

gl y_unwei ghted_fasta. fasta

>' Bacillus cereus ATCC 10987' collagen
mkerdkqnslnsnfri spnl i gptfppvptgftgi gi tgptgpggptgpggprgl qgpmgemgptgpgqg
qgi qgpvgsi gatgpegqqgppqgl rpgqgetgatgpgqvqgl qgpagptgatgaqgi qgi qgl qgpi gat
gpegpqi qgvqgl ppatgpgqi qgaqgmqgl qgpgsntgatgatqgi tgptgvtgptgi tg-----

-----psggppgptgptgatgpgggpsgstgatgatgntgatgstgvt
gstgvtgatgstgptgstgaqgl qgi qgi qgpi gptgpegpqi qgi pgptgvtgeqgi qgvqgi qgatg
atgdqgppqi qgai gpqgatgatgdqgppqi qgvpgpsgatgpgqvqgl qgpmgdi gptgpegpegl qgp
qgi qgvpgpvatgpegpqi qgi qgpvatgpgqppqi qgi qgvqgi tgatgvqatgi qgi qgei gat
gpegpqqvqgaqgi gptgpmgpgqvqgvqgi qgatgaqgvqppqi qgi qgi qgi qgptgatgdtgatg
atgeggtgptgvtgptgpgsgpaggptgptgpgspagvtgpgsgppgptgatgatgvtgdtgatgstgvtg
atgetgatgvtgl qgpqi qgvqgei gptgpgqvqppqi qgvtgatgdqgppqvqppqdi gptgpgqi
qgpqgsqi qgatggtgaqppqi qgpqgdvgtgpgqptgi qgi qgei gptgpegpegl qgpqi qgvq
gpvgatgpegpqi qgi qgvqgatgsqppqi qgi qgvqgi tgatgaqatgi qgi qgei gatgpegpqq
vqgvqgei gptgpmgpgqvqgvqgi qgatgaqgvqppqi qgi qgptgatgetgatgatgeggtgptgvt
gptgvtgpgsgpaggptgptgpgspagvtgpgsgppgptgatgatgatgvtgdtgatgstgvtgatgetga
tgvtgl qgpqi qgvqgei gptgpgqi qgpqi qgvtgatgaqppqi qgpqgdi gptgpgqi qgpqppq
gi qgatgatgaqppqi qgpqgei gptgpgqppqi qgpqi qgptgatgatgatgl qgi qgpqi qgpqg
i qgptgatgatgatgl qgi qgpqi qgpqi qgptgatgatgatgl qgi qgpqi qgpqi qgptgatga
tgatgsqgptgdtgptgagatgatgatgvsttatyafanntsgtai svl l ggtni pl pnnqi gpgi tvs
ggntvftvasagnyyi ayti nl tagl l vssri tvngspl agti naptvatgsfsati i anl pagaai sl q
l fgl vai atl sttptgat l ti i rl s

>' Dani o reri o (zebrafish)' type I alpha 1
mfsfvdi rllal l satvl l argqgeddrtggsctl dgqvyndrdvkwpepcqi cvcdsgtvmcdevi ced
tsdcpnvi phdeccpvcppdddfqepsvegprgsppgdkgergpagppgndgi pgqpgl pggpg-----

-----ppgppgl ggnfspqmsggfdeks-spmavpgpmgpmgprgagp
ppgpgpqqgftgppgepgeagapgpmgprgaagppgkngedgsgkprpgergpppgqgargfpgtpgl
pgi kghrgfsgl dgakgdagpagpkgepgagpengtngamgprgl pgergragppgaagargndgaagaa
gppgptgpagppgfpggpgskgevpqgsrgaegpqqargeagnpgpagpagngadgapgakgapg
apgi agapgfpgprgppgaagaagapgpkgntgeagapgakgeagakgeagaqgvqppppgpegkrp
rgepgaggargptgergapgarfpgadgaagprgapergpgvvpkgatgepgrngepgmpgskgmt
gspgspgpdgktgl agapqdgprgppgpvgargqpgvmgfpkgkaageagkpgergvmgai gatgapg
kdgdvgapgapppagpgergeqgaagppgfqgl ppgqgatgepgksgeqgapgeagapppsgsrgdrgf
pgergapppagpvgarpsgsagndgakgesgaagapgaqppgl qgmpgergaagl pgl kgdrgdqgak
gadgaagkdg rmtgpi gppgpagapgdkgesgaqgl vgptgargppgergetgapppagfagppgadg
l pgakgepgdngakgdagapppagatgagppqpvatgpkgar gaagppgatgfpgaagrvgppgsgn
sgpppppagkegkgnrgetgpagrtgevgaaagppgagpekgnpaegatgpagi pppqi ggrgi v
gl pgrgergfpgl ppgsgei gkqpsgspgergppgpmgppgl agppgepgergtgpnegsagrdaag
pkdrgetgpgstgpagppgaagpi gpagktgdrgetgpagvppagpsgprgpgspagargdkgetge
agergmkghrgftgmpgppgppgsgesgpagasgpagprgpagsagsagkdgmsgl ppgi gppgprgrn
gei gpagppppppgagppsgggfdi gfi aqp-qekapdpfr-----

-----hfraddanvmrdrl
evdttl ksl sqqi esi i spdgtkknpartcrdl kmchpdwksgeywi dpdqcnq dai kvycnmetgetc
vnptesai pkknwytskni kekhhwfwgeamtdgfqfeygsegskpedvni ql tfl rl msteasqi tyh
cknsi aymdqasnl kkal l l qgsnei ei raegnsrftysvtedgctsthtgawgktvi dykttktsrl pi
i di apmdvgapnqefgi evgpvcfl

>' Dani o reri o (zebrafish)' type I alpha 3
mfsfvdi rllgl l aavvl avrgqgeddi sfgsctl egqnyndkdvkwpepcqi cvcdsgtvmcdevi ced
tsdcanpei pdgeccpi cpdeevtaapeppyvdtvgpagp-----rgdpgpvppgrdgtngen-----

-----gl pggppgpgtsgl gggsfl sqmaygsekssgppqgppgpmgprgl pg
pagspgpqqgftgppgepgepagpmpsrppgppgkngedgagkagrpgergaagpqqgargfpgtpgl

gl y_unwei ghted_fasta. fasta

pgi kghrgfpgl dgakgdagi agpkgeagspgeagtpgamargl pgergrpppppsgargndgntgaa
gppgptgpagppgfpggagakgetgaaggrgseppqgprgepgnppagaagpaggppsdtppgkgsppg
aagi agapgfpgarppppagpngagppkgnngdpgl qpkgdagpkgeppvppqgl agnageegkrga
rgepgaagpvpppgargapgnrgl pgsdgappl gapgergangpmaaggasgesgrppgepl pgskgmt
gapgsppgdgkagpsgapqddgrpppppagsrgapgvmgfpgpkgt dgesgkppgergl agptgpl gapg
kdgdi gspgapgpapagekgeqgsagpsgfqgl pgaagatgetgkppgeqgl pgeagpsgspgprgdrf
pgergapppgptgprgapsppndgakgeppgaaganggvppagmqmpgergagmpgargdrdggpk
gpdgapgkdgvrgmtgpi gppgspgapgdkeppgavpppl tprgppgergevgapppagfagppggng
qqgakgetgdsapkgdagapgpavgasppqppagatgpkgarggagppgatgfpgaagrvgppgppsga
sgppgptgpagkeqrgargekppagrppgeagaagpppsgergsppgdgpl gl agapppgvi gsrgi v
gl pqrgergfagl agpsgeppkqgppsgpvgerppppmpppl sgapgeagregnpghdgvsgdrdppg
pkdrgesgnsppgagppagppgi gpsgktdrgeagppagpsppagargal gpagprgdrgeage
agergmkghrgl sgmpgvppgptsggeqppagatgpagprgpagsngsppgkdgmnvgppvppppgrgs
gemppagapgl pppppppgsagppgepl pmpqqekgpdpl rgg-----

-----yraddasvrdrds
evdtsl ksl sqki eni rspegtqanparmcrdl rmchpewksgsywvdpnqgspl dai kvfcnmetgetc
vspyhdsi prknwytskni rekkhwfgesmpngfqqfygsdgsdpedvni ql tfmrl msnqgsqni tyh
cknsi aymdetsgnl kkal l l qgsndi ei raegnsrftyrvsedgctsthtgtwgtkvi dykttktsrl pi
i di apmdi gapdqefgvevppvcfl

>'Drosophila melanogaster (fruit fly)' type IV alpha 1
ml pfwkrl l yaavi agal vgadaqfwktagtagsi qdsvkhynrnepkfpi ddsydi vdsagvargdl pp
knctagyagcvpkci aekgnrgl pppl gptgl kgemgfpgmegpsgdkgqkdppgyqrgdkgergspg
l hgqagvppvqppagnppagpi ngkdgcddgdgi pgl egl sgmpgprgyagql gskgekgepakengdya
kgekgeppwrgtagl agppgfpgekgergdsppygakpprgehgl kgekgascygpmkppagpi kgekge
passfpvkphtvmgprgdmgqkgepl vgrkgeppgedtgl dgqkgekgl pggppgdrgrqgnfppgs
tqkkgdrgepl ngl pgnppqkgeppragatgepl l gppppppgggrtppppgkpprgyvgapppqgl
ngvdgl pppqyngqkdaggl pprpneppppgkgektagl ngpkgsi gpi ghpppppegqkdagl p
gygi qgskgdagi pgyppgl kgskgergfkgnagapgdskl grpgtppaagapqkkgdagrpgtppqkdm
gi kgdvggkcsscragpkgdktsgl pgi pgkdgarppppgergyppgerghdgi ngqtgppgekgedgrtg
l ppatgeppkpal cdl sl i epl kgdkgypgagpakgvqgfkaegl pgi pppkgefgekgl sgapgn
dgtpragrddgyppi pqqsi kgeppfhrdgdakgdksfgrsgekgeppscal dei kmpakgnkgeppqt
gmppppedgspgergytgi kgntgppqppgvepprgl ngprgekgnqgavgvppgnppkdgl rgi pgrng
pppprgeppi srpppmpppl ngl qgekgrgptgpi gfpagadsvgyppgdrdagl pgvsgppgi vgek
gdvppi gpagvappppvppi dgvrgrdgakgeppsgpl vmppgnkgdrgapndgpkfagvtgapgkrg
papi pgvsgakgdkgasgl tndgppvgrppppagpl mgi kgdqgl agapqqgl dmpgekgnqgfpl
dpppl pgdasekqkgepppsgl rgdtgpagtppwpekgpl pgl avhgragppgekdgqrsi dgrdg
i ngekgeqgl qvvgppgeksgsvgapi ppagmdgl pgaagapgvgyppgdrgdkepl gl pgl kge
tppvpl qgftgapppkgergi rqqpl patvpdi rgdkgsqgergytgekgeqgergl tppagvagakgd
rgl qpppgasgl ngi ppakgdi gprgei gypgvti kgekgl pprpgrnrgl i gappl i gergl pgl pe
srl vgl pppi gpagskgergl agspgppqdgfppagpl kgdtgppqgfkggergl ngfegqkdkgdrgl q
gpsgl pgl vqkgtgyppl ngndgppvppgergftgpkgrdgrdgtpl ppqkgeppml pppgkgepp
ppgrnppkgepprgergl i gi qqel gekgergl i getgnvppgpkgdrgeppgergygai gl i gqkge
ppagapaa dyl tli i trhsqsetvpacsaghtel wtgysl l yvdgndyahnqdl gscvprfstl pvl s
cgqnnvcnyasrndktfwl ttnaai pmpveni ei rqi srcvceapanvi avhsqti evpdcppngweg
l wi gysfl mhtavnggggqal qspgsl edfratpfi ecngakgtchfyetmtsfwmyl essqperp
qqqt i kagerqshvsrcqvcmkns

>'Gallus gallus (chicken)' type I alpha 2
ml sfvdtri l l l l avtsyl atsqhvseasagrkgprgdkgqgergpppppprdgedgppppppppppp
l ggnfaaqydpaskaadfgpppmgl mgprpppgasgppppppgfqgvppgeppgqtgppqpprppppppgka
ge-----

-----dghpgkpprppergvaggppargfpgtppg
pgfkgi rghngl dgl tqppagptkgeppagpengtppqppgargl pgergri gapppagarsdgsagpt
gpa-xxxxxxxxxxxxxxxxxgei gpagnegptgpapprgei gl pgsppvppppnppangl ppakgaag
l pvvagapgl ppprgi pppppagpsgargl vgepppagakgesgnkgeppgaagpppppppsgeegkrgs
ngeppsagppppagl rpeppsgl ppadragvmppagnrgasppvagakppngdagrppgepl mpprxxx
xxxxxxxxxxxxxxxxxgfpagdrvgpi gpagnrgeppni gfpgkgtgeppkpekgenvgl agpprapp
pegnngaqqppgvtgnqqakgetgpagppgfqgl ppsppageagkppgergl hgefvgppagppgergl
ppesgavppagi gsrppsgpppdgnkgeppnvgpapppagppgi pgergvagvppgkgekagpl r

gl y_unwei ghted_fasta. fasta

gdtgatgrdgargl pgai gapgpaggagdrgeggpagpagargi pgergepgvpvpsgfagppgaag
qpgakgergpkpkgetgptgai gpi gasgppgvgaagpagprgdagppgmtgfpagaagrvppppagi
tgppppppagkdgrgl rgdvvpvgrtgeqqi agppgfagekgpsgeagaagppgtppqqi l gapi l
gl pgsrgergl pgi agatgeppgl gvsppgargpsgpvgsppngapgeagr dgnpgndgppgrdgapp
fkgergapgnppsgal gapgphqvpsgkpgnr gdpvvpvpgafgprgl agppgrgekgppgd
kghrgl pgl kghngl qgl pgl agqhdqgppgnngpagprpppsgppgkdgrngl pgi gpavrgsh
gsqgpagppppppppppppnggyevgfdaeyyradqpsl rp-----

-----kdy
evdatl ktl nnqi etl l tpegscknpartcrdl rl shpewssgyfi dnnqgctadai raycdfatgetc
i hasl edi ptktwyvsknkdkkhi wfgeti nggtqfeyngevttkdmatl afmrl l anhasqni tyh
cknsi aymdeetgnl kkavi l qgsndvel raegnsrftfsvl vdgcskknkkgkti i eyrtknpsrl pi
l di apl di ggadqefgl hi gpvcfk

>' Gallus gallus (chicken)' type II alpha 1
mhgrprpsaalllllllltaaaaaq-----
-----rdl rppgpkqkgepgdi kdvvprgppppqppageqqrgdrgek
gekgapgp-----

-----rgrdgepgtpgnppppppppppppgl ggnfaaqmaggfdekaggaqmgvmqgpmgmprgppp
ptgapppqgfqgnpgepggegaagpmprgppppgkpgddgetgkpgksgergppppqgargfpgtpgl
pgvkghrgypgl dgakgeagpagkgesgsppngsppmgprgl pgergrppsgaagargndgl pppa
ppppvpagapgfpgapgskgeagptgargpegaqpprgesgtppsgppagapgnpgtdgi pgakgsag
apgi agapgfpprgppppqatgpl gpkqgtgepgi agfkgeqqpkgetgpapppagppageegkrga
rgepgaagvpvppgergagprgrfpgqdg l agpkgapgerppagl agpkgatgdpprppgepgl pgargl t
grpgdagppqkvgtgappedrppppppqargqppvmgfpgpkgangepgkagekgl ppapgl rgl pg
kdgetgaagpppagvpgergeqgappsgfql pppppppgesgkpgdqgvpeagapgl vprgergf
pgergsppaagl qpprgl pgtpgtdgpkgatgpagpngaqqppgl qmpgergaagi agl kgdrdvgek
gpegapgdgargl tgpi ppppagpngekgesppppsgaagargapgergeppagppagfagppgadg
qpgakgeqqepgqkdagapppqpsgapppqptgvtgpkgar gaqppga tgfpaagrvppppngn
pppppppsagkdgpkgvrgdagppragdpgl qppagppgekgeppedgpapppppppqgl agqrgi v
gl ppprgergfpl ppsgepgkqgapsagdrppppvpvppgl tppageppregnpgadgl pgrdgaag
vkdrgetgvpvappagppagppvgtgkqdrgetgaqppmgsppagargmppppppprgdkgetge
agergl kghrgftgl qgl pppppsgdqgaagppagpprgppppvpsgkdgshgmpgpi gppgprgrs
geppagppgnppppppppppgtgi dmsafagl qktekppdi -----

-----rymradeaaggl rghd
vevdatl ksl nnqi esi rspegscknpartcrdl kl chpewksdyfi dnnqgctl dai kvfcmetget
cvyptpsi prknwtsktdkhhvwfaeti nggfhsygdnl spntasi qmtfl rl l stegsqnvtih
cknsi aymdeetgnl kkai l i qgsndvei raegnsrftysvl edgctkhtgkkgkti vi eyrl qktsrl si
vdtapmdi ggadqefgvdv gpvcfl

>' Gallus gallus (chicken)' type V alpha 2
al pgsppgdppppgl gppgl pgl kgdsgpkgekghpgl i gl i gppgeqqekgdrgl pppqgsagpkgegg
i tgpsgpi gppppppgl pppppkagkssgptgpkgesgl pppppppppgevi qpl pi qsskrtrni d
asq-----

gl y_unwei ghted_fasta. fasta

-----I vddgnadnymd
yadgmeei fgsi nsl kl ei eqmkhpl gtqhnpartckdl ql chpdfpdgeyvwdpnqgcsrdsfkvycnf
taggetci fpdkksegari tswpkenpgswfsefkrkgl l syvdsdgnpi gvvqmtfl r l l sasahqni t
yncyqsvawhdattdsyd kai rfl gsndeemsydnpyi raafdgc aakkgyqktvl ei ntpkveqvpi v
di mfndfgeasqkfgfevgpacfmg

>'Homo sapiens (human)' type I alpha 1
mfsfvdl r l l l l aatal l thgqeegqveggdedi ppi tcvqngl ryhdrdvwkpepcri cvcdngkvl c
ddvi cdetkncpgaevpegeccpvcpdgsesptdqettgvegpkgdtgprgprgpapppgrdgi pggppl

-----pgpppppppppppgl ggnfapql sygydekstggi svpgpmgppsgprgl pg
ppgappqqgfqpppgepgepgasgpmgprgppppgkngddgeagkpprgergppppqgargl pgtagl
pgmkghrgfsgl dgakgdagpagpkgepgsgengapqmgprgl pgergrpgapppagargndgatgaa
gpppgtgpagppgfpgavgakgeagppqprgseppqgvrgpgeppppagaagpagnpgadgppgakgang
appi agapgfpgargpsppqppppgpkngsgeppagpskgdtgakgeppvgvqpppppageegkr ga
rgepgptgl ppppgerggpgrgfpagadvagpkpgergsppgpagpkgspeagrpeagl pgakgl t
gsppsgppdgktgppppagqdrppppppgargqagvmgfpgpkgaaageppkagergvppppgavgpag
kdgeagaqqppppagpgergeqqpagspfgql ppppppgeagkpgeqqvpgdl gapppsgargergf
pgergvqppppagprgangapndgakgdagapagpsqgagpl qgmpgergaagl pppkgrgdagpk
gadgsppkdgvr gl t gpi gppppagapgdkgesgppagptgargapdrgepppppagfagppgadg
qpgakgeppdagakgdagppppagpappgpi gnvgapgakgargsagppgatgfpgaagrvgpppsgn
agpppppppagkeggkgrgetgpagrpevgpppppppagekgsppadgpapagtpgpqgi agqrgvv
gl ppprgergfpgl pppsgppkqppsgasgerppppmpppgl agppgesregappaegspprdgspg
akgrgetgpapppagpappagppvppagksgrgetgpapagppvgpvargpappppgrgdgkgetge
qgdr gi kghrgfsgl qppppppgspeqqpsgasppagprgppgsagapgdgl ngl pppi gpppprgt
gdagpvpppppppppppppsagfdfsfl ppppkekahdgg

-----yrraddanvvrdrdl
evdttl ksl sqqi eni rspegrknpartcrdl kmchsdwksgeywi dppnqcnl dai kvfcmetgetc
vyptqpsvaqknwyi sknpkdkrhvwfgesmtdgqfeyggqgsdpadvai ql tfl r l msteasqni tyh
cknsvaymdqqtgnl kkal l l qgsnei ei raegnsrftysvtvdgctshgawgkvti eykttktsr l pi
i dvapl dvgapdqefgfdvgpvcfl

>'Homo sapiens (human)' type II alpha-1
mi r l gapqsl vl l tl l vaavl rcqqqdvqeagscvqdqgryndkdvwkpepcri cvcdtgvtl cddi i ce
dvkdcl spei pfgeccpi cptdl atasgppgpkqkgeppgi kdi vgpkgppppqpppageqqpprgdrdk
gekgapgp

-----rgrdgepgtgnpppppppppppgl ggnfaaqmaggfdekaggaql gvmqgppmmpgprgppg
pagappqqgfqgnpgepgepgvsgpmgprgppppgkpgddgeagkppkagergppppqgargfpgtpgl
pgvkghrgypgl dgakgeagapgvkgesgspgengspgmpgprgl pgergrtpagaagargndgppgpa
gppppvppagppgfpgappakgeagptgarpegaqpprgepgtpgsppagasgnpgtdgi pgakgsag
appi agapgfpprgppppqgagtpgl gpkqqtgeppi agfkgeqqpkgeppagppqgappageegkr ga
rgeppgvpi gppgergppnrgfpgqdg l agpkgapgergppgl agpkgangdpprpepgl pgargl t
grpgdagppqkvpsgapgedgrppppppqargqppgvmgfpgpkgangeppkagekgl pgapgl rgl pg
kdgetgaagppppagpgergeqqpagspfgql pppppppgeggkpgdqgvpeagapgl vpprgergf
pgergsppaagl qpprgl pgtgtdgpkgasppppgaqpppgl qgmpgergaagi agpkgrgdvgek
gpegappkdgggl t gpi gppppagangekgevgppppagsagargapgergetgppppagfagppgadg
qpgakgeqgeagqkdagapppqpsgapppqptgvtgpkgar gaqpppgatgfpgaagrvgpppsngn
pppppppppsgkdgpkargdsgppgragepgl qppagppgekgeppgddgppsgaegpppppgl agqrgi v
gl ppprgergfpgl pppsgppkqgappgasdrppppvpppgl tppageppregspgadpppprdgaag
vkgdrgetgavgappppppsppagptgkqgdrgeagaqppmppsppagargi qpppppprgdkgeage
ppgerl kghrgftgl qgl ppppppsgdqgasppagppppppppvpppsgkdgangi pppi gpppprgs
getgpapppnpppppppppppgpi dmsafagl gprekdpdl

-----qymradqaaggl r qhd
aevdatl ksl nnqi esi rspegrknpartcrdl kl chpewksdywi dppnqct l damkvfcmetget
cvypnpanvpkknwsskskekhi wfgeti nggfhfsgygdnl apntanvqmtfl r l l stegsqni tyh
cknsi ayl deaagnl kkal l l qgsndvei raegnsrftytal kdgctkhtgkwgkvti eyrsqktsr l pi
i di apmdi ggpqefgvdi gpvvcfl

gl y_unwei ghted_fasta. fasta

>'Homo sapiens (human)' type III alpha 1
mmsfvqkgswl l l al l hpti i l aqceaveggcshl gqsyadrdrvkwpepcqi cvcdsgsvl cddi i cddg
el dcpnpei pfgeccavcpqptaptrppngqppqgpkgdppgppi pgrngdpgi pggpgspspgppgi
cesc-----

-----ptgpqnyspqydsydvksgvavvgl agypgpagpppppppgtsg
hpgspspgyqgppgepgqagpsgppppgai gpsgpagkdgesgrpgrpgergl pppgi kgpagi pgf
pgmkghrgfdgrngekgetgapgl kgengl pgengapgmprgapergrpgl pgaagargndgargsd
gpppppppgtagfpgspgakgevgpagspsngapqrgpqpqghagaqppppppi ngspgkgemg
pagi pgapgl mgargpppagangapgl rggagepgkngakgepprgergeagi pgvpgakgedgkds
pgepgangl pgaagergapgrfpgapngi pgekpgagergapppagprgaagepgrdgvppgmrgmp
gsppggsdkgpppsqgesrpppppsgrgppgvmgfpgpkgndgapgknger gppppppppppg
kngetppppppgtgppgdkgdtgppppqgl qgl pgtgpppgengkpgeppkgdagapppgkdaga
pgerppgl agapgl rggagpppeggkgaagppppgaagtgl qmpgergl gspgkdkgeppg
gadgvpkgdgrgptgpi gpppagppgdkggeggapgl pgi agrgspgergetgpppagfpgapng
epgkgergapgekgeggppgvagppgsgpagppppqgkgergsppggaagfpgargl ppppsngn
pppppsgsdkgppppagntgapsgvsgpkgdagppgekgsppaagppgagpl gi agi tgargl a
gppmpgprgsppgkvkgesgkpgangl sgerppppqgl pgl agtagepgrdgnpsdgl pgrdgspp
gkdrngensppagppgppgvpagksdrgesppagpagppagsrgappppprgdkgetge
rgaagi kghrgfpgnpgapsppagqgai gspgpagprgvpgsppgkdgtsghppgi gppprgpr
gergsegspghppppppppgagpccgvgvaaai agi ggekag-----

-----gfapyygdepmdfki nt
dei mtsl ksvngqi esl i spdgsrknparncrdl kfchpel ksgeywdpnqgckl dai kvfcnmetget
ci sanpl nvprkhwwtdssaekkhvwfgesmdggfqsygnpel pedvl dvhl afl r l l ssrasqni tyh
cknsi aymdqasgnvkkal kl mgsnegefkaegnskftytlv edgctkhtgewsktvfeyrtrkavr l pi
vdi apydi gpdqefgvdvpgvcfl

>'Mus musculus (house mouse)' type I alpha 1
mfsfvd l r l l l l gatal l thgqedi pevsci hngl rvpngetwkpevcl i ci chngtavcddvqcneel
dcpnpqrregeccafcpceeyvspnsedvgvegpkgdppgpprgrpvpppprdgi pggpgl pg-----

-----pppppppppppgl ggnfasqmsygydeksagvsvpmpgmpgprgl pg
ppgagppqgfqppgepgpessgpmprgppppgkngddgeagkprpgergppppqgargl pgtagl
pgmkghrgfsgl dgakgdagpagpkgepgspgengapqmgprgl pgergrpppgtagargndgavgaa
gpppgtptgppgfpavgakgeagppqargseppqvrgepppppagaagpagnpgadgppgakgang
aggi agapgfpgargpsppqpsgppgkngsgeppagpnkgdtgakgepgatgvqppppageegkrga
rgeppsgl pppgergppgsrgfpgadgvagpkpsgergapppagpkgspeagrpeagl pgakgl t
gspsppgdgktppppagqdrpppagppgargqagvmgfpgpkgtagepgkagergl pppgavgpag
kdgeagaqgappagpagergeqppagsppgfqgl pppagppgeagkpgeqgvpgdl gappsgargergf
pgergvqppppagprgngapngndgakgdtgapgapsqgagpl qmpgergaagl pppkdrgdagpk
gadgspgkdgargl tpi gpppagapgdkeagpsppppgtgargapdrgeagpppagfagppgadg
qppakgepgdtgvkgdagppppagppppgi gnvagppkprgaagppgatgfpagaagrvpppsgn
agppppppvkeggkprgetgpagrpgevgppppgpagekgsppgadgpagspgtppqgi agqrgv
gl pqrgergfppl ppsgepgkqpsgssgergppppmpppl agppgesregspgaegspgrdgag
akdrgetgpagppgagppagppvppagkngdrgetgpagpagpi gpagargpaggpprgdkgetge
qgdrgi kghrgfsgl qpppspspsgeqpsgasppagpppsagspgkdgl ngl pppi gppprgrt
gdsgpagpppppppppppppsgydfsfl ppppkeksqdggr-----

-----yyraddanvvrdrdl
evdttl ksl sqqi eni rspegsrknpartcrdl kmchsdkwsgeywi dnpqcnl dai kvycnmetgqtc
vfptqpsvpqknwyi spnpkekhhvwfgesmtdgfpfeygsegspadvai ql tfl r l msteasqni tyh
cknsaymdqqtgnl kkal l l qgsnei el rgegnsrftystl vdgctstgtwtgktvi eykttktsr l pi
i dvapl di gapdqefgl di gpacfv

>'Mus musculus (house mouse)' type II alpha 1
mi r l gapqsl vl l tl l i aavl rcqqdaqeagscl qngqrykdkdvkpscri cvcdtgnvl cddi i ce
dpdcl npei pfgeccpi cpadl atasgrkl gpkqkgepgdi rdgagpprgrdgepgtgnpppagp---

-----pgpppppgl sagnfaaqmaggydekaggaqmgvmqgppmmpgprgppg
pagagppqgfqgnpgepgpvsppmprgppppagkpgddgeagkpgksgergl pppmgarfgpgtpl

gl y_unwei ghted_fasta. fasta

pgvkghrgypgl dgakgeagapgvkgesgspgengspgmpgrgl pgergrtgpaagaargndgqppga
gppgpvpgagppgfpagpagkgeagptgarpegaqgsrgepgnpgspgpagasgnpgtdgi pgakgsag
aggi agapgfpgprgppgpgqatgpl gpkqagepgi agfkdgqpkgetgpapqpagppageegkrga
rgepggagpi gppgergapgnrgfpgqdl agpkgapgergpsgl agpkgangdpgrpgepgl pgargl t
grpgdagpqqkvpsgapgedgrppppgqgarqppgvmgfpgpkgangepgkagekgl agapgl rgl pg
kdgetgaagppgspgagergeqgagppsgfagl pppppppgeggkqgdqi pgeagapgl vgrgergf
pgergsppaagl qprgl pgtpgtdgpkgaagpdgppgaqppgl qmpgergaagi agpkdrgdvgek
gpegapgdggrgl tgpi gppgpagangekgeagppgsgstgargapgepgetgppgpagfagppgadg
qqgakgdqgeagqkdgadagppqgpgspgagppgtgvtgpkgarqaqppgagtgfpgaagrvgppgan
ppagpppgagkdgpkgvrgdsgppragdpgl epgapgekegpgddgpsgl dgppppgl agrgi v
gl pqrgergfpgl ppsgepgkqgagpgasgdrppgpvppgl tgpapeggregspgadgppgrdgaag
vkgdrgetgal gapgppgspgpagptgkqgdrgeagaqgmpgspgpagari agppgrgdkgesge
qgergl kghrgftgl qgl pppppsgdqgasgpagpsgrppppvpsgkdgsgni pgi gppgprgrs
getgpvppgspgppppppppgi dmsafagl qprekdpdm-----

-----qymradeadstl rhd
vevdatl ksl nnqi esi rspdgsrknpartcqd kl chpewksdywi dpnqgctl damkvfcmetget
cvypnpatvprknwsskskekhi wfgetmnggfhsygdgnl apntanvqmtfl rll stegsqni tyh
cknsi ayl deaagnl kkal l l qgsndvemraegnsrftytal kdgctkhtgkwgktvi eyrsqktsrl pi
i di apmdi ggaeqefgvdi gpvcfl

>'Oncorhynchus myki ss (rainbow trout)' type I alpha 1
mfsfvdi rll alll satvll argqgeddrtagstcl dgqfyndrdvkwpepcqi cvcdsgtvmcdevi ced
tsdcpnvi phdecpci cpddgfqepkvegpgqdrgakgepgpagfpgndgi pqqpl pppp-----

-----ppppgl ggnfspqmsggfdeksggmsmpgmpgmpgrppp
ppgssgppqgftgppgepgeagssgmpgrpagppgkngddgesgkprpgergasgpggargfpgtpgl
pgi kghrgfsgl dgakgesgpagpkgeggasgengaagamgrgl pgergragpngaagargndgaagaa
gppgptgpagapgfpgppgkgevgaaqgarggegpqsrgaenpgpagaagpagngadgnpgtkgapg
ssgi agapgfpgprgppgpgaggagppkngtgevngakgeagakgesgpagvqppagpageegkrge
rgepggagargapgergapsrgfpgsdgasgpkggpgerggagvagakngtgepgrngepgmpgskgmt
gspgspgdgktpgsagggdgrppppgvargppgvmgfpgpkgaagegkpergvmgpgsavgapg
kdgdvgapgapgvagpsgergeqgagppgfagl sgpqgai getgkpgeqgl pgeggapgsagsrgdrf
pgergappsgpagarspsagndggkgeagaagapggqppgl qmpgergaggl pgl kgdrgdqgvk
gadgaggkdgvrgmtgpi gpngpagspgdgetgapgavgpgsargapgergesgagpagfagppggd
qqgakgeagdnakgdggaqppagptgapppqpagntgakgargaagppgagtgfpgaagrfgppgsgn
ngppgtppgggkegqkngretgpagrpgel gaagppgpkgekqppgddgngpsgtppqgi gqrgi v
gl pqrgergfpgl aggl gepkqppgpfgerppgmpppgl agapgepgrtppnegssgrdgaag
pkgergesgvagasgappppgpagvpgksgdrgesgpagagi agpagprgspgpagargdkgesge
agergmkgghrgftgmqppgpgsgsgesgpagaprgpsgsagaagkdgmsgl pgi gppgprgrs
gemgsgtppppppppppppggfdmgfi aqpaqekapdfr-----

-----hfraddanvmrdrl
evdttl ksl sqqi eni rspegtkknpartcrtl kmchpdwksgeywi dpdqgctq dai kvycnmetgetc
vypteadi pkkswytskni kekhhvwfgeamtdgfqfeygsegsnakdvni ql tfl rll mateasqni tyh
cknsi aymdqqsgnl kksl l l qgsnei ei raegnsrftysvtedgctsthtgawgktvi dykttktsrl pi
i di apmdvgapnqefgi evgpvcfl

>'Oncorhynchus myki ss (rainbow trout)' type I alpha 2
ml sfdnrllll avtsll ascqs-----ggk kgprgakprgdrppqgnrdgkagl pgi agppppp
l ggnfaaqfdggkgsdppgmpgl mgsrgppngppgagppqgftghagepgepgqtsi gargptgsagkp
ge-----

-----dgnngrpgkpgdrppgtqgargfpgtpgl
pgmkghrgyngl dgrkgesgtagakgetgahgangsppagsrgl ngergragpagpagargadgstgpa
gpapgl gaagppgfpgapppkgei ggagsngpsgppggrgepgi ngavgpvgpvgngngi ngakgaag
l pgvagapgfpgprgppgppqgstgargl gdpgsgqkdgsgakgepghsgvqgaagpageegkrge
tgevgtgpagl rgarggagtrgl pgl egrgpi gmpgargatppgi rgapgdagragesgl tgargl p
gnsgqggppgkegppgaagl dgrtppppgtgprgppni gfpgkpggeagkggdkgptgatgl rggp
adgnngapppagvvgntgekgeqpagapgfagl pppagpageagkagnqgmpgdqgl pppagvkgern
sgpagsagsqgai gargpagtppdggkgepgsvgi vgaaghqppgmpgergaggtppkgekgegggr

gl y_unwei ghted_fasta. fasta

gl egnmrgdgarppgppsgpppsgangekgesgsfgpagpagl rgpsgergeggpagl pgfagppgsdg
qsgprgekgpaggkdvvpagpagppsgqsgpsgasgpagppggrgdagpsgl tgfpgaagrvgppgpagi
agppgsagpagkdgprgl rgdppgppqgeqqvvpagi sgdkgpsgesgpppaggtagppgvl gpsgfv
gl pgsrgdkgl pggpgavgeprl gpagasprgpagni gmpgmtgtqgeagregnsngndgppgrppaag
fkdrgeppsgal gssgppgngpagsagrppnr gesgptngggpvavgar gappagprgekggage
kgdrmgkgl rghggl qgmpgngpsgetgsagi tgpagprgpagphgppgkdgragghgai gpvghrgsp
ghl gpagpppsgl pppagpaggydqsggydeyradqpsfra-----

-----kdy
evdati ksl nsqi enl l tpegscknpartcrdi rl shpdwssgyfiw dnpqgci adai kaycdfstghtc
i hphpesi arknwy--rssenkkhwfgeti nggtefayndetl spqsmatql afmrl l anqatqni tyh
ckns vaymdgengl kkal l qgsndvel raegnsrftfnvl edgctrhtgqsktvi eyrtnkpsrl pi
l di apl di geadqefgl di gpvcfk

>' Oncorhynchus myki ss (rainbow trout)' type I alpha 3
mfsfvdi rl al l l saavl l argqgeddsgfgscsl dgqyndkdvwkpepcqi cvcdsgtvmcdevi ced
ttdcadpi i pdgeccpi cpddsesedpql pegppgppggdgpkgddgpegvpgndgvdgppgl -----

-----pgppppppppgl gggv fspqmhhgydqkssggqaqmqmmpmprgl ag
ppgntgppghhggpgepgeppgsgpmger gamgphgkngdngpgrsgrpgergsagppggrgtptgpl
pgi kghrgfngmdgakdggpagpkgeggspgssngsgamgargl pgergrppagssgargndngsgpa
gsngptgpsgagpfpaggakgetgpqggrgsdgpqgsrgeppnppagaagpsgapggdggaggkgs tg
aagi sgapgfpgtrgpagppsgggngpkgnhgehgaqqpkgepgvkgepgsagppgppggsgeegkr ga
rgeppgaggvppgergmpgmrgfpgsdgaaggkggppgergangpmaagga teagrppgeppmgskgmt
gspgspgpdgkvpagvtgndgrsppgssgargqppgvmgfpgpkgppgdngkppergeagavvggl gapg
kdgdsgapppqgpagsgqekgeggpagpngfql pgtqgatgepgkageqqgpggevvgppgsgsrgdrgf
pgerggi gsagpagprgangspgndgargesgaagapggmgapgl qgmpgergsngsgakgergdggpk
gadggpgkdgmrgmtgpi gppgptgahgekggggl gppgptggrgsppgergehgapppagfagppgadg
qpknkgetgnngpkgeagapppgppvgapppagpagnsgtkgtrgapppgasgmpgppggrvppggsa
pgsagpppagkegqkgrgetgi agrpgeagaagppgsgasgakngdgmgapgtppgppi agqrgi v
ggppgrgpsgtagi pgtagepgkqppgppvgergsppmppgl sgapeagregstghdgvsgrdgppg
pkdrgengnagsppagppgagppsgasgktsr gesgpagpagnppagrggl gpagargdkgeage
ggerghkghrgl ngmpgl ppspsgesgpagl ngpsgrgphgssgstgkdgmngl pgpi gppgprgrn
gkmaagapgspl pppppagggfhvapi ynqekgpdpmrggg-----

-----yhyradepdmmnrgdm
dvdttl ksl sqkveni rspegsqknparmcrdi rmchpewksqyvw dnpqgspl davqvycnmetgetc
vyqsqdtvpmknwytskni rekkhwfsesmdngfqqfygsegsnpedvni ql tfmrl msneasqnvtyh
cknsi ayndeatgnl kkal l l qganei ei raegnsrftysvsedgctsthtgtwgktvi dyktsktsrl pi
i di apmdvgapdqefgvevgpvcfv

>' Paralichthys olivaceus (bastard halibut)' type I alpha 1
mfsfvdl rl al l l saavl l vraqgeddraggscti dgqvfadr dvwkpepcqi cvcdsgtvmcdevi ced
ttdcpnvi phdecipi cpddgfqdsqvegpkgergpkdrgl pppgndgi pqqngl ppppg-----

-----pgpppgl ggnf spqmsggyddks-pgmsvppgmppmprgppg
pagssgppqgftgppgepgeagasgpmprgaagppgkngedgesgkpprggergppsgppgargfpgtgpl
pgi kghrgfsgl dgakgdsgpagpkgeagtpgengtpgamgprgl pgergrtgatgsagargndgaagaa
gppgptpagppgfpgppgakgeggpqsrgpegpagar gepgsagpagaagpagnpgtdgapgakgl pg
sagvagapgfppgrppgaggaagatgpkngtgeagapgakgesgakgesgvagvqppgppgsgeegkr ga
rgepgaagvrpppgerggppgrgfpgsdgsagakgapgergapgl l gpkgstgepprtgeagl pgakgmt
gspgspgpdgkmaagapqdgprppgsvgarqppgvmgfpgpkgaaageggkppgergvmgptgpvgapg
kdgdvgaqqpsgppagergeqgaagppgfql pppqgavgetgkppgeql pgeagatgpagargdrgf
pgergapi ngpagarpsppsgndgakgdagapagaagppgl qgmpgergsagl pgl rgdrgdqgak
gadgagpkdgrgmgtagi gl pgaagatgdkgeppaggi vppagargppgergesgpppagfagppgadg
ppgakgeagdngakgdagppgsgptgaagppgppvngtgpkgargpaggppgatgfpgsagrvgppgsgn
agpagppgatgkegqkgrgetgpagrpgemgaagl ppsgekgnpgeagppgssi pppqi nggrgi v
gl pqrgergfpl ppsgevgkhgppngngppmppgl agappeagregapgnegsagr dgaag
pkdrgesgpagasgppgagppvpagksdrgesgpagpagsagpagprgpagapgl rgdkgesge
agergmkgghrgftgmqppppagasdqqpagaagpgrgppsagsagkdgmngl ppgtgpppgrgrs
gemgpagpppppagapagppggfdl gfi aqp-qekapdpyr-----

gl_y_unwei ghted_fasta. fasta

-----mfraddanvl rdrdl
evdstl ksl sqqi eqi rspdgtrknpartcrdl kmchpdwksgeywi dpdqgctq dai kvycnmetgetc
vspsqrevakknwyvskni kek khvwfgeamndgfqfeygsegsapedvni qmtfl rl msteasqnvtyh
cknsvaymdasvgnl kkal l l qgsnei ei raegnsrftysvl edgctshgtwtgkvti dyktsktsrl pi
i di apmdvgapdqefgl evgpcvfl

>'Paralichthys olivaceus (bastard halibut)' type I alpha 2
ml sfvdtri l l l l avtsyl ascqs-----mgprgdkgprgdr gpkpdpdndgkpgl ppppppppppp
l ggnfaaqydgskgpdagppgmimgargppppppppppqghtghagepgepggtgal gprgpppppgka
ge-----

-----dgnngrpgkpgdr gapvqgargfpgtpgl
pgmkghrgygtl dgrkgepgsagakgesgahgaagspgl agsrgmagergragpagaagar gadgnvga
gpsgpl gaagppgfppgppkgei gpagatgpsgaqgsrgepgangavpvpgagnppangl ngakgaag
tpgvagapgfppgrgppppppqgaagprgl agdpgaqqvkgdggpkgepgnsppqgspgphgeegkrgp
tgepgatgpagsrgargaagsrgmpgaegrtgpvmppgargstgsggprgppgdagragepgsagl rgl p
gspgssppgkegpapagqdgrtgp ppptgprgppgni gfppkgasgeagkpkgekgatgptgl rgtpp
pdgnngatgamppagppgekeqgppsgapgfqgl pppagaageggkpgdr gi pddqgl apgakgergi
ngvagasvqgavgar gpapagppdgskepgi tgaaggpghgppgmpger gaagapggkgekeggghr
gpegnagrdgarmppagppgtgangdkgesgfgpapgprgasgergevgpagapgfagppgadg
qpgargergpagi kgevpspsgpagqsgpagnppagppgargdngpsgl tgfppagrvgtpgpagi
vgppgptgaagkdgprgprgdvpgggsgeqgmvgppgvgekppsgesppgagpagtggpl gl qgfv
gl pgar drgspggaggl gepgrvppppsgargppgni gl pgmtgpqgeagregnpndgppprgpagp
fkgrgepppagamgl agappagpsgaagrppnr gesppgaagsvppagargaagpagsrgekvgvgd
kgergmkl rghpgl qgmpgspgsgdtgaagpagssgprgpagphgpagkdgragghgti gspgarpp
gyvgpapppppgl pppagpsgggydvsg-ydeyradqpal ra-----

-----kdy
evdati ksl ntqi dnl l spegsrknpartcrdi kl shpewssgyfiwi dpnqgcsndai kvfcdfttretc
i yahpesi arknwf--rsteskkhvwfgeti nggtefayndetl spqsmatql afmrl l snqasqi tyh
cknsi aymdgergnmkkavvl qgsndvel raegnsrftfsvl edgctthtgewskvti eyrtnkpsrl pi
l di apl di ggadqefgl di gpcvfk

>'Rana catesbeiana (bullfrog)' type I alpha 1
mfsfvdtri l l l vi aati l vakcqeddl gysgcvvdgrtyndkdvwkpeacqi cvcdeg ti l cdevi ced
i gdcnppei pmgeccpvcgeggyqtgsvvegpkggetgprgergppgagprdgi ppppgi ppp-----

-----ppppppagl ggnfapqmsygydeksagi smpgpmgpmgprgppp
psgspgppqgfqpppgepgepgaagamgprgppppgkngedgagkpprgergppppqgargl pgtagl
pgmkghrgfngl dgakgdtgpagpkgepgnpgengappqvgrgl pgergppgspgpagargndgtpgaa
gppgptgptgppgfppgvpgkdagppgsrgpdgppgar gepagppagpagsagnpgtdgqpgakgatg
apgi agapgfpgargapppqppgspgpkgngepgaagkngepgakgesgpagsgppppppgeegkrgs
rgeppsgppppager gapgsrgfpgadgaggpkppgergvpvsagpkgspsgesrppgepl pgakgl t
gspgspgpdgktpagapggdgrppppppgargqsgvmgfpgpkaagepgkppgergvagppgavgapg
kdgeagaqpppppagpgergeqppagppgfqgl pgsppaggesgkpeqgagpdvpsgpagrgergf
ppergai gpppppprgangapndgakgeagapagppgppsgl qgmpgergaggl pgakgdrgdqgpk
gadgapgkdgvrgl tgi pppppgagpdgkeagpagpagptgsrgapgergepppsgpagfagppgadg
ppgakgeqgdagpkgdagppgaagptgapppagavgatgpkgar gpagppgstgfpgaagrvgpppsgn
agpppsgpagkegqkprgetgpagrppgepgaagpppsgekgspgsdgpagapgi ppppgi agtrgtv
gl pqrgergfpgl ppptgepgkqgssgspgergpppsgppgl agppgeppregspgsegspprdgsag
pkgdr gesppagppgagpagppagpagkngdrgetgpagpagpagpagargpsppagargdkgeage
qgergmkgghrgfndl pppppagphageqppsgasppagprpppgssgspgkdgsgnl ppgi gppppgrt
gdvpgappppagppppppppgggfdfsfmpqppqeksh-----

-----hyraddanamrdm
evdttl ksl skqi eni rspegtrknpartcrdl kmchsdwksgeywi dpnqgctl dai kvfcnmetgetc
vyptqsti dqknwyi snnprek khvwfgetmsdgf qfdygsegsdpadvni ql tfl rl mateasqi tyh
cknsvaymdqetgnl kkal l l qgsnei ei raegnsrftysvi edgctqhtgqwgkvti eyktpktsrl pi
tdvapmdi gapdqefgvei gpcvfv

gl y_unwei ghted_fasta. fasta

>' Rana catesbeiana (bullfrog)' type I alpha 2
ml sfvdl rsvl l l avtl yl vtcqevrrgprgdkgppgeqgppgi pgrdgedgl pgl ppppgvpgl ggnfa
aqydpksaepgq-----qgi mgprgppgppgspsqgfqgl pgengepgqtgpvgsrgpsgapgka
ge-----

-----dghpgksgrpgergpvpgqgargfpgtpgl
pgfkgi rghtgsdqkqgapgaagvkgengangdngsppqagargl pgergri gpagsagsrgsdgssgpv
gpapgi gsagapgl pgapgakgel gpagnngptgaaggrgepppgsl gpagppgnpgtngvngakgtag
l pggvgapgl pggri pppagpagpsgargl agdpgi agkgdtgskgepgsvqqppagpsgeegkrpp
ngeagssgpsnagi rgvpgtrgl pppdgraggi gpagsrgssgppgargpndagrppgepl l gargl p
gfsngppqgkegpapqgi egrsgaagpagargepgai gfpgpkgpngepgkngdkngqgsgnrgap
pdgnngaqqpagl ggatgekgeqgppsgapgfqgl ppppppgevgkppgergapgdfgppgsagtrgerga
pgeggagphgpsgrpsgagppdgqkgeppaagl nggl gpsgpagi pgergtagtpgtkgekdgagns
gdygnpgrdgargpagaagapppagppgdr gesgpagpsgvagprgapergeagpagptgfagppgaag
htgakgdrgakgpkgeagsppgl gahgsagpagnpagstgargdagpsgatgfpagragagppppgn
vgpsgptghpgkdgsrgprgdsppvgrpgeqqghgpvpl agdkgpsgeagpappgaagpsgl gargl l
gl pgrtrgergl pppgngnepgpgsl agssgprgppgsvgspgvghsgeagr dghpndgppgrdgl pg
akgergypgntgpgsl agapgpagsagpagksnrgeggpsgpagi tgpsgprgpagppgvr gdkgeage
rgargl dgrkghngl sgl pppsgtpgetgpgsgvppvgrgpgsgpsppgkegrsghpgamgpvgrgpa
gftgpagppppppghagpsgggydggdgyeyradqperkp-----

-----kdy
evdatl ksl nqqi evi l tpegsrknpartcrdl rl shpewtsgfywi dnpqgctsdai rvfcdffssetc
i hanpdei tqknwyi ntsnkdkkhl wfgei l nggtqfeyhdegl takdmatql afmrl l anqasqni tyh
cknsi aymdeetgnl kkavi l qgsndvel raegntrftysvl edgctkhtgewgktvi eyrtnkpsrl pi
l di apl di gghdqei gfei gpvcfk

>' Shrimp white spot syndrome virus' collagen
mayi dqgafgaksvmqqqyl l ppsnr l skqqpl asssl qpssnkprstsrvt di fvvi tcvvl i aafi
snsfsvtknvkl skeqteki l ekdl pdkvykl fenl kdgtfgi gedeeereergegeeel etkni rl k
rkwkemteqgdqgi kvrl hgrgergetgpagavgpapqgergai gpagkdgavgpapqgergai gp
agkdgavgpqppgergengrpprdgavg-----

-----ppqgergai gpagkdgavgpqgergai gpagkdgavgpapqg
ergengrpprdgavgpapppgergai gpagrdgavgpapppgergatgi pgrdgvdgsvgppgergei gr
pprdgavgpappprrgatgragkdgavgpapqgekgeagkdgsi gpqgi qpprgetgpprdgtaer
gergfppppgetgppgkdgvdgsegppqkrgetgpvpprgepl agl pgrdgai gpagppgergatgl pg
rngvdgsi ppqrrgatgragkdgavgpapppgergatgi pgrdgvdgsvgppgergetgpagrdgsvgp
agppgerengrpprdgatgi gpagppgekengrpprdgatgi pprgetgamgkngvdgsmgppgrr
gatgragkdgavgpapppgergetgpagrdgsvgpapqgetgl tgspprdgatgi gpagppgekeng
pprdgatgi gpagppgekengrpprdgatgi gpagppgetgl tgrpprdgatgi pprgetgamgk
ngvdgstgpprrgatgragkdgavgpapppgergengrpprdgatgi gpagppgetgl agl pgrdgai
gpqgekengrppkdgatgpmppgergetgi gpagppgatgl pgrdgvdgsvgppqkrgl i grtgrdg
ai gpvvpagpkgetgl agl pgi dgkdgsvgppgai gpi pprgergetpprdgedgstgpmppqgl rga
tgappqgergl kprgkdgetppprqrdgi mgprgl rkegapndgl egpegrdgapppagpi ppq
gi rgl kgi qgrpprdgemppagkdgi egprqdggtgagkprgl rfgqgrtgetgaqgsrgekdrgl tg
pprdgppgeegppgl rgergpprpprgi rrgspqgsngvqpprpprgtkgrtgi qgl tgi egprgp
rgi qgkegrmgki ghrgekgdkgdrgeqgi agadgekpprgl rgi rgpi gapgkpgtegvrgprgvr
gypgaqgel ppqptgppppagppppmgtgdtgpmppgavpprgekgrgkngpkgadgkdavni
i qkysi thara-----ei mwegnei geayi grsygttdi pvmi enri gmtned
kkneyci qvmtmhsi ttrgrtsgvfvsnktdy l l vt l l mpesvscrt dvstnarservnavreresk
yrfi rpsdqsi gthsrski avmypoasmsysvdtl dadvarrettsvl l l aeti hgekdrgyadrgrtv
grl mvppteel l vl qadpdi tpti adpi tvaevgnpvvri ktpscdel nkfmgfstnfsfl kyvkrfma
qassnsrndmsafl l dl i fekl i efl l stgwnreal ggttgai l mi rtatcr l aqsyyp l petvgmdfl
qtpgfdyksl pnneryi yefqatm

>' Streptococcus equi ' collagen
mi nkrkkkl i yrygvcsaavi l aal asl gtckeakal sgppgypl trdfsrnfl eentakyl dql rehl q
hrf sel esl trkl ekeggtrgpappppagpkgergeppvgpmpprgeqqpkl dgargpegqqgkppg
vgppppagpagkgetgpppprgdkgtgetgkqppagergpapppgprgdkgengapgeqgpi gpkget
gpkgdkgergekdgqqrgekdgqqrge-----

gl y_unwei ghted_fasta. fasta

-----I ygdnpnagkts
krytkedl kkdetafeal vei geai eafknptgtraapartckdl al ahpemedgvyvwdpnqgspi dai
evychi kthqtcvfakpsqvfkgsywsgpqkyvwfgedmdngfqfkykadrvqmtfl ql l steayqni ty
hcqnsapfynmqtqstdnaakfmtsndl el qanhkkftysvkl decqyki kkwasttfefrttktrrl pi
l di apydi ggpdaqafgl ei gpvcfs

>' R i f t i a p a c h y p t i l a ' c o l l a g e n a l p h a c h a i n
yragpryi q-----

-----aavvpi gprgppppgsppqqyyqgl r
gepgdsgpmgpi gkrppppagi agksdgdgrdgepprggi gpmgprgaggmpgmpgpxghrgfrgl sg
sxgeqgksgnqgpdggppagpsgpi gprgqtgergrdgksl pgl rgvdgl agppppppgi gstgspgf
pgtpgskgdrqsgl xgaqgl qgpvvl sqppgvagenghpgmpgndgangepgasgesgl pgpsgfpgpr
gmpgtagsppqagaxgdggptgeqgrpgagpvxgssgppgdvgapghageagkrsgpsgppagsgppqg
drgl pgsrgl pgmtgasgamgi pgekpsgepgakgptgdtgrqgnqgtpgi agl pgnpsdgrpgkdgr
pgi rgkdgkqgeqppqppgl agl qragppgargepgkngapgepgahgeqgdagkdgetgaagppgaa
gptgargppgprgqqgfqgl agaagtpgeagktgergavgatgpsgpagppgergapdrgnvgrmpg
ergatgpagptgspgvagakgqgppppagl vgl pgergpkgvvgsxgsrgdi gprgkagergkdgerge
rgengl pgpsgl aasxgergdmgsppgergsppagergpagsqi qgppppgdagpagtxgdi gfpger
gtrgatgkqgargprgl agkrgl rgaggsrgetgaqgei gl pgsppqgl pgpsgpppsgpagtagkqg
vxgargspgl vgkqgdrgsdgepprdgtxgergedgppgvsgptgapgqqgergmpgmvgi rgetgpmgg
qgmxdggpppsgdrgergnagppgptgpsgqagapggqegapgdgl pgl agrpgergepgvagragsq
gl agl mgqrgl pgaagppgdrgergeaggqvqgppvgapgsqppagi mgmxgeaggkaxgdkgw-----

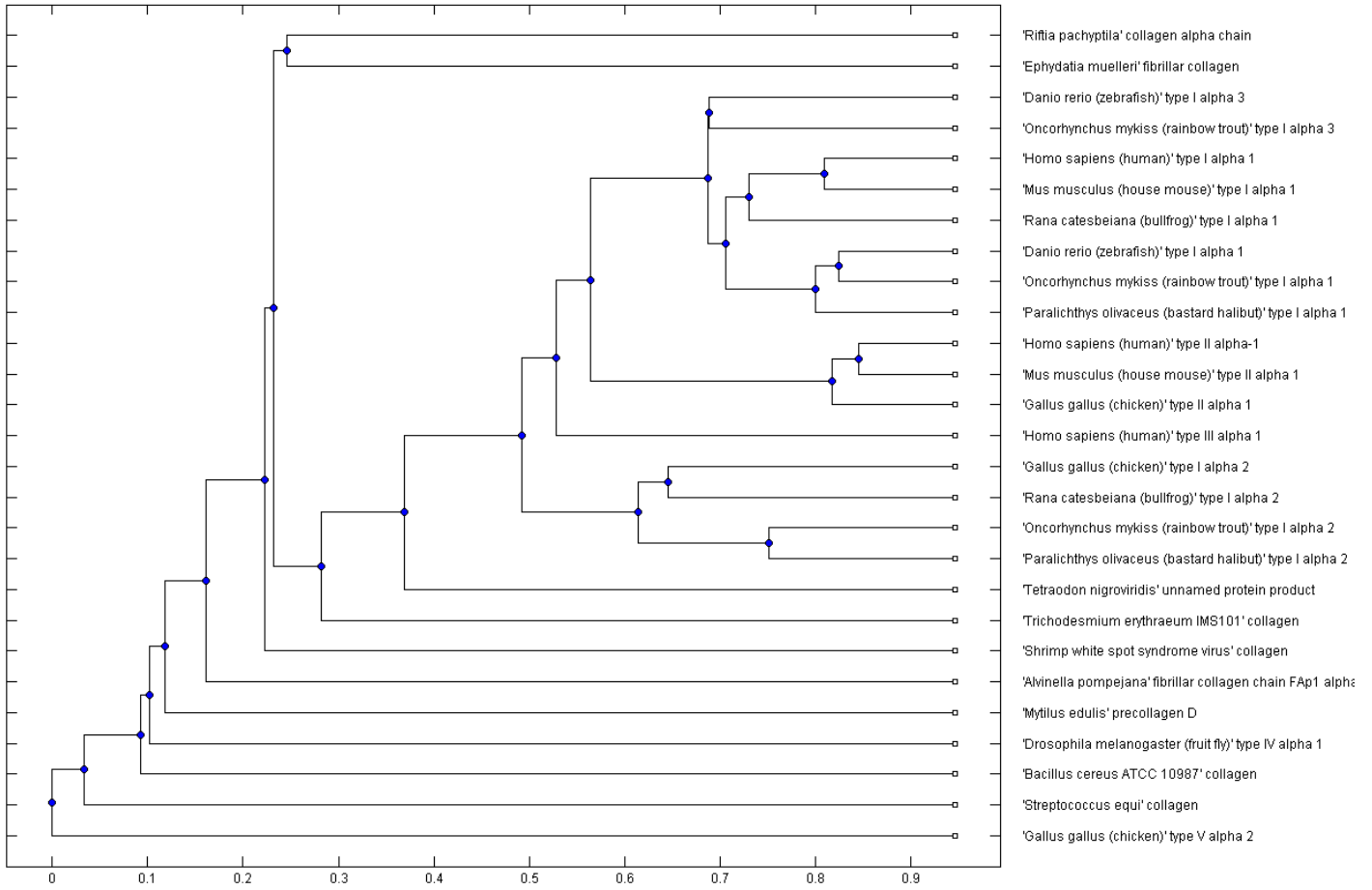
-----tgl pgl qgl qgtpghsesgppgagprgar geaggrgsqppgkdqppgsgrvgrgps
gddgrsgppppppppppnsdyga

>' E p h y d a t i a m u e l l e r i ' f i b r i l l a r c o l l a g e n
mgp-----

-----ppngapanpvppppppgagdpfsgl dgk
ngkpgnqgprgdtgppgsvgnagl pgapai mppgffdpkgpapel sgvkargspgl sgpsgpagprgdh
gdtgqpgi pghpgnqgaggi l qppgvpgvpgdtgpmgl pgvppgdgmpgqgsggqsfkgpl gpkgtvg
dqsqgdagppgaqgdqrrgappgakggvgal grpsagai gl pppageagpagtgpagpkgakggkgl
vgppgl pgl agppgdqgdagstgl tgappagl ngrdgapgdgl ngasgsi gi tgdi gl tgdaysgpi
gpnvppppppgdtgal qppgatgapgkegakqvfgppagpdrgl qgdmgmkgvkgesgdmgpagpag
vi gapgmkgppgkpgi pgl dgrl gakgtpgsnpprgpi gl pgvpgvdgkagaagapgvtagpkegrpgs
tgl sgaageppdgvgqsi gdtgsagpvgapgl tgptgskgpagaegtppmpgaagpagvtgdhgnvgqq
gpegprgpagptgppgi vgvpgdqqdvgvppqagaqppsgpagdvdrgl qgeagdagl pggsgqpgl pg
epgkegpnqgeagdnppqgaagepgl pgpvpl gapprgadgi agdagvpgi agl kgppppgnrga
l gpqgyqppppppppgpi gnrgpi gpsgptgpagpvndgnqqppgvpl pgdagmagl sgpmgedgda
gapypgeqgl pgtagl pspfa i qgdagaqgdqghppgi gapgatgpvpl pgi qvppngdvvpagptg
pagl dgapgaqppdgepgl pgl pgsqsksgasgppgvppvgaagkpgsi rgppppppppgd l grpgerg
akgvrtpgagpvdgvagi agai gfpgmpgdgaagpsgyppgfdgvagkppppgamgpkqgagerppgt
pgtqgskgvvpgkvvppqgdsgdtgdagqkgargtagsvgaktvpl pgnqppqppagl kvvk-----

-----ekgevgdkgi l gpdgdkgptgmsgdagpagpi gdagi qppppqdgptgaqpprggqgkpg
agavgdvgrgstgpagppppgt

unweighted tree



gl_y_wei ghted_fasta.fasta

>' Bacillus cereus ATCC 10987' collagen
mkerdkqnslnsnfri spnl i gptfppvptgftgi gi tgptgpqgptgpqgprgl qgpmgemgptgpqgv
qqi qgpvgsi gatgpegqqqppqgl rqpqgetgatgpqgvqgl qppagptgatgaqqi qgi qgl qgpi gat
gpegpqqi qgvqgl ppatgpqqi qgaqgmqgl qgpsngtgatga-----

-----tgqgi tgptgvtgptgi tgpsg
pppgptgptgatgppggpsgstgatgatgntgatgstgvtgvtgvtgatgstgptgstgaqgl qgi qgi q
gpi gptgpegpqqi qgi ppgtgvteqgi qgvqgi qgatgatgdqppqgi qqai gpqgatgatgdqppqg
i qgvppsgatgpqgvqgl qgpmgdi gptgpegpegl qppqgi qgvppvgatgpegpqqi qgi qgpvga
tgpqppqgi qgi qgvqgi tgatgvqatgi qgi qgei gatgpegpqqvqgaqqgi gptgpmgpqgvqgvq
gi qgatgaqgvqppqgi qgi qgi qgi qgptgatgdtgatgatgeggtgptgvtgptgpsggpagptgptg
psgpagvtgpsg-----gpppgptgatgatgvtgdtgatgstgvtgatgetgatgv
tgi qgpqqi qgvqgei gptgpqgvqppqgi qgvtgatgdqppqgvqppqgdi gptgpqqi qgpqgsqgi q
gatggtgaqppqgi qgpqgdvgptgpqgptgi qgi qgei gptgpegpegl qgpqqi qgvqgpvgtgpeg
pqqi qgi qgvqgatgsqppqgi qgi qgvqgi tgatgaqatgi qgi qgei gatgpegpqqvqgvqgei gp
tgpmpqgvqgvqgi qgatgaqgvqppqgi qgi qgptgatgetgatgatgeggtgptgvtgptgvtgpsg
gpagptgptgpsgpagvtgpsggppgptgatgatgatgvtgdtgatgstgvtgatgetgatgvtgi qgpq
gi qgvqgei gptgpqqi qgpqqi qgvtgatgaqppqgi qgpqgdi gptgpqqi qgpqppqgi qgatgatg
aqppqgi qgpqgei gptgpqppqgi qgpqqi qgptgatgatgatgi qgi qgpqqi qgpqqi qgptgatga
tgatgi qgi qgpqqi qgpqqi qgptgatgatgatgi qgi qgpqg-----

-----i qgpqqi qgptgatga
tgatgsqgptgdtgptgagatgatgatgvsttatyafanntsgtai svl l ggtni pl pnnqni gpgi tvs
ggntvftvasagnyyi ayti nl tagl l vssri tvngspl agti naptvatgsfsati i anl pagaai sl q
l fgl vai atl stttpgatl ti i rl s

>' Dani o reri o (zebrafi sh)' type I alpha 1
mfsfvdi rllal ll satvll argqgeddrtggsctl dgqvyndrdvkwpepcqi cvcdsgtvmcdevi ced
tsdcpnvi phdeccpvcppdddfqepsvegprgsppgdkgergpagppgndgi pgqpgl ppppg-----

-----ppgppgl ggnfspqmsggfdeks-spmavppgmpgmpgprgagp
ppgpsgpqgftgppgepgeagapgpmgprgaagppgkngedgsgkprpgergppgpqgar gfgtpgl
pgi kghrgfsgl dgakgdagpagpkgepgapgen tpgamgprgl pgergragppgaagar gndgaagaa
gppgptgpagppgfpggpgskgevpqgsrgaegpqqargeagnppagpagpagngadgapgakgapg
apgi agapgfpprgppgaagaagapgpkgntgeagapgakgeagakgeagaqgvqppppgpegkrp
rgepgaggar gptgergapgar gfpgadgaagprgaper gppgvvpgk gatgepgrngepmpgskgmt
gspgspgdgktgl agapgdgrppppgvargqppvmgfpgpkgaaageagkpgergvmgai gatgapg
kdgdvgapgapppagpagergeqgaagppgfqgl pppqgatgepgksgeqgapgeagapppsgsr gdrgf
pgergapppagpvgar gspgsagndgakgesgaagappaqppgl qmpgergaagl pgl kgdrgdqgak
gadgaagkdg i rgmtgpi gppgpagapgdkgesgaqgl vgptgarppgergetgapgpagfagppgadg
l pgakgepgdngakgdagapppagatgappqppvgtgpkgar gaagppgatgfpgaagr vppgppsgn
sgppppppagkegkgnrgetgpagrtgevgaaagppgapekgnppaegatgpagi pppqgi ggrgi v
gl pqrgergfppl ppsgei gkqpsgspsergppppmppgl agppgepgergtgnegsagr dgaag
pkdrgetgpsgtppagppgaagpi gpagktgdrgetgpagvppagpsgprgpsgpagargdkgetge
agergmkg hrgftgmpgppppppsgesgpagasgpagprgpagsagsagkdgmsgl ppgi gppgprgrn
gei gpagppppppppgagppsgggfdi gfi aqp-qekapdpfr-----

-----hfraddanvmrdrl
evdttl ksl sqqi esi i spdgtkknpartcrdl kmchpdwksgeywi dpdqcnq dai kvycnmetgetc
vnptesai pkknwytskni kek khwvfgeamtdgfqfeygsegskpedvni ql tfl rl msteasqni tyh
cknsi aymdqasnl kkal ll qgsnei ei raegnsrftysvtedgctsthtgawgktvi dykttktsrl pi
i di apmdvgapnqefgi evgpvcfl

>' Dani o reri o (zebrafi sh)' type I alpha 3
mfsfvdi rllgl ll aavvl avrgqgeddi sfgsctl egqnyndkdvkwpepcqi cvcdsgtvmcdevi ced
tsdcanpei pdgeccpi cpdeevtaapeppyvdtvgpagp-----rgdpgpvppgrdgtpgen-----

-----gl pppppppgtsgl gggsfl sqmaygsekssgppqppppgmpgprgl pg
pagsppqgftgppgepgepagpmpsrppppgkngedgagkagrppgergaagppqgar gfgtpgl

gl_y_wei ghted_fasta. fasta

pgi kghrgfpgl dgakgdagi agpkgeagspgeagtpgamargl pgergrpppppsgargndgntgaa
gppgptgpagppgfpggagakgetgaaggrgseppqgprgepgnppagaagpaggppsdtppgkgsppg
aagi agapgfpgargpppgapngagppkgnngdpgl qpkgdagpkgeppvppqgl agnageegkrga
rgepgaagpvpppgargapnrgl pgsdgappl gapgergangpmaaggasgesgrppgepl pgskgmt
gapgsppgdgkagpsgapqddgrpppppagsrgapgvmgfppkgt dgesgkppgergl agptgpl gapg
kdgdi gspgapgpapagekgeqgsagpsgfqgl pgaagatgetgkppgeqgl pgeagpsgppgrgdrf
pgergapppgptgprgapsppndgakgeppgaaganggvppagmqmpgergagmpgargdrdggpk
gpdgapgkdgvrmtgpi gppgspgapgdkeppgavpppl tprggppgergevgappagfagppgng
qqgakgetgdsapkdgagppagpvgasgppqpagatgpkgarggagppgatgfpgaagrvgppgpsga
sgppgptgpagkeqrgargekgpaprpgeagaagpppsgergsppdpgl gl agapppgvi gsrgi v
gl pqrgergfagl agpsgepgkqgppsvgerppppmpppl sgapgeagregnpghdgvsgdrdppg
pkdrgesgnsppgagppagppi gpsgktdrgeagpagpspagargal gpagprgdrgeage
agergmkghrgl sgmpgvppgptsggeqppagatgpagprgpagsngsppgkdgmnvgppvppppgrs
gemgpagapgl pppppppgsagppgepl pmpqqekgpdpl rgg-----

-----yraddasvrdrds
evdtsl ksl sqki eni rspegtqanparmcdrdl rmchpewksgsywvdpnqgspl dai kvfcnmetgetc
vspyhdsi prknwytskni rekkhwfgesmpngfqqfygsdgsdpedvni ql tfmrl msnqgsqni tyh
cknsi aymdetsgnl kkal l l qgsndi ei raegnsrftyrvsedgctshtgtwgtkvi dykttktsrl pi
i di apmdi gapdqefgvevppvcfl

>'Drosophila melanogaster (fruit fly)' type IV alpha 1
ml pfwkrl l yaavi agal vgadaqfwktagtagsi qdsvkhynrnepkfpi ddsydi vdsagvargdl pp
knctagyagcvpkci aekgnrgl ppgl gptgl kgemgfpgmegpsgdkgqkdpypqrgdkgergspg
l hgqagvppvqppagnppagpi ngkdgcddgdgi pgl egl sgmpgprgyagql gskgekgepakengdya
kgekgeppwrgtagl agppgfpgkeggersdsgpygakpprgehgl kgekgascygpmkppagpi kgekge
passfpvkpthtvmgprgdmqkgepgl vgrkgeppgedtgl dgqkgekgl pggpgdrgrqgnfppgs
tqkkgdrgepgl ngl pgnppqkgeppragatgepgl l gppppppgggrtppppgkpprgyvgappqgl
ngvdgl pppqyngqkdaggl pprpneppppgkgektagl ngpkgsi gpi ghpppppegqkdagl p
gygi qgskgdagi pgypl kgskgergfkgnagapgdskl grpgtpgaagapqkdgdrpgtppqkdm
gi kgdvggkcsscragpkgdktsgl pgi pgkdgarppppgergyppgerghdgi ngqtgppgekgedgrtg
l ppatgeppkpal cdl sl i epl kgdkgypgagkqvqgfkaegl pgi pppkgefgekgl sgapgn
dgtpragrddgypgi ppgsi kgeppfhrdgdakgdksfgrsgekgeppscal dei kmpakgnkeppgt
gmpppgedgspgergytgi kgntgppppgvepprgl ngprgekgnqgavgvppgnpgkdgl rgi pgrng
pppprgeppi srppmppppgl ngl qgekgrgptgpi gfpgadgsvgyppgdrdagl pgvsgppgi vgek
gdvpi gpagvappppvpi dgvrgrdgakgeppsgl vmpgnkgdrgapndgpkfagvtgapgkrg
papi pgvsgakgdkgasgl tndgppvgrppppagpi mgi kgdgl agapqaggl dmpgekgnqgfpgl
dpppl pgdasekqkgepppsgl rdtgpagtppwpekggl pgl avhgragppgekdgqrsi dgrdg
i ngekgeqgl qvvgppgeksgvppgi ppagmdgl pgaagapavgyppgdrgdkepgl sgl pgl kge
tppvl qgftgapppkgergi rqppl patvpi rdkgsqgergytgekgeqgergl tppagvakgd
rgl qpppgasgl ngi ppakgdi gprgei gypvgti kgekgl pprpgrnrgl i gapgl i gergl pgl pe
srl vgl pgi gpagskgergl agspgppqdgfppagpi kgdtpqgfkgergl ngfegqkdkgdrgl q
gpsgl pgl vqkgtgyppl ngndgppvppgergftgpkgrdgrdgtpl ppqkgeppml pppgkgepp
ppgrnppkgepprgergl i gi qqel gekgergl i getgnvppgkdrgeppgergygai gl i gqkge
ppagapal dyl tgl i trhsqetvpacsaghtel wtgysl l yvdgndyahnqdl gscvprfstl pvl s
cgqnnvcnyasrndktfwl ttnaai pmpveni ei rqi srcvceapanvi avhsqti evpdcppngweg
l wi gysfl mhtavnggggqal qspgsl edfratpfi ecngakgtchfyetmtsfwmyl essqperp
qqqt i kagerqshvsrvcvcmkns

>'Gallus gallus (chicken)' type I alpha 2
ml sfvdtri l l l l avtsyl atsqhvseasagrkgprgdgkqgergpppppprdgedgppppppppppp
l ggnfaaqydpaskaadfgpppmgl mgprpppgasgppppppgfqgvppgeppggtgpppprppppppgka
ge-----

-----dghpgkpprppergvaggargfpgtppg
pgfkgi rghngl dgl tqppagptkgeppagpengtppqppargl pgergri gapgpagarsdgsagpt
gpa-xxxxxxxxxxxxxxxxxgei gpagnegptgpapprgei gl pgsppvppppnppangl ppakgaag
l pvvagapgl ppprgi ppppppagpsargl vgepppagakgesgnkgeppgaagpppppppsgeegkrsg
ngeppagppppagl rpepprgl ppadragvmpagnrgasppvagakppngdagppgepl mpprxxx
xxxxxxxxxxxxxxxxxgfpadgrvpi gpagnrgeppni gfpkgtgeppkpekgvnl agpprapp
pegnnaqqppgvtgnqgakgetgpagppgfqgl ppspageagkppgergl hgefvpvppagppgergl
ppesgavpagpi gsrpsppppdgnkgeppnvgpapppppgi pgergvagvppgkgekagpl r

gl y_wei ghted_fasta. fasta

-----i dasql vddgnadnymd
yadgmeei fgsi nsl kl ei eqmkhpl gtqhnpartckdl ql chpdfpdgeyvwdpnqgcsrdsfkvycnf
taggetci fpdkksegari tswpkenpgswfsefkrkgl l syvdsdgnpi gvvqmtfl r l l sasahqni t
yncyqsvawhdattdsyd kai rfl gsndeemsydnpyi raafdgc aakkgyqktvl ei ntpkveqvpi v
di mfndfgeasqkfgfevgpacfmg

>'Homo sapiens (human)' type I alpha 1
mfsfvdl r l l l l aatal l thgqeegqveggdedi ppi tcvqngl ryhdrdvwkpepcr i cvcdngkvl c
ddvi cdetkncpgaevpegeccpvcpdgsesptdqettgvegpkgdtgprgprgpapppgrdgi pggppl

-----pgpppppppppppgl ggnfapql sygydekstggi svpgpmgppsgprgl pg
ppgappqqgfqpppgepgepgasgpmgprgppppgkngddgeagkpprpggergppppqgargl pgtagl
pgmkghrgfsgl dgakdagpagpkgepgsgengapqmgprgl pgergrpgapgpagarndgatgaa
gpppgtgpagppgfpgavgakgeagpqqprgseppqgvrgpgeppppagaagpagnpgadgppgakgang
appi agapgfpgargpsgppqppppgpkngsgepgapgskgdtgakgeppvgvqpppppageegkr ga
rgepgptgl ppppgerggpgrgfpgadgvagpkpgergsppgpagpkgspeagrpeagl pgakgl t
gspsgppdgktgppppagqdrppppppgargqagvmgfpgpkaagepgkagergvppppgavgpag
kdgeagaqqppppagpgergeqqpagspfgql pppppgeagkpeqqvpgdl gapgppsgargergf
pgergvqpppppagprgangapngdgakgdagapagsqgagpl qgmpgergaagl pppkgrgdagpk
gadgsppkdgvr gl t gpi gppppagapgdkgesgppagptgargapdrgepppppagfagppgadg
ppgakgepgdagakgdagppppagpappgpi gnvgapgakgargsagppgatgfpgaagrvgpppsgn
agpppppppagkeggkgrgetgpagrpevgpppppppagekgsppgadgpapagtpgpqgi agqrgvv
gl ppprgergfpgl pppsgepgkqppsgasgerppppmpppgl agppgesregapgaegspgrdgspp
akgdrgetgpagppgagpappgpvgpagksgrgetgpagpavgpvgargpappppgrgdgkgetge
qgdr gi kghrgfsgl qppppppgspeqqpsgasppagprgppgsagapgdgl ngl p gpi gppppgrt
gdagpvpppppppppppppsagfdfsfl ppppkekahdgr

-----yrraddanvvrdrdl
evdttl ksl sqqi eni rspegrknpartcrdl kmchsdwksgeywi dppnqcnl dai kvfcmetgetc
vyptqpsvaqknwyi sknpkdkrhvwfgesmtdgdfqfeyggqgsdpadvai ql tfl r l msteasqni tyh
cknsvaymdqqtgnl kkal l l qgsnei ei raegnsrftysvtvdgctshgtawgkvti eykttktsr l pi
i dvapl dvgapdqefgfdvvpvcfl

>'Homo sapiens (human)' type II alpha-1
mi r l gapqsl vl l tl l vaavl rcqqqdvqeagscvqdqgryndkdvwkpepcr i cvcdtgvtl cddi i ce
dvkdcl spei pfgeccpi cptdl atasgppgpkqkgepgdi kdi vgpkgppppqpppageqqprgdr gk
gekagpprgrdgepgtpg

-----npppppppppppppgl ggnfaaqmaggfdekaggaql gvmqgppmmpmmpgppp
pagappqqgfqgnpgepgepgvsgpmgprgppppgkpgddgeagkppkagergppppqgargfpgtpgl
pgvkghrgypgl dgakgeagapgvkgesgspgengspgmpgprgl pgergrtpagaagargndgqppga
gppppvppagppgfpgapgakgeagptgarpegaqpprgepgtpgsppagasnpgtdgi pgakgsag
appi agapgfpprppppqgagtpgl gpkqgtgepgi agfkgeqqpkgeppagppqgappageegkr ga
rgeppgvpi gppgergppnrgfpgqdg l agpkgapgergppgl agpkgangdpprpepgl pgargl t
grpgdagppqkvpsgapgedgrppppppqargqppgvmgfpgpkgangeppkagekgl pgapgl rgl pg
kdgetgaagppppagpgergeqqpagspfgql pppppppgeggkpgdqgvpeagapgl vpprgergf
pgergsppaagl qppr gl pgtgtdgpkgasppppgaqpppgl qgmpgergaagi agpkgrgdvgek
gpegappkdgg rgl t gpi gppppagangekgevgppppagsagargapgergetgppppagfagppgadg
ppgakgeqgeagqkdagapppqpsgapppqgptgvtgpkgar gaqpppgatgfpgaagrvgpppsngn
pppppppppsgkdgpkargdsgppgragepgl qppagppgekgepgddgppsgaegpppppgl agqrgi v
gl ppprgergfpgl pppsgepgkqgappgasdrppppvpppgl tgpapeppregspgadpppprdgaag
vkgdrgetgavgappppppsppagptgkqgdrgeagaqppmppsppagargi qppppgrgdgkgeage
ppgerl kghrgftgl qgl pppppsgdqgasppagppppppppvpppsgkdgangi p gpi gppppgrs
getgpagpppppppppppppppgi dmsafagl gprekpgdpl

-----qymradqaaggl r qhd
aevdatl ksl nnqi esi rspegrknpartcrdl kl chpewksdywi dppnqct l damkvfcmetget
cvypnpanvpkknwsskskekhi wfgeti nggfhfsgygdnl apntanvqmtfl r l l stegsqni tyh
cknsi ayl deaagnl kkal l l qgsndvei raegnsrftytal kdgctkhtgkwgkvti eyrsqktsr l pi
i di apmdi ggpqefgvdi gpvvcfl

gl y_wei ghted_fasta. fasta

>'Homo sapiens (human)' type III alpha 1
mmsfvqkgswl l l al l hpti i l aqceaveggcshl gqsyadrdrvkwpepcqi cvcdsgsvl cddi i cddg
el dcpnpei pfgeccavcpqptaptrppngqppqgpkgdppgppi pgrngdpgi pggpgspgspgppi
ce-----

-----scptgppqnydqydsydvksgvavagl agypgpagpppppppgtsg
hpgspgspgyqppgepgqagpsgppppgai gspgpagkdgesgrpgrpgergl pppgi kgpagi pgf
pgmkghrgfdgrngekgetgagpl kgengl pgengapgmprgagpgerppl pgaagargndgargsd
gpppppppgtagfpgspgakgevgpagspagsngapqrgpgeppqghagaqppppppi ngspgkgemg
pagi pgapgl mgargpppagangapgl rggagepgkngakgepprgergeagi pgvpgakgedgkds
pgepgangl pgaagergagpfrgpagpni pgekpgagergagpagprgaagepgrdgvppgmrgmp
gspgpgsdgkpppgsqesgrppppgsprgppgvmgfpgpkgndgapgknger gppppppppppg
kngtgpqpppgtgpaggkdgdtgppppqgl qgl pgtgpppgengkpgeppkgdagapppgkdaga
pgerpppl agapgl rggagpppeggkgaagppppgaagtpl qmpgergl gspgkdkgeppg
gadgvpkgdgrgptgpi gpppagppgdkgaggapgl pgi agrgspgergetgpppagfpgagqng
epggkgergagpekgeggppgvagppgsgpagppppqgvkgergsppggaagfpgargl ppppsngn
pppppsgspgkdgpppagntgapsgpvsppkdgagppgekgsppaagppgagpl gi agi tgargl a
gppmpgprgspgppgkgesgkpgangl sgerppppqgl pgl agtagepgrdgnpgsdgl pgrdgspg
gkdrngenspgagpghppppvpgpagksdrresgpagpagpagsrgappppprgdkgetge
rgaagi kghrgfpgnpgapsppagqgai gspgpagprgvpgpsppgkdgtsghppgi gppprgpr
gergsegspghppppppppgagpccggvgaai agi ggek-----

-----aggfapyygdepmdfki nt
dei mtsl ksvngqi esl i spdgsrknparncrdl kfchpel ksgeywdpnqgckl dai kvfcmnetget
ci sanpl nvprkhwwtdssaekkhvwfgesmdggfqsygnpel pedvl dvhl afl r l l ssrasqni tyh
cknsi aymdqasgnvkkal kl mgsnegefkaegnskftytvl edgctkhtgewsktvfeyrtrkavr l pi
vdi apydi gpdqefgvdvpgvcfl

>'Mus musculus (house mouse)' type I alpha 1
mfsfvd l r l l l l gatal l thgqedi pevsci hngl rvpngetwkpevcl i ci chngtavcddvqcneel
dcpnpqrregeccafcpceyvspnsedvgevgpkgdppgppgprgvpvppgrdgi pggpl pg-----

-----pppppppppppl ggnfasqmsygydeksagvsvpmpgmpgprgl pg
ppgagppqgfqppgepgpessgpmprgppppgkngddgeagkprpgergppppqgargl pgtagl
pgmkghrgfsgl dgakgdagpagpkgepgspgengapqmgprgl pgergrpppgtagargndgavgaa
gpppgtptgppgfpagvagakgeagppgargseppqvrgepppppagaagpagnpgadgppgakgang
aggi agapgfpgargpsppqpsppgkngsgeppagpnkgtgakgepgatgvqppppageegkrga
rgeppsgl pppgergppgsrgfpgadgvagpkpsgergagppagpkgspeagrpeagl pgakgl t
gspgspgdgktppppagqdrpppagppgargqagvmgfpgpkgtagepgkagergl pppgavgpag
kdgeagaqgagppagpgergeqppagsppgfqgl pppagppgeagkpgeqgvpgdl gapgspgargergf
pgergvqppppagprgngagpndgakgdtgapgapsqgagpl qmpgergaagl pppkdrgdagpk
gadgspgkdgargl tpi gpppagapgdkeagpspppgtgargapdrgeagpppagfagppgadg
qpgakgepgdtgvkdgadpppagppgpi gnvagppkprgaagppgatgfpgaagrvgpppsgn
agppppppvgeggkprgetgpagrpgevgppppgpagekgsppgagpagspgtppqgi agqrgv
gl pqrgergfppl ppsgepgkqpsgssgergppppmpppl agppgesgspgaegspgrdgag
akdrgetgpagppgagpaggppvpgagkngdrgetgpagpagpi gpagargpaggpprgdkgetge
qgdrgi kghrgfsgl qpppspspgeqpsgasppagprpppsagspgkdgl ngl pppi gppprgrt
gdsgpagpppppppppppppsgydfsfl ppppeksqdggr-----

-----yyraddanvvrdrdl
evdttl ksl sqqi eni rspegsrknpartcrdl kmchsdwksgeywi dnpqcnl dai kvycnmetgqtc
vfptqpsvpqknwyi spnpkekhhvwfgesmtdgfpfeygsegspadvai ql tfl r l msteasqni tyh
cknsaymdqqtgnl kkal l l qgsnei el rgegnsrftystl vdgctstgtwtgktvi eykttktsr l pi
i dvapl di gapdfeqgl di gpacfv

>'Mus musculus (house mouse)' type II alpha 1
mi r l gapqsl vl l tl l i aavl rcqqdaqeagscl qngqrykdkdvkpscri cvcdtgnvl cddi i ce
dpdcl npei pfgeccpi cpadl atasgrkl gpkqkgepgdi rdgagprgrdgepgtgnppgagp---

-----pgpppppppl sagnfaaqmaggydekaggaqmgvmqgppmmpgprgppg
pagagppqgfqgnpgepgpvsppmprgpppagkpgddgeagkpgksgergl pppmgarfgpgtpl

gl_y_weighted_fasta.fasta

pgvkghrgypgl dgakgeagapgvkgesgspgengspgmpgrgl pgergrtgpaagaargndgqppga
gppgpvpagppgfpagpagkgeagptgarpegaqgsrgepgnpgspgpagasgnpgtdgi pgakgsag
aggi agapgfppgrppppqgattgl gpkqageggi agfkdgqpkgetgpapppagppageegkrge
rgeppgagpi gppgergapgnrgfpgqdl agpkgapgergpsgl agpkgangdpgrpgepgl pgargl t
grpgdagppgkvpsgapgedgrppppqgarqppgvmgfpgpkgangepgkagekgl agapgl rgl pg
kdgetgaagpppspgergeqgagppsgfagl pppppppgeggkqgdqgi pgeagapgl vgrgergf
pgergsppaagl qpprgl pgtptdgpkaagpdpaggppgl qmpgergaagi agpkdrgdvgek
gpegapgdggrgl tpi gpppagangekgeagpppsgstgargapgepgetgppgpafagppgadg
qpgakgdqgeagqkdagappppsgapppgtgtvtpkgargaqppgatgfpgaagrvgppgan
ppagpppagkdgpkgvrgdsppgragdppl egpagapekgepgddgpsgl dgppppgl agrgi v
gl pqrgergfpl ppsgepgkqgapgasgdrppppvppgl tgpapeggregspgadgppgrdgaag
vkgdrgetgal gapgpppsppagptgkqgdrgeagaqgmpgspagargi agppgrgdkgesge
qgergl kghrgftgl qgl ppppsgdqgasppagppgrppppvpsgkdgsgni pppi gppgrgrs
getgpvpppsppppppppppi dmsafagl qprekdpdm-----

-----qymradeadstl rhd
vevdatl ksl nnqi esi rspdgsrknpartcqd kl chpewksdywi dnpqgctl damkvfcmetget
cvypnpatvprknwsskskekhi wfgetmnggfhsygdgnl apntanvqmtfl rll stegsqni tyh
cknsi ayl deaagnl kkal l l qgsndvemraegnsrftytal kdgctkhtgkwgktvi eyrsqktsrl pi
i di apmdi ggaeqefgvdi gpvcfl

>'Oncorhynchus myki ss (rainbow trout)' type I alpha 1
mfsfvdi rll al l satvl l argqgeddrtagstcl dgqfyndrdvwkpepcqi cvcdsgtvmcdevi ced
tsdcpnvi phdecpci cpddgfqepkvegpqgdrgakgeppagfpgndgi pppgl pppp-----

-----ppppgl ggnfspqmsggfdeksggmsmpgmpgmpgrppp
ppgssgppqftgppgepgeagssgmpgrpagppgkngddgesgkpprgergasgppgargfpgtppgl
pgi kghrgfsgl dgakgesppagpkgeggasgengaagamgrgl pgergragpngaagargndgaagaa
gppgptgpagapgfppppgkgevgaaqgargggeppqsrgeagnppagaagpagngadgnpgtkgapg
ssgi agapgfppgrppppqgaggapppkngtgevngakgeagakgesppagvqppagpageegkrge
rgeppgagargapgergapsrgfpgsdgasgpkggppgerggagvagakngtgepprgeppgmpgskgmt
gspgspgdgktpgsagggdgrppppvgarqppgvmgfpgpkgaageggkpergvmgpsgavgapg
kdgdvgapgavagpsgergeqgagppgfagl sgpqgai getgkpeqgl pgeggapgsagsrgdrf
pgergappsgpagargspgsagndggkgeagaagapggqppgl qmpgergaggl pgl kgdrgdqgvk
gadgaggkdvrgmtgpi gpngpagspgdgetgapgavgpsgargapgergesgapgafagppggdg
qpgakgeagdnakgdggaqppagptgapppqpagntgakgargaagppgatgfpgaagrfgpppsgn
ngppgtppgggkqgnrgetgpagrpgel gaagppgpkgekqppgdgngpsgtppqgi gqrgi v
gl pqrgergfpl aggl gepkqppppfgerppppmppgl agapgepgegtppnegssgrdgaag
pkgergesvagasgppppgavpagksgdrgesppagpagi agpagppgspagargdkgesge
agergmkgghrgftgmqpppsgqsgesppagasppagppgsagsagaagkdgmgl pppi gppgrgrs
gemgsgtppppppppppppppgfdmgfi aqpaqekapdfr-----

-----hfraddanvmrdrl
evdttl ksl sqqi eni rspegtkknpartcrl kmchpdwksgeywi dpdqgctq dai kvycnmetgetc
vypteadi pkkswytskni kekhhvwfgeamtdgfqfeygsegsnakdvni ql tfl rll mateasqni tyh
cknsi ayndqqsgnl kksl l l qgsnei ei raegnsrftysvtedgctsthtgawgktvi dykttktsrl pi
i di apmdvgapnqefgi evgpvcfl

>'Oncorhynchus myki ss (rainbow trout)' type I alpha 2
ml sfdnrll lll avtsl l ascqs-----ggl kgprgakpgrgdrppqppngrdgkagl pgi agppppp
l ggnfaaqfdggkgsdppppmgl mgsrgppppgppgqftghageppgeppqtsi gargptgsagkp
ge-----

-----dgngprgkpgdrppgtqgargfpgtppgl
pgmkghrgynl dgrkgesgtagakgetgahgangsppagsrgl ngergragpagpagargadgstgpa
gpapgl gaagppgfppagppkgei ggagsngpsppqpprgeppi ngavgpvppvgnppngi ngakgaag
l pgvagapgfppgrppppqppqgstgargl gdpgpsgqkdgsgakgepphsgvqgaagpageegkrge
tgevgtgpagl rgarggagtrgl pgl egrgppi gmpgargatppgi rgapgdagragesgl tgargl p
gnsgqppppgkeppgaagl dgrtppppgtpprgppni gfpgpkgppgeagkggdkgptgatgl rggp
adgnngapppagvvgntgekgeqppagppgfagl pppagpageagkagnqmpgdqgl pppagvkgern
sgpagsagsqgai gargpagtppdggkgeppsvgi vgaaghqppgmpgergaggtppkgekgegggr

gl y_wei ghted_fasta. fasta

gl egnmrgdgarggppgsgppgsgangekgesgsfgpagpagl rgpsgergeggpagl pgfagppgsdg
qsgprgekgpaggkdvvpagpagpssgqsgpsgasgpagppggrgdagpsgl tgfpgaagrvggppgpagi
agppgsagpagkdgprgl rgdppgppqgeqqvvgpagi sgdkgpsgesgpppagpagttagppgvl gpsgfv
gl pgsrgdkgl pggpgavgeprgl gpagasprgpagni gmpgmtgtqgeagregnsngndgppgrpgaag
fkgrgepgspgal gssgqppngpagsagrppnr gesgptngggpvavgar gappagprgekggage
kgdrgmkg l rghggl qgmpgpnpsgetgsagi tgpagprgpagphgppgkdgragghgai gpvghrgsp
ghl gpagpppsgl ppgpaggggydqsggydeyradqpsfra-----

-----kdy
evdati ksl nsqi enl l tpegscknpartcrdi rl shpdwssgyfiw dnpqgci adai kaycdfstghtc
i hphpesi arknwy--rssenkkhwfgeti nggtefayndetl spqsmatql afmrl l anqatqni tyh
ckns vaymdgengl kkal l qgsndvel raegnsrftfnvl edgctrhtgqwsktvi eyrtnkpsrl pi
l di apl di geadqefgl di gpvcfk

>' Oncorhynchus myki ss (rainbow trout)' type I alpha 3
mfsfvdi rl al l l saavl l argqgeddsgfgscsl dgqyndkdvwkpepcqi cvcdsgtvmcdevi ced
ttdcadpi i pdgeccpi cpddsesedpql pegpqqpqqgdgpkddgpegvpgndgvdgqpgl -----

-----pgppppppppgl gggv fspqmhhgydqkssggqaqmqgmppmprgl ag
ppgntgppghhgppgepgepggsgpmger gamgphgkngdngepgrsgrpgergsagpqqgrgtptgpl
pgi kghrgfngmdgakgdgppagpkgeggspgssngsgamgargl pgergrppagssgargndgngspa
gsngptgpsgagpfpaggagakgetgpqqgrgsdgpqgsrgepgnppagaagpsgapggdggaggkgs tg
aagi sgapgfpgtrgpagppsgggngpkgnhgehgaqqpkgepgvkgepgsagppgppgsggeegkr ga
rgepggaggvppgergmpgmrgfpgsdgaaggkggppgergangpmaagga teagr pgepgmpgskgmt
gspgspgpdgkvgpagvtgndgrsppgssgargqppgvmgfpgpkgppgdngkppergeagavvg l gapg
kdgdsgapppqgpagsgqekgeggpagpngfql pgtqgatgepgkageqqgpggevvgppgsgsr gdrgf
pgerggi gsagpagprgangspgndgargesgaagapggmgapgl qgmpgergsngngsakgergdggpk
gadggpgkdgmrgmtgpi gppgptgahgekggg l gppgptggrgspgergehgapppagfagppgadg
qpknkgetgnngpkgeagapppgppvgapppagpagnsgtkgtrgapppgasgmpgppgvrppgsga
pgsagpppagkegqkgrgetgi agrpgeagaagppgsgasgakgndgmpagp gtpgggi agqrgi v
ggppgrgpgstagi pgtagepgkqppgppvgergsppmppgl sgapgeagregstghdgvsgrdgppg
pkdrgengnagspgapppgagppsgasgktsr gesgpagpagnppagrggl gpagargdkgeage
ggerghkghrgl ngmpgl ppgpsgspgespagl ngpsgrgphgssgstgkdgmngl pgpi gppgprgrn
gkmaagapgspl pppppagggfhvapi ynqekgpdpmrgg-----

-----ggyhyradepdmmngrdm
dvdttl ksl sqkveni rspegsqknparmcrdi rmchpewksqywwdnpqgspl davqvynmetgetc
vyqsqdtvpmknwytskni rekkhwfsesmdngfqqfygsegsnpedvni ql tfmrl msneasqnvtyh
cknsi ayndeatgnl kkal l l qganei ei raegnsrftysvsedgctsthtgtwgktvi dyktsktsrl pi
i di apmdvgapdqefgvevgpvcfv

>' Paralichthys olivaceus (bastard halibut)' type I alpha 1
mfsfvdl rl al l l saavl l vraqgeddraggsct l dgqvfadr dvwkpepcqi cvcdsgtvmcdevi ced
ttdcpnvi phdecipi cpddgfqdsqvegpkgergpkdrgl pppgndgi pqqngl ppppg-----

-----pppppgl ggnf spqmsggyddks-pgmsvppgmpmmpgppg
pagssgppqgftgppgepgeagasgpmprgaagppgkngedgesgkpprggergpgspqgargfpgtgpl
pgi kghrgfsgl dgakgdsgpagpkgeagtpgengtpgamgprgl pgergrtgatgsagargndgaagaa
gppgptgpagppgfpggpgakgeggpqsrgpegpagar gepgsagpagaagpagnpgtdgapgakgl pg
sagvagapgfppgrppgaggaagatgpkngtgeagapgakgesgakgesgvagvqppgpgsgeegkr ga
rgepgaagvrppgerggppgrgfpgsdgsagakgapgergapgl l gpkgstgepgrtgeagl pgakgmt
gspgspgpdgkmaagapqdgprppgsvgarqppgvmgfpgpkgaaageggkppgergvmgptgpvgapg
kdgdvgaqqpsgspgagergeqgaagppgfql pppqgavgetgkppgeql pgeagatgpagargdrf
pgergapi ngpagar gspgsgndgakgdagapagaagppgl qgmpgergsagl pgl rgdrgdqak
gadgagpkdgrgmtgai gl pgaagatgdkgeppaggi vmpagargppgergesgppgpagfagppgadg
ppgakgeagdngakgdagppgsgptgaagppgpgvntgpkgar gpagppgatgfpgsagrvgppgsgn
agpagppgatgkegqkgrgetgpagrpgemgaagl ppgsgekgnpgeagapgsagi pppqi nggrgi v
gl pqrgergfpl ppgsgevghkpgspnngergppgmpgppgl agapgeagregapgnegsagr dgaag
pkdrgesgpagasgppgagppvpgagksdrgesgpagpagsagpagprgpagapgl rgdkgesge
agergmkgghrgftgmqqpppagasgdqppagaagpagprgpsgsagsagkdgmsgl ppgtgpppgrgrs
gemgpagppppgpagapagppgfdl gfi aqp-qekapdpyr-----

gl y_wei ghted_fasta. fasta

-----mfraddanvl rdrdl
evdstl ksl sqqi eqi rspdgtrknpartcrdl kmchpdwksgeywi dpdqgctqdai kvycnmetgetc
vpsqrevakknwyvskni kekhhvwfgeamndgfqfeygsegsapedvni qmtfl rl msteasqnvtyh
cknsvaymdasvgnl kkal l l qgsnei ei raegnsrftysvl edgctshgtwtgkvti dyktsktsrl pi
i di apmdvgapdqefgl evgpcvfl

>'Paralichthys olivaceus (bastard halibut)' type I alpha 2
ml sfvdtri l l l l avtsyl ascqs-----mgprgdkgprgdrpkpdpdndgkpgl ppppppppppp
l ggnfaaqydgskgpdagppgmimgargppppppppppqgghtghagepgepgqtgal gprgppppgka
ge-----

-----dgnngrpgkpgdrgapgvqgargfpgtpgl
pgmkghrgygtl dgrkgepgsagakgesgahgaagspgl agsrgmagergragpagaagaradgnvga
gpsgpl gaagppgfppgppkgei gpagatgpsgaqgsrgepgangavpvgpagnppangl ngakgaag
tpgvagapgfppgrgppppppqgaagprgl agdpgaqqvkgdggpkgepgnsppqgspgphgeegkrpp
tgepgatgpagsrgargaagsrgmpgaegrtgpvmppgargstgsggprgppgdagragepgsagl rgl p
gspgssppgkegpapagqdgrtgpmpgtgprgppgni gfpgpkgasgeagkpkgekgatgptgl rgtpp
pdgnngatgamgpaggpgekeqgppsgapgfqgl pppagaageggkppgdrgi pddqgl apagakgergi
ngvagasvqgavgarppagapgdgskgeppi tgaaggpghgppgmpgergaagapggkgekeggghr
gpegnagrdgarmppagppgtgangdkgesgsfppagpagprgasgergevgpagapgfagppgadg
qpgargergpapi kgevpspsgpagqsgppagppagppgargdngpsgl tgfppagrvtpppapi
vgppgptgaagkdgprgprgdvppggsgeqgmvgppgvgekppsgesppgagpaggtpgl gl qgfv
gl pgaragdrgspggaggl gepgrvppppsgargppgni gl pgmtgpqgeagregnpndgppprppagp
fkgrgepppagamgl agapppagpsgaagrppnrgepsppgaagsvppagargaagpagsrgekvggd
kgergmkl rghpgl qgmpgspgsgdtgaagpagssgprgpagphgpagkdgragghgti gspgarpp
gyvgpaggppppgl pppagpsgggydvsg-ydeyradqpal ra-----

-----kdy
evdati ksl ntqi dnl l spegsrknpartcrdi kl shpewssgyfiwi dppqgsndai kvfcdfttrec
i yahpesi arknwf--rsteskkhvwfgeti nggtefayndetl spqsmatql afmrl l snqasqi tyh
cknsi aymdgergnmkkavvl qgsndvel raegnsrftfsvl edgctthtgewskvti eyrtnkpsrl pi
l di apl di ggadqefgl di gpcvfk

>'Rana catesbeiana (bullfrog)' type I alpha 1
mfsfvdtri l l l vi aati l vakcqeddl gysgcvvdgrtyndkdvwkpeacqi cvcdegiti l cdevi ced
i gdcnppei pmgeccpvcgeggyqtgsvvegpkggetgprgergppgagprdgi pppppi ppp-----

-----ppppppagl ggnfapqmsygydeksagi smpgpmgmpgprgppp
psgspgppqgfqpppgepgepgaagamgprgppppgkngedgeagkpprgergppppqgargl pgtagl
pgmkghrgfngl dgakgdtgpagpkgepgnpgengappqvprgl pgergppgspgpagargndgtppaa
gppgptgptgppgfppgvppkgdaggppgsrppdppqgargeppagqagpagsagnpgtdgppgakgatg
apgi agapgfpgargapppppgppgkgnngpeggaqgnkgepgakgesgpagsqppppppgeegkrps
rgeppspgpppgergapsrgfpgadgaggpkpppgergppvgsagpkgspsgesrppgepl pgakgl t
gspgspgpdgktpagapggdgrppppppgargqsgvmgfpgpkgaaagepgkppgergvagppgavgapg
kdgeagaqpppppagpgergeqppagppgfqgl pgsppaggesgkpeqgagpddvpsgpagsrgergf
ppergai gpppppprgangapngdgaageagppagppgppsgl qgmpgergaggl pgakgdrgdqgpk
gadgapgkdgvrgl tppi gppppgagpdkgeagpagpagptgsrgapgergepppsgpagfagppgadg
ppgakgeqgdagpkgdagppgaagptgapppagavgatgpkgarppagppgstgfppaagrvgppgppsgn
agppppspgpagkegkqprgetgpagrppgeppgaagppgspgkgspsgdgpagapgi pppppi agtrgtv
gl ppprgergfpl ppptgepgkqgssgppgergppppgpppl agppgeppregspgsegspprdgsag
pkgdrgesppagppgagpaggpagpagkngdrgetgpagpagpagpagargpsppagargdkgeage
qgergmkgghrgfndl pppppagphageqppsgasppagppppgssgspgkdgsgnl pppi gppppgrt
gdvvpagppppagppppppppgggfdfsfmpqppqeksh-----

-----hyraddanamrdrdm
evdttl ksl skqi eni rspegtrknpartcrdl kmchsdwksgeywi dppqgcti dai kvfcnmetgetc
vyptqsti dqknwyi snprekhhvwfgetmsdgfqqfdygsegsdpadvni ql tfl rl mateasqi tyh
cknsvaymdqetgnl kkal l l qgsnei ei raegnsrftysvi edgctqhtgqwgkvti eyktpktsrl pi
tdvapmdi gapdqefgvei gpcvfv

gl y_wei ghted_fasta. fasta

>' Rana catesbeiana (bullfrog)' type I alpha 2
ml sfvdl rsvl l l avtl yl vtcqevrrgprgdkgppgeqqppgi pgrdgedgl pgl ppppgvpgl gg---
-----nfaaqydpksaepgqqgi mgprgppgppgspsqgfqgl pgengepgqtgpvgsrgpsgapgka
ge-----

-----dghpgksgrpgergpvpgqgargfpgtpgl
pgfkgi rghtgsdqkqgapgaagvkgengangdnsgppqagargl pgergri gpagsagsrgsdgssgpv
gpapgi gsagapgl pgapgakgel gpagnngptgaaggrgepppgsl gpagppgnpgtngvngakgtag
l pvggagpl pggri pppagpagpsgargl agdpgi agkgdtgskgepgsvqqppagpsgeegkrpp
ngeagssgpsnagi rgvpgtrgl pppdgraggi gpagsrgssgppgargpndagrppgepl l gargl p
gfsngppqgkegpapqgi egrsgaagpagargepgai gfpgpkgpngepgkngdkngqgsgnrgap
pdgnngaqqpagl ggatgekgeqqpsgapgfqgl ppppppgevgkppgergapgdfgppgsagtrgerga
pgeggagphgpsgrpsgapppdgqkgeppaagl nggl gpsgpagi pgergtagtpgtkgekdgagns
gdygnpgrdgargpagaagapppagppgdr gesgpagpsgvagprgapgergeagpagptgfapppgaag
htgakgdrgakgpkgeagsppgl gahgsagpagnpagstgargdagpsgatgfpagpagragapppgn
vpgspgtghpgkdgsrgprgdsppvgrpgeqqqhgpvpl agdkgpsgeagpagppgaagpsgl gargl l
gl pgrtrgergl pppgnggeppsgl agssgprgppsvgsppvghsgeagr dghpndgppgrdgl pg
akgergypntgpsgl agapgpagsagpagksnrgeggpsgpagi tgpsgprgpagppgvr gdkgeage
rgargl dgrkghngl sgl pppsgtpgetgpsgsvgppgprgspgspgppgkegrsghpgamgpvgrgpa
gftgpagppppppghagpsgggydggdgyeyradqperkp-----

-----kdy
evdatl ksl nqqi evi l tpegrknpartcrdl rl shpewtsgfywi dnpqgctsdai rvfcdffssetc
i hanpdei tqknwyi ntsnkdkkhl wfgei l nggtqfeyhdegl takdmatql afmrl l anqasqni tyh
cknsi aymdeetgnl kkavi l qgsndvel raegntrftysvl edgctkhtgewgktvi eyrtnkpsrl pi
l di apl di gghdqe i gfei gpvcfk

>' Shrimp white spot syndrome virus' collagen
mayi dqgafgaksvmqqqyl l ppsnr l skqqpl asssl qpssnkprstsrvt di fvvi tcvvl i aafi
snsfsvtknvkl skeqteki l ekdl pdkvykl fenl kdgtf gi gedeeereergegeeel etkni rl k
rkwkemteqqdqi kvrl hgrprgergetgpagavgpapppgergai gpagkdgavgpapppgergai gp
agkdgavgpqppgergengrp-----

-----grdgavgpqgergai gpagkdgavgpqgergai gpagkdgavgpapppg
ergengrpgrdgavgpapppgergai gpagr dgavgpapppgergatgi pgrdgvdgsvgpqgergei gr
pgrdgavgpappprrgatgragkdgavgpapppgekgeagkdgsi gpqgi qpprgetgppgrdgtaer
gergfppppgetgppgkdgdvdgsegppqkrgetgpvgrgepgl agl pgrdgai gpagppgergatgl pg
rngvdgsi ppprrgatgragkdgavgpapppgergatgi pgrdgvdgsvgpqgergetgpagr dgsvgr
agppqgergengrpgrdgatgpi gpagppgekengrpgrdgatgpi gprgetgamgkngvdgsmppqrr
gatgragkdgavgpapppgergetgpagr dgsvgpapppgetgl tgspprdgatgpi gpagppgekeng
rpgrdgatgpi gpagppgekengrpgrdgatgpi gpagppgetgl tgrpgrdgatgpi gprgetgamgk
ngvdgstgpprrgatgragkdgavgpapppgergengrpgrdgatgpi gpagppgetgl agl pgrdgai
gpqgekengrpdkdgatgpmppgergetgpi gpagppgatgl pgrdgvdgsvgpqgkrgl i grtgrdg
ai gpvvpagpkgetgl agl pgi dgkdgsvppqai gpi gprgergetgrpgrdgedgstgpmppqgl rga
tgappqgergl kprgkdgetgppgrqrdgi mgprgl rgekagpndgl egpegrdgapppagpi gpq
gi rgl kgi qgrpgrdgemgpagkdgi egprgdgttgakgprgl rfgqgrtgetgaqgsrgekgrgl tg
ppgrdgppgeegppgl rgergpprgprgi rrgspqgsngvqpprgprgtrkgrtgi qgl tgi egprgp
rgi qgkegrmgki ghrgekgdkgdrgeqgi agadgekprgl rgi rgpi gapgkpgtegrgprgvr gvp
gypgaqgel gpqptgpqppagppgmrtgdtgpmppgavgp-----rgek
ggrgrkngkngpkgadgdavni i qkysi tharaei mwegnei geayi grsygt di pvmi enri gmtnd
kkneyci qvmtmhsi ttrgrtsgvfvsnktdyi l l vt l l mpesvscrt dvstnarservnavreresks
yrfi rpsdqi gthsrski avmypoasmsysvdtl dadvarrettsvl l l aeti hgekdrgyadrgrtv
grl mvppteel l vl qadpdi tpti adpi tvaevgnpvvri ktpscdel nkfmf stnfsfl kyvkrfma
qassnsrndmsaf l l dl i fekl i efl l stgwnreal ggttgai l mi rtatcr l aqsyyp l petvgmdfl
qtpgf dynksl pnneryi yefqatm

>' Streptococcus equi ' collagen
mi nkrkkkl i yrygvcsaavi l aal asl gtckeakal sgppgypl trdfsrnfl eentakyl dql rehl q
hrf sel esl trkl ekeggtrgpappppagpkgergeppvpmgprgeqqpkl dgargpegqqgkppg
vg-----

gl_y_wei_ghted_fasta.fasta

-----rpkgpdpI ygddpnagkts
krytkedI kkdetafeal vei geai eafknptgtraapartckdl al ahpemedgvywvdpnqgspi dai
evychi kthqtcvfakpsqvfkgsywsgpqkyvwfgedmdngfqfkykadrvqmtfl qI I steayqni ty
hcqnsapfynmqtqstdnaakfmtsndl el qanhkkftysvkl decqyki kkwasttfefttktrrl pi
I di apydi ggpdaqafgl ei gpvcs

>'Ri fti a pachypti I a' collagen alpha chain
yra-----

-----gpriy qaavgpi gprgppppgsppqqyyqgl rgepgdsgpmgpi gkrppppagi a
gksdgdgrdgepprggi gpmgprgaggmpgmpgxhrgfrgl sgxgeqgksnqgpdggppagpsg
pi gprgqtgergrdgksl pgl rgvdgl agppppppgi gstgspgfpgtpgskdrgqsgl xgaqgl qgp
vgl sgqpgvagenghpmpgndgangepgasgesl ppgspgfpgrgmpgtagspgqagaxgdggptgeq
grpgagpvxgssgppgdvgapghageagkrsgpspgpagspgqgdrgl pgsrgl pgmtgasgamgi pg
ekgspgepgakgptgdtgrqgnqgtpgi agl pgnpsdgrppgkdgrppi rgkdgkqgeqppqppgl agl
qragppgargepgkngapgepgahgeqgdagkdgetgaagppgaagptgarpppprgqqfql agaq
gtpgeagktgergavgatgpsgpagppgergagpdrngvprgmpgergatgpagptgspgvagakggg
ppppagl vgl pgergpkvggsxgsrgdi gprgkagergkdgergergengl ppgsl aasxgergdmgs
pgergspgagergpagsqi qqqppppgdagpagtxgdi gfpgergtrgatgkqgarpprgl agkrgl r
gaggsrgetgaqgei gl pgsppqgl ppgspqppsgpagtagkqgvxgargspgl vgkqgdrgsdgepg
rdgtxgergedgppgvsgptgappqqgergmpgmvgI rgetgpmggqgmxdggpppsgdrgergnagp
qqptgspgqagapqegapgkdgl pgl agrpgergepgvagragsqgl agl mgqrgl pgaagppgdrger
geaggqvvgpvgapgsqppagi mgmxgeaggkxagdkgwgt-----

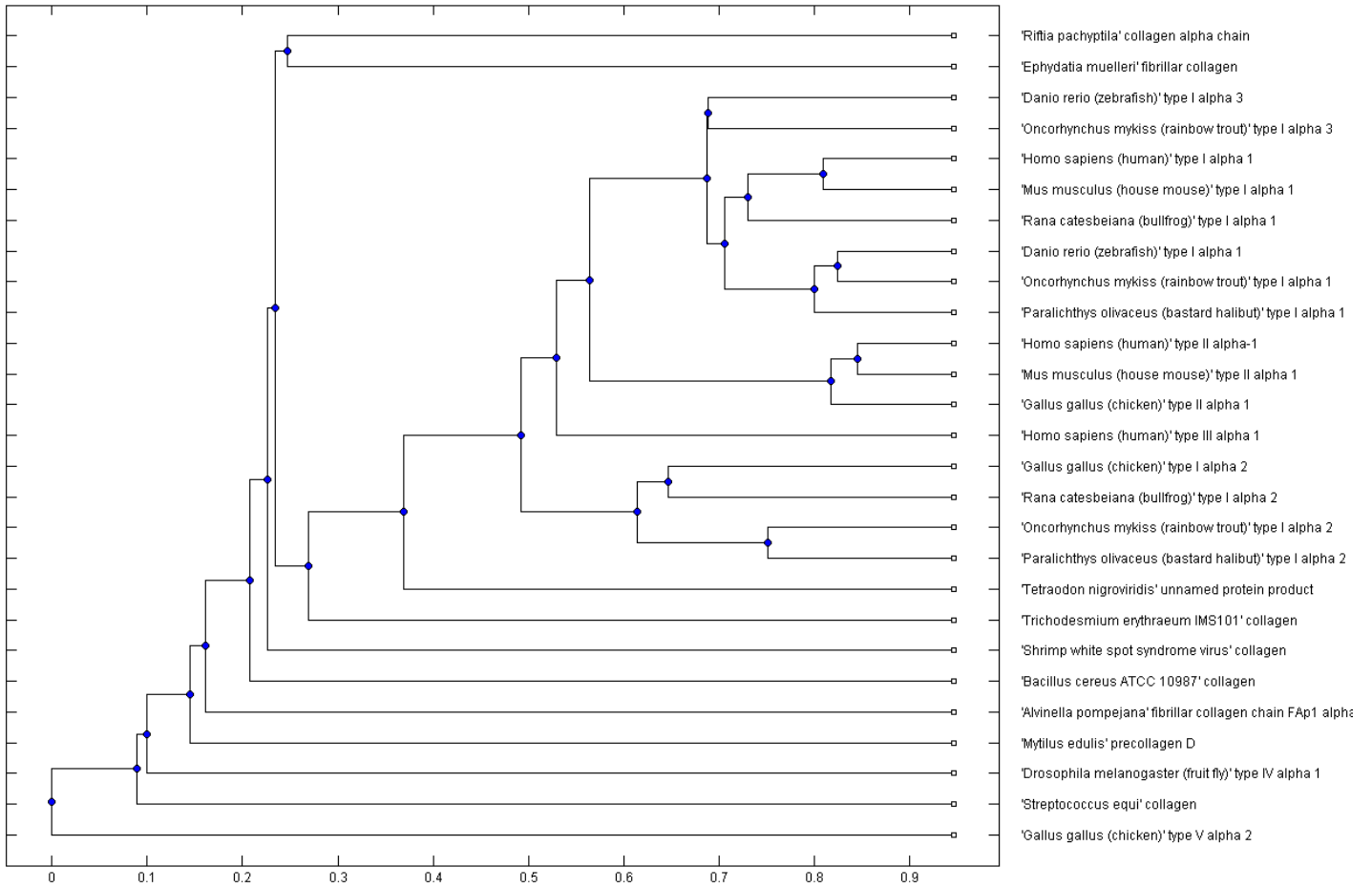
-----I pgl qgl qgtpghsesgppgagprgar geaggrgsqppgkdqppgsgrvgrgps
gddgrsgppppppppppnsdyga

>'Ephydati a muelleri' fibrillar collagen
mgp-----

-----ppngenganpvppppppgagdpgfsgl dgkngkpgnqgprgdtgppgsvgnagl
pgapai mppgffdpkgpapel sgvkgarpsgl sgpsgpagprgdhgdgqpgi pghpgnqgaggi I gqp
gvpgvpgdtgpmgl pgvpgpdgmpgqsgsgqsfkgl gpkgtvgdqgsqgdagppgaqgdqrrgagp
akggvgal grpsagai gl pppageagpagqtgpagpkgakggkgl vgpqgl pgl agpqqdqqdagstgl
tgappagl ngrdgapgkdgl ngasgsi gi tgdI gl tgdagsqgi ppgvppqppgpgdtgal qppgat
gapgkegkqvgfpgapgdrgl qgdmgmkgvkgesgdmgpagpagvi gapgmkgppgkpgi pgl dgrI g
akgtpgsnprgpi gl pgvpgvdgkagaagapgvtagpkegrpgstgl sgaageppgdgvqgsi gdtgs
agpvgapgl tgptgskgpagaegtppmpgaagpagvtdghngvqqqpegprgpagptgppgi vgvpgdq
gdvgvpgqagaqpsgpagdvdrgl qgeagdgl pggsgqgl pgepgkegvpvgnqgeagdngppqaag
epgl pgpvpl gapprgadgi agdagvpi agl kppppppnrgal gpqgyqppppqppmpgi gnrgp
i gpsgptgpagpvndgnqqppvpl pgdagmagl sgpmgedgdagapgyppgegl pgtgl pspfaI q
gdagaqgdqghppgi gapgatgpvl pgi qvppngdvpgagptgpagl dgapgaqppdgepgl pgl pg
qsgksgasgqppvppvgaagkpgsi rgqppppppgdI grpgergakgvrtpgagpvdgvagi agai g
fppmpgdgaagpsgyppfdgvagkppqgamgpkqagerppqgtpgtqgskgvvpgkvvppqgdsgd
tdagqkgargtagsvgakgtvl pgnqppqppagl kvvkgk-----

-----gevgdkgi I gpdgdkgptgmsgdagpagpi gdagi qppppqdgptgaqpprggqppkpg
agavgdvdrgstgpagppppgpt

weighted tree



GenPeptID	Species	Type	Chain	Sequence	Length
NP_981161	'Bacillus cereus ATCC 10987'	collagen		'mkerdkqnslnsnfrispnligptfppvpt	1321
AAH63249	'Danio rerio (zebrafish)'	type I	alpha 1	'mfsfvdirllllsatvllargqgeddrtggs	1447
AAH58045	'Danio rerio (zebrafish)'	type I	alpha 3	'mfsfvdirlllgaavvlavrgqgeddisfgs	1449
AAA28404	'Drosophila melanogaster (fruit fly)'	type IV	alpha 1	'mlpfwkrllyaaviagalvgadaqfwkta	1775
AAB64082	'Drosophila melanogaster (fruit fly)'	type IV	alpha 2	'mlprdrhlsglpggvvylglsvsvtladgki	1761 DELETED
P02467	'Gallus gallus (chicken)'	type I	alpha 2	'mfsfvdrlllllavtsylatsqhvseasagr	1362
AAK98621	'Gallus gallus (chicken)'	type II	alpha 1	'mhgrpprsaallllllltaaaaaqdrdrqr	1420
Q90689	'Gallus gallus (chicken)'	type V	alpha 2	'alpgspgpdgppgplgppgplgkgsdsc	390
P02452	'Homo sapiens (human)'	type I	alpha 1	'mfsfvdrlllllaatallthgqeegqvegqd	1464
AAC41772	'Homo sapiens (human)'	type II	alpha-1	'mirlgapqslvlltllvaavlrccgqdvqea	1487
P02461	'Homo sapiens (human)'	type III	alpha 1	'mmsfvqkgsdlllallhptiilaqqeaveg	1466
P11087	'Mus musculus (house mouse)'	type I	alpha 1	'mfsfvdrlllllgatallthgqedipevscihr	1453
AAA68100	'Mus musculus (house mouse)'	type II	alpha 1	'mirlgapqslvlltlliaavlrccgqdaqea	1459
BAB55661	'Oncorhynchus mykiss (rainbow trout)'	type I	alpha 1	'mfsfvdirllllsatvllargqgeddrtags	1449
BAB55663	'Oncorhynchus mykiss (rainbow trout)'	type I	alpha 2	'mfsfvdrlllllavtsllascqsgglkpgrg	1356
BAB55662	'Oncorhynchus mykiss (rainbow trout)'	type I	alpha 3	'mfsfvdirllllsaavllargqgeddsfgs	1458
BAB03287	'Pagrus major (red seabream)'	type V	alpha 1	'memqrwttrwtkrwlwdssltaavtlivi	1820 DELETED
CAE53096	'Paracentrotus lividus (common urchin)'	type	alpha 5	'mysfvdqslqhrklllilvalatsavicqgq	2795 DELETED
BAD77968	'Paralichthys olivaceus (bastard halibut)'	type I	alpha 1	'mfsfvdrllllsaavllvraqgeddragg	1447
BAD77969	'Paralichthys olivaceus (bastard halibut)'	type I	alpha 2	'mfsfvdrlllllavtsylascqsmgprgdk	1352
O93251	'Rana catesbeiana (bullfrog)'	type I	alpha 1	'mfsfvdrlllviaatilvakcqeddlgysg	1445
O42350	'Rana catesbeiana (bullfrog)'	type I	alpha 2	'mfsfvdrsvllavtlylvtcqevrrgprgdk	1355
AAK77699	'Shrimp white spot syndrome virus'	collagen		'mayidqgafgaksvmqqqyllppsnrls	1684
AAQ91575	'Streptococcus equi'	collagen		'minkrkkkliyrygvcsaavilaalasgtc	491
CP000393	'Trichodesmium erythraeum IMS101'	collagen		'medfldqivfptdssfqstpintlvpgesitv	1340
CAG02093	'Tetraodon nigroviridis'	unnamed protein product		'MFSFVDLRLALLLSAQGLLVRL	1399
AAB96638	'Mytilus edulis'	precollagen D		'MVYKLLTVCLVASLLEICLADY	922
AAC35289	'Alvinella pompejana'	fibrillar collagen chain FAp1 alpha		'MGPAGAPGLQGLTGPAGPPG	890
AAB24972	'Riftia pachytila'	collagen alpha chain		'YRAGPRYIQAQVGPVIGPRGPF	1027
1920343A	'Ephydatia muelleri'	fibrillar collagen		'MGPPGPNGEPANPVGPPGPF	1095

```
%Script for loading and analyzing data from excel file
```

```
[loaded_num_data loaded_text_data]=xlsread('LAYTON_TRICH_DATA.xls',1);  
loaded_text_data=loaded_text_data(2:end,:); %first line is headers
```

```
%create alternate scoring matrix  
unweighted_scoring=BLOSUM(80);  
%set (8,8) to zero .... remove gly weighting  
unweighted_scoring(8,8)=0;
```

```
names=cell(length(loaded_text_data),1);  
treated_seqs=cell(length(loaded_text_data),1);
```

```
for i=1:length(loaded_text_data)  
    names{i}=[loaded_text_data{i,2}, ' ', loaded_text_data{i,3}, ' ', loaded_text_data{i,4}];  
    temp=loaded_text_data{i,5};  
    treated_seqs{i}=lower(temp(isletter(temp)));  
end
```

```
seq_struct=struct('Header',names,'Sequence',treated_seqs);
```

```
mask=~ismember(1:length(seq_struct),[5 18 17]);
```

```
weighted_multialign_mat=multialign(seq_struct(mask),'verbose',true,'gapopen',@(sm,sx,len1,len2) 100*sm);  
unweighted_multialign_mat=multialign(seq_struct(mask),'ScoringMatrix',unweighted_scoring,'verbose',true,'gapopen',@(sm,sx,len1,len2) 100*sm);
```

```
weighted_distances=seqpdist(weighted_multialign_mat,'method','p-distance');  
unweighted_distances=seqpdist(unweighted_multialign_mat,'method','p-distance');
```

```
weighted_tree_dig=seqlinkage(weighted_distances,'single',seq_struct(mask));  
unweighted_tree_dig=seqlinkage(unweighted_distances,'single',seq_struct(mask));
```