNRRFQQTQYLALPERAELAASLGLTQTQVKIWFQNKRSKFKKLMKQGGAALEGSALANGRALSAAGSP
PVPPGWNPNSSSGKSGGNAGSYIPSYTSWYPSAHQEAMQQPQLM
```

**11.** Uiteindelijk willen we dus een hele lange Fasta-file, met de sequenties van Distalless uit heel veel dieren. Laten we eerst maar eens dicht bij huis blijven en nog een zoogdier uitzoeken. Ik heb de koe genomen ("cow", *Bos taurus*), maar je kunt natuurlijk ook de muis nemen. Van typische modellen of veelgebruikte organismen, zoals de muis en de koe zijn de genoomsequenties natuurlijk beter en beter geannoteerd. Neem de menselijk sequentie (of die van de chimp), ga weer naar de blast pagina <http://blast.ncbi.nlm.nih.gov/Blast.cgi> en herhaal stap 7, maar kies nu "Cow" (Taxid 9913). En herhaal het hele riedeltje. Plak de uiteindelijke eiwitsequentie van de koe weer in je fasta-file.

**12.** Zullen we meteen de zoogdieren maar verlaten? We gaan naar vogels, amfibieën en vissen. Zoek de kipsequentie ("chicken" taxid 9031), de *Xenopus-laevis* (taxid 8355)sequentie. Pas op hier! In *Xenopus* zit nog een ander gen: Dlx-1 like. Dit is de eerste hit. Maar dit is vermoedelijk een kopie. Hoewel die kopie als eerste BLAST-hit verschijnt, is deze sequentie nogal afwijkend en gooit de uiteindelijke boom door elkaar. De tweede hit is hetzelfde. De beste hit is Dll-1. *Xenopus* is sowieso een beetje raar; voor Pax6 bijvoorbeeld heb ik hem er helemaal uit gelaten. En zoek de zebravis-sequentie (taxid 7955). Plak ze allemaal (kip, koe, kikker, zebravis) in je fasta-file.

**13.** Hoe nu verder? Je kunt bij "Organism" ook in hele groepen zoeken, zoals "Amphibians", of hele groepen uitsluiten, zoals "chordates". Laten we eerst binnen de deuterostomen blijven. Ik heb de zeester *Pisaster ochraceus* genomen (taxid:7612). Het makkelijkst is om gewoon direct op dit organisme te zoeken. Voor Pax6 heb ik de zeester *Acanthaster planci* genomen (taxid:133434). Plak de *Pisaster ochraceus* sequentie in je fasta-file (het eerste stukje van deze sequentie ontbreekt, maar dat is niet erg).

**14.** Tja, en dan naar de protostomen. Voor de lophotrochozoa heb ik gekozen voor de annelide *Platynereis dumerilii* (taxid:6359) en de mollusk (oester) *Pinctada fucata* (taxid:50426). Voor de ecdysozoa heb ik het insect *Drosophila melanogaster* (taxid:7227) genomen. Bij *Drosophila* heb ik isoform A gekozen, maar zoals gezegd zal het niets uitmaken als je isoform B neemt. Voor Pax6 heb ik als oester *Crassostrea gigas* (taxid:29159) genomen. Door een duplicatie in insecten, heeft *Drosophila* twee gelijkende genen: eyeless en Twin of eyeless. Het maakt voor de uiteindelijke boom niet uit welke van de twee je neemt.

**15.** Zo, we hebben de tweezijdig symmetrische dieren (waarvan de meeste met pootjes) nu even grof gesampled. Maar zouden radiaal symmetrische dieren ook distalless hebben? Kijk eens naar de kwal *Clytia hemisphaerica* (taxid:252671) Neem de eerste hit en plak deze als laatste in je fasta file. Voor Pax6, heb ik in de zeeanemoon *Nematostella vectensis* gekeken.

Voor de zekerheid nog even de unieke identifiers:

```
>NP_835221.2 homeobox protein DLX-1 isoform 1 [Homo sapiens]
>XP_009442039.1 homeobox protein DLX-1 [Pan troglodytes]
>NP_001092487.1 homeobox protein DLX-1 [Bos taurus]
>NP_001039307.2 homeobox protein DLX-1 [Gallus gallus]
>NP_001090284.1 homeobox protein DLL-1 [Xenopus laevis]
>NP_571380.1 homeobox protein Dlx1a [Danio rerio]
>ALH43434.1 homeodomain protein Dlx, partial [Pisaster ochraceus]
>CAJ38799.1 distalless, Dlx-1 [Platynereis dumerilii]
>ASY93124.1 DLX [Pinctada fucata]
>NP_523857.1 Distal-less, isoform A [Drosophila melanogaster]
>ACM62726.1 Dll1 distalless homeodomain protein [Clytia hemisphaerica]
```

Mijn fasta-file ziet er nu zo uit:

```
>Mens
MTMTTTPESLNSPVSGKAVFMEFGPPNQMQSPSPMSHGHYSMHCLHSAGHSQPDGAYSSASSFSRPLGYPYVNSVSSHASSPYISSVQSYPGSA
SLAQSRLEDPGADSEKSTVVEGGEVRFNGKGKKIRKPRTIYSSLQLQALNRRFQQTQYLALPERAELAASLGLTQTQVKIWFQNKRSKFKKLMK
QGGAALGSGALANGRALSGSPVPVPGWNPNSSGKGSGGNAGSYIPSYTSWYPSAHQEAMQQPQLM

>Chimpansee
MTMTTTPESLNSPVSGKAVFMEFGPPNQMQSPSPMSHGHYSMHCLHSAGHSQPDGAYSSASSFSRPLGYPYVNSVSSHASSPYISSVQSYPGSA
SLAQSRLEDPGADSEKSTVVEGGEVRFNGKGKKIRKPRTIYSSLQLQALNRRFQQTQYLALPERAELAASLGLTQTQVKIWFQNKRSKFKKLMK
QGGAALGSGALANGRALSGSPVPVPGWNPNSSGKGSGGSAGSYIPSYTSWYPSAHQEAMQQPQLM

>Koe
MTMTTTPESLNSPVSGKAVFMEFGPPNQMQSPSPMSHGHYSMHCLHSAGHSQPDGAYSSASSFSRPLGYPYVNSVSSHASSPYISSVQSYPGSA
SLTQSRLEDPGADSEKSTVVEGGEVRFNGKGKKIRKPRTIYSSLQLQALNRRFQQTQYLALPERAELAASLGLTQTQVKIWFQNKRSKFKKLMK
QGGAALGSGALANGRALSGSPVPVPGWNPNSSGKGSGGSAGSYIPSYTSWYPSAHQEALQQPQLM

>Kip
MTMTTTPESLNSPVSGKAVFMEFGPPNQMQSPSPMSHGHYSMHCLHSAGHSQPD SAYSTASSFSRPLGYPYVNSVSSHSGNPYISSVQYPNNS
GLAQPRLEETGAESEKSTVVEGGEVRFNGKGKKIRKPRTIYSSLQLQALNRRFQQTQYLALPERAELAASLGLTQTQVKIWFQNKRSKFKKLMK
QGGAALGSGALANGRALSGSPVPVAVWNTSSASGKASSGSAGTYIPSYTSWYPSAHQEAMQQPQLM

>Kikker
MMTMTTMADGLEAQDSSKSAFMEFGQQQSHSQSSPVMAAGHYPSLHCLHSGSHHPQHQHDTNYSGSNSYSRSLAAYPYMSHSQHSPYLQSCN
SNTTTQSRAEEDQQKTTVIENGEIRFNGKGKKIRKPRTIYSSLQLQALNRRFQQTQYLALPERAELAASLGLTQTQVKIWFQNKRSKYKLIK
QGNPLEIDQLAGTVALSPRSPAIPVVDVSASKGVSMAPNSYMPGYSHWYSSPHQDTMQRSQMM

>Zeester
GKKLRKPRTIYSSLQLQKLNHRFTKTQYLALPERADLAASLGLTQTQVKIWFQNNRSKYKKILKQHGSTPTSSTGSQDPEQDPPQSHTPASQQG
SLSPAPPPTSAPHPQGTPTMSGPSSTGLGSACYSSQQHVPSHHHHHHHPHHQMPTDGSSSTHSSPVPPWDYVGDIFTPPAVRSSSGTVDNNYPFH
QQHYQHHPYFWFGGQDGTSGMSAQQHMLASAPVDKTIQLAKECDFLELKNAFHTNA

>Annelide
MTLEKGYSMNLNMGSGEGLGPDQDVSKSAFMEIQQQQMNASMPNPAIRGGYPGNAGQMEGFPGQQTRGHIGYPFSMTPMGSHGTYNPTSYPHFS
APGYQTSAPSVTPPTSRESMYDLNPNYSSYDKLENKYIYRKGLFEDSPNSDEKPSMDELRVNGKGKKMRKPRTIYSSLQLQKLNKRFQRTQYL
ALPERAELAASLGLTQTQVKIWFQNNRSKYKKLMKQNP GIGPGGAQNGPPMDQGGMSPPPTHTPTSTGQAPTPSPQGAQGGPPXNGQPNP
HHGHAGAHGAGHGANPATPSMMGGPPPNMSMPPVSWGASDFQLKSEINCSSNSTNSTPTPTHTNTYMSQYPWYSQNPLAAQQHSLLT
```



```
>Oester
MLNVGSGVEGMEQEMAGKSAFMELQQSGGMPMGHPAYPMRSSYQPPHHGGQHGESVFSNPQGRGPLAGYPFHMNAMSPTAYNPSSGHHFSMPPYQ
SPSPTRDDKSQMDLRLINGKGKMKRPRTIYSSLQLQQLNRRFQRTQYLALPERAELAASLGLTQTQVKIWFQNRSSKAKKIMKQGGTTPPGPNQ
QPPTPVTSPPANQQMQPSSPVQHSPTSTNTHAHTPHQQLPPNGLKMEAPEQQHNLVSPSSSASPEPENHWNDHMSGMTSSHHAHASQPYMSM
PSMPSSSMMPSGMPMSYHSAWYSQPMVNQQSCLT

>Fruitvlieg
MDAPDAPHTPKYMDGGNTAASVTPGINIPGKSAFVELQQHAAAGYGGIRSTYQHFGPQGGQDSGFSPSPRS
ALGYFPFPMHQNSYSGYHLGSYAPPCASPPKDDFSISDKCEDSGLRVNGKGKKMRKPRTIYSSLQLQQLN
RRFQRTQYLALPERAELAASLGLTQTQVKIWFQNRSSKYKKMMKAAQGPNTSGMPLGGGGPNPGQHSPN
QMHSGGNNGGSGNSGSPSHYLPBGHSPTSPSTPVSELSPEFPPTGLSPPTQAPWDQKPHWIDHKPPPQMT
PQPPHPAATLHPQTTHHNPPPMGGYVPQYWYQPETNPSLVTVPVPAV

>Zebravis
MTMSTIPESLNSPVSGKSVFMEFGPPSQMSPSSMTHGHYSMHCLHSSGHPQHDSAYSAPASFPRSLPYP
YVNSVSGSHSSSPYLSTVQTYPNNSALAQTRLEDPAPESEKNTVVEGGEVRFNGKGKKIRKPRTIYSSLQL
QALNRRFQRTQYLALPERAELAASLGLTQTQVKIWFQNRSSKFKKLMKQGGGTIDTNALANGRLSTGSP
SVAPVWNTTATVKTSTPTSYIPSYSYPTAHQDTMQQPQLM

>Kwal
MIEMKMSPHQQAVLSQQVPNMSSFLPPSKAPEAYRPLNYTPTSFVTNFKEMSQQIDSNNNNNNNNNNPHES
EGKFSRPRPATQFSMPGDLRSHPTQDTHPFKTFDDSKSFLNNNTPNGSMPEQDSISSSLQSFHSSGLPTS
YTSSYNNNSMQTTLSSPSSITSNMSRSFSSSYVKQESDEEDEGDLTQSKNGKGKLPKRPRTIFTSQQL
RELNRAFERTHYLSLPERAELAHALGLTQTQIKIWFQNRSSKFKKI KANGGQMPPSSSLVSGGNPGLWP
SGYGEKSFGGPPMHLSPFSPPPSSSHSVPGDSWYYRMQSAGGYGSHEMTFPYGGTSGTDVRTTRPNFSY
HM
```

16. Zo, we hebben wat we nodig hebben. Nu gaan we een alignment en stamboom maken. Ga daarvoor naar <https://www.ebi.ac.uk/Tools/msa/clustalo/>. De hele fastafile kan daar zo, met alle sequenties tegelijk, in het input venster gecopy-paste worden (rode pijl in screenshot hieronder).

Tools > Multiple Sequence Alignment > Clustal Omega

## Multiple Sequence Alignment

Clustal Omega is a new multiple sequence alignment program that uses seeded guide trees and HMM profile-profile techniques or more sequences. For the alignment of two sequences please instead use our pairwise sequence alignment tools.

**Important note:** This tool can align up to 4000 sequences or a maximum file size of 4 MB.

**STEP 1 - Enter your input sequences**

Enter or paste a set of

PROTEIN

sequences in any supported format:

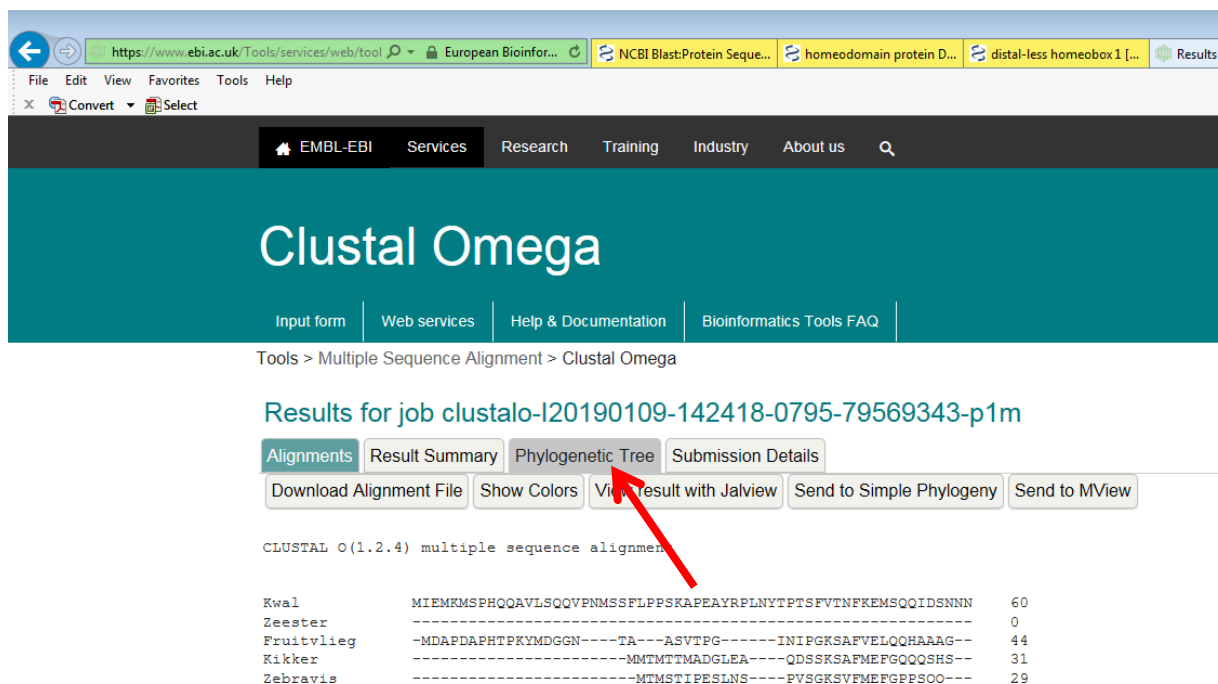
```
>Mens
MTMTTPESLNSPVSGKAVFMEFGPPNQMSPPMSHGHYSMHCLHSAGHSQPDGAYSSASSFSRPLGYP
YVNSVSSHASSPYISSVQSYPGSASLAQSRLEDPGADSEKSTVVEGGEVRFNGKGKKIRKPRTIYSSLQL
QALNRRFQRTQYLALPERAELAASLGLTQTQVKIWFQNRSSKFKKLMKQGGGALEGSALANGRLSAGSP
PVPPGWNPNSSSGKSGGNAGSYIPSYSYPTAHQEQAMQQPQLM

>Chimpansee
MTMTTPESLNSPVSGKAVFMEFGPPNQMSPPMSHGHYSMHCLHSAGHSQPDGAYSSASSFSRPLGYP
```

Or, upload a file:  [Use a example sequenc](#)

17. Druk op “submit” en wacht even. Bekijk het alignment. Striking toch, dat je zoveel gelijkenis met een kwal vertoont?

18. Klik op phylogenetic tree (rode pijl) en bekijk de stamboom.



Results for job clustalo-l20190109-142418-0795-79569343-p1m

Alignments | Result Summary | **Phylogenetic Tree** | Submission Details

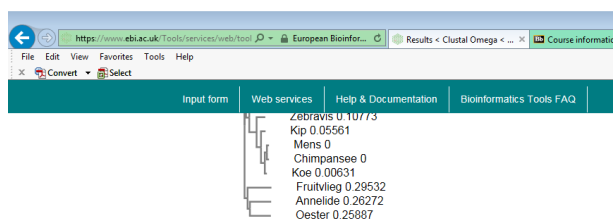
Download Alignment File | Show Colors | **View result with Jalview** | Send to Simple Phylogeny | Send to MView

CLUSTAL O (1.2.4) multiple sequence alignment

| Kwal       | MIEMKMSPHQQAVLSQQVPMSSFLPPSKAPEAYRPLNYTPTSFVTNFKEMSQQIDSNNN | 60 |
|------------|---|----|
| Zeester    | -----   | 0  |
| Fruitvlieg | -MDAPDAPHTPKYMDGNN---TA---ASVTPG-----INIPGKSAFVELQQHAAAG--  | 44 |
| Kikker     | -----MMTMTTMADGLEA---QDSSKSAFMFPGQQQSHS--                   | 31 |
| Zebravis   | -----MTMSTIPESLNS---PVSGKSVFMFPGPPSOO---                    | 29 |

19. Als je op “Real” klikt, geven de lengtes van de lijnen de genetische afstand weer.

20. Helaas ziet de boom er niet altijd even mooi uit, terwijl de boom op zich wel goed is. Bij Pax6 bijvoorbeeld, lijkt de boom nergens op, maar hij is alleen verkeerd geworteld. Verder is ie bijna perfect! Met de MEGA software kunnen we dit soort dingen aanpassen. Ga naar <https://www.megasoftware.net/> en installeer het programma op je computer. Om de tree-file met MEGA te openen, heb je alleen de “Tree data” nodig. Die staat onder je stamboom van Clustal Omega (rode pijl hieronder). Die kun je zo in een tekst-file copy-pasten (bijvoorbeeld Notepad). Sla de tekstfile op. Dit heet een “newick file”.



Tree Data

Tree Data

```
(
  Kwal:0.37164,
  (
    Zeester:0.22611,
    (
      Kikker:0.21523,
      (
        Zebravis:0.10773,
        (
          Kip:0.05561,
          (
            Mens:0.00000,
            Chimpanse:0.00000)
          :0.00545,
          Koe:0.00631)
        :0.04635)
      :0.05462)
    :0.09040)
  :0.04030)
  :0.01735,
  (
    Fruitvlieg:0.29532,
    (
      Annelide:0.26272,
      Oester:0.25887)
    :0.03869)
  :0.03671);
```



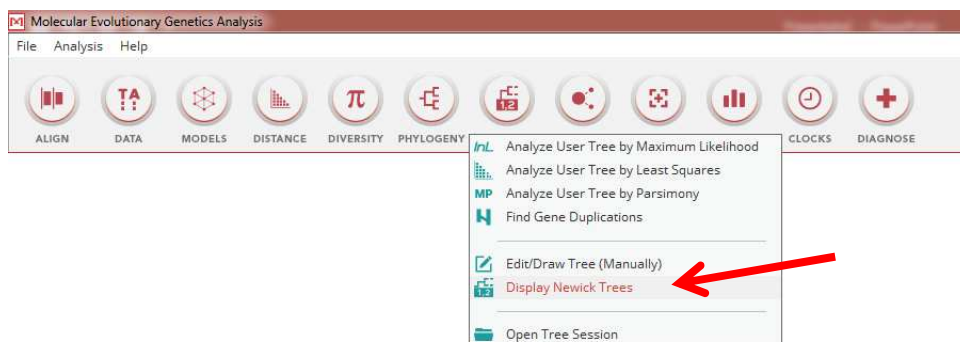
De enige informatie die de boom bevat staat in deze Newick file en is dus:

```
(  
Kwal:0.37164,  
(  
Zeester:0.22611,  
(  
Kikker:0.21523,  
(  
Zebravis:0.10773,  
(  
Kip:0.05561,  
(  
(  
Mens:0.00000,  
Chimpansee:0.00000)  
:0.00545,  
Koe:0.00631)  
:0.04635)  
:0.05462)  
:0.09040)  
:0.04030)  
:0.01735,  
(  
Fruitvlieg:0.29532,  
(  
Annelide:0.26272,  
Oester:0.25887)  
:0.02869)  
:0.03671)
```

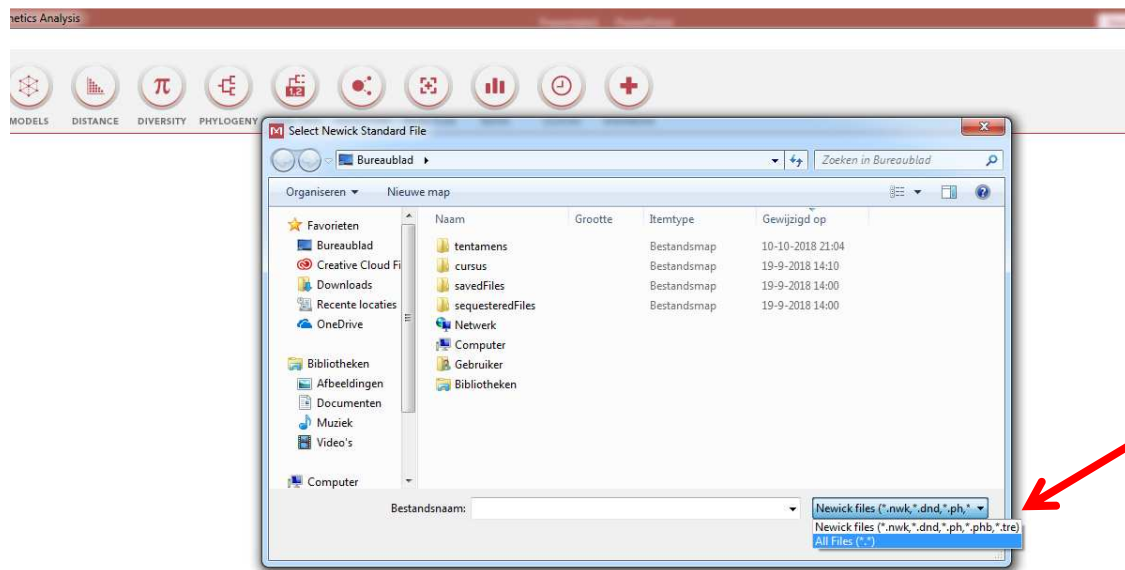
**21.** Open nu MEGA en klik op “User Tree” (rode pijl).



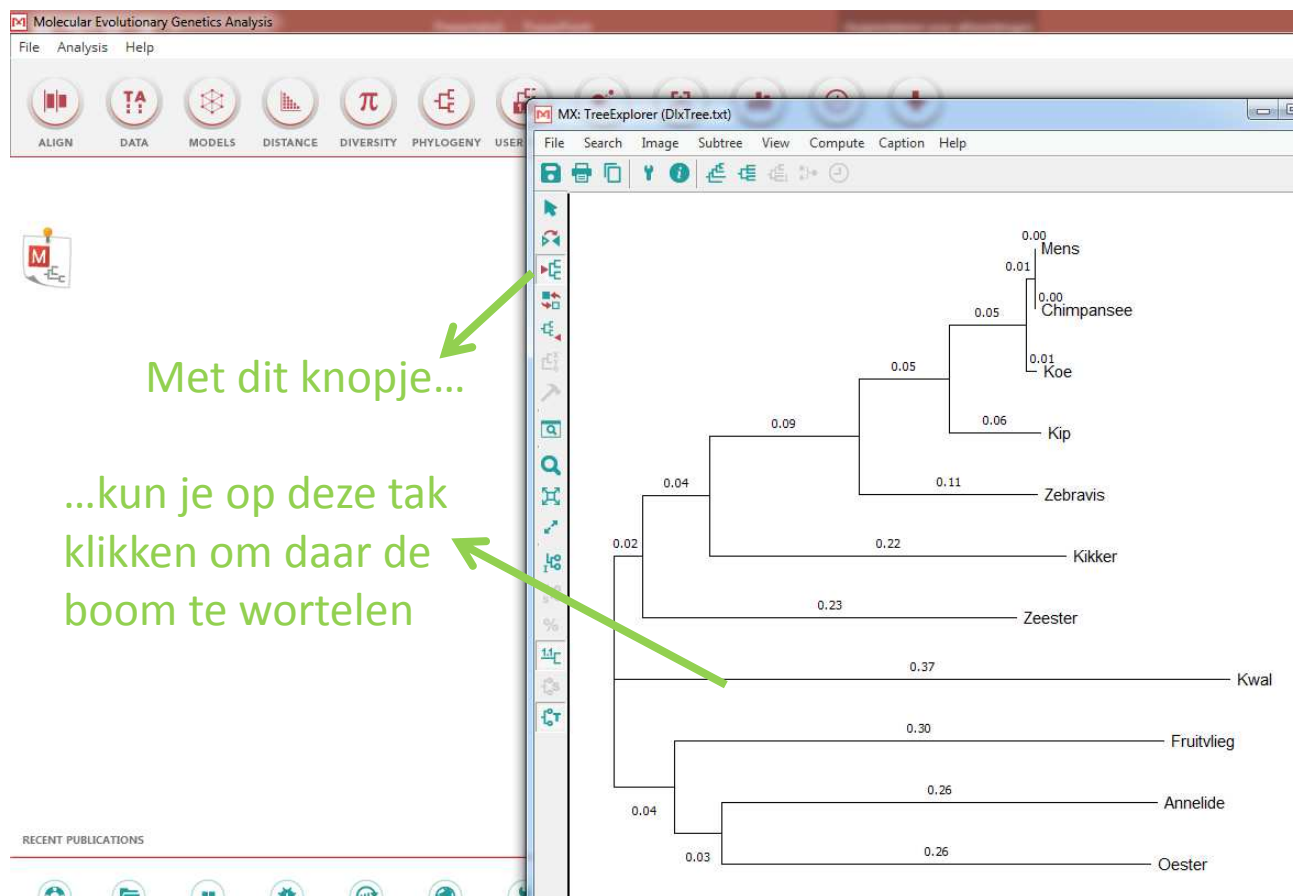
en selecteer “Display Newick Trees”.



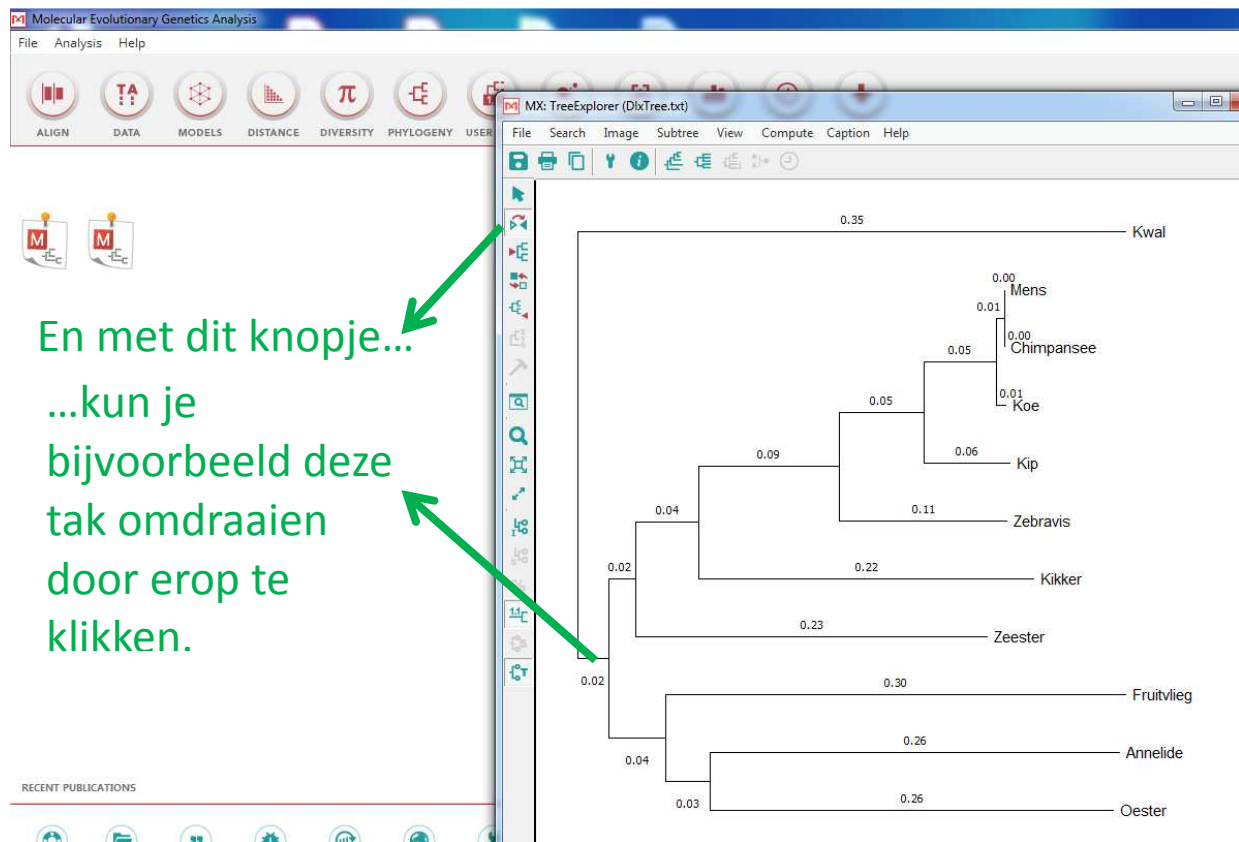
Denk eraan dat je even “All files” selecteert als je je bestand gewoon hebt opgeslagen als .txt en je je file niet weergegeven ziet (rode pijl). Selecteer je file en open hem.



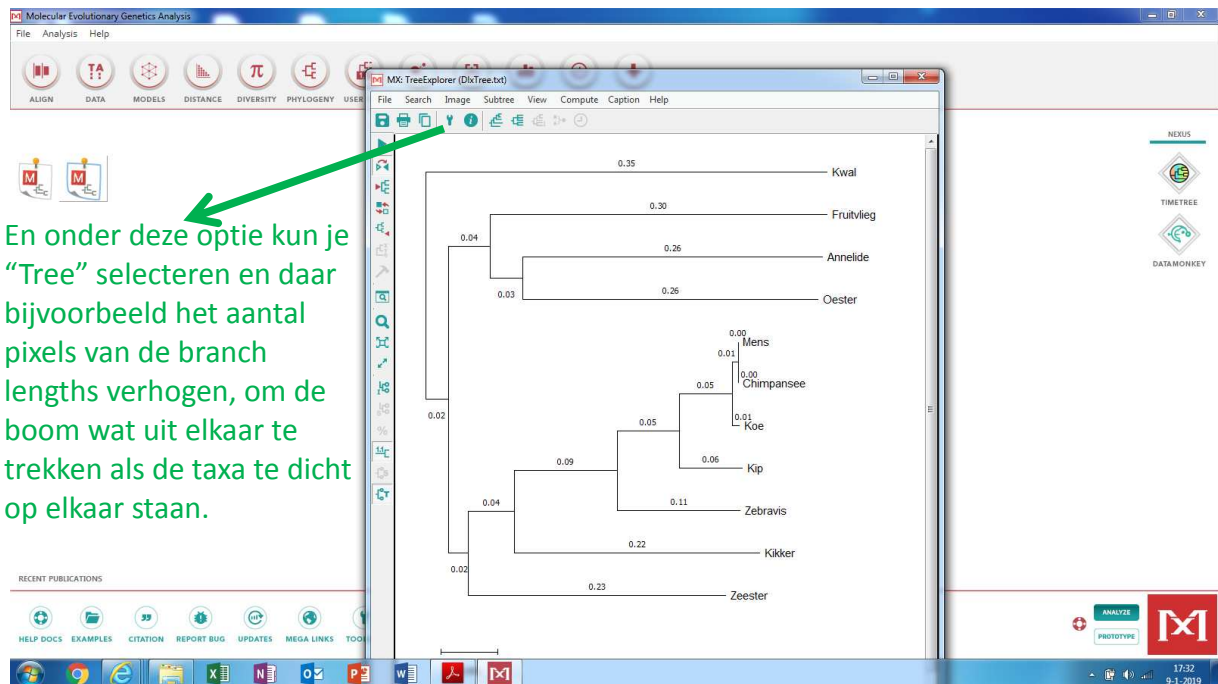
22. Nu kun je je tree gaan editen, zonder dat de boom qua informatie verandert.



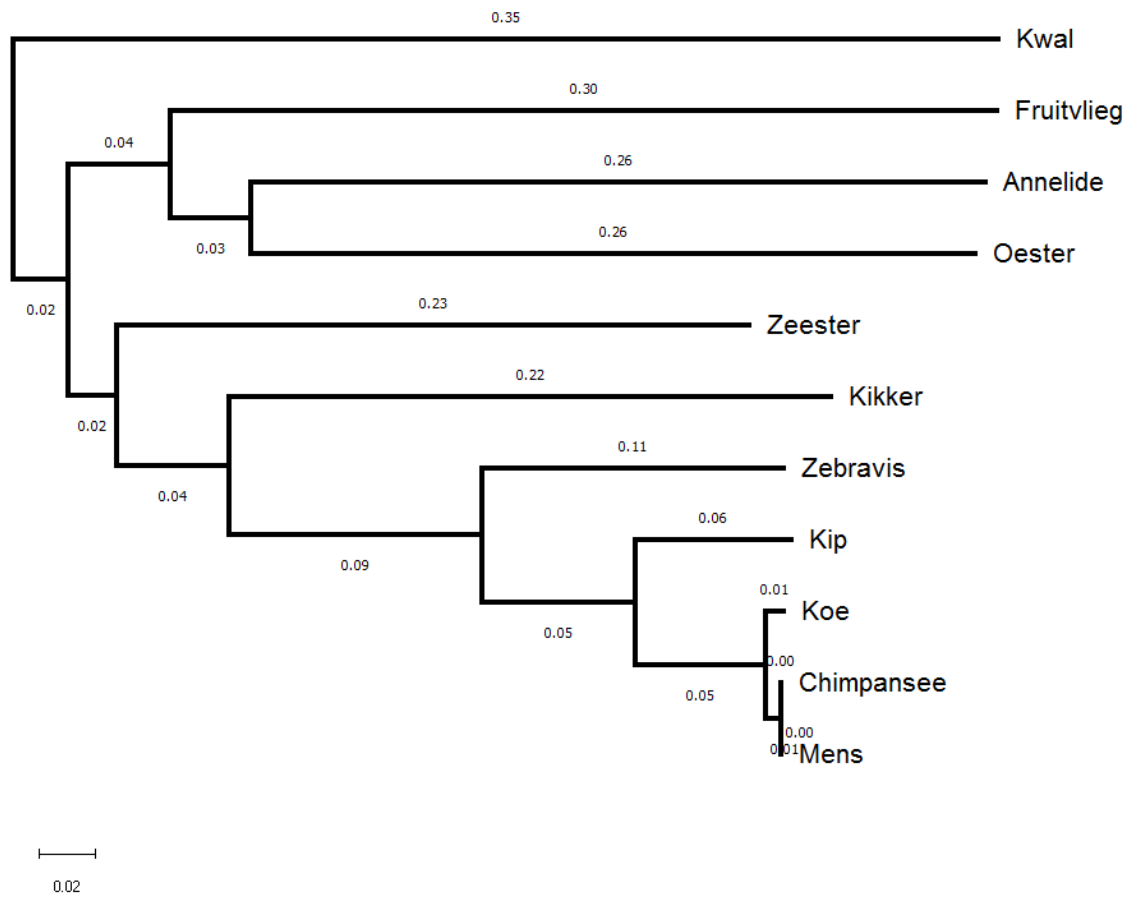
Dit is het resultaat. Ziet er een stuk beter uit, hoewel er wezenlijk niets veranderd is. Voor Pax6, moet je hier dus op de zee-anemoon klikken. Meteen geweldig resultaat. Je kunt nog meer doen:



Dat ziet er dan zo uit als hieronder. En je kunt nog meer:



Uiteindelijk ziet mijn boom er na wat gesleutel zo uit:



Je eigen bio-informatica-project!