

MAKING AN EVOLUTIONARY TREE

THEORY

The relationship between different species can be derived from different information sources. The connection between species may turn out by similarities in structure and/or embryonic development.

Another prove can be found in the molecular biology, in particular, in DNA. DNA molecules contain genetic information, this information says something about the characterizations of an organism.

The more organisms are related to each other, the more DNA matches. The opposite is also true, when two organisms are totally different less similarities in the DNA of those organisms can be found.

The same phenomenon takes place in the relationship within a species. The DNA of two sisters looks quote the same, while the DNA of two grandnieces are less similar. This comparison can also be made between different species. For example you can compare the DNA of a human, a rhesus monkey and a chimpanzee. The DNA of human is for 97.3% complementary to the DNA of chimpanzees, and for 93% to that of the rhesus monkey. The human is, when only looking to the DNA, apparently closer related to the chimpanzees that to the rhesus monkey.

By comparing the DNA of different species you can create a kind of family tree of life on earth. The tree gives the relationship between species. The less species are related, the earlier the 'branches' are separated.

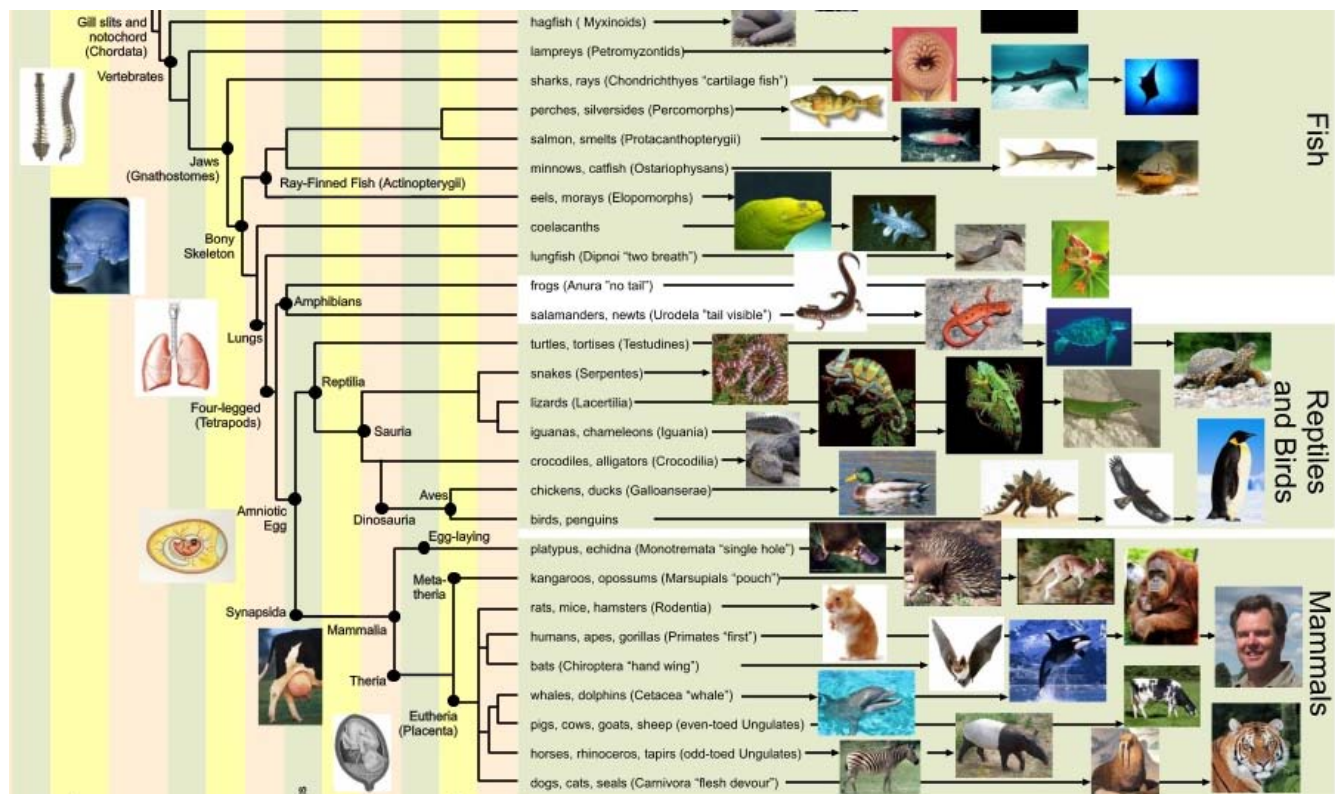


Figure 1: An evolutionary tree based on DNA analysis.

A family tree is based on information from DNA. DNA molecules contain four different nucleobases (adenine (A), guanine (G), Thymine (T) en cytosine (C)). The sequence of these nitrogenous bases forms a code, and this code is used by organisms to create many different proteins.

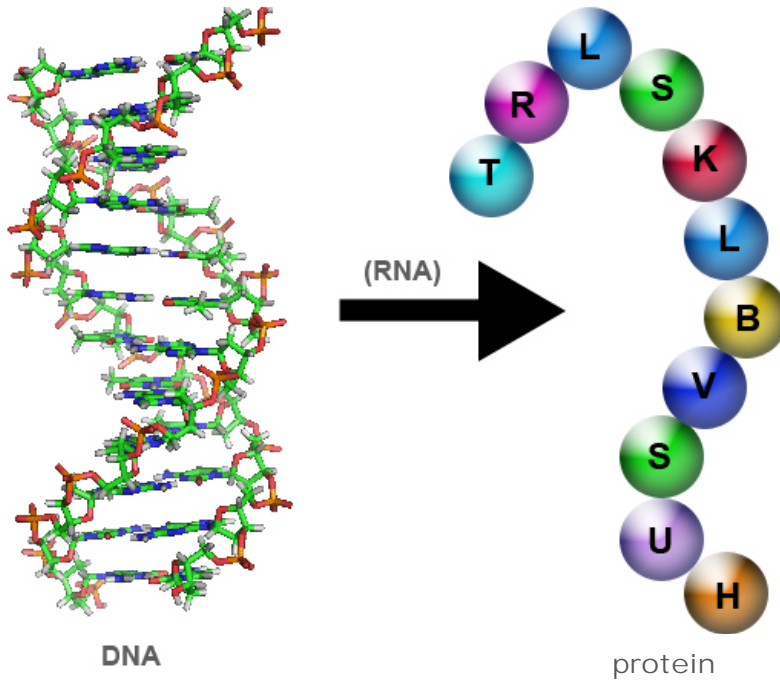


Figure 2: DNA gives information about the order of amino acids in a protein.

Proteins are built of building blocks. Such as a cord with different beads. Each bead is an amino acid. Humans have 20 different amino acids to build proteins. Each amino acid has a name which is often shortened to a single letter. This is called a single-letter code. Different proteins have different amino acid sequences which are represented as one-letter codes. These amino acid sequences can be compared with each other. In this lesson you will compare proteins that are present in different species and fulfill the same function, for example the code for hemoglobin in humans and chimpanzees. Remember, if the organisms are highly related to each other, their amino acid sequences, and thus, their DNA, will have little differences. Here the same story about relationships counts as for DNA. Only now you compare a protein code instead of a DNA code. Comparing proteins on the basis of their amino acid sequence is referred to as aligning. This means putting the amino acid codes of different proteins on top of each other and observing them for similarities or differences.

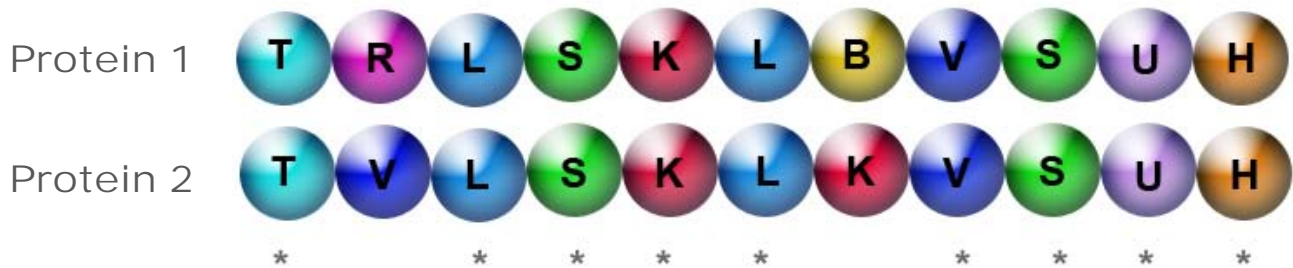
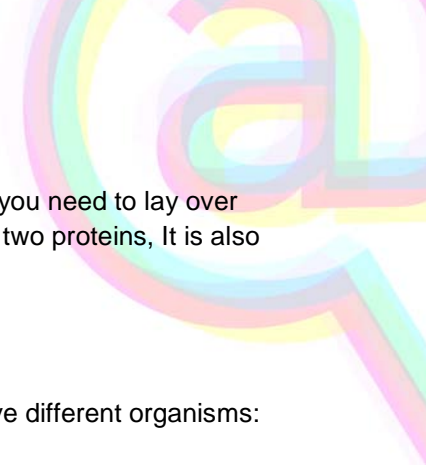


Figure 3: Aligning of two different proteins.



The proteins should be similar if you want to align, otherwise you do not know what parts you need to lay over each other. If you align the protein codes you can distinguish the differences between the two proteins, It is also possible to align multiple proteins at the same time.

Exercise 1:

Below you can see the amino acid sequences in one letter code of a small protein from five different organisms: yeast cell, fern, pig, monkey and human.

1. **VWERNRKLRSFNLSAIEKHG** - Fern
2. **VWERNRKLRSFNLTAEKHG** -
.....
3. **VWERNRKLKSFNLSALEKHG** -
.....
4. **VWEKNRKLKSFNLSALEKHG** -
.....
5. **VWERNRKLKSFNLSAIEKHG** -
.....

- A. Mark the differences between the amino acid sequences (for example with a marker)
- B. Which two proteins have the most differences in the amino acid sequence?

- C. Line the proteins in order of similarity. The proteins of exercise 1B are the beginning and the end of the set.

- D. Determine which protein belongs to which organism, the protein of fern is already given.

E. explain, that a human and yeast are the least related based on this protein.

This protein does not exist. However bioinformatics find in a similar manner similarities and differences between species.

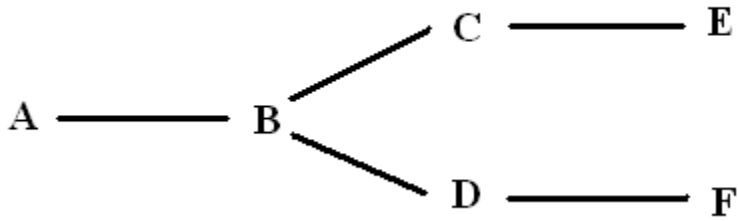
F. Name a protein that exists in humans and chimpanzees.

G. Name protein that is present in humans. But not in chimpanzees.

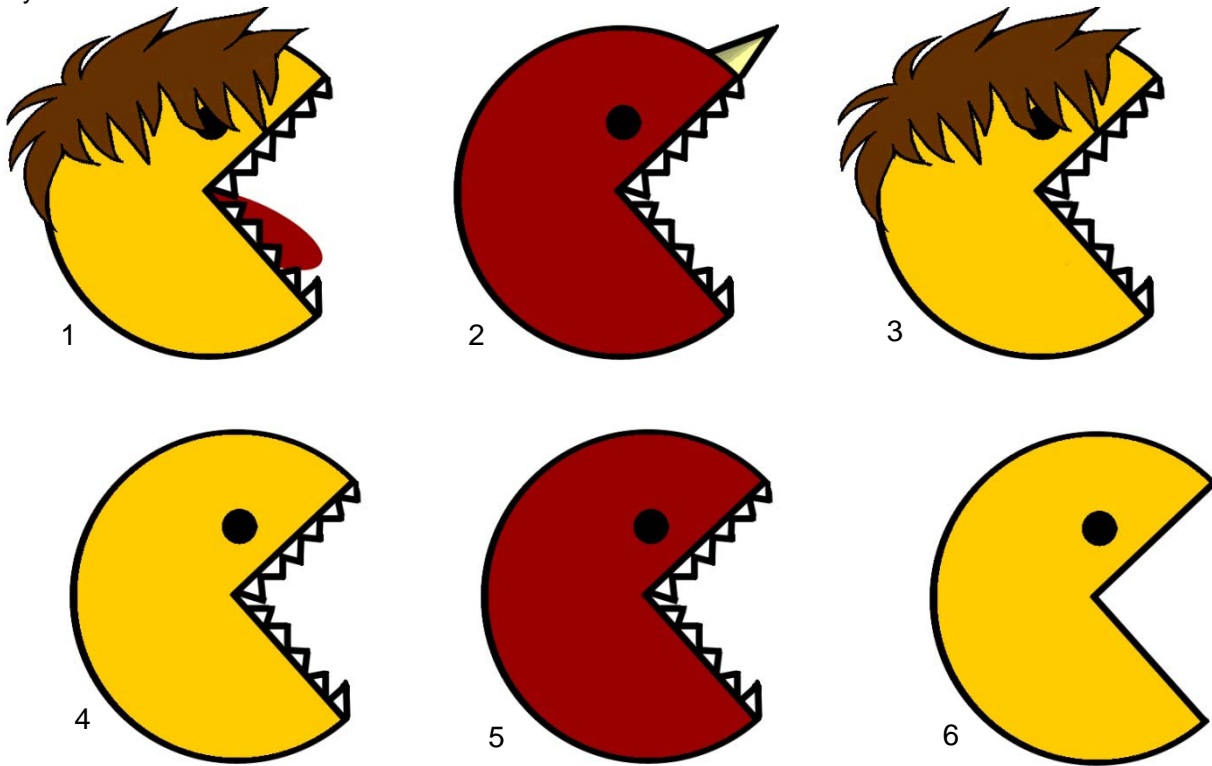
Only 2.7% of the DNA codes for differences in proteins between humans and chimpanzees.

exercise 2:


An evolutionary tree is never linear, as in exercise 1. A slightly more complex tree is seen below.



This tree illustrates the evolution of Pacman. The letters in this tree represent different subspecies of Pacman. They look like this:



Complete the table below. Therefore you first determine the differences in amino acid sequence of the proteins. Analyze which proteins have the most differences in amino acid sequence. Those proteins are the beginning and the end of the tree. Complete the evolutionary tree using the results of the table below.

Proteins (amino acid sequences)	Pacman subspecies	Place on the evolutionary tree
SRQPALSAACAEEKRRLS		
SRQPALSAACGEKQRLS		
SRQPALSAACVEKYRLS		
SRQPALSAACVEKRRLS		
SRQPALSAACGEKRRLS		
SRQPALSAACAEEKRRLS		c

THEORIE

The following four proteins are approximately equal. They represent the protein Acyl-peptide hydrolase, a protein that has a function in the digestion. It cleaves other proteins. The proteins are from four different organisms: bovine, human, pig and rat.

```
MERQVLLSEP EEAALYRGL SRQPALSAAC LGPEVTTQYG GRYRTVHTEW TQRDLER MEN IRFCRQYLVF HDGDSVVFAG PAGNSVETRG ELLSRESPPSG TMKAVLRKAG STGEEKQFLE
VWEKNRKLKS FNLSALEKHG PVEYEDDCFG CLSWSHSETHL LYVAEKKRPK AESEFFQTKAL DISGSDDEMA RPKKPDQAIAK GDQFLFYEDW GENMVKSGSP VLVCLVDIESG NISVLEGVPE
SVSPGQAFWA PGDTGVVVFAG WWHEPFRLLGI RFCTNRRSAL YYVDLTTGGNC ELLSDDSLAV TSPRLSPDQC RIVYLQPPSL VPHQQCGQLC LYDWYTRVTV VVVVVVPRQL GENFSGIYCS
LLPLGCSWAD SQRVVFDTAQ RSRQDLFAVD TQMGTVTPLT AGGSGGSKWL LTIDRDLMVA QFSTPNLPPC LKVGFLPPAG MEQEVVWVSL EEAEPDPDIS WSIRVLQPPP EQEHAQYVGL
DFEAILIQPS NPPDKTQVPM VVMPHGGPHS SFVTSWMLLP AMLCKMGFAA LLVNYRGSTG FQGDSILSLP GNVGSQDVKD VQFAVEQVQL EEFHDAGRVA LLGGSHGGFL SCHLIGQYPE
TYGACVVRNP VINIASMMGS TDIPDWCVE AGYLYSSDCL PDPNVWSEML NKSPYKYPQ VKTPLVLLMLG QEDRRVPPKQ GMEYRALKR RNVFVRLLLY PKSTHSLSEV EVESDSFMNA
VIWMCTHLGH
```

```
MERQVLLSEP EEAALYRGL SRQPALSAAC LGPEVTTQYG GQYRTVHTEW TQRDLER MEN IRFCRQYLVF HDGDSVVFAG PAGNSVETRG ELLSRESPPSG TMKAVLRKAG GTGPGEKQF
LEVWEKNRKL KSFNLSALEK HGPVYEDDCF GCLSWSHSET HLLYVAEKKR PKAESFFQTK ALDVSASDDE IARLKKPDQA IKGDQFVFE DWGENMVKSP IPVLCVLDVE SGNISVLEGV
PENVSPPGQAF WAPGDAGVVF VGMWHEPFRLL GIRFCTNRRS ALYYVDLIGG KCELLSDDSL AVSSPRLSPD QCRIVYLQYP SLIPHHQCSQ LCLYDWYTKV TSVVVDVPR QLGENFSGIY
CSLLPLGCSW ADSQRVVFDS AQRSRQDLFA VDTQVGTVTS LTAGSGGSGW KLLTIDQDLM VAQFSTPSILP PTLKVGHFLPS AGKEQSVLWV SLEEAEPDPD IHWGIRVLQP PPEQENVQYA
GLDFEAILLQ PGSPDPKTQV PMVMPHGGP HSSFVTAAML FPAMLCKMGF AVLLVNYRGS TGFQDSILS LPGNVGHQDV KDVQFAVEQV LQEEHFDASH VALMGGSHGG FISCHLIGQY
PETYRACVAR NPVINIASML GSTDIPDWCV VEAGFPFSSD CLPDLVSWAE MLDKSPIRYI PQVKTPLLLM LGQEDRRVVF KQGMYYRAL KTRNVPRLL LYPKSTHALS EVEVESDSFM
NAVLWLRTHL GS
```

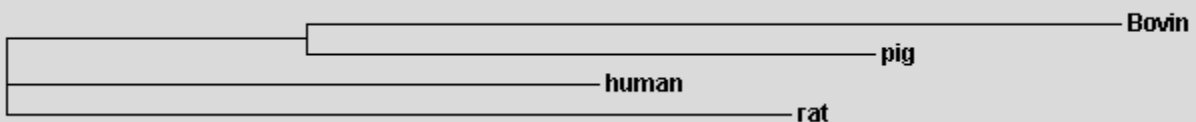
```
MERQVLLSEP EEAALYRGL SRQPALSAAC LGPEVTTQYG GRYRTVHTEW TQRDLER MEN IRFCRQYLVF HDGDSVVFAG PAGNSVETRG ELLSRESPPSG TMKAVLRKAG GTGTAEKQF
LEVWEKNRKL KSFNLSALEK HGPVYEDDCF GCLSWSHSET HLLYVAEKKR PKAESFFQTK ALDVTGSDDA MARTKPPDQA IKGDQFLFYE DWGENMVKSP TPVLCVLDIE SGNISVLEGV
PEVSPPGQAF WAPGDAGVVF VGMWHEPFRLL GIRFCTNRRS ALYYVDLIGG KCELLSDES AVTSPRLSPD QCRIVYLQYP SLVPHQQCGQ LCLYDWYTRV TSVVVDIVPR QLGEDFSGIY
CSLLPLGCSW ADSQRVVFDS PQRSRQDLFA VDTQMGVSTS LTAGSGGSGW KLLTIDRDLM VVQFSTPSVP PSLKVGFLPP AGKEQAVSWV SLEEAEPDPD ISWSIRVLQP PPQEQHVQYA
GLDFEAILLQ PSNSPEKTQV PMVMPHGGP HSSFVTAAML FPAMLCKMGF AVLLVNYRGS TGFQDSILS LPGNVGHQDV KDVQFAVEQV LQEEHFDAGR VALMGGSHGG FLSCHLIGQY
PETYSACVVR NPVINIASMM GSTDIPDWCV VEAGFSYSSD CLPDLVSWAA MLDKSPIKYA PQVKTPLLLM LGQEDRRVVF KQGMYYRVL KARNVPRLL LYPKSTHALS EVEVESDSFM
NAVLWLCTHL GS
```

```
MERQVLLSEP QEAAALYRGL SRQPSLSAAC LGPEVTTQYG GLYRTVHTEW TQRDLER MEN IRFCRQYLVF HDGDSVVFAG PAGNSVETRG ELLSRESPPSG TMKAVLRKAG GTVSGEEKQF
LEVWEKNRKL KSFNLSALEK HGPVYEDDCF GCLSWSHSET HLLYVAEKKR PKAESFFQTK ALDISASDDE MARPKPDQA IKGDQFVFE DWGETMVKSP IPVLCVLDID SGNISVLEGV
PENVSPPGQAF WAPGDAGVVF VGMWHEPFRLL GIRYCTNRRS ALYYVDLGGG KCELLSDGSL AICSPRLSPD QCRIVYLQYP CLAPHHQQCSQ LCLYDWYTKV TSVVVDIVPR QLGEFSGIY
CSLLPLGCSW ADSQRVVFDS AQRSRQDLFA VDTQTSITS LTAAGSAGSW KLLTIDKDLM VAQFSTPSLP PSLKVGFLPP PGKEQSVSWV SLEEAEPDPD IHWGVRVLRH PPDQENVQYA
DLDFEAILLQ PSNPPDKTQV PMVMPHGGP HSSFVTAAML FPAMLCKMGF AVLLVNYRGS TGFQDSILS LPGNVGHQDV KDVQFAVEQV LQEEHFDARR VALMGGSHGG FLSCHLIGQY
PETYSACIAR NPVINIASMM GSTDIPDWCV VETGFPYSNS CLPDLNVWEE MLDKSPIKYI PQVKTPVLLM LGQEDRRVVF KQGMYYRAL KARNVPRLL LYPKSNHALS EVEAESDSFM
NAVLWLHTHL GS
```

Alignment of these proteins is a bit harder! On average, proteins have a length of 300 amino acids. Thankfully, computer programs aid in aligning large proteins. When such a program aligns the four proteins shown above, you will get the following result (only a small part is shown).

QYPSLIPHHQCSQLCLYDWYTKVTSVVVDVVPRLGGENFSGIYCSLLPLGCSWADSQRVVFDSAQRSRQD	human
QYPCLAPHHHQCSQLCLYDWYTKVTSVVVDIVPRQLGESFSGIYCSLLPLGCSWADSQRVVFDSAQRSRQD	Rat
QFPSLVPHQQCGQLCLYDWYTRVTVVVVDVVPRLGGENFSGIYCSLLPLGCSWADSQRVVFDTAQRSRQD	bovine
RFPSLVPHQQCGQLCLYDWYTRVTSVVVDIVPRQLGEDFSGIYCSLLPLGCSWADSQRVVFDPQRSRQD	pig

The colors make it easier to identify differences. These computer programs are also helpful for creating evolutionary pedigrees. For the four proteins above, the pedigree is as follows:



Summary

We can tell something about the evolutionary process by comparing amino acid sequences. Molecular Biology aids in creating evolutionary pedigrees, which provide more information and knowledge about evolution. Proteins of different species which are not related will show great differences in their amino acid sequences. The opposite is also true: closely related proteins will have almost the same amino acid sequences.

Appendix 2 – Answer model

Exercise 1:

Below you can see the amino acid sequences in one letter code of a small protein from five different organisms: yeast cell, fern, pig, monkey and human.

1. **VWERNRKLRSFNLSAIEKHG** - fern
2. **VWERNRKLRSFNLTATIEKHG** - yeast cell
3. **VWERNRKLKSFNLSALEKHG** - monkey
4. **VWEKNRKLKSFNLSALEKHG** - human
5. **VWERNRKLKSFNLSAIEKHG** - pig

- A. Mark the differences between the amino acid sequences (for example with a marker)
- B. Which two proteins have the most differences in the amino acid sequence?

Protein 2 and protein 4 have the most differences, 4 differences in total.

- C. Line the proteins in order of similarity. The proteins of exercise 1B are the beginning and the end of the set.

The order of similarity is 2, 1, 5, 3 en 4.
Of course, the order can also be reversed.

- D. Determine which protein belongs to which organism, the protein of fern is already given.

E. explain, that a human and yeast are the least related based on this protein.

The amino acid sequences between the yeast and humans have relatively the most differences from each other.

This protein does not exist. However bioinformatics find in a similar manner similarities and differences between species.

F. Name a protein that exists in humans and chimpanzees.

Chimpanzees and humans share many proteins. Examples are hemoglobin or amylase.

G. Name protein that is present in humans. But not in chimpanzees.

Real examples of proteins available in humans, but not in chimpanzees, are hard to find. Many human proteins are slightly changed when compared to their chimpanzee counterparts. Therefore, their biological function can be also be slightly changed. That's why humans can talk, thanks to a mutation in the FOX2-gene. Apparently, chimpanzees are also immune to HIV

Only 2.7% of the DNA codes for differences in proteins between humans and chimpanzees.

Background:

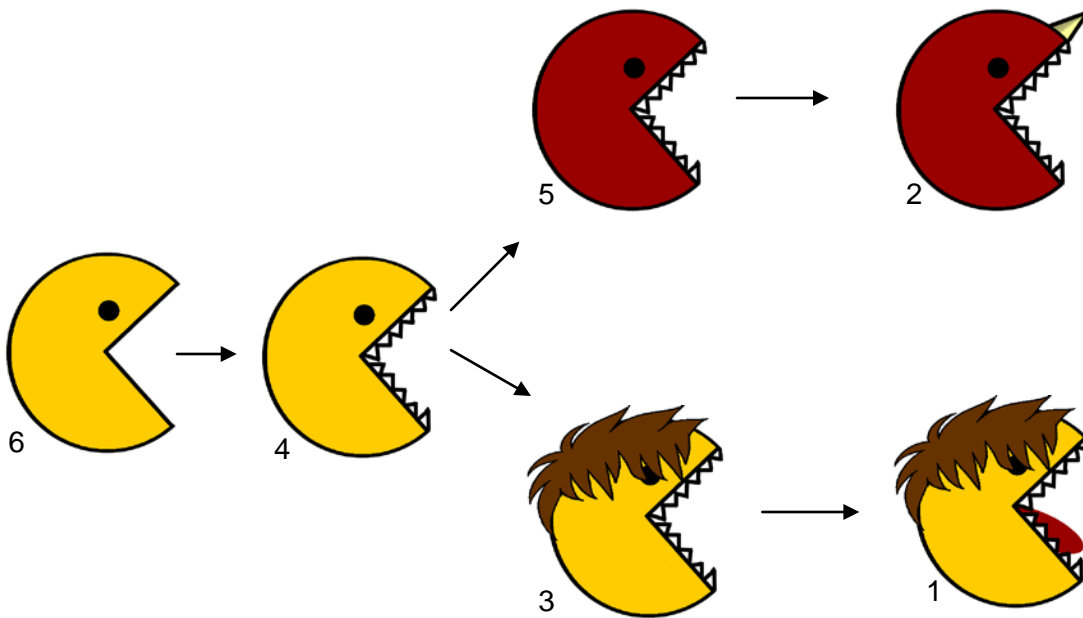
Researchers have found that 29% of the genes found in humans and chimpanzees code for exactly the same proteins. The rest of the genes are only slightly different. Genes of humans and chimpanzees coding for slightly different proteins only have 2 different amino acids on average in their amino acid sequences.

Exercise 2:

Complete the table below. Therefore you first determine the differences in amino acid sequence of the proteins. Analyze which proteins have the most differences in amino acid sequence. Those proteins are the beginning and the end of the tree. Complete the evolutionary tree using the results of the table below.

Proteins	Pacman subspecies	Place on the evolutionary tree
SRQPALSAACA AE KKRLS	2	E
SRQPALSAAC GE K R RLS	1	F
SRQPALSAAC VE K Y RLS	6	A
SRQPALSAAC VE K R RLS	4	B
SRQPALSAAC GE K R RLS	3	D
SRQPALSAACA AE K R RLS	5	C

The evolutionary tree looks as follows:
Of course, C+E can be interchanged with D+F



Appendix 3 – Amino acid sequences

> bovi n

```
MERQVLLSEP EEAAALYRGL SRQPALSAAC LGPEVTTQYG GRYRTVHTEW TQRDLERMEN
IRFCRQYLVF HDGDSVVFAG PAGNSVETRG ELLSRESPSG TMKAVLRKAG STGEEKQFLE
VWEKNRKLKS FNLSALEKHG PVYEDDCFCG LSWSHSETHL LYVAEKRRPK AESFFQTKAL
DISGSDDEMA RPKKPDQAI K GDQFLFYEDW GENMVSCKGSP VLCVLDI ESG NISVLEGVPE
SVSPGQAFWA PGDTGVVFAG WWHEPFRLGI RFCTNRRSAL YYVDLTGGNC ELLSDDSLAV
TSPRLSPDQC RIVYLQFPSP VPHQQCGQLC LYDWYTRVTV VVVVVVPRQL GENFSGI YCS
LLPLGCWSAD SQRVVFDTAQ RSRQDLFAVD TQMGTVTPLT AGGSGGSKWL LTI DRDLMVA
QFSTPNLPPC LKVGFLPPAG MEQEVVWVSL EEAPEI PDI S WSI RVLQPPP EQEHAQYVGL
DFEAILIQPS NPPDKTQVPM VVMPHGGPHS SFVTSWMLLP AMLCKMGFAA LLVNYRGSTG
FGQDSLISLP GNVGSQDVKD VQFAVEQVLQ EEHFDAGRVA LLGGSHGGFL SCHLIGQYPE
TYGACVVRNP VINIASMMGS TDI PDWCVVE AGYLYSSDCL PDPNVWSEML NKSPI KYTPQ
VKTPVLLMLG QEDRRVPFKQ GMEYYRALKA RNVPVRLLY PKSTHSLSEV EVESDSFMNA
VIWMCTHLGH
```

> human

```
MERQVLLSEP EEAAALYRGL SRQPALSAAC LGPEVTTQYG GQYRTVHTEW TQRDLERMEN
IRFCRQYLVF HDGDSVVFAG PAGNSVETRG ELLSRESPSG TMKAVLRKAG GTGPGEKQF
LEVWEKNRKL KSFNLSALEK HGPVYEDDCF GCLSWHSET HLLYVAEKRR PKAESFFQTK
ALDVSASDDE IARLKKPDQA IKGDFVDFYE DWGENMVSKS IPVLCVLDVE SGNISVLEGV
PENVSPPGQAF WAPGDAGVVF VGWWHEPFRL GIRFCTNRRS ALYYVDLIGG KCELLSDDSL
AVSSPRLSPD QCRI VYLQYP SLI PHHQCSQ LCLYDWYTKV TSVVVVVVPR QLGENFSGI Y
CSLLPLGCWS ADSQRVVFDS AQRSRQDLFA VDTQVGTVTS LTAGGSGGWS KLLTI DQDLM
VAQFSTPSLP PTLKVGFLPS AGKEQSVLWV SLEEAPEI PD IHWGI RVLQP PPEQENVQYA
GLDFEAILLQ PGSPPDKTQV PMVMPHGGP HSSFVTAWML FPAMLCCKMGF AVLLVNYRGS
TGFGQDSLIS LPGNVGHQDV KDQFAVEQV LQEEHFDASH VALMGGSHGG FISCHLIGQY
PETYRACVAR NPVINIASML GSTDI PDWCV VEAGFPFSSD CLPDLVWAE MLDKSPI RYI
PQVKTPLLLM LGQEDRRVPF KQGMEYYRAL KTRNVPVRL LYPKSTHALS EVEVESDSFM
NAVLWLRTHL GS
```

> pi g

```
MERQVLLSEP EEAAALYRGL SRQPALSAAC LGPEVTTQYG GRYRTVHTEW TQRDLERMEN
IRFCRQYLVF HDGDSVVFAG PAGNSVETRG ELLSRESPSG TMKAVLRKAG GTGTAEKQF
LEVWEKNRKL KSFNLSALEK HGPVYEDDCF GCLSWHSET HLLYVADKKR PKAESFFQTK
ALDVTGSDDDE MARTKKPDQA IKGDFVDFYE DWGENMVSKS TPVLCVLDIE SGNISVLEGV
PESVSPGQAF WAPGDAGVVF VGWWHEPFRL GIRFCTNRRS ALYYVDLTGG KCELLSDES
AVTSPRLSPD QCRI VYLRFPSL VPHQQCGQ LCLYDWYTRV TSVVVVDI VPR QLGEDFSGI Y
CSLLPLGCWS ADSQRVVFDS PQRSRQDLFA VDTQMGVTS LTAGGSGGWS KLLTI DRDLM
VVQFSTPSVP PSLKVGFLPP AGKEQAVSWV SLEEAEPFPD ISWSI RVLQP PPQEHVQYA
GLDFEAILLQ PSNPEKPTQV PMVMPHGGP HSSFVTAWML FPAMLCCKMGF AVLLVNYRGS
TGFGQDSLIS LPGNVGHQDV KDQFAVEQV LQEEHFDAGR VALMGGSHGG FLSCHLIGQY
PETYSACVVR NPVINIASMM GSTDI PDWCM VEAGFSYSSD CLPDLVWAA MLDKSPI KYA
PQVKTPLLLM LGQEDRRVPF KQGMEYYRVL KARNVPVRL LYPKSTHALS EVEVESDSFM
NAVLWLCTHL GS
```

> rat

```
MERQVLLSEP QEAAALYRGL SRQPSLSAAC LGPEVTTQYG GLYRTVHTEW TQRDLERMEN
IRFCRQYLVF HDGDSVVFAG PAGNSVETRG ELLSRESPSG TMKAVLRKAG GTVSGEEKQF
LEVWEKNRKL KSFNLSALEK HGPVYEDDCF GCLSWHSET HLLYVAEKRR PKAESFFQTK
ALDISASDDE MARPKKPDQA IKGDFVDFYE DWGETMVSKS IPVLCVLDID SGNISVLEGV
PENVSPPGQAF WAPGDAGVVF VGWWHEPFRL GIRYCTNRRS ALYYVDLSGG KCELLSDGSL
AICSPRLSPD QCRI VYLQYP CLAPHHQCSQ LCLYDWYTKV TSVVVVDI VPR QLGESFSGI Y
CSLLPLGCWS ADSQRVVFDS AQRSRQDLFA VDTQTGSI TS LTAAGSAGSW KLLTI DKDLM
VAQFSTPSLP PSLKVGFLPP PGKEQSVSWV SLEEAPEI PG IHWGVRVLP PPDQENVQYA
DLDFEAILLQ PSNPPDKTQV PMVMPHGGP HSSFVTAWML FPAMLCCKMGF AVLLVNYRGS
TGFGQDSLIS LPGNVGHQDV KDQFAVEQV LQEEHFDARR VALMGGSHGG FLSCHLIGQY
PETYSACIAR NPVINIASMM GSTDI PDWCM VETGFPYSNS CLPDLNVWEE MLDKSPI KYI
PQVKTPVLLM LGQEDRRVPF KQGMEYYRAL KARNVPVRL LYPKSNHALS EVEAESDSFM
NAVLWLHTHL GS
```

