



Genomanalyse

8. Übungsblatt WS 2009

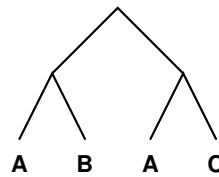
Abgabe: Dienstag 8.12.2009 vor der Übung, Besprechung: in der Übung

Bitte geben Sie auf Ihren abgegebenen Lösungen Name und Matrikelnummer an und arbeiten Sie in Gruppen von 2 oder 3 Personen zusammen.

Aufgabe 16: [2+3+4=9 Punkte] Fitch versus Sankoff-Cedergren

Wenden Sie auf den gegebenen phylogenetischen Baum an:

- a) Algorithmus von Fitch
- b) Algorithmus von Sankoff-Cedergren, wobei alle Substitutionen 1 kosten sollen ("ungewichtet")
- c) Algorithmus von Sankoff-Cedergren mit der nebenstehenden Substitutionsmatrix



	A	B	C
A	0	1	4
B	1	0	2
C	4	2	0

Tragen Sie beim Algorithmus von Sankoff-Cedergren alle Traceback-Kanten ein.

Aufgabe 17: [4+3=7 Punkte] Fitch versus Sankoff-Cedergren

Unter dem "ungewichteten" Sankoff-Cedergren Algorithmus wollen wir verstehen, dass wir die folgende Substitutionsmatrix verwenden: $s(x, x) = 0$, $s(x, y) = 1$ für $x \neq y$.

- a) Zeigen Sie: Wenn ein phylogenetischer Baum vom Fitch-Algorithmus gefunden wird, dann kann man ihn auch als Lösung vom ungewichteten Sankoff-Cedergren Algorithmus erhalten.
- b) Gilt die Umkehrung auch, das heißt, *jede* Lösung, die der ungewichtete Sankoff-Cedergren Algorithmus produziert, kann man auch mit dem Algorithmus von Fitch erhalten? Begründen Sie Ihre Antwort.

Aufgabe 18: [6 Punkte] Multiples Sequenzalignment

Auf der Webseite finden Sie die Datei `globins.fasta`. Erstellen Sie mit Werkzeugen im Internet (z.B. EMBOSS) ein sinnvolles multiples Alignment und drucken Sie es schön bunt formatiert aus. Auf der Rückseite finden Sie zwei Beispiele.

Wichtig: Bei dieser Aufgabe muss jede Gruppe eine erkennbar *eigene*, am besten *verschiedene* Lösung abgeben. (Z.B. Formatierung oder Farben – aber bitte nicht per Hand!)

globins

clustalw alignment / emboss

```

HBB-HUMAN  VHLPEEKAMALKVNVDVMEALRLLMVGESIDIDAMNFKVKAHKKVLVADLHLDNLKSTIAALE 90
HBB-HORSE  VQLPEEKAMALDKYNEEVEALRLLMVGESIDIDSNPAMNFKVKAHKKVLVADLHLDNLKSTIAALE 90
HBA-HUMAN  VYLADKMYKAAIKYAAALAEALERLSTIRKIPHLPHSHACQVKAHKKVDALNAVAVHVDLPAALKLD 85
HBA-HORSE  VYLADKMYKAAIKYAAALAEALERLSTIRKIPHLPHSHACQVKAHKKVDALNAVAVHVDLPAALKLD 85
MYG-PHYCA  VLEAEELCLVLMWAKVFAVAHQCLLRLKSHPELEKDRKHKVTEAEVKAEDLKFHVTMTALAIKKKGFHAEKFLAQ 91
GLB5-PETMA PIVDTGSAVPLAAEKTKIRAAATMTEETSIMLMKLTSTIAAEPKPKGELADQLKKADVRHERRI 100
LGB2-LUPLU  RLQSGVSKVADALRWKALVKEEENANPKHHRITLMEIATAAKDLSFKGHLVYVQKIFELQAHAKM 90
Consensus  LS-AEKT-VKAAW-KV-AN-EYG-EAL-RLFV-YP-TQ-FF-F-L-T-D-M-GN-EVKAHGKVKLVNALSDAV-HLDDLK-LS-LSD 100

```

```

HBB-HUMAN  LFLCDKLVDFENRLLMLVYLAHCKEELAAAGMMAVANALAHKH 146
HBB-HORSE  LFLCDKLVDFENRLLMLVYLAHCKEELAAAGMMAVANALAHKH 146
HBA-HUMAN  LLAHLRVDVYVYKLLSHLLVLAHFAEETAVAALELAVAVTMTLSKR 141
HBA-HORSE  LLAHLRVDVYVYKLLSHLLVLAHFAEETAVAALELAVAVTMTLSKR 141
MYG-PHYCA  SATHKIPYKLELSEALIKVLSHRHPDADAGQAVNKALELFRKIDAAKELGYQG 153
GLB5-PETMA  LSGKHAKSGMVAHMKVLAAMADTVAAGDAGEKLVNCLLRK 149
LGB2-LUPLU  RLQSGVSKVADALRWKALVKEEENANPKHHRITLMEIATAAKDLSFKGHLVYVQKIFELQAHAKM 153
Consensus  LH-KLRVDP-NFKLL-H-LL-VLA-H-DFTPEV-ASL-K-LA-YA-L-KY 166

```

```

1
HBB_HUMAN  -----VH LTPEEKSAVT ALWGKVN--V DEVGGEALGR LLVVYPWTQR FFESFGDLST PDAVMGNPKV
HBB_HORSE  -----VQ LSGEKAAVL ALWDKVN--E EEVGEALGR LLVVYPWTQR FFDSFGDLSN PGAVMGNPKV
HBA_HUMAN  -----V LSPADKTNVK AAWGKVGAAH GEYGAEALER MFLSFPTTKT YFPFH-DLS ----HGSAQV
HBA_HORSE  -----V LSAADKTNVK AAWSKVGGAH GEYGAEALER MFLGFPTTKT YFPFH-DLS ----HGSAQV
GLB5_PETMA PIVDTGSAVPL SAAEKTIR SAWAPVYSTY ETSGVDILVK FFTSTPAAQE FFPKFKGLTT ADQLKKSADV
MYG_PHYCA  -----V LSEGEWQLVL HVWAKVEADV AGHGQDILIR LFKSHPETLE KFDRFKHLKT EAEMKASEDL
LGB2_LUPLU  -----GA LTESQAALVK SSWEEFNANI PKHTRFFIL VLEIAPAAKD LFSFLKGTSE VP--QNNPEL

```

```

71
HBB_HUMAN  KAHGKKVLGA FSDGLAHLDN -----LKGTF ATLSELHCDK LHVDPENFRL LGNVLVCVLA HHFGKEFTPP
HBB_HORSE  KAHGKKVLHS FGEVGHLDN -----LKGTF AALSELHCDK LHVDPENFRL LGNVLVVLA RHFGKDFTP
HBA_HUMAN  KGHGKKVADA LTNVAHVDD -----MPNAL SALSDLHAKK LRVDPVNFKL LSHCLLVTLA AHLPAEFTPA
HBA_HORSE  KAHGKKVGDA LTLAVGHLDD -----LPGAL SNLSDLHAKK LRVDPVNFKL LSHCLLSTLA VHLPNDFTPA
GLB5_PETMA RWHAERIINA VNDAVASMD T--EKMSMKL RDLSGKHAKS FQVDPQYFKV LAAVIADTVA AG-----
MYG_PHYCA  KKHGVTVLTA LGAILKKKGH -----HEAEL KPLAQSHATK HKIPIKYLEF ISEAIHVLH SRHPGDFGAD
LGB2_LUPLU  QAHAQGVFKL VYEAAIQLQV TGVVVVTDATL KNLGSVHVS K G-VADAHFPV VKEAILKTIK EVVGAKWSEE

```

```

141
HBB_HUMAN  VQAAYQKVVA GVANALAHKY H-----
HBB_HORSE  LQAAYQKVVA GVANALAHKY H-----
HBA_HUMAN  VHASLDKFLA SVSTVLTISKY R-----
HBA_HORSE  VHASLDKFLS SVSTVLTISKY R-----
GLB5_PETMA -DAGFEKLMS MICILLRSAY -----
MYG_PHYCA  AQQAMNKALE LFRKDIAAKY KELGYQG
LGB2_LUPLU  LNSAWTIAVD ELAIVIKKEM NDAA---

```

```

>HBB_HUMAN Sw:Hbb_Human => HBB_HUMAN
VHLTPEEKSAVTALWGKVNVDDEVGGEALGRLLVVYPWTQRFFESFGDLSTPDAVMGNPKV
KAHGKKVLGAFSDGLAHLNKGTFATLSELHCDKLHVDPENFRLLGNVLVCVLAHFGK
EFTTPVQAAAYQKVAVGANALAHKYH
>HBB_HORSE Sw:Hbb_Horse => HBB_HORSE
VQLSGEKAALVLAWDKVNVEEVGGEALGRLLVVYPWTQRFFDSFGDLSNPGAVMGNPKV
KAHGKKVLHSGFGEVGHLDNKGTFAAALSELHCDKLHVDPENFRLLGNVLVVLARHFGK
DFTPELQASVYKVVAGVANALAHKYH
>HBA_HUMAN Sw:Hba_Human => HBA_HUMAN
VLSPADKTNVKAANGKVGAAHGEYGAEALERMFLSFPTTKTYPPHFDLSHGSAQVKGHGK
KVADALTNVAHVDDMPNALSALSDLHAKKLRVDPVNFKLLSHCLLVTLAAHLPAEFTPA
VHASLDKFLASVSTVLTISKYR
>HBA_HORSE Sw:Hba_Horse => HBA_HORSE
VLSAADKTNVKAANSKVGGAHGEYGAEALERMFLGFPTTKTYPPHFDLSHGSAQVKAHGK
KVGDALTLAVGHLDDLPGALSNSDLHAKKLRVDPVNFKLLSHCLLSTLAVHLPNDFTPA
VHASLDKFLSSVSTVLTISKYR
>MYG_PHYCA Sw:Myg_Phyca => MYG_PHYCA
VLSGEWQLVLHVWAKVEADVAGHGQDILIRLFKSHPETLEKFDRFKHLKTEAMKASED
LKKHGVTVLTLGAILKKKGHEAELKPLAQSHATKHKIPIKYLEFISEAIHVLHSHRHP
GDFGADAGQAMNKALELFRKDIAAKYKELGYQG
>GLB5_PETMA Sw:Glb5_Petma => GLB5_PETMA
PIVDTGSAVPLSAAEKTIRSAWAPVYSTYETSVDILVKFFTSTPAAQEFPPKFKGLTT
ADQLKKSADVRWHAERIINAVNDAVASMDTEKMSMKLRDLSGKHAKSFQVDPQYFKVLA
AVIADTVAAGDAGPEKLSMILLRSAY
>LGB2_LUPLU Sw:Lgb2_Luplu => LGB2_LUPLU
GALTESQAALVKSWEFNANIIPKHTRFFILVLEIAPAAKDLFSFLKGTSEVPQNNPEL
QAHAQGVFKLVYEAAILQVVTATLKNLGSVHVS KGVADAHFPVKEAILKTIKE
VVGAKWSEELNSAWTIAVDELAIVIKEMNDAA

```