

1)

```
>sp|Q03LF7|CAS9A_STRTD CRISPR-associated endonuclease Cas9 1 OS=Streptococcus
thermophilus (strain ATCC BAA-491 / LMD-9) OX=322159 GN=cas9-1 PE=1 SV=1
MSDLVLGLDIGIGSVGVGILNKVTGEIIHKNSRIFPAAQAENNLVRRTNRQGRRRLARRKK
HRRVRLNRLFEEGLITDFTKISINLNPYQLRVKGLTDELSNEELFIALKNMVKHRGISY
LDDASDDGNSSVGDYAQIVKENSQKLETKTPGQIQQLERYQTYGQLRGDFTVEKDGGKKHRL
INVFPTSAYRSEALRILQTQQEFNPQITDEFINRYLEILTGKRKYHGPNGNEKSRTDYGR
YRTSGETLDNIFGILIGKCTFYPDEFRAAKASYTAQEFNLLNDLNNLTVPTETKKLSKEQ
KNQIINYVKNEKAMGPAKLFKYIAKLLSCDVADIKGYRIDKSGKAEIHTFEAYRKMKTLE
TLDIEQMDRETLDKLAYVLTINTEREGIQEALHEFADGSFSQKQVDELVQFRKANSSIF
GKGWHNFSVKLMMELIPELYETSEEQMTILTRLGKQKTTSSSNKTKYIDEKLLTEEIYNP
VVAKSVRQAIKIVNAAIKEYGDFDNIVIEMARETNEDDEKKAIQKIQKANKDEKDAAMLK
AANQYNGKAELPHSVFHGHKQLATKIRLWHQQGERCLYTGKTISIHDLINNSNQFEVDHI
LPLSITFDDSLANKVLVYATANQEKQRTPYQALDSMDDAWSFRELKAFVRESKTLNKK
KEYLLTEEDISKFDVRKKFIERNLVDTRYASRVVLNALQEHFRAHKIDTKVSVVRGQFTS
QLRRHWGIEKTRDTYHHHAVDALIIAASSQLNLWKKQKNTLVSYSEDQLLDIETGELISD
DEYKESVFKAPYQHFVDTLKSKEFEDSILFSYQVDSKFNRKISDATIYATRQAKVGKDKA
DETYVLGKIKDIYTQDGYDAFMKIYKKDKSKFLMYRHDPQTFEKVIEPILENYPNKQINE
KGKEVPCNPFLKYKEEHGYIRKYSKKNGPEIKSLKYYSKLGHNHIDITPKDSNNKVVLQ
SVSPWRADVVFYFNKTTGKYEILGLKYADLQFEKGTGTYKISQEKYNDIKKKEGVDSSEFK
FTLYKNDLLLVKDTETKEQQLFRFLSRTMPKQKHVELKPYDKQKFEGGEALIKVLGNVA
NSGQCKKGLGKSNISIIYKVRTDVLGNQHIIKNEGDKPKLDF
```

1a)

No new sequences were obtained after the 2nd iteration.

2)

I chose 10 homologous sequences.

A) Streptococcus thermophilus

Length: 1121

Identity: 100%

>Q03LF7.1 RecName: Full=CRISPR-associated endonuclease Cas9 1; AltName: Full=Cas9**
[Streptococcus thermophilus LMD-9]

```
MSDLVLGLDIGIGSVGVGILNKVTGEIIHKNSRIFPAAQAENNLVRRTNRQGRRRLARRKKHRRVRLNRLFEEGLITDFT
KISINLNPYQLRVKGLTDELSNEELFIALKNMVKHRGISYLDASDDGNSSVGDYAQIVKENSQKLETKTPGQIQQLERYQ
TYGQLRGDFTVEKDGGKKHRLINVFPTSAYRSEALRILQTQQEFNPQITDEFINRYLEILTGKRKYHGPNGNEKSRTDYGR
YRTSGETLDNIFGILIGKCTFYPDEFRAAKASYTAQEFNLLNDLNNLTVPTETKKLSKEQKNQIINYVKNEKAMGPAKLF
KYIAKLLSCDVADIKGYRIDKSGKAEIHTFEAYRKMKTLETLDIEQMDRETLDKLAYVLTINTEREGIQEALHEFADGS
```

FSQKQVDELVQFRKANSSIFGKGWHNFSVKLMMEIPELYETSEEQMTILTRLGKQKTTSSSNKTKYIDEKLLTEEIYNP
VVAKSVRQAIKIVNAAIKEYGDFDNIVIAMARETNEDEKKAIQKIQKANKDEKDAAMLKAANQYNGKAELPHSVFHG
HK
QLATKIRLWHQQGERCLYTGKTISIHDLINNSNQFEVDHILPLSITFDDSLANKVLVYATANQEKGQRTPYQALDSMDD
A
WSFRELKAFVRESKTLNKKKEYLLTEEDISKFDVRKKFIERNLVDTRYASRVVLNALQEHFRAHKIDTKVSVVRGQFTS
QLRRHWGIEKTRDTYHHHAVDALIAASSQLNLWKKQKNTLVSYSEDQLLDIETGELISDDEYKESVFKAPYQHFDVTLK
SKEFEDSILFSYQVDSKFNKRISDATIYATRQAKVGKDKADETYVLGKIKDIYTQDGYDAFMKIYKKDKSKFLMYRHPDQ
TFEKVIEPILENYPNKQINEKGKEVPCNPFLKYKEEHGYIRKYSKKGNNGPEIKSLKYYDSKLGNHIDITPKDSNNKVVLQ
SVSPWRADVFNKTTGKYEILGLKYADLQFEKGTGTYKISQEKYNDIKKKEGVDSSEFKFTLYKNDLLLVDKTETKEQQ
LFRFLSRTMPKQKHVELKPYDKQKFEGGEALIKVLGNVANSQGCKKGLGKSNISYKVRTDVLGNQHIIKNEGDKPKLD
F

B) *Pasteurella multocida*

Length: 1056

Identity: 23.97%

>Q9CLT2.1 RecName: Full=CRISPR-associated endonuclease Cas9 [*Pasteurella multocida* subsp. *multocida* str. Pm70]

MQTTNLSYILGLDLGIASVGWAVVEINENEDPIGLIDVGVRIFERA EVPKTGESLALSRRRLARSTRRLIRRRHRLL
LAK
RFLKREGILSTIDLEKGLPNQAWELRVAGLERRLSAIEWGAVLLHLIKHRGYLSKRKNESQTNNKELGALLSGVAQN
HQL
LQSDDYRTPAELALKKFAKEEGHIRNQRGAYTHTFNRLDLLAELNLLFAQQHQFGNPHCKEHIQQYMTLLMWQKPA
LSG
EAILKMLGKCTHEKNEFKAAKHTYSAERFVWLTKLNNLRILEDGAERALNEEERQLLINHPYEKSKLTYAQVRKLLG
LSE
QAIFKHLRYSKENAESATFMELKAWHAIRKALENQGLKDTWQDLAKKPDLLDEIGTAFSLYKTDEDIQQYLTNKVPN
SVI
NALLVSLNFDKFIELSLKSLRKILPLMEQGKRYDQACREIYGHYGEANQKTSQLLPAIPAQEIRNPVVLRTLSQAR
KVI
NAIRQYGSPARVHIETGRELGKSKERREIQKQQEDNRTKRESAVQKFELFSDFSSEPKSKDILKFRLYEQQHGK
CLY

SGKEINIHLRNEKGYVEIDHALPFSRTWDDSFNNKVLVLASENQKGNQTPYEWLQGKINSERWKNFVALVLGSQCS
AAK
KQRLLTQVIDDNKFIDRNLNDTRYIARFLSNYIQENLLLVGKNKKNVFTPNGQITALLRSRWGLIKARENNNRHHAL
DAI
VVACATPSMQQKITRFIRFKEVHPYKIENRYEMVDQESGEIISPHFPEPWAYFRQEVNIRVFDNHPDVLKEMPLPDR
PQA
NHQFVQPLFVSRAPTRKMSGQGHMETIKSAKRLAEGISVLRIPLTQLKPNLLENMVNKEREPALYAGLKARLAEFNQ
DPA
KAFATPFYKGGGQVKAIRVEQVQKSGVLVRENNGVADNASIVRTDVFICKNNKFFLVPIYTWQVAKGILPNKAIVAH
KNE
DEWEEMDEGAKFKFSLFPNDLVELKTKKEYFFGYIIGLDRATGNISLKEHDGEISKGKDGVYRVGVKLALSFEKYQV
DEL
GKNRQICRPQQRQPVR

C) *Corynebacterium diphtheriae*

Length: 1,084

Identity: 17.07%

>Q6NKI3.1 RecName: Full=CRISPR-associated endonuclease Cas9; AltName:
Full=SaCas9 [*Corynebacterium diphtheriae* NCTC 13129]

MKYHVGIDVGTFVGLAAIEVDDAGMPIKTLSLVSHIHDSGLDPDEIKSAVTRLASSGIARRTRRLYRRKRRRLQQL
DKF

IQRQGWPIELEDYSDPLYPWKVRAELAASYIADEKERGEKLSVALRHIARHRGWRNPYAKVSSLYLPDGPSDAFKA
IRE

EIKRASGQVPETATVGQMVTLCELGTLKLRGEGVLSARLQQSDYAREIQEICRMQEIGQELYRKIIDVVFAAESP
KGS

ASSRVGKDPLQPGKNRALKASDAFQRYRIAALIGNLRVRVDGEKRILSVEEKNLVFDHLVNLTPKKEPEWVTIAEIL
GID

RGQLIGTATMTDDGERAGARPPTHDTNRSIVNSRIAPLVDWWKTASALEQHAMVKALSNAEVDDFDSPEGAKVQAFF
ADL

DDDVHAKLDSLHLPVGRAAYSEDTLVRLTRRMLSDGVDLYTARLQEFGIEPSWTPPTPRIGEPVGNPAVDRVLKTVS
RWL

ESATKTWGAPERVIIEHVREGFVTEKRAREMDGDMRRRAARNAKLQEMQEKLNVQKGKPSRADLWRYQSVQRQNCQC
AYC

GSPITFSNSEMDHIVPRAGQGSTNTRENLVAVCHRCNQSKGNTPFAIWAKNTSIEGVSVKEAVERTRHWVTDTGMRSTDF
KKFTKAVVERFQRATMDEEIDARSMESVAWMANELRSRVAQHFASHGTTVRVYRGS LTAEARRASGISGKLKFFDGV
GKS
RLDRRHHAIDAAVIAFTSDYVAETLAVRSNLKQSQAHRQEAPQWREFTGKDAEHRAAWRVWCQKMEKLSALLTEDLR
DDR
VVMSNVRLRLGNGSAHKETIGKLSKVKLSSQLSVSDIDKASSEALWCALTREPGFDPKEGLPANPERHIRVNGTHV
YAG
DNIGLFPVSAGSIALRGGY AELGSSFHARVYKITS GKKPAFAMLRVYTIDLLPYRNQDLFSVELKPQTMSMRQAEK
KLR
DALATGNAEYLGWLVDDELVDTSKIATDQVKAVEAELGTIRRWVVDGFFSPSKLRLRPLQMSKEGIKKESAPELS
KII
DRPGWLPVANKLFS DGNVTVRRDSLGRVRLESTAHLPVTWKVQ

D) *Campylobacter jejuni*

Length: 984

Identity: 25.06%

>Q0P897.1 RecName: Full=CRISPR-associated endonuclease Cas9 [Campylobacter jejuni subsp. jejuni NCTC 11168 = ATCC 700819]

MARILAFDIGISSIGWAFSENDELKDCGVRIFTKVENPKTGESLALPRRLARSARKRLARRKARLNHLKHLIANEFK
LNY

EDYQSFDESLAKAYKGSLISPYELRFRALNELL SKQDFARVILHIAKRRGYDDIKNSDDKEKGAILKAIKQNEEKLA
NYQ

SVGEYLYKEYFQKF KENSKEFTNVRNKKESYERCIAQSFLKDELKLIFKKQREFGFSFSKKFEEEVLSVAFYKRALK
DFS

HLVGNC SFFTDEKRAPKNSPLAFMFVALTRIINLLNNLKNTEGILYTKDDLNALLNEVLKNGTLTYKQTKKLLGLSD
DYE

FKGEKGYFIEFKKYKEFIKALGEHNLSQDDLNEIAKDITLIKDEIKLKKALAKYDLNQNQIDSLSKLEFKDHLNIS
FKA

LKLVTPLMLEGKKYDEACNELNLKVAINEDKKDFLPAFNETYKDEV TNPVVLRAIKEYRKVLNALLKKYGVHKIN
IEL

AREVGKNHSQRAKIEKEQNENYKAKKDAELECEKLGKINSKNILKLRLFKEQKEFCAYSGEKIKISDLQDEKMLEI
DHI

YPYSRSFDDSYMKNVLVFTKQNQEKL NQTPFEAFGNDSAKWQKIEVLAKNLPTKKQKRILDKNYKDKEQKNFKDRNL
NDT

RYIARLVLNYTKDYLDLPLSDDENTKLNDTQKGSKVHVEAKSGMLTSALRHTWGFSAKDRNNHLHHAIDAVIIAYANNS

IVKAFSDFKKEQESNSAELYAKKISELDYKNKRKFFEPFSGFRQKVLDKIDEIFVSKPERKKPSGALHEETFRKEEEFYQ

SYGGKEGVLKALELGKIRKVNGKIVKNGDMFRVDIFKHKKTNKFYAVPIYTMDFALKVLPNKAVARSKKGEIKDWILMDE

NYEFCFSLYKDSLILIQTKDMQEPEFVYYNAFTSSTVSLIVSKHDNKFETLSKNQKILFKNANEKEVIAKSIGIQNLKVF

EKYIVSALGEVTKAEFRQREDFKK

E) *Listeria innocua*

Length: 1334

Identity: 19.52%

>Q927P4.1 RecName: Full=CRISPR-associated endonuclease Cas9 [*Listeria innocua* Clip11262]

MKKPYTIGLDIGTNSVGWAVLTDQYDLVKRKMKIAGDSEKKQIKKNFWGVRLFDEGQTAADRRMARTARRRIERRNRIS

YLQGIFAEEMSKTDANFFCRLSDSFYVDNEKRNSRHPFFATIEEEVEYHKNYPTIYHLREELVNSSEKADLRLVYLA LAH

IIKYRGNFLIEGALDTQNTSVDGIYKQFIQTYNQVFASGIEDGSLKKLEDNKDVAKILVEKVTRKEKLERILKLYPG EKS

AGMFAQFISLIVGSKGNFQKPFDLIEKSDIECAKDSYEEDLESLLALIGDEYAELFVAAKNAYSAVVLSSIIITVAET ETN

AKLSASMIERFDTHEEDLGELKAFIKLHLPKHYYEIFSNTTEKHGYAGYIDGKTKQADFYKYMKMTLENIEGADYFIA KIE

KENFLRKQRTFDNGAIPHQLHLEELEAILHQQAKYYPFLKENYDKIKSLVTFRIPYFVGPLANGQSEFAWLTRKADG EIR

PWNIEEKVDFGKSAVDIEKMTNKDTYLPKENVLPKHSLCYQKYLVYNELTKVRYINDQGKTSYFSGQEKEQIFNDL FKQ

KRKVKKKDLELFLRNMSHVESPTIEGLEDSFNSSYSTYHDLLKVGIKQEILDNPVNTEMLENIVKILTVFEDKRMIEQL

QQFSDVLDGVVLKKLERRHYTGWGRLSAKLLMGIRDKQSHLTILDYLMNDDGLNRNLMQLINDSNLSFKSIEKEQV TTA

DKDIQSIVADLAGSPAIIKKGILQSLKIVDELVSVMGYPPQTIVVEMARENQTTGKGKNNSRPRYKSLEKAIKEFGSQ ILK

EHPTDNQELRNNRLYLQNGKDMYTGQDLDIHNLSDYDIDHIVPQSFITDNSIDNLVLTSSAGNREKGDDVPPL
IVR
KRKVFWEKLYQGNLMSKRKFDYLTKAERGGLTEADKARFIHRQLVETRQITKNVANILHQRFNYEKDDHGNTMKQVR
IVT
LKSALVSQFRKQFQLYKVRDVNDYHHAHDAYLNGVVANTLLKVYPQLEPEFVYGDYHQFDWFKANKATAKKQFYTNI
MLF
FAQKDRIIDENGEILWDKKYLDTVKKVMSYRQMNIVKKTEIQKGEFSKATIKPKGNSKLIPRKTNWDPMKYGG LDS
PNM
AYAVVIEYAKGKNKLVFEKKIIRVTIMERKAFKDEKAFLEEQGYRQPKVLAKLPKYTLYECEEGRRRMLASANEAQ
KGN
QQVLPNHLVTLHHAANCEVSDGKSLDYIESNREMFAELLAHVSEFAKRYTLAEANLNKINQLFEQNKEGDIKAIQA
SFV
DLMAFNAMGAPASFKEFFETTIERKRYNNLKELLNSTIIYQSITGLYESRKRLDD

F) *Actinomyces naeslundii*

Length: 1101

Identity: 15.81%

>J3F2B0.1 RecName: Full=CRISPR-associated endonuclease Cas9; AltName:
Full=AnaCas9 [Actinomyces naeslundii str. Howell 279]

MWYASLMSAHHLRVGIDVGTHSVGLATLRVDDHGTPIELLSALSHIHDSGVGKEGKKDHDTRKKLSGIARRARRLLH
HRR
TQLQQLDEVLRDLGFPIPTPGEFLDLNEQTDPIRVWRVRARLVEEKLPEELRGPAISMAVRHARHRGWRNPYSKVE
SLL
SPAEESPFMKALRERILATTGEVLDDGITPGQAMAQVALTHNISMARGPEGILGKLHQSDNANEIRKICARQGVSPDV
CKQ
LLRAVFKADSPRGSAVSRVAPDPLPGQGSFRRAPKCDPEFQRFRIISIVANLRISSETKGENRPLTADERRHVVTFLT
EDS
QADLTWVDVAEKLGVHRRDLRGTA VHTDDGERSAARPPIDATDRIMRQTKISSLKTWWEADSEQRGAMIRYLYEDP
TDS
ECAEIIAELPEEDQAKLDSLHLPAGRAAYSRESLTALSDHMLATTDDLHEARKRLFGVDDSWAPPAEAINAPVGNPS
VDR
TLKIVGRYLSAVESMWGTPEVIHVEHVRDGF TSEMA DERDKANRRRYNDNQEAMKKIQRDY GKEGYISRGDIVRLD
ALE
LQGCACLYCGTTIGYHTCQLDHIVPQAGPGSNNRRGNLVAVCERCNRSKSNTPFVWAQKCGIPHVGVKEAIGRVRG
WRK

QTPNTSSIDLTRLKKEVIARLRRTQEDPEIDERSMESVAWMANELHHRIAAAYPETTVMVYRGSITAAARKAAGIDS
RIN
LIGEKGRKDRIDRRHHAVDASVVALMEASVAKTLAERSSLRGEQRLTGKEQTKWQYTGSTVGAREHFEMWRGHMLHL
TEL
FNERLAEDKVYVTQNIRLRLSDGNAHTVNPSKLVSHRLGDGLTVQQIDRACTPALWCALTREKDFDEKNGLPAREDR
AIR
VHGHEIKSSDYIQVFSKRKKTDSDRDETPFGAIAVRGGFVEIGPSIHHARIYRVEGKKPVYAMLRVFTHDLLSQRHG
DLF
SAVIPPQSISMRCAPKLRKAITTTGNATYLGWVVVGDELEINVDSFTKYAIGRFLEDFPNTTWRICGYDTNSKLT
KPI
VLAAEGLNPSSAVNEIVELKGWRVAINVLTKVHPTVVRDALGRPRYSSRSNLPTSWTIE

G) *Neisseria meningitidis*

Length: 1082

Identity: 23.35%

>C9X1G5.1 RecName: Full=CRISPR-associated endonuclease Cas9 [*Neisseria meningitidis* 8013]

MAAFKPNSINYILGLDIGIASVGWAMVEIDEEENPIRLIDLGVRFERA EVPKTGDSLAMARRLARSVRRLTRRRRAH
RLL
RTRRLKREGVLQAANFDENGLIKSLPNTPWQLRAAALDRKLTPLEWSAVLLHLIKHRGYLSQRKNEGETADKELGA
LLK
GVAGNAHALQTGDFRTPAELALNKFEEKESGHIRNQSDYSHTFSRKDLQAELILLFEKQKEFGNPHVSGGLKEGIET
LLM
TQRPALSGDAVQKMLGHCTFEPAEPKAAKNTYTAERFIWLTCLNNLRILEQGSERPLTDTERATLMDEPYRKSCLTY
AQA
RKLLGLEDTAFFKGLRYGKDNAEASTLMEMKAYHAISRALEKEGLKDKKSPLNLSPELQDEIGTAFSLFKTDEDITG
RLK
DRIQPEILEALLKHISFDKFVQISLKALRRIVPLMEQGKRYDEACAEIYGDHYGKKNTEEKIYLPPIPADEIRNPVV
LRA
LSQARKVINGVRRYGGSPARIHIETAREVGKSFKDRKEIEKRQEENRKDREKAAAKFREYFPNFVGEPKSKDILKLR
LYE
QQHGKCLYSGKEINLGRLENEGYVEIDHALPFSRTWDDSFNNKVLVLGSENQKNQTPYEYFNGKDNSREWQEFKA
RVE
TSRFPRSKKQRILLQKFDEDDGFKERNLNDTRYVNRFLCQFVADRMRLTGKGKKRVFASNGQITNLLRGFWGLRKVRA
END

RHHALDAVVVACSTVAMQQKITRFVRYKEMNAFDGKTIDKETGEVLHQKTHFPQPWEFFAQEVMIRVFGKPDGKPEF
EEA

DTLEKLRTLLEAKLSSRPEAVHEYVTPLFVSRAPNRKMSGQGHMETVKSARKLDEGVSVLRVPLTQLKLDLEKMTN
RER

EPKLYEALKARLEAHKDDPAKAFAPFYKYDKAGNRTQQVKAVRVEQVQKTGVWVRNHNGIADNATMVRVDVFEEKD
KYY

LVPIYSWQVAKGILPDRAVVQGKDEEDWQLIDDSFNFKFSLHPNDLVEVITKKARMFGYFASCHRGTNINIRIHD
LHK

IGKNGILEGIGVKTALSFQKYQIDELGKEIRPCRLKKRPPVR

H) *Streptococcus mutans*

Length: 1345

Identity: 19.19%

>Q8DTE3.1 RecName: Full=CRISPR-associated endonuclease Cas9 [*Streptococcus mutans* UA159]

MKKPYSIGLDIGTNSVGWAVVTDDYKVPKKMKVLGNTDKSHIEKNLLGALLFDSGNTAEDRRLKRTARRRYTRRRN
RIL

YLQEIFSEEMGKVDDSFHRLSDSLVTEDKRGERHPIFGNLEEEVKYHENFPTIYHLRQYLADNPEKVDLRLVYLA
LAH

IIKFRGHFLIEGKFDTRNNDVQRLFQEF LAVYDNTFENSSLQEQNVQVEEILTDKISKSAKKDRVLKLPNEKSNR
FAE

FLKLIVGNQADFKKHFELEEKAPLQFSKDTYEEELVLLAQIGDNYAELFLSAKKLYDSILLSGILTVDVGTKAPL
SAS

MIQRYNEHQMDLAQLKQFIRQKLSKDYNEVFSVSKDGYAGYIDGKTNQEAIFYKYLKGLLNKIEGSGYFLDKIERE
FLR

KQRTFDNGSIPHQIHLQEMRAIIRQAIFYPFLADNQDRIEKLTLFRIPYYVGPLARGKSDFAWLSRKSADKITPWN
FDE

IVDKESSAEAFINRMTNYDLYLPNQKVLPKHSLLEYKFTVYNELTKVKYKTEQGKTAFDANMKQEIFDGVFKVYRK
VTK

DKLMDFLEKEFDEFRIVDLTGLDKENKVFNASYGTYHDLCKILDKDFLDNSKNEKILEDIVLTLTLFEDREMIRKRL
ENY

SDLLTKEQVKKLERRHYTGWGRLSAELIHGIRNKESRKTILDYLIIDGNSNRNFMQLINDDALSFKKEEIAKAQVIGE
TDN

LNQVWSDIAGSPAIIKKGILQSLKIVDELVKIMGHQPENIIVVEMARENQFTNQGRNSQQRLKGLTDSIKEFGSQILK
EHP

VENSQLQNDRLFLYYLQNGRDMYTGEELDIDYLSQYDIDHIIPQAFIKDNSIDNRVLTSSKENRGKSDDVPSKDVVR
KMK

SYWSKLLSAKLITQRKFDNLTKAERGGLTDDDKAGFIKRQLVETRQITKHVARILDERFNTETDENNKKIRQVKIVT
LKS

NLVSNFRKEFELYKVVREINDYHHAHDAYLNAVIGKALLGVYPQLEPEFVYGDYPHFHGHKENKATAKKFFYSNIMNF
FKK

DDVRTDKNGEIIWKKDEHISNIKKVLSYPQVNIVKKVEEQTGGFSKESILPKGNSDKLIPRKTCKFYWDTKKYGGFD
SPI

VAYSILVIADIEKGKSKKLKTVKALVGVTIMEKMTFERDPVAFLEKGYRNVQEENIIKLPKYSLFKLENGRKRLLA
SAR

ELQKGNEIVLPNHLGTLTYHAKNIHKVDEPKHLDYVDKHKDEFKELLDVVSNFSSKKYTLAEGNLEKIKELYAQNGE
DLK

ELASSFINLLTFTAIGAPATFKFFDKNIDRKRYTSTTEILNATLIHQSIITGLYETRIDLNLKLGDD

I) *Staphylococcus aureus*

Length: 1053

Identity: 34.87%

>J7RUA5.1 RecName: Full=CRISPR-associated endonuclease Cas9; AltName:
Full=SaCas9 [*Staphylococcus aureus*]

MKRNYILGLDIGITSVGYGIIDYETRDVIDAGVRLFKEANVENNEGRRSKRGARRLKRRRRRHRIQRVKKLLFDYNLL
TDH

SELSGINPYEARVKGLSQKLSEEEFSAALLHLAKRRGVHNVNEVEEDTGNELSTKEQISRNSKALEEKYVAELQLER
LKK

DGEVRGSINRFKTS DYVKEAKQLLKVQKAYHQLDQSFIDTYIDLLETRRTYYEGPGEGSPFGWKDIKEWYEMLMGHC
TYF

PEELRSVKYAYNADLYNALNDLNNLVITRDENEKLEYEKFQIIENVFKQKKKPTLKQIAKEILVNEEDIKGYRVTS
TGK

PEFTNLKVYHDIKDITARKEIENAELLDQIAKILTIYQSSEDIQEELTNLNSLTQEEIEQISNLKGYTGTHNLSL
KAI

NLILDELWHTNDNQIAIFNRLKLVPKKVDLSQQKEIPTTLVDDFILSPVVKRSFIQSIKVINAIIKKYGLPNDIIIE
LAR

EKNSKDAQKMINEMQKRNRQTNERIEEIIIRTTGKENAKYLIEKIKLHDMQEGKCLYSLEAIPLEDLLNNPFPNYEVDH
IIP

RSVSFDNSFNKVLVKQEENSKGNRTPFQYLSSSDSKISYETFKKHILNLAKGKGRISKTKKEYLLEERDINRFSV
QKD

FINRNLVDTRYATRGLMNLRSYFRVNNLDVKVKSINGGFTSFLRRKWKFKKERNKGYKHHAEDALIIANADFIFKE
WKK
LDKAKKVMENQMFEEKQAESMPEIETEQEYKEIFITPHQIKHIKDFKDYKYSHRVDKKPNRELINDTLYSTRKDDKG
NTL
IVNNLNGLYDKDNDKLLKLINKSPEKLLMYHHDPPQTYQKLKLIMEQYGDEKNPLYKYYEETGNYLTKYSKKDNGPVI
KKI
KYYGNKLNAHLDITDDYPNSRNKVVKLSLKPYRFDVYLDNGVYKFVTVKNLDVIKKENYYEVNSKCYEEAKKLLKIS
NQA
EFIASFYNNDLIKINGELYRVIGVNNDLLNRIEVMIDITYREYLENMNDKRPPRIIKTIASKTQSIKKYSTDILGN
LYE
VKSCKHPQIIKKG

J) *Streptococcus pyogenes*

Length: 1368

Identity: 19.14%

>Q99ZW2.1 RecName: Full=CRISPR-associated endonuclease Cas9/Csn1; AltName:
Full=SpCas9; AltName: Full=SpyCas9 [*Streptococcus pyogenes* serotype M1]

MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKN
RIC
YLQEIFSNEMAKVDDSFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTKADLRLIYLA
LAH
MIKFRGHFLIEGDLNPDNSDVKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRENLIAQLPGEKKNGL
FGN
LIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLNLLAQIGDQYADFLAAKNLSDAILLSDILRVNTEITKAPL
SAS
MIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNRED
LLR
KQRTFDNGSIPHQIHLGELHAILRRQEDFYFPLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWN
FEE
VVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNR
KVT
VKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLK
TYA
HLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMLIHDDSLTFKEDIQKAQVSGQG
DSL

HEHIANLAGSPAIIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILK
EHP

VENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYVDVHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVK
KMK

NYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVIT
LKS

KLVSDFRKDFQFYKVVREINNYHHAHDAYLNAVVGTAIIKKYPKLESEFVYGDYKVYDVRKMIKSEQEIGKATAKYF
FYS

NIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVL SMPQVNIKKTEVQTGGFSKESILPKRNSD
KLI

ARKKDWDPKKYGGFDSPTVAYSVLVAKVEKGKSKKLKSVKELLGITIMERSSEKNPIDFLEAKGYKEVKKDLIIK
LPK

YSLFELENGRKRMLASAGELQKGNELALPSKYVNFYLYASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFS
KRV

ILADANLDKVL SAYNKHDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVL DATLIHQ SITGLYE
TRI

DLSQLGGD

3)

- a) Gap Open Penalty=10, Gap Extension Penalty=0.05
Weight Transition: NO
Hydrophilic Residues for Proteins: GPSNDQERK
Hydrophilic Gaps: YES

Weight Matrix- BLOSUM

- b) CLUSTAL 2.1 Multiple Sequence Alignments

```
Sequence type explicitly set to Protein
Sequence format is Pearson
Sequence 1: sp|Q03LF7|CAS9A_STRTD 1121 aa
Sequence 2: sp|Q9CLT2|CAS9_PASMU 998 aa
Sequence 3: Q6NKI3.1 1084 aa
Sequence 4: Q0P897.1 984 aa
Sequence 5: Q927P4.1 1334 aa
Sequence 6: J3F2B0.1 1101 aa
Sequence 7: C9X1G5.1 1082 aa
Sequence 8: Q8DTE3.1 1345 aa
Sequence 9: J7RUA5.1 1053 aa
Sequence 10: Q99ZW2.1 1368 aa
Start of Pairwise alignments
```

Aligning...

Sequences (1:2) Aligned. Score: 15.7315
Sequences (1:3) Aligned. Score: 11.9926
Sequences (1:4) Aligned. Score: 15.5488
Sequences (1:5) Aligned. Score: 12.2212
Sequences (1:6) Aligned. Score: 12.0799
Sequences (1:7) Aligned. Score: 15.8965
Sequences (1:8) Aligned. Score: 11.5968
Sequences (1:9) Aligned. Score: 22.887
Sequences (1:10) Aligned. Score: 13.4701
Sequences (2:3) Aligned. Score: 13.8277
Sequences (2:4) Aligned. Score: 24.2886
Sequences (2:5) Aligned. Score: 10.02
Sequences (2:6) Aligned. Score: 12.8257
Sequences (2:7) Aligned. Score: 63.1263
Sequences (2:8) Aligned. Score: 10.7214
Sequences (2:9) Aligned. Score: 19.5391
Sequences (2:10) Aligned. Score: 11.4228
Sequences (3:4) Aligned. Score: 13.1098
Sequences (3:5) Aligned. Score: 10.1476
Sequences (3:6) Aligned. Score: 40.9594
Sequences (3:7) Aligned. Score: 13.6784
Sequences (3:8) Aligned. Score: 11.9926
Sequences (3:9) Aligned. Score: 11.9658
Sequences (3:10) Aligned. Score: 11.5314
Sequences (4:5) Aligned. Score: 12.3984
Sequences (4:6) Aligned. Score: 11.4837
Sequences (4:7) Aligned. Score: 25.1016
Sequences (4:8) Aligned. Score: 12.5
Sequences (4:9) Aligned. Score: 18.8008
Sequences (4:10) Aligned. Score: 13.4146
Sequences (5:6) Aligned. Score: 10.3542
Sequences (5:7) Aligned. Score: 12.5693
Sequences (5:8) Aligned. Score: 56.8216
Sequences (5:9) Aligned. Score: 13.2953
Sequences (5:10) Aligned. Score: 44.6027
Sequences (6:7) Aligned. Score: 13.586
Sequences (6:8) Aligned. Score: 10.5359
Sequences (6:9) Aligned. Score: 10.4463
Sequences (6:10) Aligned. Score: 10.2634
Sequences (7:8) Aligned. Score: 11.2754
Sequences (7:9) Aligned. Score: 18.5185
Sequences (7:10) Aligned. Score: 11.4603
Sequences (8:9) Aligned. Score: 14.7198
Sequences (8:10) Aligned. Score: 61.7844
Sequences (9:10) Aligned. Score: 14.7198
Guide tree file created: [\[clustalw.dnd\]](#)

There are 9 groups
Start of Multiple Alignment

Aligning...

Group 1:		Delayed
Group 2: Sequences:	2	Score:14385
Group 3:		Delayed
Group 4:		Delayed
Group 5: Sequences:	2	Score:19252
Group 6: Sequences:	3	Score:10964
Group 7: Sequences:	5	Score:7442
Group 8: Sequences:	2	Score:7983
Group 9: Sequences:	7	Score:6654
Alignment Score		20693

c) Conserved Sequences

DSG

R

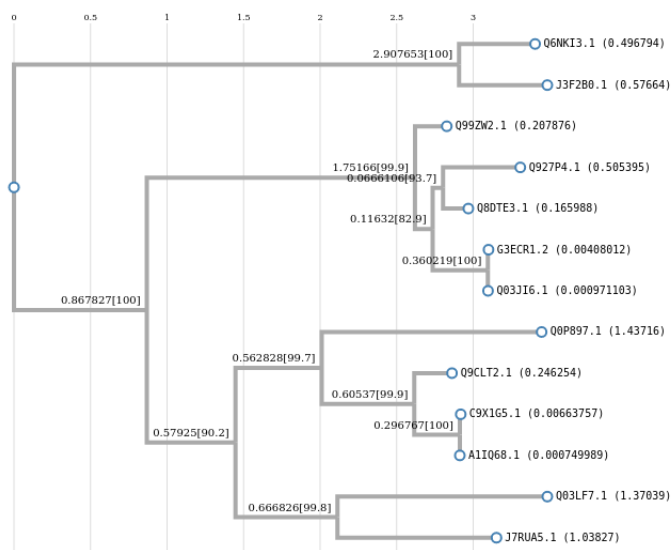
S

P, G and E

R

Q, Y, DH, P, N, V and K

HHA and DA



d)

e) CLUSTAL 2.1 Multiple Sequence Alignments

Sequence type explicitly set to Protein
Sequence format is Pearson

Sequence 1: Q03LF7.1 1121 aa
Sequence 2: Q9CLT2.1 1056 aa
Sequence 3: Q6NKI3.1 1084 aa
Sequence 4: Q0P897.1 984 aa
Sequence 5: Q927P4.1 1334 aa
Sequence 6: J3F2B0.1 1101 aa
Sequence 7: C9X1G5.1 1082 aa
Sequence 8: Q8DTE3.1 1345 aa
Sequence 9: J7RUA5.1 1053 aa
Sequence 10: Q99ZW2.1 1368 aa
Start of Pairwise alignments
Aligning...

Sequences (1:2) Aligned. Score: 15.4356
Sequences (1:3) Aligned. Score: 11.9926
Sequences (1:4) Aligned. Score: 15.5488
Sequences (1:5) Aligned. Score: 12.2212
Sequences (1:6) Aligned. Score: 12.0799
Sequences (1:7) Aligned. Score: 15.8965
Sequences (1:8) Aligned. Score: 11.5968
Sequences (1:9) Aligned. Score: 22.887
Sequences (1:10) Aligned. Score: 13.4701
Sequences (2:3) Aligned. Score: 13.3523
Sequences (2:4) Aligned. Score: 24.2886
Sequences (2:5) Aligned. Score: 10.7955
Sequences (2:6) Aligned. Score: 13.3523
Sequences (2:7) Aligned. Score: 62.3106
Sequences (2:8) Aligned. Score: 11.0795
Sequences (2:9) Aligned. Score: 18.7085
Sequences (2:10) Aligned. Score: 11.7424
Sequences (3:4) Aligned. Score: 13.1098
Sequences (3:5) Aligned. Score: 10.1476
Sequences (3:6) Aligned. Score: 40.9594
Sequences (3:7) Aligned. Score: 13.6784
Sequences (3:8) Aligned. Score: 11.9926
Sequences (3:9) Aligned. Score: 11.9658
Sequences (3:10) Aligned. Score: 11.5314
Sequences (4:5) Aligned. Score: 12.3984
Sequences (4:6) Aligned. Score: 11.4837
Sequences (4:7) Aligned. Score: 25.1016
Sequences (4:8) Aligned. Score: 12.5
Sequences (4:9) Aligned. Score: 18.8008
Sequences (4:10) Aligned. Score: 13.4146
Sequences (5:6) Aligned. Score: 10.3542
Sequences (5:7) Aligned. Score: 12.5693
Sequences (5:8) Aligned. Score: 56.8216
Sequences (5:9) Aligned. Score: 13.2953
Sequences (5:10) Aligned. Score: 44.6027
Sequences (6:7) Aligned. Score: 13.586
Sequences (6:8) Aligned. Score: 10.5359
Sequences (6:9) Aligned. Score: 10.4463
Sequences (6:10) Aligned. Score: 10.2634
Sequences (7:8) Aligned. Score: 11.2754
Sequences (7:9) Aligned. Score: 18.5185

Sequences (7:10) Aligned. Score: 11.4603
Sequences (8:9) Aligned. Score: 14.7198
Sequences (8:10) Aligned. Score: 61.7844
Sequences (9:10) Aligned. Score: 14.7198
Guide tree file created: [\[clustalw.dnd\]](#)

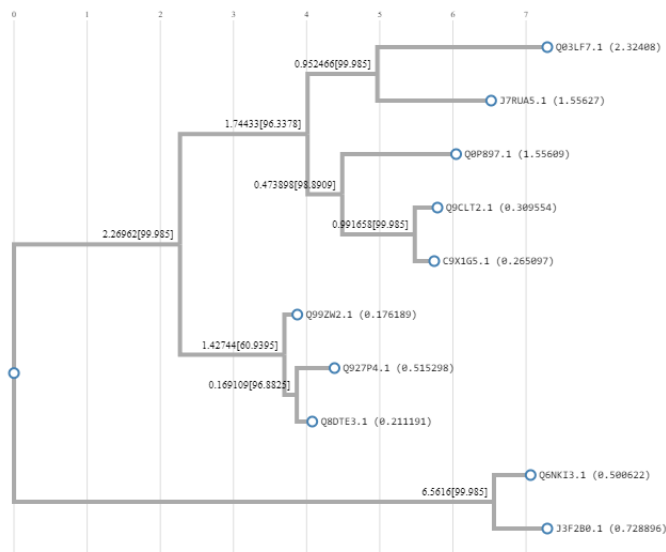
There are 9 groups
Start of Multiple Alignment

Aligning...

Group 1:		Delayed
Group 2: Sequences:	2	Score:16220
Group 3:		Delayed
Group 4:		Delayed
Group 5: Sequences:	2	Score:20539
Group 6: Sequences:	3	Score:13900
Group 7: Sequences:	5	Score:10220
Group 8: Sequences:	2	Score:10159
Group 9: Sequences:	7	Score:9423
Alignment Score 12413		

CLUSTAL-Alignment file created [\[clustalw.aln\]](#)

Using PAM as the scoring matrix the score is reduced to 12413

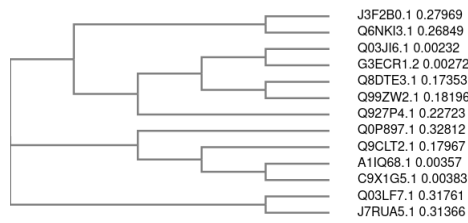


Yes, the tree is also different in both weight and shape

4)

a)Based on comparing the first paragraph, we find the same conserved amino acids.

b)



The tree of Clustal Omega has form and weight difference with that of ClustalW

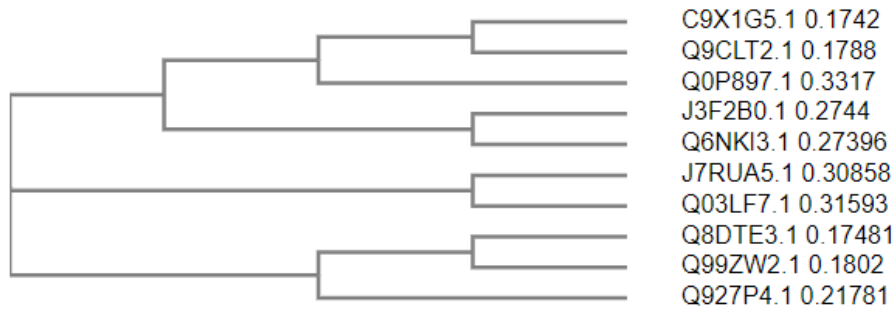
5)

Different tree is obtained.

Phylogenetic Tree

This is a Neighbour-joining tree without distance corrections.

Branch length: ☒ Cladogram ☐ Real



As observed above, the tree is different in weight and shape.

6)


```

1      10      20      30      40
sp|Q9CLT2|CAS9_PASMU  ....MQTTNLSYIIGLGLDIDTAASVQWAVVEINENEDPIGLIDVG.....VRIFERAIEVP
C9X1G5.1  .MAAFKFNSENNYIIGLGLDIDTAASVQWAVVEIDSEENPIRLIDVG.....VRVFERAEVVP
Q8DTE3.1  .MKKPYISIGLGLDIDTAASVQWAVVTDYKVPKAKMMKVLGNTDKSHIEKNLLGALLF
Q99ZW2.1  .MDKKYISIGLGLDIDTAASVQWAVITDEYKVPKSKFKVVGNTDRHSIKKNLLGALLF
Q927P4.1  .MKKPYIIGLGLDIDTAASVQWAVLTDQYDLVKRMMKVIAGDSEKKQIKKMFVGVRLF
Q6NKK13.1  .MKKPYHVGIDIDTAASVQWAAIEVDAGMPFKTSLVS.....HIHDSGLDPDEIK
J3F2B0.1  MMYASLMSAHLHRLVGIIDVGTSSVGLATLRLVDDHGTPIELLALS.....HIHDSGVGKEGKK
Q0P897.1  .MARILAFDIDTSSIGWAFSEMDLKDCCG.....VRIFTKVENP
J7RUA5.1  .MKRNYIIGLGLDIDTISVGVGIIDYEETR.....DVIDAGVRLFK
sp|Q03LF7|CAS9A_STRTD  ....MSDLVIGLGLDIDTISVGVGILNKVITG.....EIIHKNSRIFF

```

```

50      60      70      80      90      100
sp|Q9CLT2|CAS9_PASMU  KTGESLALSRLARSTARRLRARRLRLLAKRFLKREGISTIDLEKG....LPNQAWNE
C9X1G5.1  KTGDLSLAMSRLARSVRRLLRRRAHRLLRTRRLLKREGVLAQANFDENGLIKSLPNTFQW
Q8DTE3.1  DSGNTAEDRRLLKRTARSRYYTRRRNRILYLQEIFSEEMGKVDDSFHRLLEDSEFVTEDEKRG
Q99ZW2.1  DSGETAEDRRLLKRTARSRYYTRRRNRILYLQEIFSEEMAKVDDSFHRLLEDSEFVTEDEKRG
Q927P4.1  DEGQTAADRRMMARTARRRIRRRNRISYLGQIFAEEMSKTDANFFCRLSDSFYVDNEERN
Q6NKK13.1  SAVTRLASSGGIARRTRRLYRKRRLRLQQLDKFIQRQGWV.....TELEDYSDPLYPNK
J3F2B0.1  DHDTRKKLSGIARRARLLHRRRTQLQQLDEVLRDLGFPITPFGFELDLNEQTDPPYRVNR
Q0P897.1  KTGESLALPRLARSARRLRARRLRLLAKRFLKREGISTIDLEKG....LPNQAWNE
J7RUA5.1  EANVENNEGRSRKRGARRLRARRLRRLRIQVRKLLFDYNLLTDHSELSG.....INPYE
sp|Q03LF7|CAS9A_STRTD  AAQAENNLVRRTRRQGRRLARRRRRRHRRVRLNRLFEESGLTDFTKISIN....LNPFYQ

```

```

110      120      130
sp|Q9CLT2|CAS9_PASMU  LRVAGLERRTSAIENGA.....VLLHLIKHRCYLSKRKN
C9X1G5.1  LRAAALDRRLITPLENSA.....VLLHLIKHRCYLSQRKN
Q8DTE3.1  ERHPIFGNLEEEVYHENFPTIYHLRQYLADNPEKVDLRLVYLALAHIIKFRGHFLIEGK
Q99ZW2.1  ERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSDDKADLRLIYLAALAHMIKFRGHFLIEGD
Q927P4.1  SRHPPFATIEEEVEYHKNYPTIYHLREELVNSSEKADLRLVYLALAHIIKYRGNFLIEGA
Q6NKK13.1  VRAELAAASYIADEKERGE.....KLSVALRHIAHRCWRNRPYAK
J3F2B0.1  VRAELVEEKLP.EELRGP.....AISMVVRHIAHRCWRNRPYSK
Q0P897.1  ISPYELRFRALNELLKSKQ.....DFARVILHIAHRCYDDIKNS
J7RUA5.1  ARVKGLSQKLSEEFSA.....LLHLAKRCVRNVNEV
sp|Q03LF7|CAS9A_STRTD  LRVKGLTDELSEELFIA.....LKNHVKHRCISYLLDA

```

```

140      150      160
sp|Q9CLT2|CAS9_PASMU  ESQTNKELGALLSGVAQNHQLLQSDDYR.....
C9X1G5.1  EGETADKELGALLKGVVAGNAHALQTDGFR.....
Q8DTE3.1  FDRNRNDVQRLFQEFLLAVYDNTFENSSLO.....EQNVQVEEILTDKISKSAKKDRVLK
Q99ZW2.1  LNPDNDSVDKFLIQLVQTYNQLFEEENPIN.....ASGVDAKAILSAKLSRRLENLIA
Q927P4.1  LDTQNTISVDGIYKQFIQTYNQVFASGIEDGSLKKLEDNKKDVAKILVKKVTRKEDFLK
Q6NKK13.1  VSSLYLPDGPDSAFKAIREEIKRASGQF.....
J3F2B0.1  VESLLSPAEESEPFMKALREILATIGEV.....
Q0P897.1  DDKEKGAALKAIKQNEKKLANYSQSVGEYLY.....
J7RUA5.1  EEDTGNELSSTKEQISRSKALEEKYVAELQ.....LLHLAKRCVRNVNEV
sp|Q03LF7|CAS9A_STRTD  SDDGNSGVGDYAQIVKENSQKLETKTFGQIQLER.....YQTYGQLR

```

```

170      180      190
sp|Q9CLT2|CAS9_PASMU  .TPAELALIKFAKEEGHIRNQRGAYTHTFNRLD.....
C9X1G5.1  .TPAELALINKFEKESGHIRNQSDYSHTFSRKD.....
Q8DTE3.1  .LFPNEKSNRFAEFLKLIVGNQADFKKHFELEEKAPLQFSKDTYEELEVLALQIGDNYA
Q99ZW2.1  .QLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLNLLAQIGDQYA
Q927P4.1  .LYPGEKSAQMPAQFISLIVSGKGNFQKPFDLIEKSDIECAKDSYEEEDLSLALIGDEYA
Q6NKK13.1  .VPESTATVQMVILCELTGLKRGEGGVLSARLQ.....
J3F2B0.1  .LDDGITPQAMAQVALTNRNISMGRPEGILQKLLH.....
Q0P897.1  .KEYVQKFKENSKFTNVRNKKESYERCAIQSFLK.....
J7RUA5.1  .LERLKKDGEVRGSTRNRFKTSDYVKEAKQLKVQK.....
sp|Q03LF7|CAS9A_STRTD  GDFTVEKDQKKHRLINVFPTSAYRSEALRLIQTQQEFN.....

```

```

sp|Q9CLT2|CAS9_PASMU  .....
C9X1G5.1  .....
Q8DTE3.1  ELFLSAKKLYDSILLSGILTVTDVGTAKPLSASMIQRYNEHQMDLAQLKQFIRQKLSDKY
Q99ZW2.1  DLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDERHQDILLKALVYQQLPKRY
Q927P4.1  ELFVAAKNAYSASVVLSSITVAETETNAKLSASMIERFDTHREEDLGEELKAFIKHLHPKH
Q6NKK13.1  .....
J3F2B0.1  .....
Q0P897.1  .....
J7RUA5.1  .....
sp|Q03LF7|CAS9A_STRTD  .....

```