**Part : 1 - Run BLASTp to identify all homologs of BBA18 in the reference genome ("ref").**

**-cp ../../bio425/data/BBA18.pep BBA18.pep**

>BBA18

MKTLKKKSPNTTKRVKKTTMNFQHNLIVLISTLNFINLNFKKYTQKNILYLLNKNLERNKQKLIKLKTLQNYLYILEKKFKVTLNYCKHLGKNSGSETYYKLKYEKEKCYLIINTYFKEKIINKINEFTQRIKKFNQINSSVKWECINNTNNIYKYKEYRNIHKNSKKTTNNEIIKKYLSKCNFKTEIPALIMNLKTTNSTKIYHLRNLKHIENDLKEIDPKKIEKHISNVIKENINNPGYLCKFFRNSGYKRIIHKIKEPEKKYKNKNEILKKILEEKIKELEKEQYKKEDLEKFFGKTYEIYKIKPHFIIEHKKYPDLEKLVKRAKTEIPKIQDKMVRQKSIKNNIFSILLEQLRHKVDNDKLIPTLKKFIENEPDLKYSKVFDNSYYNNLIKIVS

**-blastp -query BBA18.pep -db ref -evalue 1e-3 -outfmt 6 | cut -f2 > BBA18.homologs**

**-cat BBA18.homologs**

BBA18

BBF26

BBU04

BBP30

BBG06

BBS33

BBR31

BBM30

BBL30

BBO30

BBN30

BBQ38

BBI19

BBE21

BBT04

BBC01

BBU07

BBD06

BBI01

BBU01

BBB10

**-blastdbcmd -db ref -entry\_batch BBA18.homologs > BBA18.homologs.pep**

**cat BBA18.homologs.pep**

>lcl|BBA18 unnamed protein product

MKTLKKKSPNTTKRVKKTTMNFQHNLIVLISTLNFINLNFKKYTQKNILYLLNKNLERNKQKLIKLKTLQNYLYILEKKF

KVTLNYCKHLGKNSGSETYYKLKYEKEKCYLIINTYFKEKIINKINEFTQRIKKFNQINSSVKWECINNTNNIYKYKEYR

NIHKNSKKTTNNEIIKKYLSKCNFKTEIPALIMNLKTTNSTKIYHLRNLKHIENDLKEIDPKKIEKHISNVIKENINNPG

YLCKFFRNSGYKRIIHKIKEPEKKYKNKNEILKKILEEKIKELEKEQYKKEDLEKFFGKTYEIYKIKPHFIIEHKKYPDL

EKLVKRAKTEIPKIQDKMVRQKSIKNNIFSILLEQLRHKVDNDKLIPTLKKFIENEPDLKYSKVFDNSYYNNLIKIVS

>lcl|BBF26 unnamed protein product

MKSIKTSTNKYQHKLIILISTLNYVNSNLEQYTQNDILYYFNNNIRRNCQKEVKLKTLQNYLYKLEKVFKVTSNYHRHLG

INMGTEVHYKLKYPKKVCYHLINKHFKEIKKKKYKSRVDSYFNKNCIENSSVKNRECFYNISNNKEEKRNSRKSIEKLQI

RKYIEKCNFKSNIFSLILNLGLKKDATIEVCKMIKKAENLIEKNIFEKLNVVKVIEVNKLKNKQQKLKRILKEVKIKLEN

ENYKSEQLEIEIKDIYEQYKNKPHFIIEKNKYEDLEKITEKLKNKFKQEKINIKENDTNIRNNIFNILIEQLRHKIKIEI

LIPILKDYLDKQKRLEYSKVFSNYYYYEILETVEPNKWCSQLEACK

>lcl|BBU04 unnamed protein product

MLISFVKESLPMHKIKTTNNNPRNCYNKVQYKLIVLISTICYLNKTHKKYTQKTILYYFNENLRKNGQTISTLRTMQKYI

YRLQKEIKVTTNYYQHMGVNSGTEIYYKLNYPKKDCYHKINQHFKEKKETRFQNRVTNYFNKNSDSKMGSVQCESCNSNK

NNIKEERKINEIEKYQVINYFNKCNFSCKEILSILLNLNVDKDTMIKIIKTIKRTDIKAKNKNIYFPKSCSKEKQEKLKK

ILCNTQKELEKSGYNSEQLETNFQKIYENYKYKPHFIIENHKYSDLSYIKRKLEKSIERKKENSKQDYKNLRTNIFNILI

EQLKKETNIEILKPIIKEYLNNQKKIEYKKVFRIYYSELLEIINGKHYSNLKKFRKKSVG

>lcl|BBP30 unnamed protein product

MKGFSNTTKNPTCHNKHQHKLIYLASTLDFLNKKDKKYTQQNILYYYNENLKRNGLAPTTLRTMQNYLYKLEKVLKVTTN

YYQHMGVNCGTEIYYKLKYPKKECYQKINKYFKERKNSRFKSRVNNHFKDNVSKNSSVNSVECLSNKNNIKEERKINEIE

KYQVIKYFNKSNFLCKEILPFLLTLNVDKDTMIKIIKNIKRVENKLLKNTNLNKSCFKEKQEKLKKILNNTQKEFEQNGY

NPEQLKTNLQKVYENYKFKPHFIIERHKYNDLNNIKRKLEKAIEREKQNSQQNYQNLKENIFNILIEQLKKETNIEILKP

IIKKYLNNQKKIEYNKVFGIYHLELSEIIKNEKNSLTTEEFSIKAV

>lcl|BBG06 unnamed protein product

MKKVFTFLKKLCIIYNINPIRSSTMINNSKKPNCHNKLQQKLIVLLSTLAYVNSKYNKYTQKNILYCFNENLKRNGQPTT

TLRTMQNYLYKLEKVFKVTTNYYKHLGINFGTEIYYKLNYSKKECYLKINQHFREKKDYRFKARVDNYLNDKFNKNGSVD

LVECLNNKNNNIKEERKIIQIEKYQVIKYFNKCNFSLLKEILPILNLDINKDELIKILKIIKRIEINLTKNKNTYLNVSY

FKEKQNKLKKILSNTQQQLEKNGYNPEQLEISFQKIYENYKTKPHFIIENHKYKDLSHIKRKLEKSIERKKENLQKDCEN

MRTNIFNILIEQLKKEVKIDVLKPILKIYLNSKNKLEYNKVFNNNYYYELLEIIKKEKNLQLKEVV

>lcl|BBS33 unnamed protein product

MKGFSNTTKNPTCYNKHQHKLIYLTSTVDFLNKKDKKYTQQNILYYYNKNLKRNGLAPTTLRTMQNYLYKLEKVLKVTTN

YYQHMGVNCGTEIYYKLKYPKKECYQKINKYFKERKNSRFKSRVNNHFKDNVSKNSSVNSVECLSNKNNIKEERKIKEIE

KYQVIKYFNKSNFLCKEILPFLLTLNVDKDTMIKIIKNTKRVENKLLKNTNLNKSCFKEKQEKLKKILNNTQKEFEQNGY

NPEQLKTNLQKVYENYKFKPHFIIERHKYSDLNNIKRKLEKSIEREKQNSQQNYQNLKENIFNILIEQLKKETNIEILKP

IIKKYLNNQKKIEYNKVFGIYHLELLEIIKNEKNSLTTEEFSIKAV

>lcl|BBR31 unnamed protein product

MKDFLITTKNPTCHNKHQHKLIYLTSTVDFLNKKDKKYTQQNILYYYNKNLKRNGLAPTTLRTMQNYLYKLEKVLKVTTN

YYQHMGVNCGTEIYYKLKYPKKECYQKINKYFKERKNSRFKSRVNNHFKDNVSKNSSVNSVECLSNKNNIKEERKIKEIE

KYQLRNYFNNCNFKTEEALSILYLNTDKDTKIEAINILKQNEIALIKKFNIKKSCMKEKQKKLKEILYNTRKKLEENGYN

PKQLEINLQKVYENYKYKPHFIIENHKYNDLGYIKRKLEKSIERKKENSQQNYQNLKENIFNILIEQLKKEVNIEILKPI

IKEYLNNQKKIEYNKVFGIYHLELLEIIKNEKNSLTTEEFSIKAV

>lcl|BBM30 unnamed protein product

MKDFLNNIKSLTCYNKHQHKLISLTSTLDFLNKKDKKYTQQNILYYFNENLKRNGLAPTTLRTMQNYLYKLEKVLKVTTN

YYQHMGVNCGTEIYYKLKYPKKECYQKINKYFEERKNSRFKSRVNNHFKDDISKNGSVNSVECLSNKNNIKEERKIKEIE

KYQVIKYFNKSNLLCKEILPFLLTLNVDKDTMIKIIKNIKRVENKLLKNTNLNKSCFKEKQEKLKRILNNTQKEFKQNGY

NPEQLKINLQKVYESYKFKPHFIIENHKYSDLNNIKSKLEKSIERKKENPQQNYQNLKENIFNILIEQLKKETNIEILKP

IIKKYLNNQKKIEYNKVFGTYHLELLEIIKNEKNSLTTEEFNIKAV

>lcl|BBL30 unnamed protein product

MKGFLNNAKSPTCYNKHQHKLISLTSTLDYLNKKDKKYTQQNILYYFNENLKRNGLAPTTLRTMQNYLYKLEKVLKVTTN

YYQHLGVNCGTEIYYKLKYPKKECYQKINKYFKERKNSRFKSRVNKHLKDNVSKNGSVNSVECLSNKNNIKEERKIKEIE

KYQLRNYFNNCNFKTEEALSILYLNIDKDTKIEAINILKQNEIALIKKFNIKKSCMKEKQKKLKEILYNTRKKLEENGYN

PKQLEINLQKVYENYKYKPHFIIENHKYSDLNNIKRKLEKSIKRKKENSQQNYQNLKENIFNILIEQLKKETNIEILKPI

VKEYLNNQKKIEYNKVFGTYHLELSEIIKNEKNSLTTEEFSIKAV

>lcl|BBO30 unnamed protein product

MKGFPKNTKSPTCHNKHQHKLISLTSTLDYLNKKDKKYNQKNILYYYNENLKRNGLAPTTLRTMQNYLYKLEKVLKVTTN

YYQHMGVNCGTEIYYKLKYPKKECYQKINKYFKERKNSRFKSRVNTHFKDNVSKNSSVNSVECLSNKNNIKEERKIKEIE

KYQLRNYFNNCNFKTEEALSILYLNIDKDTKIEAINILKQNEIALIKKFNIKKSCMKEKQKKLKEILYNTRKKLEENGYN

PKQLEINLQKVYENYKYKPHFIIENHKYSDLNNIKRKLEKSIERKKENSQQNYQNLKENIFNILIEQLKKETNIEILKPI

VKEYLNNQKKIEYNKVFGTYHLELSEIIKNEKNSLTTEEFSIKAV

>lcl|BBN30 unnamed protein product

MQNFPNKTKIPTCHNKHQHKLISLTSTLDFLNKKDKKYTQQNILYCFNENLKRNGLAPTTLRTMQNYLYKLEKVLKVTTN

YYQHMGVNCGTEIYYKLKYPKKECYQKINKYFKERKNSRFKSRVNNHFKDNVSKNSSVNSVECLNNKNNNTKEERKIKEI

EKYQLRNYFNNCNFKTEEALSILYLNIDKDTKIEAINILKQNEIALIKKFNIKKSCMKEKQKKLKEILHNTRKKLEENGY

NPKQLEINLQKVYENYKYKPHFIIENHKYSDLNNIKRKLEKSIERKKENSQQNYQNLKENIFNILIEQLKKETNIEILKP

ITKKYLNNQKKIEYNKVFGTYYLELLEIIRKQKNSLTTEEFNIKAI

>lcl|BBQ38 unnamed protein product

MKDFPNITKIPTCHNKHQHKLISLTSTLDYLNKKDKKYTQQNILYCFNENLKRNGLAPTTLRTMQNYLYKLEKVLKVTTN

YYQHMGVNCGTEIYYKLKYPKKECYQKINKYFKEQKNSRFKSRVNKHFKDNVSKNSSVNSVGCLSNKNNIKEERKIKEIE

KYQLRNYFNNCNFKTEEALSILYLNIDKDTKIEAINILKQNEIALIKKFNIKKSCMKEKQKKLKEILYNTRKKLEENGYN

PKQLEINLQKVYENYKYKPHFIIENHKYSDLNNIKRKLEKSIERKKENSQQNYQNLKANIFNILIEQLKKETNIEILKTI

VKEYLNNQKKIEYNKVFGIYYLELLEIIKNEKNSLTTEEFSIKAV

>lcl|BBI19 unnamed protein product

MLLFKYCAINVFIKIKITRSINMQNAFTNRKNTPCQNKLQHKLIVFISTLKYINSKYKKYTQSNILYYFNENLKRNGQPT

VKLKTMQNYLYKLEKEIKVTTNYHKHLGVNFGTEIYYKLNYPKKECYLKINQHFKEKKDIRFQARVNNYLKDKFNKKGNL

NLEECNNNNNNKEEEEEDIRNNKIEKCQIKKYFNKCNFLSEEAKSILELNISKNKTIEIIKIIKKIETDLTKNKNKVCFK

KKQKMLKEILSKTKKQLEKKGYDTKQLKLKIENIYKSYKTKPHFIIENKKYKDLDKIRLKLEKSIEIKKESITKKYIHIK

VNIFNILIEQLKKELEIKTLKPIIKNYLNSKKTLEYDKVFNTYYYELLETIKKTKNSLYKDCQTKMLYKDNV

>lcl|BBE21 unnamed protein product

MLYYVRVYKKPSEVEEAEVKVNKSLQIQSKYQHKLIALIATLEYINKNKKKYNQSDILYCFNSNLRRNGQKEVSIKTLRN

YFYKLEKLNITINYYRHLGINMGTEIYYALRHSKKDCYNLLNQHFRNKKTERFQRRVNVYIKINYDKKDNVKNGECLNNK

YKKEERETERKKRINKLKLKKYAKKCNFDNEISSFIINLNLKKETTIKLFKFIIKEKYYFKKENKCNLQKTLQNKKRDLI

SILRKTQKILIKEGCDKKKIKTQIQNTYQKYKNKPHFILESNKYKDFDQIIKKIKDDTNKTEPQKHKDNIETNIYNILLD

QLHRKTNTTNLRSKIKEYLNKQNKLEYKKIFNNQYYNEIIKLIESQNIYKNSYIN

>lcl|BBT04 unnamed protein product

MNSKTTNKTNRNCYNKVQHKLIVLISTICYLNKTHKKYTQKTILYYFNENLRKNGQTISTLRTMQKYIYRLQKEIKVTTN

YYQHMGVNSGTEIYYKLNYPKKDCYHKINQHFKEKKEERYQNRVANYFNKNSDSKMGSVQLGECNNNNNNIKEERKINEI

EKYQVIKYFNKCDFSCKEILPVLLTLNIDKENIIKIIKILKITEINSKNKNIRPTKSCIKKKQEKLKGILCNTQKELEEN

GYNPKQLEINFQKIYENYKYKPHFIIENHKYSDLNNIKRKLEKSIERKKENSQQDYENLKINVFNILIEQLKKETNIEIL

KPIIKEYLNNQKKIKYNKIFGIYYLELLEIIKNRKHDLNLEKFSKKVV

>lcl|BBC01 unnamed protein product

MENSKKNTLCQNKTQHKSIVLISTLEYINKTHNKYTQKNILYYFNENLKRNGQLPVKIKTLQNYLYKLEKEIKVTTNYHK

HLGVNCGTEIYYQLNFSKKECYQKIYKYFQEKKDLRFQNRATRGLKDRFTKNGSVDLKECLNNKNNIKEERKINEIEKYQ

VRNYFNKCNFLCKKILSIFLTILFNLDIDKDNIIKILKIIKIIEIKLLKNKNIHFTKSCMKEKQEKLKKILCNTQKEFEK

NEYNPKQLEISFQKIYENYKFKPHFIIESHKYSDLNNIKRKLEKSIERKKENSQQNYQDLKTNIFNILIEQLKKEVNIEL

LKPIIKEYLNNQKKIEYNKVFCTYYCELLELIKNQKSLLNLKELDRKAI

>lcl|BBU07 unnamed protein product

MIEILKTIKRTEIKAKNKNIHFTKSCSKEKQEKLKEILCNTQKELEKSGCNSEQLETNFQKIYENYKYKPHFIIENHKYS

DLSYIKRKLEKSIEIKKENPQKDYESLKINIFHIFIEQLKKEINIETLKPLVKEYLNNQKKIKYTKVFDTYYTR

>lcl|BBD06 unnamed protein product

MQKYIYRLQKEIKVTTNYYQHMGVNSGTEIYYKLNYPKKDCYHKINQHFKEKKEKRFQNRVANYFNKNSDTKMGSVQLEN

CNNNIKEERKINEIEKYQVIKHFNKCDFLCKEIISILLTLNIDKENMIKIIKILKITEIKSKNKNIRFTKSCIA

>lcl|BBI01 unnamed protein product

MENYQYFFNLFFIFKIIFIYLYIMYKSVKEQQEKGIDHACRILILTETIFEINLILENYSQKTLLKKYNENLKNKNLPPS

NISTMKKYLKQLEKEIKIIAKFYFKNDQSLIYYKLNYTLEKIWLKLIELFYKELKQFIQKNTTT

>lcl|BBU01 unnamed protein product

MENYQYFFNLFFIFKIIFIYLYIMYKSVKEQQEKGIDHTCRILILTETIFEINLILENYSQKTLLKKYNENLKNKNLPPS

NISTMKKYLNQLEKEIKIIAKFYFKNDQSLIYYKLNYTLEKIWLKLIELFYKELKQFIQKNTTT

>lcl|BBB10 unnamed protein product

MYQIKTNKMPFNKVVDRRLKIFWVIQKLSANYFISKKKYSLSNVVAMTNSILEKKGFKRVTKRTIQNDIKIFETLGLIKS

HFNPLGKNNGSFTYYTINKALEKLAKKIISTAYFIDKKTKHEKSKNKQLKKIKIIEESQKYKISHQITSHVLSNNISKKY

KNSKYSFRRKNQNIKKTINFLEKEIKKKSKSINLEEIKKITENDITYKNSLWNLKDFMEELKEYEEKKIIKFYKKNLEKK

KQKIWFMAKKFKNTDFDKLIKKFKIKNKMEREKNYENENQIHTSNNIKNAIVLMKTLIKKQKYDKKIKK

-muscle -in BBA18.homologs.pep -out BBA18.homologs.aligned

MUSCLE v3.8.31 by Robert C. Edgar

http://www.drive5.com/muscle

This software is donated to the public domain.

Please cite: Edgar, R.C. Nucleic Acids Res 32(5), 1792-97.

BBA18.homologs 21 seqs, max length 398, avg length 326

00:00:00 24 MB(-3%) Iter 1 100.00% K-mer dist pass 1

00:00:00 24 MB(-3%) Iter 1 100.00% K-mer dist pass 2

00:00:00 29 MB(-3%) Iter 1 100.00% Align node

00:00:00 30 MB(-3%) Iter 1 100.00% Root alignment

00:00:00 30 MB(-3%) Iter 2 100.00% Refine tree

00:00:00 30 MB(-3%) Iter 2 100.00% Root alignment

00:00:00 30 MB(-3%) Iter 2 100.00% Root alignment

00:00:01 30 MB(-3%) Iter 3 100.00% Refine biparts

00:00:01 30 MB(-3%) Iter 4 100.00% Refine biparts

00:00:01 30 MB(-3%) Iter 5 100.00% Refine biparts

00:00:01 30 MB(-3%) Iter 5 100.00% Refine biparts

**-cat BBA18.homologs.aligned**

>lcl|BBB10 unnamed protein product

----------------------MYQIKTNKMPFNKVVDRRLKIFWVIQKLSANYFISK--

KKYSLSNVVAMTNSILEKKGFKRVTKRTIQNDIKIFET-LGLIKSHFNPLGKNNGSFTYY

TINKALEKLAKKIISTAYFIDK---KT------------KHEKSKNKQLKKIKIIEES-Q

KYKISHQITSHVLSNNISKKY-KNSKYSFRRKNQNIKK----TINFLEKEIKKKSKSINL

EEI-----------------------------------------KKITEND----ITYKN

SLWNLKDFMEELKE------YEEKKIIKFYKKNLEKKKQKIWFMAKKFKNTDFDKLIKKF

KIKNKMEREKNYENENQIHTSNNIKNAIV--------------LMKTLIKKQK---YDKK

IKK--------------------------------

>lcl|BBI01 unnamed protein product

-MENYQYFFNLFFIFKIIFIYLYIMYKSVKEQQEKGIDHACRILILTETIF---EINLIL

ENYSQKTLLKKYNENLKNKNLPPSNISTMKKYLKQLEKEIKIIAKFYFK---NDQSLIYY

KLNYTLEKIWLKL--IELFYKE--------------------------------------

------------------------------------------------------------

------------------------------------------------------------

------------------------------------------------------------

--------------------------------------------LKQFIQKNTTT-----

-----------------------------------

>lcl|BBU01 unnamed protein product

-MENYQYFFNLFFIFKIIFIYLYIMYKSVKEQQEKGIDHTCRILILTETIF---EINLIL

ENYSQKTLLKKYNENLKNKNLPPSNISTMKKYLNQLEKEIKIIAKFYFK---NDQSLIYY

KLNYTLEKIWLKL--IELFYKE--------------------------------------

------------------------------------------------------------

------------------------------------------------------------

------------------------------------------------------------

--------------------------------------------LKQFIQKNTTT-----

-----------------------------------

>lcl|BBA18 unnamed protein product

-----------------MKTLKKKSPNTTKRVKKTTMNFQHNLIVLISTLN---FINLNF

KKYTQKNILYLLNKNLERNKQKLIKLKTLQNYLYILEKKFKVTLNYCKHLGKNSGSETYY

KLKYEKEKCYLII--NTYFKEKIINKINEFTQRIKKFNQINSSVKWECINNTNNIYKY-K

EYRNIHKNSKKTTNNEIIKKYLSKCNF--KTEIPALIM----NLKTTNSTKIYHLRNLKH

IENDLKEIDPKKIEKHISNVIKENINNPGYLCKFFRNSGYKRIIHKIKEPE----KKYKN

KNEILKKILEEKIKELEKEQYKKEDLEKFFGKTYEIYKIKPHFIIEHKKYPDLEKLVK--

RAKTEIPKIQDKMVRQ-KSIKNNIFSILLEQLRHKVDNDKLIPTLKKFIENEPDLKYSKV

FDNSYYNNLIKIVS---------------------

>lcl|BBE21 unnamed protein product

--------MLYYVRVYKKPSEVEEAEVKVNKSLQIQSKYQHKLIALIATLE---YINKNK

KKYNQSDILYCFNSNLRRNGQKEVSIKTLRNYFYKLEK-LNITINYYRHLGINMGTEIYY

ALRHSKKDCYNLL--NQHFRNK---KTERFQRRVNVYI-KINYDKKDNVKNGECLNNK-Y

KKEERETERKKRINKLKLKKYAKKCNF--DNEISSFII----NLNLKKETTIKLFKFIIK

EKYYFK----------------------------------KE--NKCNLQK----T-LQN

KKRDLISILRKTQKILIKEGCDKKKIKTQIQNTYQKYKNKPHFILESNKYKDFDQIIK--

KIKD--DTNKTEPQKHKDNIETNIYNILLDQLHRKTNTTNLRSKIKEYLNKQNKLEYKKI

FNNQYYNEIIKLIESQNI----------YKNSYIN

>lcl|BBF26 unnamed protein product

----------------------------MKSIKTSTNKYQHKLIILISTLN---YVNSNL

EQYTQNDILYYFNNNIRRNCQKEVKLKTLQNYLYKLEKVFKVTSNYHRHLGINMGTEVHY

KLKYPKKVCYHLI--NKHFKEI---KKKKYKSRVDSYF-NKNCIENSSVKNRECFYNI-S

NNKEEKRNSRKSIEKLQIRKYIEKCNF--KSNIFSLIL----NLGLKKDATIEVCKMIKK

AENLIE----------------------------------KNIFEKLNVVKVIEVNKLKN

KQQKLKRILKEVKIKLENENYKSEQLEIEIKDIYEQYKNKPHFIIEKNKYEDLEKITE--

KLKNKFKQEKINIKENDTNIRNNIFNILIEQLRHKIKIEILIPILKDYLDKQKRLEYSKV

FSNYYYYEILETVEPNKWCS---------QLEACK

>lcl|BBI19 unnamed protein product

MLLFKYCAINVFIKIKITRSINMQNAFTNRKNTPCQNKLQHKLIVFISTLK---YINSKY

KKYTQSNILYYFNENLKRNGQPTVKLKTMQNYLYKLEKEIKVTTNYHKHLGVNFGTEIYY

KLNYPKKECYLKI--NQHFKEK---KDIRFQARVNNYL-KDKFNKKGNLNLEECNNNNNN

KEEEEEDIRNNKIEKCQIKKYFNKCNF-LSEEAKS-IL----ELNISKNKTIEIIKIIKK

IETDLT----------------------------------KN--KN----K----VCFKK

KQKMLKEILSKTKKQLEKKGYDTKQLKLKIENIYKSYKTKPHFIIENKKYKDLDKIRL--

KLEKSIEIKKESITKKYIHIKVNIFNILIEQLKKELEIKTLKPIIKNYLNSKKTLEYDKV

F-NTYYYELLETIKKTKNSLYKDCQTKMLYKDNV-

>lcl|BBG06 unnamed protein product

-MKKVFTFLKKLCIIYNINPIRSSTMINNSKKPNCHNKLQQKLIVLLSTLA---YVNSKY

NKYTQKNILYCFNENLKRNGQPTTTLRTMQNYLYKLEKVFKVTTNYYKHLGINFGTEIYY

KLNYSKKECYLKI--NQHFREK---KDYRFKARVDNYL-NDKFNKNGSVDLVECLNNKNN

NIKEERKI--IQIEKYQVIKYFNKCNFSLLKEILP-IL----NLDINKDELIKILKIIKR

IEINLT----------------------------------KN--KNTYLNV----SYFKE

KQNKLKKILSNTQQQLEKNGYNPEQLEISFQKIYENYKTKPHFIIENHKYKDLSHIKR--

KLEKSIERKKENLQKDCENMRTNIFNILIEQLKKEVKIDVLKPILKIYLNSKNKLEYNKV

FNNNYYYELLEIIKKEKN----------LQLKEVV

>lcl|BBR31 unnamed protein product

----------------------MKDFLITTKNPTCHNKHQHKLIYLTSTVD---FLNKKD

KKYTQQNILYYYNKNLKRNGLAPTTLRTMQNYLYKLEKVLKVTTNYYQHMGVNCGTEIYY

KLKYPKKECYQKI--NKYFKER---KNSRFKSRVNNHF-KDNVSKNSSVNSVECLSNK-N

NIKEERKI--KEIEKYQLRNYFNNCNF-KTEEALS-IL----YLNTDKDTKIEAINILKQ

NEIALI----------------------------------KK--FNIK--K----SCMKE

KQKKLKEILYNTRKKLEENGYNPKQLEINLQKVYENYKYKPHFIIENHKYNDLGYIKR--

KLEKSIERKKENSQQNYQNLKENIFNILIEQLKKEVNIEILKPIIKEYLNNQKKIEYNKV

F-GIYHLELLEIIKNEKNSL----TTEEFSIKAV-

>lcl|BBN30 unnamed protein product

----------------------MQNFPNKTKIPTCHNKHQHKLISLTSTLD---FLNKKD

KKYTQQNILYCFNENLKRNGLAPTTLRTMQNYLYKLEKVLKVTTNYYQHMGVNCGTEIYY

KLKYPKKECYQKI--NKYFKER---KNSRFKSRVNNHF-KDNVSKNSSVNSVECLNNKNN

NTKEERKI--KEIEKYQLRNYFNNCNF-KTEEALS-IL----YLNIDKDTKIEAINILKQ

NEIALI----------------------------------KK--FNIK--K----SCMKE

KQKKLKEILHNTRKKLEENGYNPKQLEINLQKVYENYKYKPHFIIENHKYSDLNNIKR--

KLEKSIERKKENSQQNYQNLKENIFNILIEQLKKETNIEILKPITKKYLNNQKKIEYNKV

F-GTYYLELLEIIRKQKNSL----TTEEFNIKAI-

>lcl|BBQ38 unnamed protein product

----------------------MKDFPNITKIPTCHNKHQHKLISLTSTLD---YLNKKD

KKYTQQNILYCFNENLKRNGLAPTTLRTMQNYLYKLEKVLKVTTNYYQHMGVNCGTEIYY

KLKYPKKECYQKI--NKYFKEQ---KNSRFKSRVNKHF-KDNVSKNSSVNSVGCLSNK-N

NIKEERKI--KEIEKYQLRNYFNNCNF-KTEEALS-IL----YLNIDKDTKIEAINILKQ

NEIALI----------------------------------KK--FNIK--K----SCMKE

KQKKLKEILYNTRKKLEENGYNPKQLEINLQKVYENYKYKPHFIIENHKYSDLNNIKR--

KLEKSIERKKENSQQNYQNLKANIFNILIEQLKKETNIEILKTIVKEYLNNQKKIEYNKV

F-GIYYLELLEIIKNEKNSL----TTEEFSIKAV-

>lcl|BBL30 unnamed protein product

----------------------MKGFLNNAKSPTCYNKHQHKLISLTSTLD---YLNKKD

KKYTQQNILYYFNENLKRNGLAPTTLRTMQNYLYKLEKVLKVTTNYYQHLGVNCGTEIYY

KLKYPKKECYQKI--NKYFKER---KNSRFKSRVNKHL-KDNVSKNGSVNSVECLSNK-N

NIKEERKI--KEIEKYQLRNYFNNCNF-KTEEALS-IL----YLNIDKDTKIEAINILKQ

NEIALI----------------------------------KK--FNIK--K----SCMKE

KQKKLKEILYNTRKKLEENGYNPKQLEINLQKVYENYKYKPHFIIENHKYSDLNNIKR--

KLEKSIKRKKENSQQNYQNLKENIFNILIEQLKKETNIEILKPIVKEYLNNQKKIEYNKV

F-GTYHLELSEIIKNEKNSL----TTEEFSIKAV-

>lcl|BBO30 unnamed protein product

----------------------MKGFPKNTKSPTCHNKHQHKLISLTSTLD---YLNKKD

KKYNQKNILYYYNENLKRNGLAPTTLRTMQNYLYKLEKVLKVTTNYYQHMGVNCGTEIYY

KLKYPKKECYQKI--NKYFKER---KNSRFKSRVNTHF-KDNVSKNSSVNSVECLSNK-N

NIKEERKI--KEIEKYQLRNYFNNCNF-KTEEALS-IL----YLNIDKDTKIEAINILKQ

NEIALI----------------------------------KK--FNIK--K----SCMKE

KQKKLKEILYNTRKKLEENGYNPKQLEINLQKVYENYKYKPHFIIENHKYSDLNNIKR--

KLEKSIERKKENSQQNYQNLKENIFNILIEQLKKETNIEILKPIVKEYLNNQKKIEYNKV

F-GTYHLELSEIIKNEKNSL----TTEEFSIKAV-

>lcl|BBM30 unnamed protein product

----------------------MKDFLNNIKSLTCYNKHQHKLISLTSTLD---FLNKKD

KKYTQQNILYYFNENLKRNGLAPTTLRTMQNYLYKLEKVLKVTTNYYQHMGVNCGTEIYY

KLKYPKKECYQKI--NKYFEER---KNSRFKSRVNNHF-KDDISKNGSVNSVECLSNK-N

NIKEERKI--KEIEKYQVIKYFNKSNL-LCKEILPFLL----TLNVDKDTMIKIIKNIKR

VENKLL----------------------------------KN--TNLN--K----SCFKE

KQEKLKRILNNTQKEFKQNGYNPEQLKINLQKVYESYKFKPHFIIENHKYSDLNNIKS--

KLEKSIERKKENPQQNYQNLKENIFNILIEQLKKETNIEILKPIIKKYLNNQKKIEYNKV

F-GTYHLELLEIIKNEKNSL----TTEEFNIKAV-

>lcl|BBP30 unnamed protein product

----------------------MKGFSNTTKNPTCHNKHQHKLIYLASTLD---FLNKKD

KKYTQQNILYYYNENLKRNGLAPTTLRTMQNYLYKLEKVLKVTTNYYQHMGVNCGTEIYY

KLKYPKKECYQKI--NKYFKER---KNSRFKSRVNNHF-KDNVSKNSSVNSVECLSNK-N

NIKEERKI--NEIEKYQVIKYFNKSNF-LCKEILPFLL----TLNVDKDTMIKIIKNIKR

VENKLL----------------------------------KN--TNLN--K----SCFKE

KQEKLKKILNNTQKEFEQNGYNPEQLKTNLQKVYENYKFKPHFIIERHKYNDLNNIKR--

KLEKAIEREKQNSQQNYQNLKENIFNILIEQLKKETNIEILKPIIKKYLNNQKKIEYNKV

F-GIYHLELSEIIKNEKNSL----TTEEFSIKAV-

>lcl|BBS33 unnamed protein product

----------------------MKGFSNTTKNPTCYNKHQHKLIYLTSTVD---FLNKKD

KKYTQQNILYYYNKNLKRNGLAPTTLRTMQNYLYKLEKVLKVTTNYYQHMGVNCGTEIYY

KLKYPKKECYQKI--NKYFKER---KNSRFKSRVNNHF-KDNVSKNSSVNSVECLSNK-N

NIKEERKI--KEIEKYQVIKYFNKSNF-LCKEILPFLL----TLNVDKDTMIKIIKNTKR

VENKLL----------------------------------KN--TNLN--K----SCFKE

KQEKLKKILNNTQKEFEQNGYNPEQLKTNLQKVYENYKFKPHFIIERHKYSDLNNIKR--

KLEKSIEREKQNSQQNYQNLKENIFNILIEQLKKETNIEILKPIIKKYLNNQKKIEYNKV

F-GIYHLELLEIIKNEKNSL----TTEEFSIKAV-

>lcl|BBC01 unnamed protein product

-------------------------MENSKKNTLCQNKTQHKSIVLISTLE---YINKTH

NKYTQKNILYYFNENLKRNGQLPVKIKTLQNYLYKLEKEIKVTTNYHKHLGVNCGTEIYY

QLNFSKKECYQKI--YKYFQEK---KDLRFQNRATRGL-KDRFTKNGSVDLKECLNNK-N

NIKEERKI--NEIEKYQVRNYFNKCNF-LCKKILSIFLTILFNLDIDKDNIIKILKIIKI

IEIKLL----------------------------------KN--KNIHFTK----SCMKE

KQEKLKKILCNTQKEFEKNEYNPKQLEISFQKIYENYKFKPHFIIESHKYSDLNNIKR--

KLEKSIERKKENSQQNYQDLKTNIFNILIEQLKKEVNIELLKPIIKEYLNNQKKIEYNKV

F-CTYYCELLELIKNQKSLL----NLKELDRKAI-

>lcl|BBU07 unnamed protein product

------------------------------------------------------------

------------------------------------------------------------

------------------------------------------------------------

--------------------------------------------------MIEILKTIKR

TEIKA-----------------------------------KN--KNIHFTK----SCSKE

KQEKLKEILCNTQKELEKSGCNSEQLETNFQKIYENYKYKPHFIIENHKYSDLSYIKR--

KLEKSIEIKKENPQKDYESLKINIFHIFIEQLKKEINIETLKPLVKEYLNNQKKIKYTKV

F-DTYYTR---------------------------

>lcl|BBU04 unnamed protein product

----------MLISFVKESLPMHKIKTTNNNPRNCYNKVQYKLIVLISTIC---YLNKTH

KKYTQKTILYYFNENLRKNGQTISTLRTMQKYIYRLQKEIKVTTNYYQHMGVNSGTEIYY

KLNYPKKDCYHKI--NQHFKEK---KETRFQNRVTNYFNKNSDSKMGSVQCESCNSNK-N

NIKEERKI--NEIEKYQVINYFNKCNF-SCKEILSILL----NLNVDKDTMIKIIKTIKR

TDIKA-----------------------------------KN--KNIYFPK----SCSKE

KQEKLKKILCNTQKELEKSGYNSEQLETNFQKIYENYKYKPHFIIENHKYSDLSYIKR--

KLEKSIERKKENSKQDYKNLRTNIFNILIEQLKKETNIEILKPIIKEYLNNQKKIEYKKV

F-RIYYSELLEIINGKHYS-----NLKKFRKKSVG

>lcl|BBT04 unnamed protein product

----------------------MNSKTTNKTNRNCYNKVQHKLIVLISTIC---YLNKTH

KKYTQKTILYYFNENLRKNGQTISTLRTMQKYIYRLQKEIKVTTNYYQHMGVNSGTEIYY

KLNYPKKDCYHKI--NQHFKEK---KEERYQNRVANYFNKNSDSKMGSVQLGECNNNN-N

NIKEERKI--NEIEKYQVIKYFNKCDF-SCKEILPVLL----TLNIDKENIIKIIKILKI

TEINS-----------------------------------KN--KNIRPTK----SCIKK

KQEKLKGILCNTQKELEENGYNPKQLEINFQKIYENYKYKPHFIIENHKYSDLNNIKR--

KLEKSIERKKENSQQDYENLKINVFNILIEQLKKETNIEILKPIIKEYLNNQKKIKYNKI

F-GIYYLELLEIIKNRKHDL----NLEKFSKKVV-

>lcl|BBD06 unnamed protein product

------------------------------------------------------------

----------------------------MQKYIYRLQKEIKVTTNYYQHMGVNSGTEIYY

KLNYPKKDCYHKI--NQHFKEK---KEKRFQNRVANYFNKNSDTKMGSVQLENCN----N

NIKEERKI--NEIEKYQVIKHFNKCDF-LCKEIISILL----TLNIDKENMIKIIKILKI

TEIKS-----------------------------------KN--KNIRFTK----SCIA-

------------------------------------------------------------

------------------------------------------------------------

-----------------------------------

**-fasttree BBA18.homologs.aligned > tree**

FastTree Version 2.1.8 Double precision (No SSE3)

Alignment: BBA18.homologs.aligned

Amino acid distances: BLOSUM45 Joins: balanced Support: SH-like 1000

Search: Normal +NNI +SPR (2 rounds range 10) +ML-NNI opt-each=1

TopHits: 1.00\*sqrtN close=default refresh=0.80

ML Model: Jones-Taylor-Thorton, CAT approximation with 20 rate categories

Initial topology in 0.01 seconds

Refining topology: 18 rounds ME-NNIs, 2 rounds ME-SPRs, 9 rounds ML-NNIs

Total branch-length 4.813 after 0.10 sec

ML-NNI round 1: LogLk = -7869.999 NNIs 2 max delta 13.77 Time 0.37

Switched to using 20 rate categories (CAT approximation)1 of 20

Rate categories were divided by 0.854 so that average rate = 1.0

CAT-based log-likelihoods may not be comparable across runs

Use -gamma for approximate but comparable Gamma(20) log-likelihoods

ML-NNI round 2: LogLk = -7505.429 NNIs 0 max delta 0.00 Time 0.58

Turning off heuristics for final round of ML NNIs (converged)

ML-NNI round 3: LogLk = -7503.591 NNIs 0 max delta 0.00 Time 0.82 (final)

Optimize all lengths: LogLk = -7503.579 Time 0.89

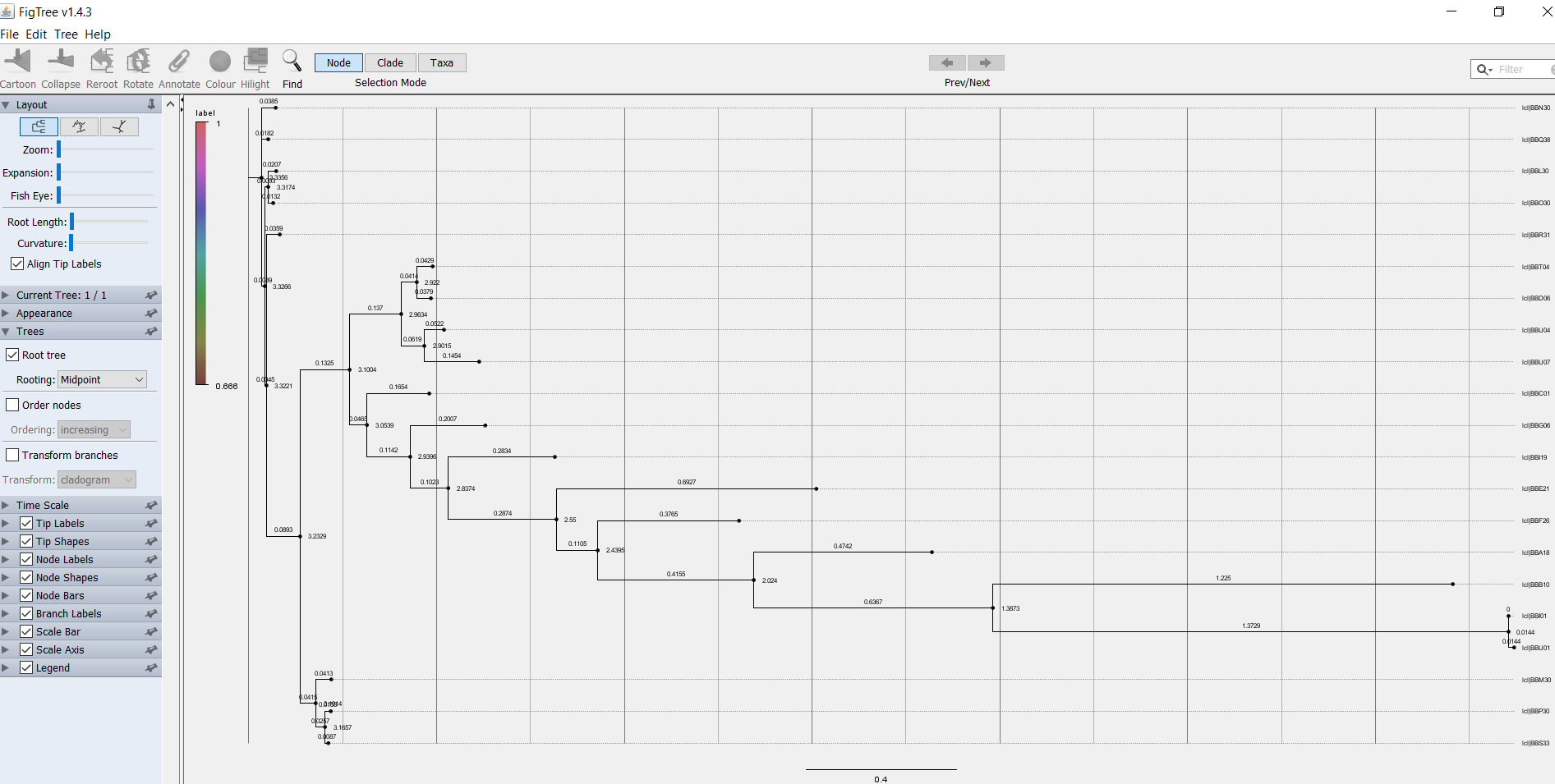
Total time: 1.10 seconds Unique: 21/21 Bad splits: 0/18

**-Cat tree**

(lcl|BBN30:0.038480212,lcl|BBQ38:0.018219773,((lcl|BBL30:0.020713486,lcl|BBO30:0.013216605)0.795:0.009283479,(lcl|BBR31:0.035938273,((((lcl|BBT04:0.042921362,lcl|BBD06:0.037888591)0.950:0.041428703,(lcl|BBU04:0.052210471,lcl|BBU07:0.145414643)0.959:0.061942864)1.000:0.136998798,(lcl|BBC01:0.165425444,(lcl|BBG06:0.200707974,(lcl|BBI19:0.283372188,(lcl|BBE21:0.692741790,(lcl|BBF26:0.376468412,(lcl|BBA18:0.474246067,(lcl|BBB10:1.225014956,(lcl|BBI01:0.000000005,lcl|BBU01:0.014422674)1.000:1.372895384)0.948:0.636726781)0.984:0.415463882)0.857:0.110480019)0.998:0.287364386)0.959:0.102282190)0.978:0.114229148)0.666:0.046548036)1.000:0.132479879,(lcl|BBM30:0.041276208,(lcl|BBP30:0.015347370,lcl|BBS33:0.008724825)0.972:0.025666837)0.978:0.041531538)1.000:0.089255043)0.711:0.004492284)0.691:0.008933705);

**EVOLVIEW :**





**Part 2 :**

**-Cat “dump-coords.pl”**

#!/usr/bin/perl

use strict;

use warnings;

use Data::Dumper; # print complex data structure

# ----------------------------------------

# Author : Muntaha Munia

# Date : March 12, 2018

# Description : Read coord file

# Input : A coord file

# Output : Coordinates and read frame for each gene

# ----------------------------------------

die "Usage:$0 <coord\_file>\n" unless @ARGV == 1;

my @genes; # declare the array

while(<>) { # this means that as long as lines come from the pipe we keep going

my $line = $\_; # a line that come from the pipe (we go line by line)

next unless $line =~ /^\d+/; # skip lines except those reporting genes

#<Your code: split the $line on white spaces and save into variables>

#<Your code: construct anonymous hash and push into @genes>

my @info = split(' ', $line);

my $hash = {

id => $info[0],

start => $info[1],

end => $info[2],

frame => $info[3],

score => $info[4],

};

push(@genes, $hash);

}

print Dumper(\@genes);

exit;

**-cat mys.coord2**

00001 16 324 +1 0.929

00002 751 308 -2 0.911

**-perl dump-coords.pl mys.coord2**

$VAR1 = [

{

'frame' => '+1',

'score' => '0.929',

'end' => '324',

'id' => '00001',

'start' => '16'

},

{

'start' => '751',

'id' => '00002',

'score' => '0.911',

'end' => '308',

'frame' => '-2'

}

];