

Svar på Multiple Alignment øvelse

Af: Rasmus Wernersson (v18013)

Spørgsmål 1

Fasta fil:

```
>pigeon_alpha-globin-D
ATGCTGACCGACTCTGACAAGAAGCTGGTCTGCAGGTGTGGGAGAAGGTGATCCGCCACCCAGACTGTG
GAGCCGAGGCCCTGGAGAGGCTGTTACACCTACCCCCAGACCAAGACCTACTTCCCCACTTCGACTT
GCACCATGGCTCCGACAGGTCGCAACACGGAAGAAGGTGTTGGCCGCTTGGGCAACGCTGTCAAG
AGCCTGGGCAACCTCAGCCAAGCCCTGTCTGACCTCAGCGACCTGCATGCCTACAACCTGCGTGTGACC
CTGTCAACTTCAAGCTGCTGGCGAGTGTTCACGTTGGTGTGGCCACACACCTGGGCAACGACTACAC
CCCGGAGGCATGCTGCCTTCGACAAGTTCCTGTCGGCTGTGTGACCGTGTGGCCGAGAAGTACAGA
TAA

>pigeon_alpha-globin-A
ATGGTGTCTGTGCCAACGACAAGAGCAACGTGAAGGCCGTCTTCGGCAAAATCGGCGGCCAGGCCGGT
ACTTGGTGGTGAAGCCCTGGAGAGGTTGTTTCATCACCTACCCCCAGACCAAGACCTACTTCCCCACTT
CGACTGTACATGGCTCCGCTCAGATCAAGGGGCACGGCAAGAAGGTGGCGGAGGCACTGTTGAGGCT
GCCAACCATCGATGACATGCTGCTGGCTCCCTTCCAAGCTGAGCGACCTCCACGCCCAAAAGCTCCGTG
TGGACCCCGTCAACTTCAAATGCTGGTCACTGCTTCTGGTGGTGTGGCCGCTCCACTTCCCCTCTCT
CCTGACCCCGGAGGTCATGCTTCCCTGGACAAGTTCGTGTGTGCCGTGGGCACCGTCTTACTGCCAAG
TACCGTTAA

>duck_alpha-globin-D
ATGCTGACCGCCGAGGACAAGAAGCTCATCGTGCAGGTGTGGGAGAAGGTGGCTGGCCACCAGGAGGAAT
TCGGAAGTGAAGCTCTGCAGAGGATGTTCTCGCCTACCCCCAGACCAAGACCTACTTCCCCACTTCGA
CCTGCATCCCGGCTCTGAACAGGTCGCTGGCCATGGCAAGAAGGTGGCGGCTGCCCTGGGCAATGCCGTG
AAGAGCCTGGACAACCTCAGCCAGGCCCTGTCTGAGCTCAGCAACCTGCATGCCTACAACCTGCGTGTG
ACCCTGTCAACTTCAAGCTGCTGGCAGAGTGTTCAGGTTGGTGTGGCCGCACACCTGGGCAAGACTA
CAGCCCGAGATGCATGCTGCCTTTGACAAGTTCCTGTCGCCGTGGCTGCCGTGCTGCTGAAAAGTAC
AGATGA

>duck_alpha-A-globin
ATGGTGTCTGTGCGGCTGACAAGACCAACGTCAAGGGTGTCTTCTCAAATCGGTGGCCATGCTGAGG
AGTATGGCGCCGAGACCTGGAGAGGATGTTTCATCGCCTACCCCCAGACCAAGACCTACTTCCCCACTT
TGACCTGCAGCAGGCTCTGCTCAGATCAAGGCCATGGCAAGAAGGTGGCGGCTGCCCTGATTGAAGCT
GTCAACCACATCGATGACATTGCGGGTGTCTCTCAAAGCTCAGTACCTCCACGCCCAAAAGCTCCGTG
TGGACCCGTCAACTTCAAATCCTGGGCCACTGCTTCTGGTGGTGGTGGCCATCCACCACCCCGCTGC
CCTGACCCAGAGGTCACGCTTCCCTGGACAAGTTCATGTGCCGTGGGTGCTGTGCTGACTGCCAAG
TACCGTTAG

>Goat_alpha-i-globin
ATGGTGTCTGTGCGCCGACAAGTCCAATGTCAAGGCCGCTGGGGCAAGGTTGGCGCAACGCTGGAG
CTTATGGCGCAGAGGCTCTGGAGAGGATGTTCTGAGCTTCCCCACCACCAAGACCTACTTCCCCACTT
CGACCTGAGCCACGGCTCGGCCAGGTCAAGGGCCACGGCGAGAAGGTGGCCGCGCTGACCAAAGCG
GTGGGCCACCTGGACGACCTGCCCGTACTCTGTCTGATCTGAGTACCTGCACGCCACAAGCTGCGTG
TGGACCCGTCAACTTAAAGTCTGAGCACTCCCTGCTGGTACCTGGCCTGCCACCTCCCAATGA
TTTACCCCCGCGGTCCACGCTCCCTGGACAAGTTCCTGGCAACGTGAGCACCGTGTGACTCCAAA
TACCGTTAA

>Goat_alpha-ii-globin
ATGGTGTCTGTGCGCCGACAAGTCCAATGTCAAGGCCGCTGGGGCAAGGTTGGCAGCAACGCTGGAG
CTTATGGCGCAGAGGCTCTGGAGAGGATGTTCTGAGCTTCCCCACCACCAAGACCTACTTCCCCACTT
CGACCTGAGCCACGGCTCGGCCAGGTCAAGGGCCACGGCGAGAAGGTGGCCGCGCTGACCAAAGCG
GTGGGCCACCTGGACGACCTGCCCGTACTCTGTCTGATCTGAGTACCTGCACGCCACAAGCTGCGTG
TGGACCCGTCAACTTAAAGTCTGAGCACTCCCTGCTGGTACCTGGCCTGCCACCTCCCAAGTGA
TTTACCCCCGCGGTCCACGCTCCCTGGACAAGTTCCTGGCAACGTGAGCACCGTGTGACTCCAAA
TACCGTTAA

>Horse_alpha-1_globin
ATGGTGTCTGTGCGCCGACAAGACCAACGTCAAGGCCGCTGGAGTAAGGTTGGCGGCCACGCTGGCG
AGTTTGGCGCAGAGGCTCTGAGAGGATGTTCTGGGCTTCCCCACCACCAAGACCTACTTCCCCACTT
CGATCTGAGCCACGGCTCGGCCAGGTCAAGGGCCACGGCAAGAAGGTGGGCGACGCGTACTCTCGCC
GTGGGCCACCTGGACGACCTGCCTGGCGCCTGTGCAATCTGAGCGACCTGCACGCACAAGCTGCGCG
TGGACCCGTCAACTTAAAGTCTGAGTCAATGCTGCTGTCACCTTGGCCGTCCACTCCCAACGA
TTTACCCCTGCGGTCCACGCTCCCTGGACAAGTTCCTGAGCAGTGTGAGCACCGTGTGACTCCAAA
TACCGTTAA

>Horse_alpha-2_globin
ATGGTGTCTGTGCGCCGACAAGACCAACGTCAAGGCCGCTGGAGTAAGGTTGGCGGCCACGCTGGCG
```

```
AGTATGGCGCAGAGGCCCTAGAGAGGATGTTCTGGGCTTCCCACCACCAAGACCTACTTCCCCACTT
CGATCTGAGCCACGGCTCCGCCAGGTCAAGGCCACGGCCAGAAGGTGGGGCAGCGGCTGACTCTCGCC
GTGGGCCACTGGACGACCTGCCTGGCGCCCTGTGCAATCTGAGCGACCTGCACGCACACAAGCTGCGCG
TGGACCCCGTCAACTTCAAGCTCCTGAGTCATTGCTGCTGCCACCTTGGCCGTCCACCTCCCAACGA
TTTACCCCTGCCGTCCACGCCTCCCTGGACAAGTTCTTGAGCAGTGTGAGCACCGTGTGACCTCCAAA
TACCGTTAA
```

>Chicken_alpha-D

```
ATGCTGACTGCCGAGGACAAGAAGCTCATCCAGCAGGCTGGGAGAGGGCCGCTTCCCACCAGGAGGAGT
TTGGAGCTGAGGCTCTGACTAGGATGTTACCACCTATCCCCAGACCAAGACCTACTTCCCCACTTTCGA
CCTTTCGCCTGGCTCTGACCAGGTCCTGGCCATGGCAAGAAGGTGTTGGGTGCCCTGGGCAACGCCGTG
AAGAACGTGGACAACCTCAGCCAGGCCATGGCTGAGCTGAGCAACCTGCATGCCTACAACCTGCGTGTTG
ACCCCGTCAATTTCAAGCTGTTGTCCAGTGCATCCAGGTGGTGTGGCTGTACACATGGGCAAAGACTA
CACCCCTGAAGTGCATGCTGCCCTTCGACAAGTTCTGTCTGCCGTGTCTGCTGTGCTGAGAAGTAC
AGATAA
```

>Chicken_alpha-A

```
ATGGTGTGTCCTGCTGACAAAGAACAACGTCAGGGCATCTTACCAAAATCGCCGGCCATGCTGAGG
AGTATGGCGCCGAGACCTGGAAAGGATGTTACCACCTACCCCAACCAAGACCTACTTCCCCACTT
CGATCTGTACACGGCTCCGCTCAGATCAAGGGGCACGGCAAGAAGGTAGTGGCTGCCTTGATCGAGGCT
GCCAACCATTTGATGACATCGCCGGCACCTCTCAAGCTCAGCGACCTCCATGCCACAAGCTCCGCG
TGGACCTGTCAACTTCAAACCTCTGGGCAATGCTTCTGGTGGTGGTGGCCATCCACCACCTGCTGC
CCTGACCCCGGAGGTCATGCTTCCCTGGACAAGTTCTGTGCGCCGTGGGCACTGTGCTGACCCCAAG
TACCGTTAA
```

BEMÆRK:

*) Det er vigtigt at bruge KORTE, beskrivende navne. Der var flere af jer der løb ind i det problem at ClustalW kun kigger på de første 30 tegn i navnene, og at der derfor gik kludder i meget lange navne (SE OGSÅ FASTA HANDOUTET FRA LEKTION 2).

*) Vær opmærksom på at i GenBank entries der indeholder flere gener (SE GENBANK HANDOUT'ET FRA LEKTION 2) findes navnet på det enkelte gen (CDS) nede i selve feature-tabellen. Når i klikker på et CDS der har "/gene_name=XYZ" eller lignende, er det derfor XYZ I skal bruge som navn i jeres FASTA fil og ikke den samlede titel for HELE GenBank entry'en (fx. "Alpha-A and Alpha-D genes ..." eller "Yeast Chromosome 2"). Jeg uploader et screenshot med en ekstra forklaring på CampusNet / Lektionsplanen.

*) Det sidste GenBank entry ("AF098919" - kylling) indeholder tre gener: "embryonic alpha-type globin pi", "adult alpha D globin" og "adult alpha A globin". Jeg har her valgt kun at tage de to sidste med, da den første kun står beskrevet som "alpha-type". Man kan godt tage "embryonic alpha-type globin pi" med, for at være påpasselig med at ikke smide for meget væk - man vil så se, at den læggers sig for sig selv i det afstands-træ ClustalW producerer. Dette er en god indikator på at den er noget anderledes. Man kan så evt. gå tilbage og fjerne den, eller skrive en bemærkning om at den ligger for sig selv. Når man bygger et "rigtigt" datasæt til et forskningsprojekt, er dette ofte en iterativ proces, hvor man 1) samler sine sekvenser, 2) luger ud i dem, 3) kører en analyse og gentager 2) og 3) indtil man er tilfreds med resultatet.

Spørgsmål 2

) "" betyder at baserne er helt ens i en given position (perfekt konserveret).

*) Hvis man ikke har "alpha-type" sekvensen med er der et enkelt område på mere end 12 baser, der er perfekt konserveret. Hvis man har tager "alpha-type" sekvensen med, er det kun 11 baser i dette område, der er perfekt konserveret.

*) Ang. "guide tree"

*1) 3 clusters (+ en "løs" gruppe, hvis man har "alpha-type" sekvensen med): En Alpha-A (kun fugle), en Alpha-D (kun fugle) og en Alpha 1 + Alpha 2 cluser (pattedyr).

*2) Ideen er her at fugle og pattedyr ikke blandes - så de ligger naturligt taxonomisk set.

*3) Alpha-A og Alpha-D ligger tydeligvis i hver sin cluster - det må nødvendigvis betyde at splittet mellem dem er gammelt. Da både alpha-a og alpha-d findes hos de tre fugle vi har med, må splittet mellem dem være ældre end den sidste fælles forfader til fuglene.

*4) Alpha-1 og Alpha-2 ser ud til at være væsentligt tættere beslægtede - husk at et guide-tree kun er et rimeligt råt estimat på slægtskabet, så hvis vi skal grave dybere i hvornår alpha-1 og

CCGGCCAGGCTTCGTGAACCAGCACCTGTGCGGCTCCACCTGGTGGAGGCGCTGTAC
CTGGTGTGCGGGAGCGCGGCTTCTTCTACACGCCAAGGCCCGTCGGGAGGCGGAGAAC
CCTCAGGCAGGTGCCGTGGAGCTGGGCGGAGGCTGGGCGGCTGCAGGCCCTGGCGCTG
GAGGGGCCCCCGAGAAGCGTGGCATCGTGGAGCAGTGTGCACCAGCATCTGTTCCCTC
TACCAGCTGGAGAACTACTGCAACTAG

>Pig_AY242098

ATGGCCCTGTGGACGCGCCTCTGCCCTGCTGGCCCTGCTGGCCCTCTGGGCGCCGCC
CCGGCCAGGCTTCGTGAACCAGCACCTGTGCGGCTCCACCTGGTGGAGGCGCTGTAC
CTGGTGTGCGGGAGCGCGGCTTCTTCTACACGCCAAGGCCCGTCGGGAGGCGGAGAAC
CCTCAGGCAGGTGCCGTGGAGCTGGGCGGAGGCTGGGCGGCTGCAGGCCCTGGCGCTG
GAGGGGCCCCCGAGAAGCGTGGCATCGTGGAGCAGTGTGCACCAGCATCTGTTCCCTC
TACCAGCTGGAGAACTACTGCAACTAG

>Pig_AY242100

ATGGCCCTGTGGACGCGCCTCTGCCCTGCTGGCCCTGCTGGCGCTCTGGGCGCCGCC
CCGGCCAGGCTTCGTGAACCAGCACCTGTGCGGCTCCACCTGGTGGAGGCGCTGTAC
CTGGTGTGCGGGAGCGCGGCTTCTTCTACACGCCAAGGCCCGTCGGGAGGCGGAGAAC
CCTCAGGCAGGTGCCGTGGAGCTGGGCGGAGGCTGGGCGGCTGCAGGCCCTGGCGCTG
GAGGGGCCCCCGAGAAGCGTGGCATCGTGGAGCAGTGTGCACCAGCATCTGTTCCCTC
TACCAGCTGGAGAACTACTGCAACTAG

>Pig_AY242101

ATGGCCCTGTGGACGCGCCTCTGCCCTGCTGGCCCTGCTGGCGCTCTGGGCGCCGCC
CCGGCCAGGCTTCGTGAACCAGCACCTGTGCGGCTCCACCTGGTGGAGGCGCTGTAC
CTGGTGTGCGGGAGCGCGGCTTCTTCTACACGCCAAGGCCCGTCGGGAGGCGGAGAAC
CCTCAGGCAGGTGCCGTGGAGCTGGGCGGAGGCTGGGCGGCTGCAGGCCCTGGCGCTG
GAGGGGCCCCCGAGAAGCGTGGCATCGTGGAGCAGTGTGCACCAGCATCTGTTCCCTC
TACCAGCTGGAGAACTACTGCAACTAG

>Pig_AY242109

ATGGCCCTGTGGACGCGCCTCTGCCCTGCTGGCCCTGCTGGCGCTCTGGGCGCCGCC
CCGGCCAGGCTTCGTGAACCAGCACCTGTGCGGCTCCACCTGGTGGAGGCGCTGTAC
CTGGTGTGCGGGAGCGCGGCTTCTTCTACACGCCAAGGCCCGTCGGGAGGCGGAGAAC
CCTCAGGCAGGTGCCGTGGAGCTGGGCGGAGGCTGGGCGGCTGCAGGCCCTGGCGCTG
GAGGGGCCCCCGAGAAGCGTGGCATCGTAGAGCAGTGTGCACCAGCATCTGTTCCCTC
TACCAGCTGGAGAACTACTGCAACTAG

>Dog_V00179

ATGGCCCTGTGGATGCGCCTCTGCCCTGCTGGCCCTGCTGGCCCTCTGGGCGCCCGG
CCACCCGAGCCTTTCGTGAACCAGCACCTGTGCGGCTCCACCTGGTAGAGGCTCTGTAC
CTGGTGTGCGGGAGCGCGGCTTCTTCTACACGCCTAAGGCCCGCAGGGAGTGGAGGAC
CTGCAGGTGAGGGACGTGGAGCTGGCGGGGCGCCTGGCAGGGCGGCTGCAGCCCTG
GCCCTGGAGGGGCCCTGCAGAAGCGAGGCATCGTGGAGCAGTGTGCACCAGCATCTGC
TCCCTTACCAGCTGGAGAATTACTGCAACTAG

>OwlMonkey_J02989

ATGGCCCTGTGGATGCACCTCTGCCCTGCTGGCGCTGCTGGCCCTCTGGGACCCGAG
CCAGCCCGGCCTTTGTGAACCAACACCTGTGCGGCCACACCTGGTGGAGCCTCTTAC
CTGGTGTGCGGGAGCGAGGTTTCTTCTACGACCCAAGACCCGCCGGAGGCGGAGGAC
CTGCAGGTGGGCGAGGTGGAGCTGGGTGGGGCTCTATCACGGGAGCCTGCCACCCTG
GAGGGTCCCATGCAGAAGCGTGGCGTCTGGATCAGTGTGCACCAGCATCTGTTCCCTC
TACCAGCTGCAGAACTACTGCAACTAG

>Human_AY138590

ATGGCCCTGTGGATGCGCCTCTGCCCTGCTGGCGCTGCTGGCCCTCTGGGACCTGAC
CCAGCCGAGCCTTTGTGAACCAACACCTGTGCGGCTCACACCTGGTGGAGCCTCTTAC
CTAGTGTGCGGGAGCGAGGCTTCTTCTACACACCCAAGACCCGCCGGAGGCGAGGAC
CTGCAGGTGGGCGAGGTGGAGCTGGGCGGGGCTGCTGTCAGGCAGCCTGCAGCCCTG
GCCCTGGAGGGTCCCTGCAGAAGCGTGGCATTGTGGAACAATGCTGTACAGCATCTGC
TCCCTTACCAGCTGGAGAACTACTGCAACTAG

>GreenMonkey_X61092

ATGGCCCTGTGGATGCGCCTCTGCCCTGCTGGCGCTGCTGGCCCTCTGGGACCTGAC
CCGGTCCCGCCTTTGTGAACCAGCACCTGTGCGGCTCCACCTGGTGGAGCCTCTTAC
CTGGTGTGCGGGAGCGAGGCTTCTTCTACACGCCAAGACCCGCCGGAGGCGAGGAC
CCGACAGTGGGCGAGGTAGAGCTGGGCGGGGCTTGGCGCAGGCAGCCTGCAGCCCTG
GCGCTGGAGGGTCCCTGCAGAAGCGCGCATCGTGGAGCAGTGTGTACAGCATCTGC
TCCCTTACCAGCTGGAGAACTACTGCAACTAG

>Human_J00265

ATGGCCCTGTGGATGCGCCTCTGCCCTGCTGGCGCTGCTGGCCCTCTGGGACCTGAC
CCAGCCGAGCCTTTGTGAACCAACACCTGTGCGGCTCACACCTGGTGGAGCCTCTTAC
CTAGTGTGCGGGAGCGAGGCTTCTTCTACACACCCAAGACCCGCCGGAGGCGAGGAC
CTGCAGGTGGGCGAGGTGGAGCTGGGCGGGGCTTGGCGCAGGCAGCCTGCAGCCCTG
GCGCTGGAGGGTCCCTGCAGAAGCGCGCATCGTGGAGCAGTGTGTACAGCATCTGC
TCCCTTACCAGCTGGAGAACTACTGCAACTAG

```

GCCCTGGAGGGTCCCTGCAGAAGCGTGGCATTGTGGAACAATGCTGTACCAGCATCTGC
TCCCTCTACCAGCTGGAGAACTACTGCAACTAG
>Chimp_X61089
ATGGCCCTGTGGATGCGCCTCTGCCCTGCTGGTGCTGCTGGCCCTCTGGGGACCTGAC
CCAGCTCGGCCTTTGTGAACCAACACCTGTGCGGCTCCACCTGGTGGAAAGCTCTCTAC
CTAGTGTGCGGGGAAACAGGCTTCTTCTACACACCCAAGACCCGCGGGAGGAGGAGC
CTGCAGGTGGGGCAGGTGGAGCTGGGCGGGGCCCTGGTGCAGGCAGCCTGCAGCCCTTG
GCCCTGGAGGGTCCCTGCAGAAGCGTGGTATCGTGGAAACAATGCTGTACCAGCATCTGC
TCCCTCTACCAGCTGGAGAACTACTGCAACTAG
>GuineaPig_K02233
ATGGCTCTGTGGATGCATCTCCTCACCGTGTGGCCCTGCTGGCCCTCTGGGGCCCAAC
ACTAATCAGGCCTTTGTGAGCCGGCATCTGTGCGGCTCCAACCTAGTGGAGACATTGTAT
TCAGTGTGTCAGGATGATGGCTTCTTCTATATACCCAAGGACCGTGGGAGCTAGAGGAC
CCACAGGTGGAGCAGACAGAAGCTGGGCATGGGCCTGGGGCAGGTGGACTACAGCCCTTG
GCACTGGAGATGGCACTACAGAAGCGTGGCATTGTGGATCAGTGTGACTGGCACCTGC
ACACGCCACAGCTGCAGAGCTACTGCAACTAG
>Mouse_X04725
ATGGCCCTGTTGGTGCACCTTCTACCCCTGCTGGCCCTGCTTGCCTCTGGGAGCCAAA
CCCACCCAGGCTTTTGTCAAACAGCATCTTTGTGGTCCCCACCTGGTAGAGGCTCTTAC
CTGGTGTGTTGGGAGCGTGGCTTCTTCTACACACCCAAGTCCCGCCGTGAAGTGGAGGAC
CCACAAGTGGAAACAACCTGGAGCTGGGAGGAAAGCCCGGGGACCTTACAGCCTTGGCGTTG
GAGGTGGCCCGCAGAAGCGTGGCATTGTGGATCAGTGTGACCAAGCATCTGCTCCCTC
TACCAGCTGGAGAACTACTGCAACTAA
>Chicken_AY438372
ATGGCTCTGTGGATCCGATCACTGCCTTCTTGGCTCTCCTTGTCTTTTCTGGCCCTGGA
ACCAGCTATGCAGCTGCCAACCCAGCACCTCTGTGGCTCCACTTGGTGGAGGCTCTTAC
CTGGTGTGTTGGAGCGTGGCTTCTTCTACTCCCCAAAGCCGACGGGATGTGAGCAG
CCCCTAGTGAGCAGTCCCTTGCCTGGCGAGGAGGAGTGTGCTCCCTTCCAGCAGGAGGAA
TACGAGAAAGTCAAGCGAGGATTGTTGAGCAATGCTGCCATAACACGTGTTCCCTCTAC
CAACTGGAGAACTACTGCAACTAG

```

Spørgsmål 5

*) Ja - fx. er der et sted med gaps af længde 1 - fuldt af en kort række nucleotider og så et gap på 5.
 Alignment algoritmen kender ikke noget til at sekvenserne er protein-kodende, så den kigger kun på DNA'et.

Human_AY138590	CTGCAGGTGGGGCAGGTGGAGCTGGGCGGGGCCCTGGTGCAGGCAGCCTGCAGCCCTTG	240
Human_J00265	CTGCAGGTGGGGCAGGTGGAGCTGGGCGGGGCCCTGGTGCAGGCAGCCTGCAGCCCTTG	240
Chimp_X61089	CTGCAGGTGGGGCAGGTGGAGCTGGGCGGGGCCCTGGTGCAGGCAGCCTGCAGCCCTTG	240
GreenMonkey_X61092	CCGCAGGTGGGGCAGGTAGAGCTGGGCGGGGCCCTGGCGCAGGCAGCCTGCAGCCCTTG	240
OwlMonkey_J02989	CTGCAGGTGGGGCAGGTGGAGCTGGGTTGGGGCTCTATCACGGGCAGCCTGCCACCCCTTG	240
Pig_AY242100	CCTCAGGCAGGTGCCGTGGAGCTGGGCGGAG-GCCTGG-----GCGGCCTGCAGGCCCTG	234
Pig_AY242101	CCTCAGGCAGGTGCCGTGGAGCTGGGCGGAG-GCCTGG-----GCGGCCTGCAGGCCCTG	234
Pig_AY242109	CCTCAGGCAGGTGCCGTGGAGCTGGGCGGAG-GCCTGG-----GCGGCCTGCAGGCCCTG	234
Pig_AY044828	CCTCAGGCAGGTGCCGTGGAGCTGGGCGGAG-GCCTGG-----GCGGCCTGCAGGCCCTG	234
Pig_AY242098s	CCTCAGGCAGGTGCCGTGGAGCTGGGCGGAG-GCCTGG-----GCGGCCTGCAGGCCCTG	234
Sheep_U00659	CCCCAGGTGGGGCGCTGGAGCTGGCCGGAG-GCCCCG-----GC-----GCGGG---TG	226
Dog_V00179	CTGCAGGTGAGGGACGTGGAGCTGGCCGGGGCCTGGCGAGGGCGCCTGCAGCCCTTG	240
Mouse_X04725	CCACAAGTGGAAACAACCTGGAGCTGGGAGGAAAGCCCGG-----GGACCTTACAGCCTTG	234
GuineaPig_K02233	CCACAGGTGGAGCAGACAGAAGCTGGGCATGGGCCTGGGGCAGGTGGACTACAGCCCTTG	240
Chicken_AY438372	CCCCTAGTGAGCAGTCCCTTGCCTGGCGAGG---CAGGAG-----TGCTGCCTTTC	228

*) Kyllig skiller sig ud - den er også den eneste der ikke er et pattedyr.

*) Ved at skifte lidt frem og tilbage mellem cladogram og phylogram (og evt. slå visning af afstande i træet til) se det at det to humane sekvenser er 100% ens (man kan altså smide den ene ud), og for grisen er følgende skvenser ens:

```

>Pig_AY044828
>Pig_AY242098

```

samt

>Pig_AY242100
>Pig_AY242101

(man kan altså smide to grise-sekvenser ud af sættet).

Sprøgsmaal 6

Sekvenserne oversættes med Virtual Ribosome, hvilket giver følgende sekvenser:

>Sheep_U00659
MALWTRLVPLLALLALWAPAPAHAFVNQHL CGSHLVEALYLVCGERGFYTPKARREVEG
PQVGALELAGGPGAGGLEGPPQKRGIVEQCAGVCSLYQLENYCN*
>Pig_AY044828
MALWTRLVPLLALLALWAPAPAQAFVNQHL CGSHLVEALYLVCGERGFYTPKARREAEN
PQAGAVELGGGLGGLQALALEGPPQKRGIVEQCCTSICSLYQLENYCN*
>Pig_AY242098
MALWTRLVPLLALLALWAPAPAQAFVNQHL CGSHLVEALYLVCGERGFYTPKARREAEN
PQAGAVELGGGLGGLQALALEGPPQKRGIVEQCCTSICSLYQLENYCN*
>Pig_AY242100
MALWTRLVPLLALLALWAPAPAQAFVNQHL CGSHLVEALYLVCGERGFYTPKARREAEN
PQAGAVELGGGLGGLQALALEGPPQKRGIVEQCCTSICSLYQLENYCN*
>Pig_AY242101
MALWTRLVPLLALLALWAPAPAQAFVNQHL CGSHLVEALYLVCGERGFYTPKARREAEN
PQAGAVELGGGLGGLQALALEGPPQKRGIVEQCCTSICSLYQLENYCN*
>Pig_AY242109
MALWTRLVPLLALLALWAPAPAQAFVNQHL CGSHLVEALYLVCGERGFYTPKARREAEN
PQAGAVELGGGLGGLQALALEGPPQKRGIVEQCCTSICSLYQLENYCN*
>Dog_V00179
MALWMRLLPLLALLALWAPAPTRAFVNQHL CGSHLVEALYLVCGERGFYTPKARREVED
LQVRDVELAGAPGEGGLQPLALEGALQKRGIVEQCCTSICSLYQLENYCN*
>OwlMonkey_J02989
MALWMHLLPLLALLALWGPEPAPAFVNQHL CGPHLVEALYLVCGERGFYAPKTRREAED
LQVGQVELGGGSITGSLPPELGPMQKRGVVDQCCTSICSLYQLQNYCN*
>Human_AY138590
MALWMRLLPLLALLALWGPDPAAAFVNQHL CGSHLVEALYLVCGERGFYTPKTRREAED
LQVGQVELGGGPGAGSLQPLALEGSLQKRGIVEQCCTSICSLYQLENYCN*
>GreenMonkey_X61092
MALWMRLLPLLALLALWGPDVPAFVNQHL CGSHLVEALYLVCGERGFYTPKTRREAED
PQVGQVELGGGPGAGSLQPLALEGSLQKRGIVEQCCTSICSLYQLENYCN*
>Human_J00265
MALWMRLLPLLALLALWGPDPAAAFVNQHL CGSHLVEALYLVCGERGFYTPKTRREAED
LQVGQVELGGGPGAGSLQPLALEGSLQKRGIVEQCCTSICSLYQLENYCN*
>Chimp_X61089
MALWMRLLPLLALLALWGPDPAFVNQHL CGSHLVEALYLVCGERGFYTPKTRREAED
LQVGQVELGGGPGAGSLQPLALEGSLQKRGIVEQCCTSICSLYQLENYCN*
>GuineaPig_K02233
MALWMHLLTVLALLALWGPNTNQAFVSRHL CGSNLVETLYSVCQDDGFFYIPKDRRELED
PQVEQTELMGLGAGGLQPLALEMALQKRGIVDQCCTGTCTRHQLQSYCN*
>Mouse_X04725
MALLVHFLPLLALLALWEPKPTQAFVKQHL CGPHLVEALYLVCGERGFYTPKSRREVED
PQVEQLELGGSPGDLQTLALEVARQKRGIVDQCCTSICSLYQLENYCN*
>Chicken_AY438372
MALWIRSLPLLALLVFSGPGTSYAAANQHL CGSHLVEALYLVCGERGFYSPKARRDVEQ
PLVSSPLRGEAGVLPFQQEEYKVKRGIVEQCCHNTCSLYQLENYCN*

Efterfølgende alinges sekvenserne med ClustalW.

*) Det sted jeg nævnt i spg. 5 med et gap på 1 + et gap på 5 lidt længere henne, er nu erstattet af et enkelt gap på 2 aminosyrer (= et gap på 6 nucleotider).

GuineaPig_K02233	PQVEQTELMGLGAGGLQPLALEMALQKRGIVDQCCTGTCTRHQLQSYCN	110
Mouse_X04725	PQVEQLELGG--SPGDLQTLALEVARQKRGIVDQCCTSICSLYQLENYCN	108
Human_AY138590	LQVGQVELGGGPGAGSLQPLALEGSLQKRGIVEQCCTSICSLYQLENYCN	110
Human_J00265	LQVGQVELGGGPGAGSLQPLALEGSLQKRGIVEQCCTSICSLYQLENYCN	110
Chimp_X61089	LQVGQVELGGGPGAGSLQPLALEGSLQKRGIVEQCCTSICSLYQLENYCN	110
GreenMonkey_X61092	PQVGQVELGGGPGAGSLQPLALEGSLQKRGIVEQCCTSICSLYQLENYCN	110

```

OwlMonkey_J02989      LQVGQVELGGGSITGSLPP--LEGPMQKRGVVDQCCTSICSLYQLQNYCN 108
Pig_AY242101          PQAGAVELGG--GLGGLQALALEGPPQKRGIVEQCCTSICSLYQLENYCN 108
Pig_AY242109          PQAGAVELGG--GLGGLQALALEGPPQKRGIVEQCCTSICSLYQLENYCN 108
Pig_AY242100          PQAGAVELGG--GLGGLQALALEGPPQKRGIVEQCCTSICSLYQLENYCN 108
Pig_AY242098          PQAGAVELGG--GLGGLQALALEGPPQKRGIVEQCCTSICSLYQLENYCN 108
Pig_AY044828          PQAGAVELGG--GLGGLQALALEGPPQKRGIVEQCCTSICSLYQLENYCN 108
Sheep_U00659          PQVGALELAG--GPG---AGGLEGPPQKRGIVEQCAGVCSLYQLENYCN 105
Dog_V00179            LQVRDVELAGAPGEGGLQPLALEGALQKRGIVEQCCTSICSLYQLENYCN 110
Chicken_AY438372      PLVSS-PLRG--EAGVLPFQQEEYEKVKRGIVEQCCHNTCSLYQLENYCN 107
.      *      *      *      **.*:** . *: :*.**

```

*) Det ses nu at på protein-niveau er alle grise-sekvenserne helt ens. Man kan altså smide fire ud.

Spørgsmål 7

ClustalW siger:

CLUSTAL W (1.83) multiple sequence alignment

```

seq2      -----KNDEWQRNMSDERTYCVNHKLIKTRSDFQ
combi     LASDFEWRYPNMYTFRECD SALKNLPRTKIKNDEWQRNMSDERTYCVNHKLIKTRSDFQ
seq1      -----LASDFEWRYPNMYTFRECD SALKNLPRTKI
              .::: . .: *: .. :*.

```

```

seq2      W
combi     W
seq1      -

```

Spørgsmål 8

T-Coffee, MUSCLE og MAFFT kan klare det, men ikke Kalign. (sidstnævnte opfører sig endnu mere mærkeligt end Clustal).

Spørgsmål 10

*) Ja - gaps går op i tre.

*) Ja - da DNA alignmentet er genereret med et protein-alignment som skabelon.

*) Ja - der er nogle korte stræk af baser, der er med små bogstaver.