

Multiple Sequence Alignment Analysis Report: *Saccharomyces cerevisiae*: Cytochrome C subunit 1 (cox1)

1) Choose a gene from your favorite bacterium. What gene did you choose and from what bacterium? Remember to format the species name correctly (Genus name capitalized, specific epithet lower case, entire name italicized).

Gene: *cox1* - cytochrome c oxidase sub unit 1

Organism: *Saccharomyces cerevisiae* S288C

2) Paste the gene sequence.

>NC_027264.1:13017-21174 cox1 [organism=*Saccharomyces cerevisiae*]
[GeneID=24573120] [chromosome=MT]

ATGGTACAAAGATGATTATATTCAACAAATGCAAAAGATATTGCAGTATTATATTTTATGTTAGCTA
TTT

TTAGTGGTATGGCAGGAACAGCAATGTCTTTAATCATTAGATTAGAATTAGCTGCACCTGGTTC
ACAATA

TTTACATGGAAATTCACAATTATATAATGTTTTAGTAGTTGGTCATGCTGTATTAATGATTTTCTGTG
CG

CCGTTTCGCTTAATTTATCACTGTATTGAAGTGTTAATTGATAAACATATCTCTGTTTATTCAATTA
ATG

AAAACTTTACCGTATCATTTTGGTTCTGATTATTAGTAGTAACATACATAGTATTTAGATACGTAAA
CCA

TATGGCTTACCCAGTTGGGGCCAACTCAACGGGGACAATAGCATGCCATAAAAGCGCTGGA
GTAAAACAG

CCAGCGCAAGGTAAGAACTGTCCGATGGCTAGGTTAACGAATTCCTGTAAAGAATGTTTAGG
GTTCTCAT

TAACTCCTTCCCCTTGGGGATTGTGATTCATGCTTATGTATTGGAAGAAGAGGTACACGAGTT
AACCAA

AAATGAATCATTAGCTTTAAGTAAAAGTTGACATTCGGAGGGCTGTACGAGTTCAAATGGAAAA
TTAAGA

AATACGGGATTGTCCGAAAGGGGAAACCCTGGGGATAACGGAGTCTTCATAGTACCCAAATT
TAATTTAA

ATAAAGTGAGATACTTTAGTACTTTATCTAAATTAATGCAAGGAAGGAAGACAGTTTAGCGTAT
TTAAC

AAAGATTAATACTACGGATTTTTCCGAGTTAAATAAATTAATAGAAAATAATCATAATAAACTTGA
AACC

ATTAATACTAGAATTTTAAAATTAATGTCAGATATTAGAATGTTATTAATTGCTTATAATAAAATTAAA
A

GTAAGAAAGGTAATATATCTAAAGGTTCTAATAATATTACCTTAGATGGGATTAATATTTTCATATTTA
AA

TAAATTATCTAAAGATATTAACACTAATATGTTTAAATTTTCTCCGGTTAGAAGAGTTGAAATTCCT
AAA

ACATCTGGAGGATTTAGACCTTTAAGTGTTGGAAATCCTAGAGAAAAAATTGTACAAGAAAGTA
TGAGAA

TAATATTAGAAATTATCTATAATAATAGTTTCTCTTATTATTCTCATGGATTTAGACCTAACTTATCTT
G

TTTAACAGCTATTATTCAATGTAAAAATTATATGCAATACTGTAATTGATTTATTAAAGTAGATTTAA
AT

AAATGCTTTGATACAATTCCACATAATATGTTAATTAATGTATTAAATGAGAGAATCAAAGATAAA
GGTT

TCATAGACTTATTATATAAATTATTAAGAGCTGGATATGTTGATAAAAATAATAATTATCATAATACA
AC

TTTAGGAATCCCTCAAGGTAGTGTTGTCAGTCCTATTTTATGTAATATTTTTTTAGATAAATTAGAT
AAA

TATTTAGAAAATAAATTTGAGAATGAATTCAATACTGGAAATATGTCTAATAGAGGTAGAAATCCA
ATTT

ATAATAGTTTATCATCTAAAATTTATAGATGTAAATTATTATCTGAAAAATTAAAATTGATTAGATTAA
G

AGACCATTACCAAAGAAATATGGGATCTGATAAAAGTTTTAAAAGAGCTTATTTTGTTAGATATG
CTGAT

GATATTATCATTGGTGTAAATGGGTTCTCATAATGATTGTAAAAATATTTTAAACGATATTAATAACTT
CT

TAAAAGAAAATTTAGGTATGTCAATTAATATAGATAAATCCGTTATTAAACATTCTAAAGAAGGAG
TTAG

TTTTTTAGGGTATGATGTAAAAGTTACACCTTGAGAAAAAAGACCTTATAGAATGATTAAAAAAG
GTGAT

AATTTTATTAGGGTTAGACATCATACTAGTTTAGTTGTTAATGCCCCTATTAGAAGTATTGTAATAA
AAT

TAAATAAACATGGCTATTGTTCTCATGGTATTTTAGGAAAACCCAGAGGGGTTGGAAGATTAATT
CATGA

AGAAATGAAAACCATTTTAATGCATTACTTAGCTGTTGGTAGAGGTATTATAAACTATTATAGATTA
GCT

ACCAATTTTACCACATTAAGAGGTAGAATTACATACATTTTATTTTATTCATGTTGTTTAACATTAG
CAA

GAAAATTTAAATTAATACTGTTAAGAAAGTTATTTTAAAATTCGGTAAAGTATTAGTTGATCCTCA
TTC

AAAAGTTAGTTTTAGTATTGATGATTTTAAAATTAGACATAAAATAAATATAACTGATTCTAATTATA
CA

CCTGATGAAATTTTAGATAGATATAAATATATGTTACCTAGATCTTTATCATTATTTAGTGGTATTTG
TC

AAATTTGTGGTTCTAAACATGATTTAGAAGTACATCACGTAAGAACATTAAATAATGCTGCCAAT
AAAAT

TAAAGATGATTATTTATTAGGTAGAATGATTAAGATAAATAGAAAACAAATTACTATCTGTAAAACA
TGT

CATTTTAAAGTTCATCAAGGTAAATATAATGGTCCAGGTTTATAATAATTATTATACTATTAAATATG
CG

TTAAATGGAGAGCCGTATGATATGAAAGTATCACGTACGGTTCGGAGAGGGCTCTTTTATATGA
ATGTTA

TTACATTCAGATAGGTTTGCTACTCTACTCTTAGTAATGCCTGCTTTAATTGGTGGATTGTTGTA
GATA

AAAATATGAAAGTAATAATAATCAAATAATAAAAAATAAATAATATAATAGTTTTAATCTAAAAT
AA

AATTATGATAATTTAGAATCTTATTTAGTTGAATTAATTGAAGGTGATGGAACATTCTAGTTCAAG
ATT

CATTAACAAAAAGTAAACAAAAAAATATAGATCTATAATTATTCTAATATTAAAATTATAATATTTTT
AA

TTAGCTAATTATTTATATAATTTAATTAATGTGGTAAAGTATATAAAAAATTAATTGTAATTATGTATT
ATGACTTATTTATGATTTAAAAGGTGTATATACATTATTAATATTATTAATGGATATATAAAAAACAC
CT

AAATATGAAACATTTGTAAGAAGTGTTTAATTTATAAATAATTATATTAATTCTCCTTTCGGGGTTC
CGG

CTCCCGTGGCCGGAACCTCCGTAACCTATAAAAATTCATCTATCTTATGGGATGAAGACTAAATAA
TAAATA

AAAAAATATAAATAATATTAAAATTAACCATTAGATATATTAGATATTGGTTCAAACGCTTAATTA
GAT

GGTATGACTGATGCGGATGGTAATTTTTCTATTAATTTAATAAATAGTAAATCGTTCTAATAAAGC
AAT

GCCTTATTATTGTTTATAATTA AAACTTATTATTACTAAAAATAATAATAATATTAATTTATCATATTTT
AATATTATGTTTATAATTGCACTATATTTAATGTTAATTTATATAATAGAGAACGTAATTTAAATTTAC
AATCTTCTCTAAAAAAGTACATATAACTATATTATAGTTATAAAGTAATATTAGCTAATCTATATAAA
AA

TATTAAGTTATAAAATACTTTAATAAATATTTTTTATTATCATCTAAACATTAGATTATTTATATTGA
TTTAGATTAGTTATTTTAATTTATAATGAAGGTCAAAGTATAAACTTAATGGTAGTTAAGAATTAGGT
CT

AAATTTACGTAAAAATTATAATAAACTATAACTACATTTAATTAATTAGATAAAAAAAATACATATT
TA

TAAAATAAATAATATATTTTCTATTTTATGGTTTTTATTATTACTTTTTCTCCACCAAACCTTGTTAATAA
TATAATAACAAGTCAATAATTTTCGTAATATAAAGATGTAACGAAATTATTAATAAAATAAGTTGCCG
TAA

TATATTGTAAAATATATTATTATTACAACACTATATGCGGGAAAATCCTAAAGTTATAATATAATATT
TT

GTGGAAATTCAATTAAGTTATGAATGTACTATAATATTTAAAAATTATTATAAATAATTATCCTCATT
AT

TATAATTATAATTATAATTCTAATTATAATATAAAAAAGTTAGGATAAATATGGACAATCCGCAGGAA
ACC

AAATAATAATAGGAATATATATTAAGATATATATATATATATTTATATATATTCTTTTTTTTTTCAATA

TATTATATTTAGTATTATTGGAAGAGGGGGTAATATATATTTAATTATTACGTAGGATCCTCAGAGA
CTA

CACGTGTTGCACCCATTATATTAATATTATATTATATTTTATTATAATGGGTTGAAGATATAGTCCAA
AT

ATAATTGAAAGATTATAATAAAAATGAACTATTTATTACCATTAATGATTGGAGCTACAGATACAG
CATT

TCCAAGAATTAATAACATTGCTTTTTGAGTATTACCTATGGGGTTAGTATGTTTAGTTACATCAACT
TTA

GTAGAATCAGGTGCTGGTACAGGGTGAAGTGTCTATCCACCATTATCATCTATTCAGGCACATT
CAGGAC

CTAGTGTAGATTTAGCAATTTTTGCATTACATTTAACATCAATTCATCATTATTAGGTGCTATTAAT
TT

CATTGTAACAACATTAAATATGAGAACAAATGGTATGACAATGCATAAATTACCATTATTTGTATG
ATCA

ATTTTCATTACAGCGTTCTTATTATTATTATCATTACCTGTATTATCTGCTGGTATTACAATGTTATTA
T

TAGATAGAACTTCAATACTTCATTCTTTGAAGTAGCAGGAGGTGGTGACCCAATCTTATATGA
GCATTT

ATTTTGATTCTTTGGTCATCCAGAAGTATATATTTTAATTATTCCTGGATTTGGTATTATTTACATG
TA

GTATCAACATATTCTAAAAAACCTGTATTTGGTGAAATTTCAATGGTATATGCTATGGCTTCAATT
GGAT

TATTAGGATTCTTAGTATGATCACATCATATGTATATTGTAGGATTAGATGCAGATACTAGAGCATA
TTT

CCTATCCGCACTGATGATTATTGCAATTCCAACAGGAATTAAAATCTTTTCTTGATTAGCCCTAA
TCTAC

GGTGGTTCAATTAGATTAGCACTACCTATGTTATATGCAATTGCATTCTTATTCTTATTCACAATG
GGTG

GTTTAACTGGTGTTCCTTAGCTAACGCCTCATTAGATGTAGCATTCCACGATATTAATTTAATA
AGTGT

CGTGCTTAAAATTCATAAAATAATATATAATAAATTATAATAATATATAAAAAAATAAAAAAA
ATA

AAAAAAATTAATATCTTATGATTAATTTTATATAAATAAAAAATTTATTAAATATTATTGGTTATATAT
ATATATATTAATAATAAAAAAATATATATATATATAGCTAACGGGGAACTCTTATAATTATTATTT
A

TATAATAATAAGACAATCCCGTGATAACTTTAATATATATATATTATATATTAAAGTATTGTAGAGA
CT

AAACGTGAATGATTTTAATATTATTTAAATATTAATAAAGAGATAGTCCAATCTTATATGTAAATA
TA

AGTTAATACCAAAAAAAAAATAATATTATTTGACTTATTATATATTAATATTATTAATAATAATTTTAA
CTAATAATAAAGTTTTTATAGAACTTTATATTATTATTTAATATTTAATTTTCAATTAATATCTCCTTT
TGGGGTTCGGTCCCTGGTCCGGCCCCCGAACTAAAGATATTAAGAATTTATATGAATCAAT
TATAAAT

AATTATATTAATATTTTAAATAAATATCTTATTAATATTAATAAAGATAATATTAATAAATTTAAATTTT
TAGATAATTATACTGAAGAAGAAAAAGGTTATTATTTATCTGGATTATTTGAAGGAGATGGTAATAT
TTA

TACGAGATGTTTTTCAATTACTTTTTCTTTAGAAGATGTTTTATTAGCTAATTATTTATGTCTTTATTTT
AAAATTGGTCATATTACAGCTAAATATAATTCTAATAAAGAATTAACAGCTGTAAATGAAATATTA
TAA

AAAAAAAAGAACAAGAAGTATTTATAAATTATATTAATGGTAAATTATTAACATATAAAAGATATGA
TCA

ATATTTTAAATATAATTTTAAATAATCGTTTAAATATTAATTTATTAACCTAAAGAATTTGATTTACTA
TTAAATCCTTGATTAACAGGTTTAAATGATGCTGATGGTTATTTTTATCTAGGTTTTCAAAAACATA
AAA

ATAGTCAATGATTAAAATTTCAATTTAGAATTATCACAAAAAGATAATTATATTTTAGATATTATTA
AA

ATATTTTAAACTTGGAGGTATTTTAAAAAGAGATTATAAATCTGGTGCTACAGCTTATATTTATA
GCT

CAATCATCAAAAGCTATAAACCTTTTATTGAATATTTTAATAATTATCAACCATTAAGTCTTAGA
GAT

ATAACAATATTTATTATTAATATTGCTTACTTATTAATAAATAAATTACATATATTACTTAATTC
TTTATTAATATTAAGAATTGATATTATTACAAAGTGTTAAAAATATATCTTTAGAAATAAAAAATG
AA

TTAAATAATAGAGTTAAAATTATTATTAATAAACTTCATTATAACAATATCGAATAATGATAATATTA
A

GAGTAAAATTCTTAAAGTGTTAATTAAATAATATTCTTTTTTTATGACTTACTACGTGGTGGGACA
TTT

TCGAGCGGTCTGAAAGTTATCATAAATAATTTACCATATAATAATGGATAAATTATATTTTATCA
AT

ATAAGTCTAATTACAAGTGTATTAAATGGTAACATAAATATGCTAAGCTGTAATGACAAAAGTAT
CCAT

ATTCTTGACAGTTATTTTATATTATAAAAAAAGATGAAGGAACTTTGACTGATCTAATATGCTCA
ACG

AAAGTGAATCAAATGTTATAAAATTACTTACACCACTAATTGAAAACCTGTCTGATATTCAATTAT
TATT

TATTATTATATAATTATATAATAATAAATAAAATGGTTGATGTTATGTATTGGAAATGAGCATACGATA
A

ATCATATAACCATTAGTAATATAATTTGAGAGCTAAGTTAGATATTTACGTATTTATGATAAAACAG
AAT

AAACCCTATAAATTATTATTATTAATAATAAAAAATAATAAATACCAATATATATATTATTTAATTTA
TTATTATTATTAATAAAATTTAATATATATTATAAATAATTATTGGATTAAGAAATATAATATTTTAT
AGAAATTTTCTTTATATTTAGAGGGTAAAAGATTGTATAAAAAGCTAATGCCATATTGTAATGATAT
GGA

TAAGAATTATTATTCTAAAGATGAAAATCTGCTAACTTATACTATAGGTGATATGCCTATCTTTATTT
AT

ATATATATTATTATTATTAATAATAAAAAAAAAAAATTAAAAAAAAAGATAGGAGGTTTATATATAACTG
AT

AAATATTTATTATATTATTTTTTTTTTATAATAAATATTTAAAAGATATTGCGTGAGCCGTATGCGATGA
AA

GTCGCACGTACGGTTCTTACCGGGGGGAAAACCTTGTAAGGTCTACCTATCGGGATACTATGTA
TTATCAA

TGGGTGCTATTTTCTCTTTATTTGCAGGATACTATTATTGAAGTCCTCAAATTTAGGTTTAAACTA
TAA

TGAAAAATTAGCTCAAATTCAATTCTGATTAATTTTCATTGGGGCTAATGTTATTTTCTTCCCAAT
GCAT

TTCTTAGGTATTAATGGTATGCCTAGAAGAATTCCTGATTATCCTGATGCTTTCGCAGGATGAAA
TTATG

TCGCTTCTATTGGTTCATTCAATTGCACTATTATCATTATTCTTATTTATCTATATTTTATATGATCAATT
AGTTAATGGATTAAATAATAAAGTTAATAATAAATCAGTTATTTATAATAAAGCACCTGATTTTGTA
GAA

TCTAATCTTATCTTTAATTTAAATACAGTTAAATCTTCATCTATCGAATTCTTATTAACCTTCTCCACC
AG

CTGTACACTCATTTAATACACCAGCTGTACAATCTTAA

3) Using BLAST, find the 3 closest relatives from different genera (each homolog should be from a different genus as in the example AND a different genus from one another, meaning you should have 4 genera total) for your gene sequence. List the three species. Which type of BLAST search did you do for this step (e.g. BLASTn, BLASTx, etc)?

Nucleotide BLASTx:

- *Saccharomyces cerevisiae* S288C

1) *Kluyveromyces lactis*

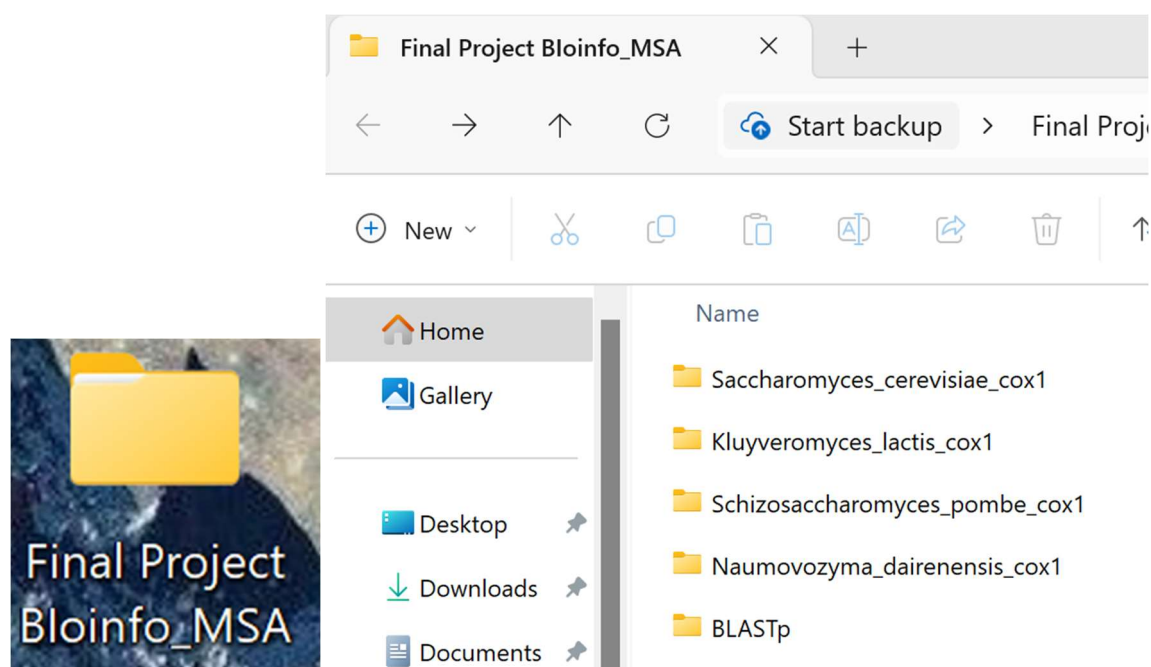
2) *Schizosaccharomyces pombe*

3) *Naumovozyma dairenensis*

13 sequences selected ?

Organism	Blast Name	Score	Number of Hits	Description
Ascomycota	ascomycete fungi		29	
• Saccharomycetaceae	budding yeasts		24	
• Saccharomyces cerevisiae	budding yeasts	105	1	Saccharomyces cerevisiae hits
• Naumovozyma dairenensis	budding yeasts	541	4	Naumovozyma dairenensis hits
• Naumovozyma castellii	budding yeasts	491	6	Naumovozyma castellii hits
• Nakaseomyces castellii	budding yeasts	491	2	Nakaseomyces castellii hits
• Kazachstania servazzii	budding yeasts	496	4	Kazachstania servazzii hits
• Kluyveromyces lactis	budding yeasts	491	2	Kluyveromyces lactis hits
• Kluyveromyces lactis NRRL Y-1140	budding yeasts	491	1	Kluyveromyces lactis NRRL Y-1140 hits
• Kazachstania sinensis	budding yeasts	489	2	Kazachstania sinensis hits
• Nakaseomyces glabratus	budding yeasts	489	2	Nakaseomyces glabratus hits
• Schizosaccharomyces pombe	ascomycete fungi	570	5	Schizosaccharomyces pombe hits

4) Create a file like the demo file for your gene and its closest homologs. Add a screenshot of the file. If you are having trouble creating a .fas file, you can download my demo file, replace my sequences with yours, then change the name.



5) Find the amino acid sequence for your gene of interest (the same gene you selected for number 1). Paste the amino acid sequence.

```
>YP_009144706.1 cox1 [organism=Saccharomyces cerevisiae] [GeneID=24573120]
```

```
MVQRWLYSTNAKDIAVLYFMLAIFSGMAGTAMSLIIRLELAAPGSQYLHGNSQLYNVLVVGHAVLMI  
FFL
```

VMPALIGGFGKYLPLMIGATDTAFPRINNIAFWVLPMLVCLVTSTLVESGAGTGWTVYPPLSSIQA
HS

GPSVDLAIFALHLTSISSLLGAINFIVTTLNMRTNGMTMHKLPLFVWSIFITAFLLLLSLPVLSAGITML
LLDRNFNTSFFEAVAGGGDPILYEHLFWFFGHPEVYILIIPGFGIISHVVSTYSKKPVFGEISMVYAMASI
GLLGFLVWSHHMYIVGLDADTRAYFTSATMIIAIPITGIKIFSWLATIYGGIIRLATPMLYIAIAFLFLFTM
GGLTGVALANASLDVAFHDTYYVVGHFHYVLSMGAIFSLFAGYYYWSPQILGLNYNEKLAQIQFWLI
FIG

ANVIFPPMHFLGINGMPRRIPDYPDAFAGWNYVASIGSFIATLSLFLFIYILDQLVNGLNKVNKNS
VI

YNKAPDFVESNTIFNLNTVKSSSIEFLTSPPAVHSFNTPAVQS

6) Using BLAST, find the 3 closest relatives from different genera (each homolog should be from a different genus as in the example) for your amino acid sequence. List the three species. Which type of BLAST search did you do for this step (e.g. BLASTn, BLASTx, etc)?

BLASTp

- ***Saccharomyces cerevisiae S288C***

1) *Naumovozyma dairenensis*

2) *Nakaseomyces delphensis*

3) *Torulaspora quercuum*

12 sequences selected

Organism	Blast Name	Score	Number of Hits	Description
Saccharomycetaceae	budding yeasts		31	
• Saccharomyces	budding yeasts		8	
• • Saccharomyces cerevisiae	budding yeasts	1066	1	Saccharomyces cerevisiae hits
• • Saccharomyces cerevisiae YJM993	budding yeasts	1066	1	Saccharomyces cerevisiae YJM993 hits
• • Saccharomyces cerevisiae YJM978	budding yeasts	1066	1	Saccharomyces cerevisiae YJM978 hits
• • Saccharomyces cerevisiae YJM981	budding yeasts	1066	1	Saccharomyces cerevisiae YJM981 hits
• • Saccharomyces cerevisiae YJM984	budding yeasts	1066	1	Saccharomyces cerevisiae YJM984 hits
• • Saccharomyces cerevisiae YJM987	budding yeasts	1066	1	Saccharomyces cerevisiae YJM987 hits
• • Saccharomyces cerevisiae YJM990	budding yeasts	1066	1	Saccharomyces cerevisiae YJM990 hits
• • Saccharomyces cerevisiae YJM996	budding yeasts	1066	1	Saccharomyces cerevisiae YJM996 hits
• Naumovozyma dairenensis	budding yeasts	993	2	Naumovozyma dairenensis hits
• Nakaseomyces delphensis	budding yeasts	992	2	Nakaseomyces delphensis hits
• Nakaseomyces nivariensis	budding yeasts	988	2	Nakaseomyces nivariensis hits
• Nakaseomyces braccarenensis	budding yeasts	988	2	Nakaseomyces braccarenensis hits
• Torulaspora quercuum	budding yeasts	977	4	Torulaspora quercuum hits
• Torulaspora globosa	budding yeasts	976	2	Torulaspora globosa hits
• Nakaseomyces glabratus	budding yeasts	974	2	Nakaseomyces glabratus hits
• Torulaspora microellipsoides	budding yeasts	974	2	Torulaspora microellipsoides hits
• Kazachstania sinensis	budding yeasts	974	2	Kazachstania sinensis hits
• Kluyveromyces lactis	budding yeasts	971	2	Kluyveromyces lactis hits
• Kluyveromyces lactis NRRL Y-1140	budding yeasts	971	1	Kluyveromyces lactis NRRL Y-1140 hits

7) Perform a multiple sequence alignment of your gene AND protein sequences using T-COFFEE. Paste a screenshot of the results here. In a couple of sentences, use your own words to describe what a multiple sequencing alignment is and how it is

generated.

```

cox1      TTAAACTATAATGAAAAATTAGCTCAAATTCATTCTGATTAATTTTCAT
cox1_1    TTAATGTACAATGAACTCTAGCTTCAATCCAATTCTGGATTTTATTCAT
cox1_2    TTATATTATAATGAAAGATTATCACAAATTCATTCTGATTAATTTTGT
cox1_3    TTATATTATAATGAAAAATTAGCACAAATTCATTCTGATTAATTTTGT
          ***  **  *****  ** *  ***  *****  *  *  **  *

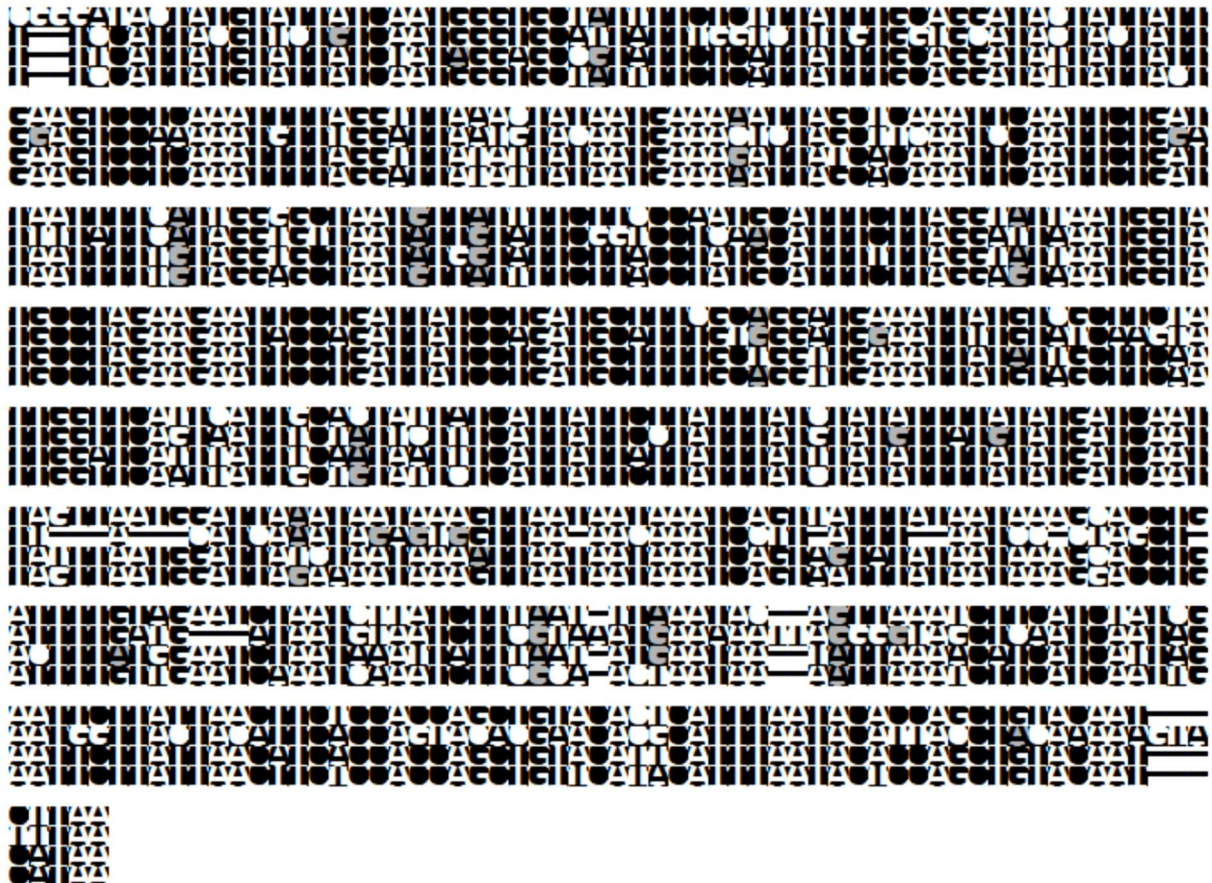
cox1      TGGGGCTAATGTTATTTTCTTCCCAATGCATTTCTTAGGTATTAATGGTA
cox1_1    AGGTGTTAATATTGTATTTCGGTCCTCAACATTTCTTAGGATTAATGGTA
cox1_2    AGGTGCTAATATGGTATTCTTACCTATGCATTTTCTTAGGTATTAATGGTA
cox1_3    AGGAGCTAATGTTATTTTCTTACCTATGCATTTCTTAGGAGTAAATGGTA
          ** *  *****  *  ***  **  *****  *****  *  *****

cox1      TGCCTAGAAGAATTCCTGATTATCCTGATGCTTTTCGCAGGATGAAATTAT
cox1_1    TGCCTAGAAGAATACCAGATTATCCAGATGCATTTGTGGGATGGAATTTT
cox1_2    TGCCTAGAAGAATTCCTGATTATCCTGATGCTTTTGTGGTTGAAATTAT
cox1_3    TGCCTAGAAGAATTCCTGATTATCCTGATGCTTTTGCAGGTTGAAATTAT
          *****  *****  **  *****  *****  ** *  **  *  *****

cox1      GTCGCTTCTATTGGTTCATTCAATTGCACTATTATCATTATTCTTATTTAT
cox1_1    GTATCAAGTATTGGTTCAGTAATTTCTATTCTTTCATTATTCTTATTTAT
cox1_2    ATTGCTTCAATTGGATCATTATTCTTCAATAATTTCAATTATTATTATTTAT
cox1_3    GTAGCTTCAATTGGTTCATTTATTGCTGTATTCTCATTATTCTTATTTAT
          *  *  *****  ** *  ***  *  *  *****  *****

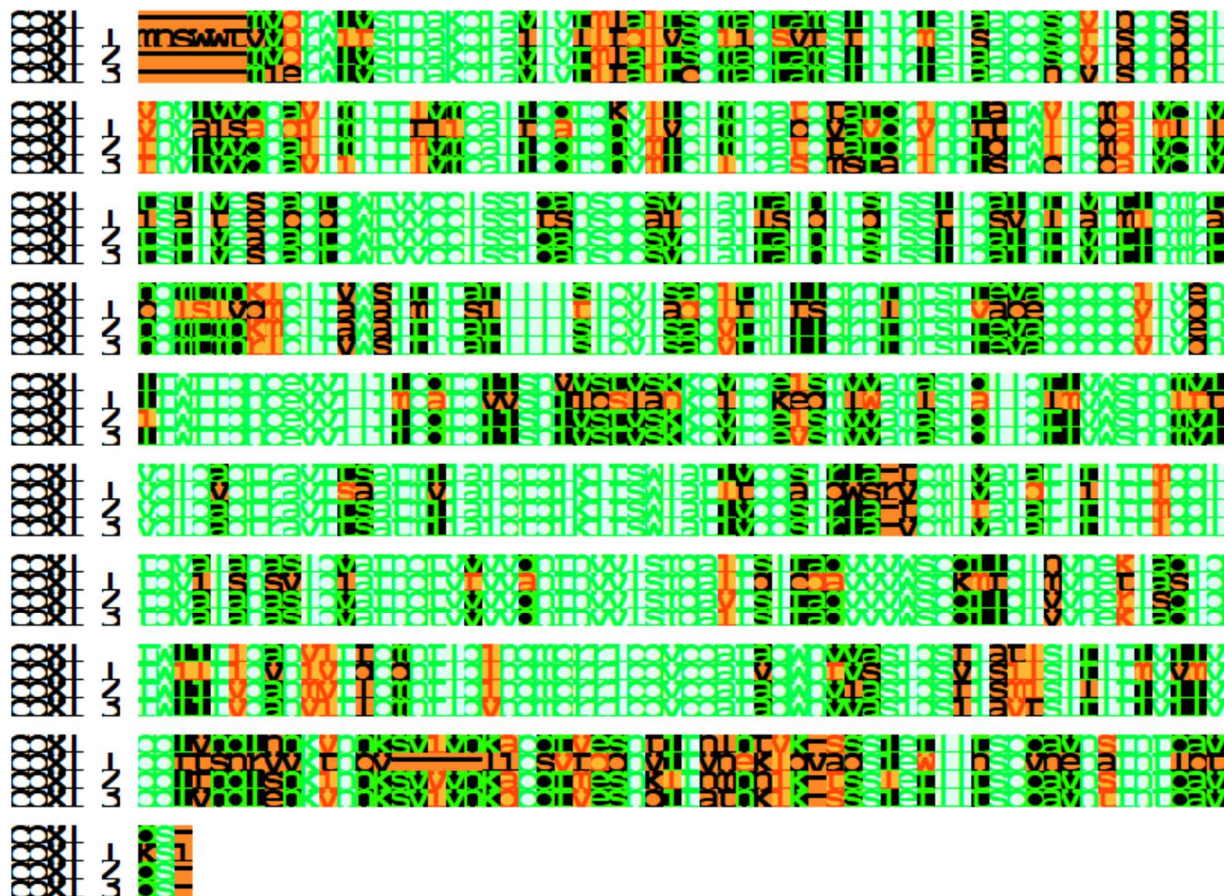
cox1      CTATATTTTATATGATCAATTAGTTAATGGATTAAATAATAAAGTTAATA
cox1_1    GTATGTTATGTATGATCAATTT--A---CATCAAATAGAGTGGTTAA-A
cox1_2    TTATATTTTATATGATCAATTATTTAATGGATTATCTAATAAAATTAATA
cox1_3    CTATATTTTATATGATCAATTAGTTAATGGATTAGAAAATAAAGTTAATA
          ***  ** *  *****  *****  *  ** *  *  *****  *

```



A multiple sequence alignment is a method used to align two or more biological sequences, such as DNA, RNA, or proteins, to identify regions of similarity and differences. It is generated by algorithms that compare the sequences and insert gaps to maximize the alignment, highlighting conserved regions and differences among the sequences.

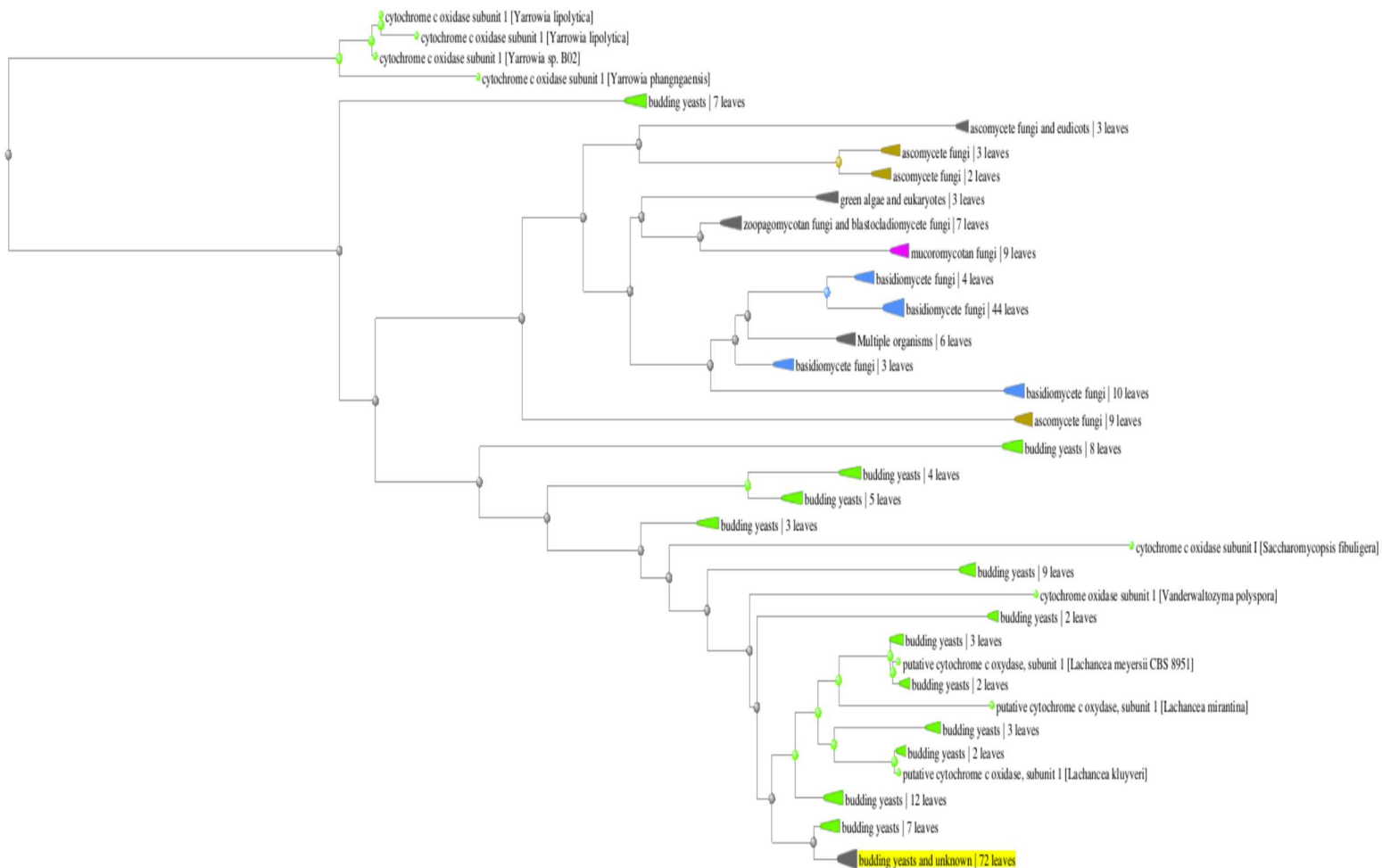
8) Generate a boxshade of your protein alignment (using whatever settings you would like!) and include a screenshot of the results here.



9) Using MEGA, please create a pairwise distance of your protein sequences. Paste a screenshot of the results here. Based on the results of your pairwise alignment, predict which two species are most closely related and which two species are most distantly related. How do you know?

In the analysis of phylogenetic trees, *Yarrowia lipolytica* and *Yarrowia phangngensis* are depicted as the most closely related species, sharing a branch early in the tree which points to a recent common ancestor and minimal evolutionary divergence. Conversely, *Saccharomyces cerevisiae* (S288C) and species from the basidiomycete fungi group, like

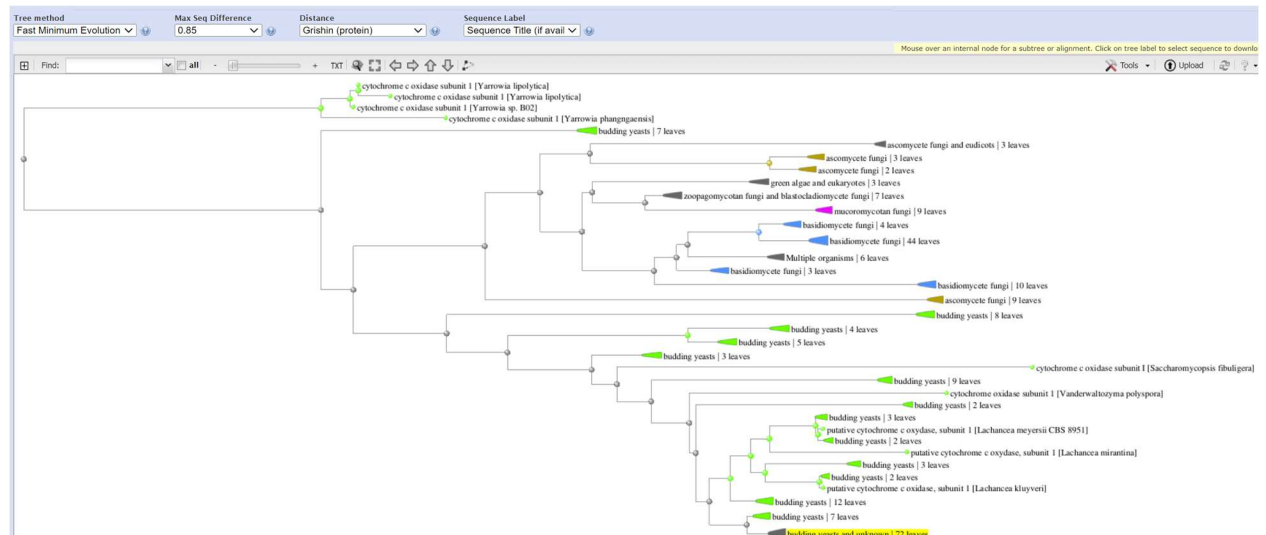
Tremella mesenterica, are found to be the most distantly related, positioned in entirely separate clades marked by significant distances and divergent evolutionary paths. In a separate query analysis, *Naumovozyma dairenensis* and *Nakaseomyces delphensis*, both members of the Saccharomycetaceae family, appear to be the closest genetically, potentially due to shared sequences within their family, whereas *Nakaseomyces delphensis* and *Torulaspora quercuum* represent the most distant relation within the same family, attributed to their belonging to different genera and evolutionary trajectories. These assessments rely on the assumption that genetic proximity and evolutionary closeness correlate with shorter connecting branches in phylogenetic trees, which can be further substantiated by detailed sequence alignments.



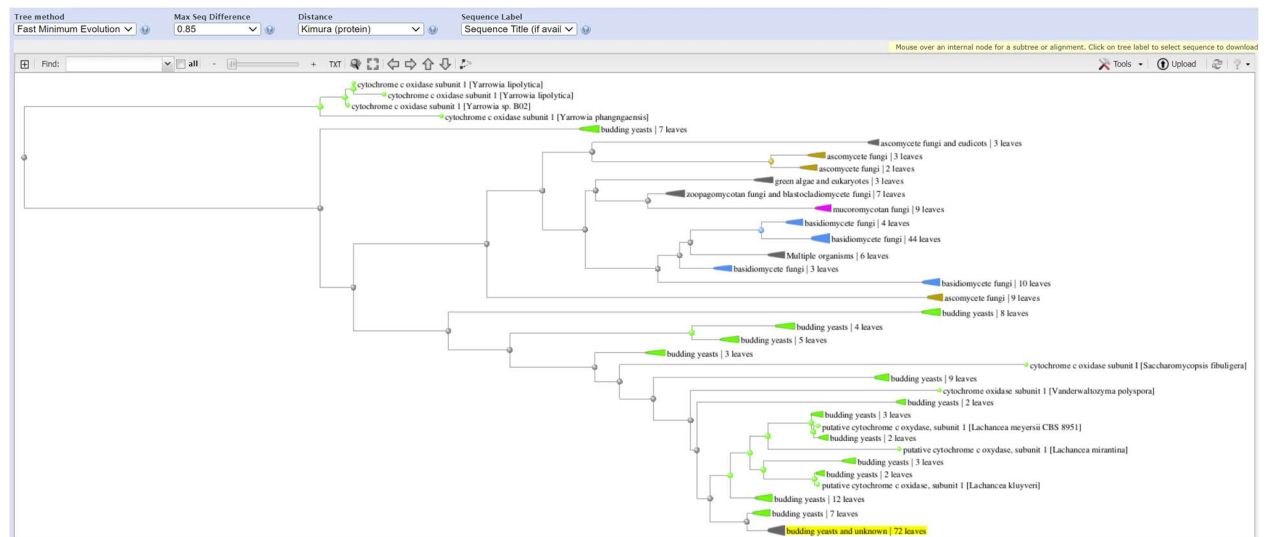
10) Generate two different trees from your protein alignment. For each tree, describe what type of tree you did, include a one sentence explanation of what this type of tree is, and how many bootstraps you did. You should also include a brief explanation (in your own words) of what bootstrapping is. Do your trees support your hypothesis about relatedness from your pairwise distances?

1. **Fast Minimum Evolution:** is a method for constructing phylogenetic trees by finding the topology that minimizes the total branch length, using bootstrapping to assess the stability of the tree topology by repeatedly resampling the data and constructing trees for each sample.

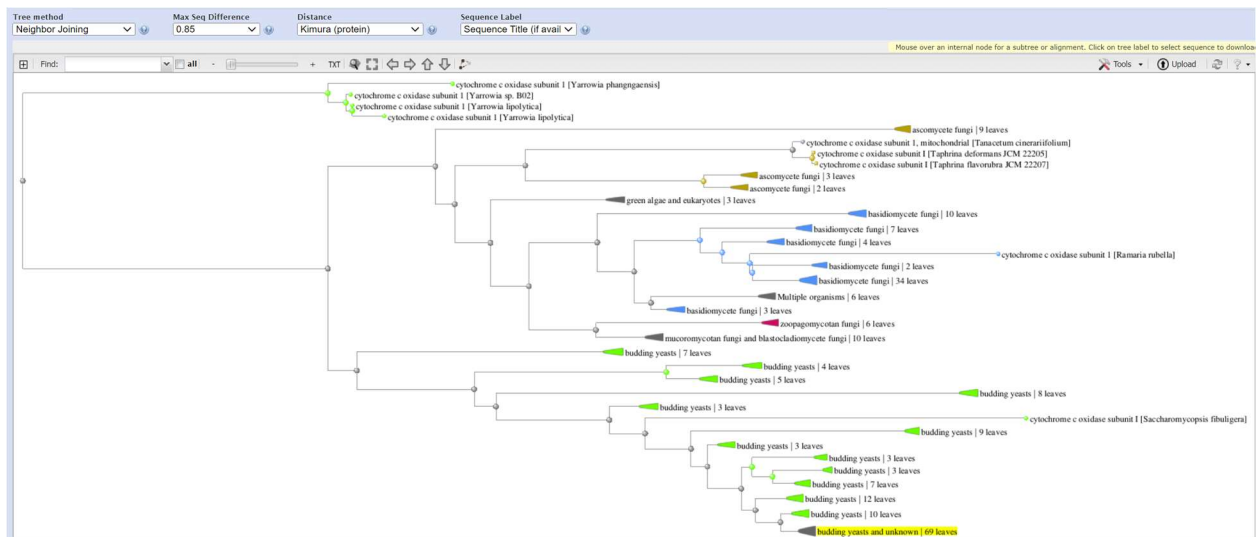
Fast Minimum Evolution: Grishin (0.85)



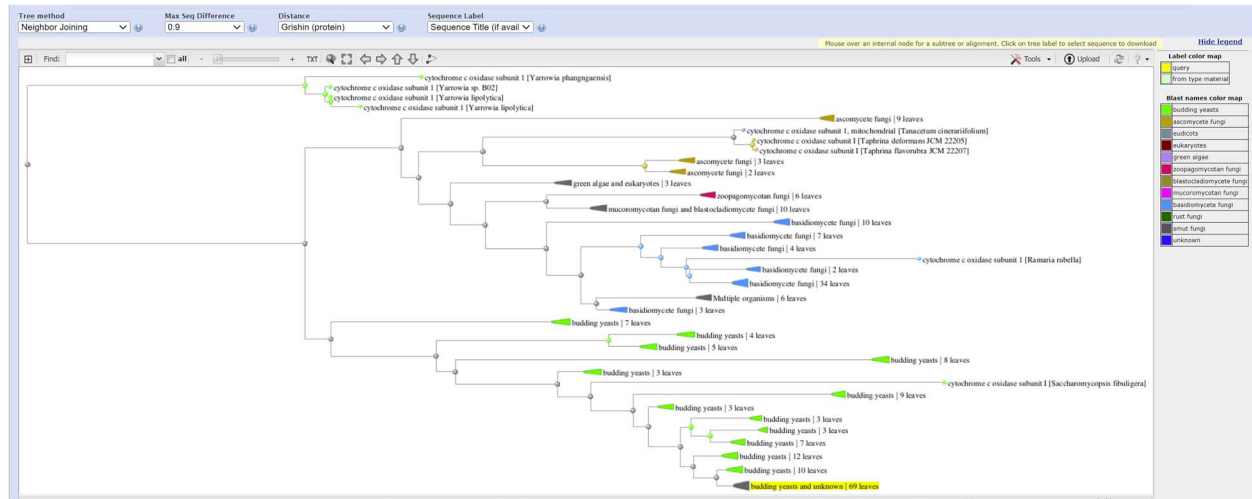
Fast Minimum Evolution: Kimura (0.85)



2. **Neighbor-Joining:** is a distance-based method for constructing phylogenetic trees that iteratively joins pairs of nodes with the smallest distance, using bootstrapping to estimate the confidence of the tree's branches by assessing the frequency of groupings across resampled datasets.



Neighbor Joining: Grishin (0.90) – test bootstrapping.



Bootstrapping: is a statistical method to estimate the reliability of phylogenetic trees. It involves repeatedly resampling the data and reconstructing the phylogenetic tree for each sample. The bootstrap values show how often a particular branch or grouping of species appears in the resampled trees, providing a measure of its support or reliability.

Yes, these data support the hypothesis that *Naumovozya dairenensis*, *Nakaseomyces delphensis*, *Torulaspora quercuum* are closely related to *Saccharomyces cerevisiae* based on the analysis of these MSA and phylogenetic trees. This phylogenetic analysis, showed that *Naumovozya dairenensis* and *Nakaseomyces delphensis* are both part of the Saccharomycetaceae family, suggesting a close phylogenetic relationship due to shared genetic markers typical for this group. These two species, belonging to different genera within the same family, show significant genetic similarities that suggest a common evolutionary path. In contrast, *Torulaspora quercuum*, also a member of the Saccharomycetaceae, is placed in a separate branch due to distinct genetic characteristics that differentiate it from *Naumovozya* and *Nakaseomyces*. *Torulaspora quercuum* shares an evolutionary pathway that coincides with *Candida* and other fungi. This separation in the phylogenetic tree focuses on the genetic and evolutionary differences that define genus-level distinctions within Saccharomycetaceae.

11) Are these trees rooted or unrooted? How do you know?

Rooted, because rooted trees have a designated common ancestor, while unrooted trees do not.