MCDB 187 Lab Introduction

Colin Farrell / Feiyang Ma

MCDB 187

April 3, 2019

Contact Info

Lab Notes: https://nuttylogic.github.io/teaching/

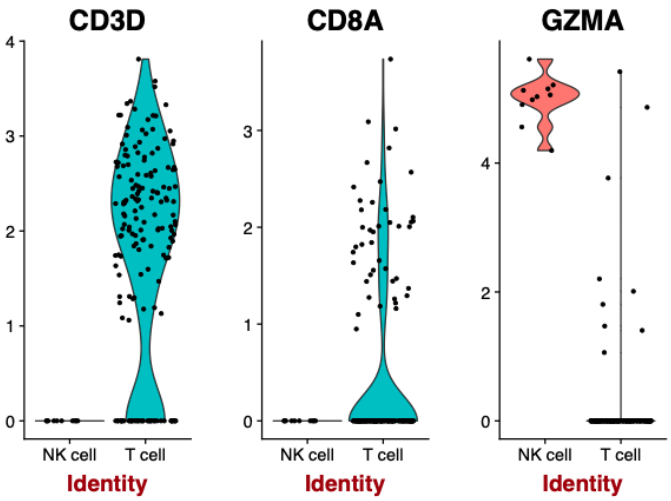
- Colin Farrell
- 3rd Year PhD Student, Human Genetics
- colinpatfarrell@g.ucla.edu

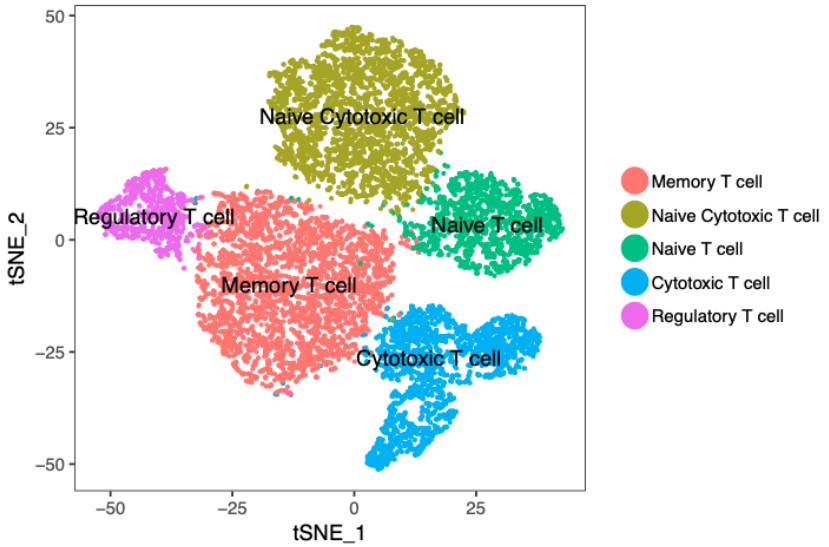
- Feiyang Ma
- 3rd Year PhD Student, Molecular Biology
- mafeiyang@g.ucla.edu

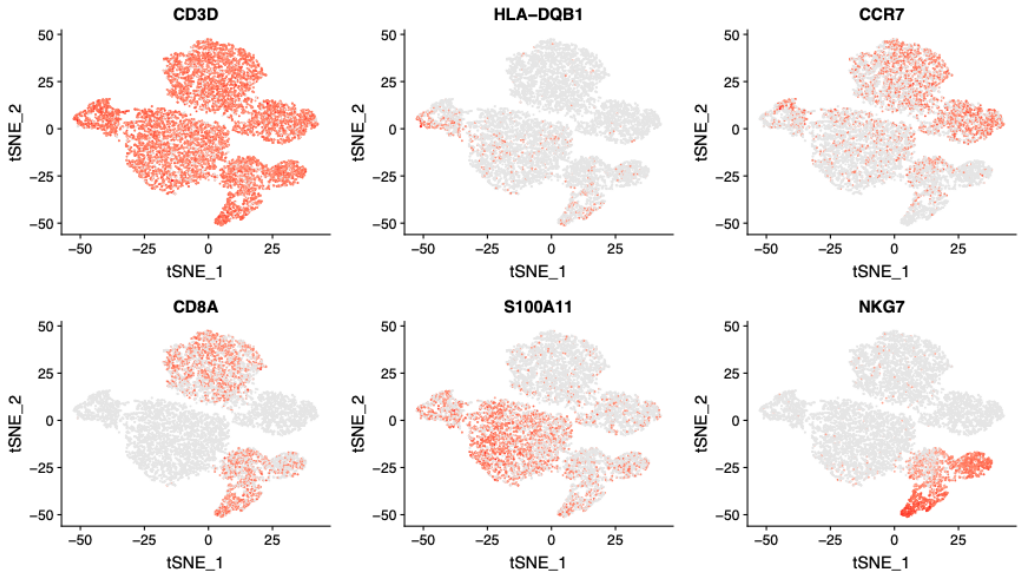
Feiyang Background

Chongqing, China
中国 重庆

Feiyang Research



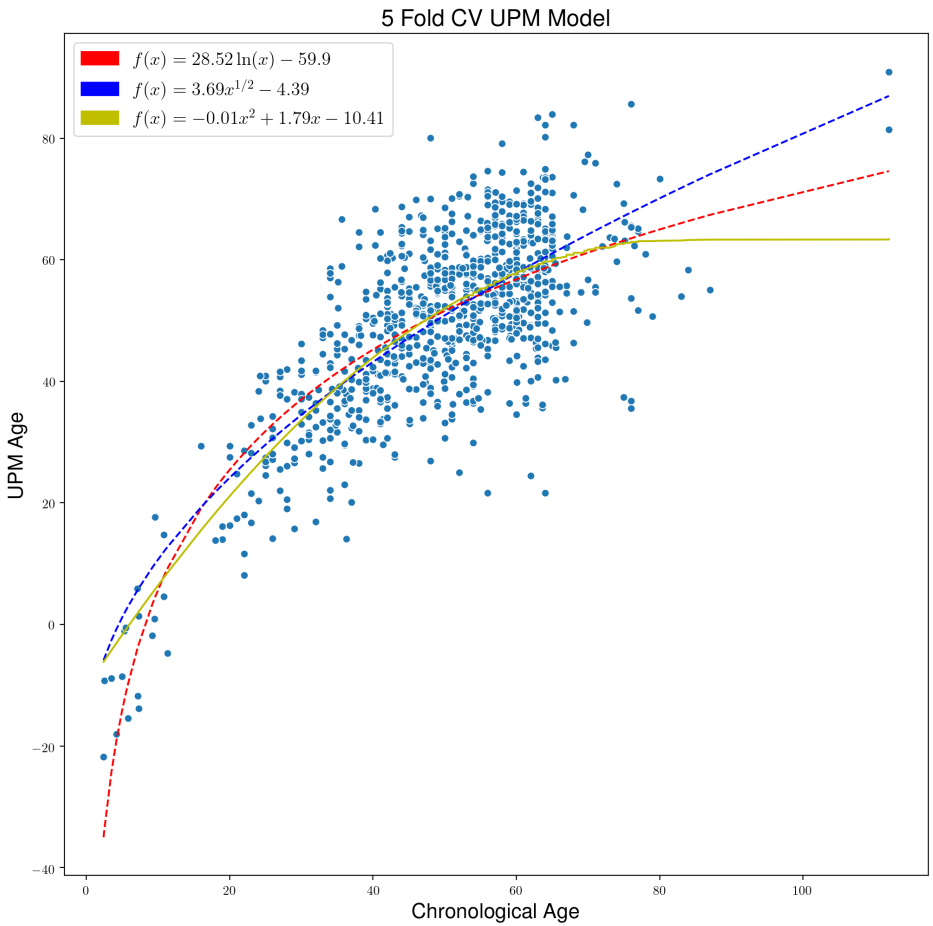


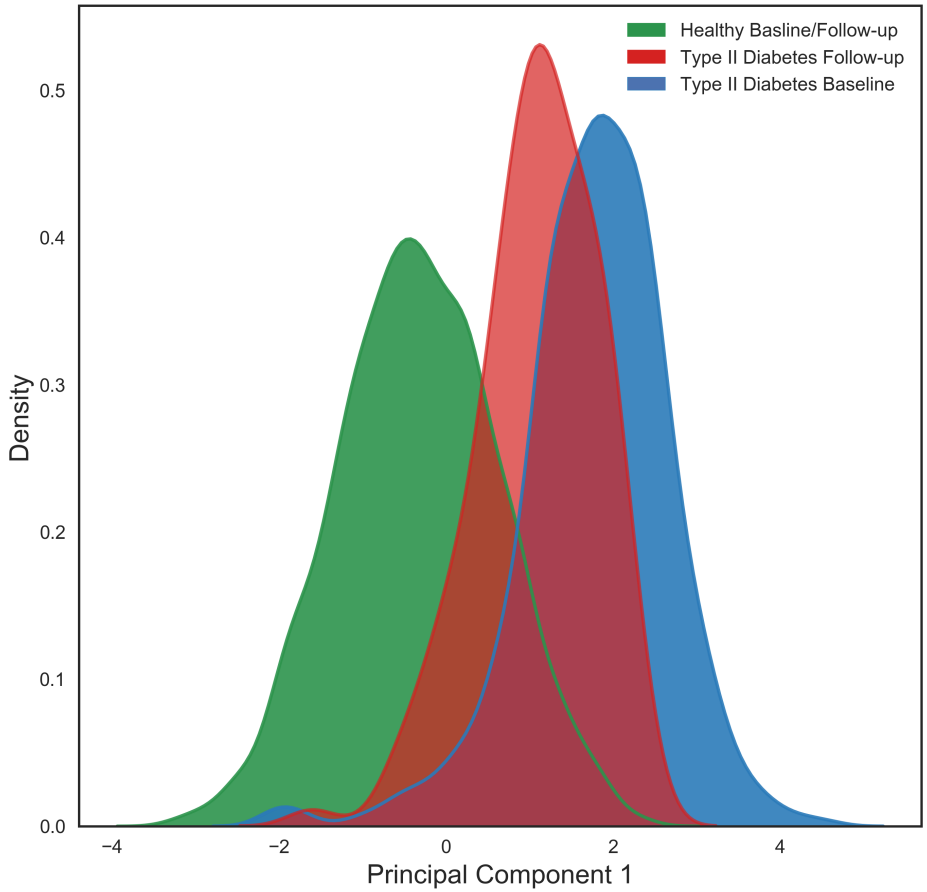


Feiyang Interests

Big fan of basketball (76ers)

Colin Research

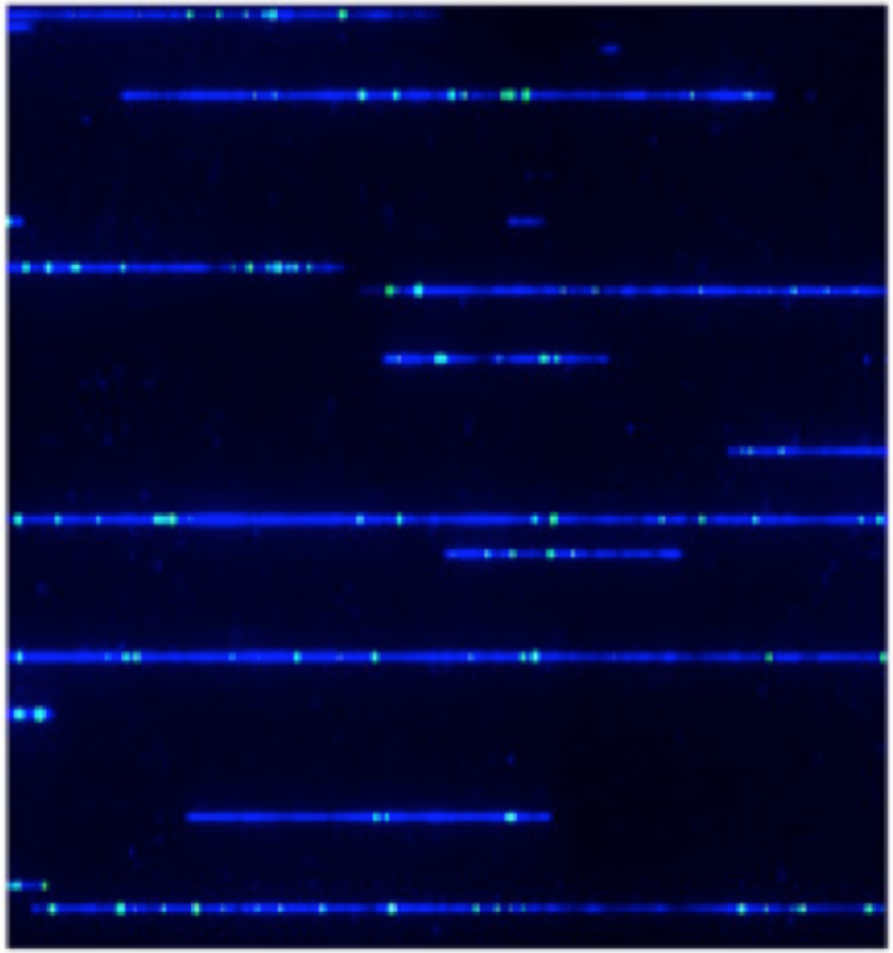




Colin Interests

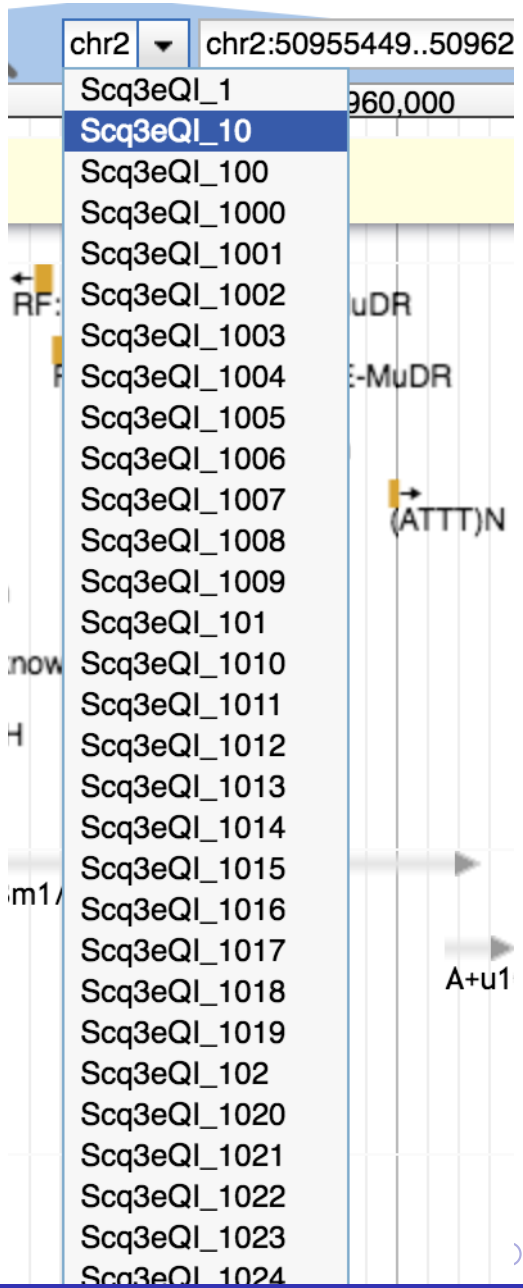
Genome Assembly

- Pull together overlapping sequence
- Consensus sequence between long stretches of highly repetitive DNA sequence is hard to link together
- This issue is complicated by the technology of choice, short read parallel sequencing
- Older methods get around this issue (BAC) as well as emerging technologies (optical mapping / long read sequencing)



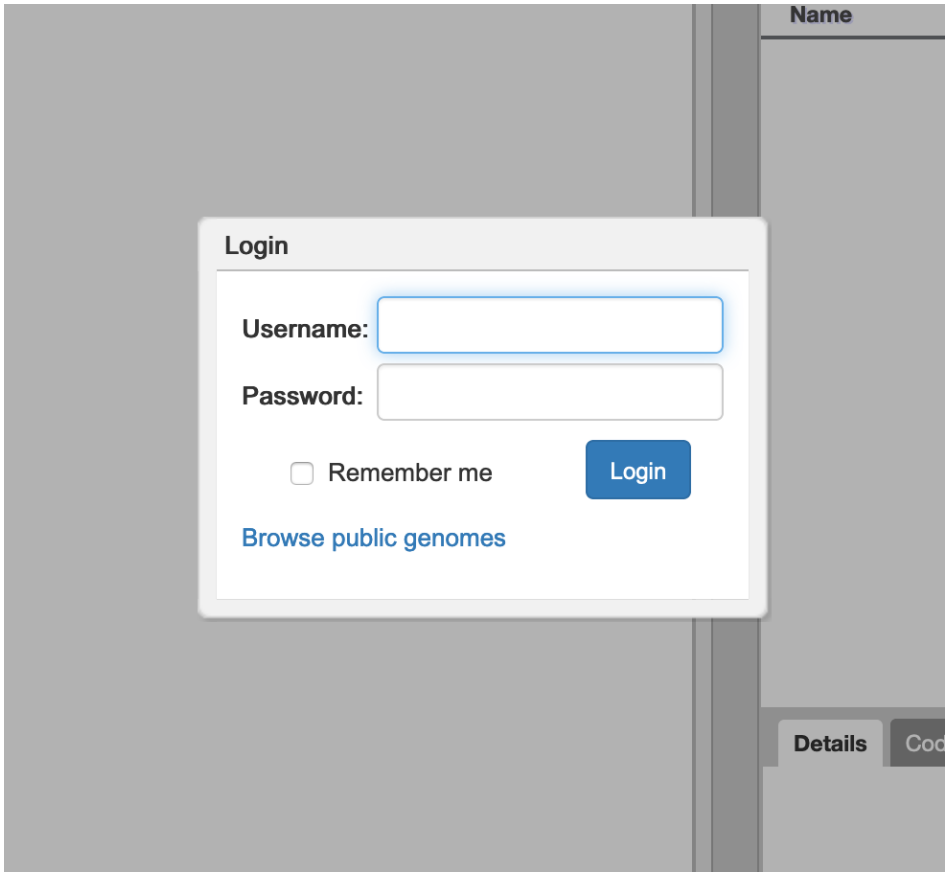
Fucus Genome

- A contig is a set of overlapping DNA sequence that represent a consensus region of DNA
- 110,288 Contigs
- 118579 - 203 base pairs
- Current genome assembly has many more contigs than chromosomes, why?
- 60 (2X) largest contigs randomly assigned to individuals in class roster
- Assigned contigs will vary in quality



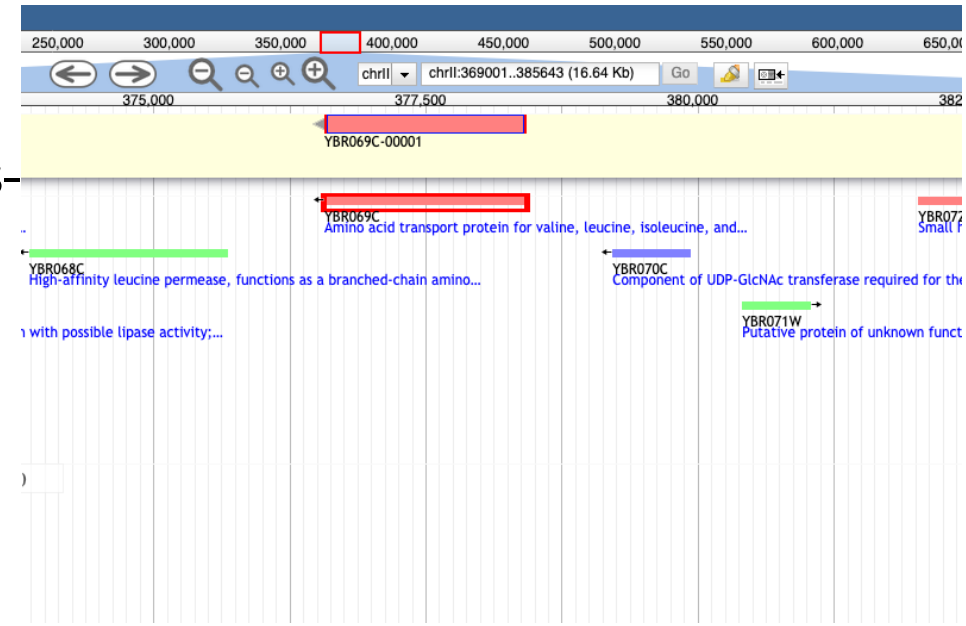
Apollo - Genome Annotation Web Server

- Apollo hosted at 159.89.132.226
- Login = My.UCLA Email (lowercase)
- Password = UCLA ID Number
- Test Server Load
- 2 Test Data Sets
  - Yeast
  - Volvex (Globe Algae)



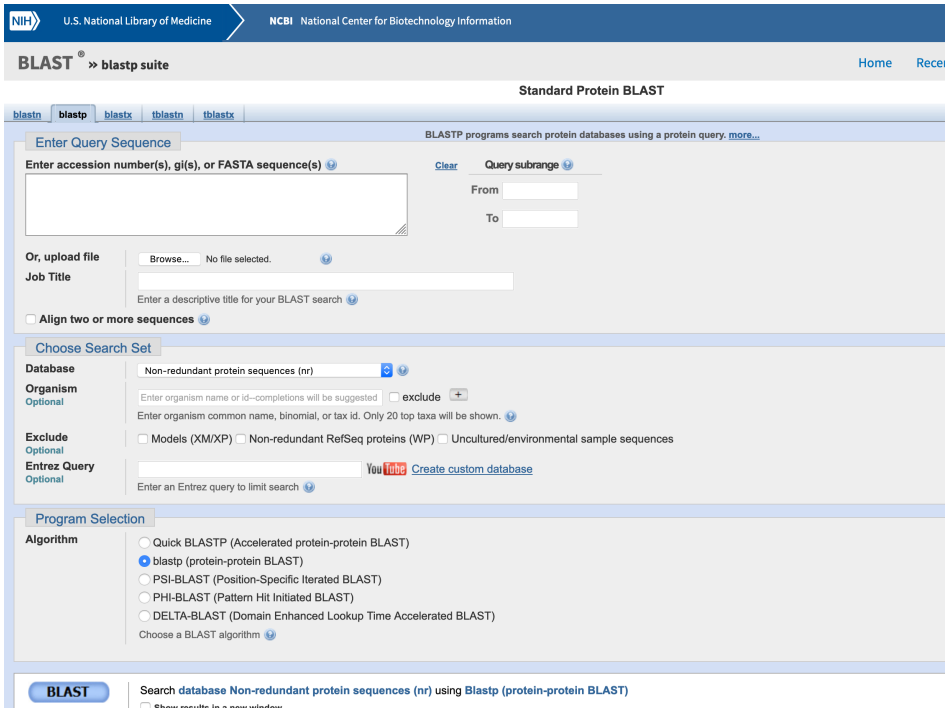
Apollo - Yeast Example

- Apollo user manual, http://genomearchitect.org/users-guide/
- Drag feature to user space to begin annotation
- User features are automatically saved



Apollo - Validating Models

- BLAST
  - blast.ncbi.nlm.nih.gov
- Multiple Sequence Alignment



BLAST

What happens if the input sequence is a truncated version of the target sequence?

Input

TGATTCAGTGTAGTGTGACATATGATAGGAAGGCCATTATC
ATAAATGGGCAAAGGAGAATTCTCTTCTCTGGCTCTATACA
TTACCCCAGAAGCACTCCTGAGATGTGGGAGGATCTGATAC
AGAAAGCAAAAGATGGAGGTTTAGATGTGGTTGAGACCTAT
ATTTTTTTGGAATGTTCATGAGCCTTCTCCTGGCAATTACAA

Target

ATGGAAATCAACTCAGTTTCCAAGTTGTTCTTTTTTATTTGGT
TTAGCGTGGTTCATTGGTTTCTTCCAGCTGATTCAGTGTAGT
GTGACATATGATAGGAAGGCCATTATCATAAATGGGCAAAGG
AGAATTCTCTTCTCTGGCTCTATACATTACCCCAGAAGCACT
CCTGAGATGTGGGAGGATCTGATACAGAAAGCAAAAGATGGA
GGTTTAGATGTGGTTGAGACCTATATTT

BLAST

Output

100% Match

TGATTCAGTGTAGTGTGACATATGATAGGAAGGCCATTATC
ATAAATGGGCAAAGGAGAATTCTCTTCTCTGGCTCTATACA
TTACCCCAGAAGCACTCCTGAGATGTGGGAGGATCTGATAC
AGAAAGCAAAAGATGGAGGTTTAGATGTGGTTGAGACCTAT
ATTTTTTTGGAATGTTCATGAGCCTTCTCCTGGCAATTACAA

Problem?

Multiple Sequence Alignment

- Clustal Omega https://www.ebi.ac.uk/Tools/msa/clustalo/
- Protein or Nucleotide sequence

>sequence_label_1
TGATTCAGTGTAGTGTGACATATGATAGGAAGGCCATTATC
ATAAATGGGCAAAGGAGAATTCTCTTCTCTGGCTCTATACA
TTACCCCAGAAGCACTCCTGAGATGTGGGAGGATCTGATAC
AGAAAGCAAAAGATGGAGGTTTAGATGTGGTTGAGACCTAT
ATTTTTTTGGAATGTTCATGAGCCTTCTCCTGGCAATTACAA
>sequence_label_2
TGATTCAGTGTAGTGTGACATATGATAGGAAGGCCATTATC
ATAAATGGGCAAAGGAGAATTCTCTTCTCTGGCTCTATACA
TTACCCCAGAAGCACTCCTGAGATGTGGGAGGATCTGATAC
AGAAAGCAAAAGATGGAGGTTTAGATGTGGTTGAGACCTAT
ATTTTTTTGGAATGTTCATGAGCCTTCTCCTGGCAATTACAA

Clustal Omega Example

>Quercus_lobata
ATGGAAATCAACTCAGTTTCCAAGTTGTTCTTTTTTATTTGGTTTAGCGT
TGATTCAGTGTAGTGTGACAT...CAACCATTAAGAACTGGCA
>Quercus_suber_beta-galactosidase_3
GACCTTTGAGTGTGGAGTCCAAAGCACAAATGGACCCTGTAGTAGTCTAC
TAGACCAAAT....CCGACCTTTTAACCGGAAATGGGC

Clustal Omega Example

Multiple Sequence Alignment

Clustal Omega is a new multiple sequence alignment program that uses seeded guide trees and HMM profile-profile techniques to generate alignments between **three or more** sequences. For the alignment of two sequences please instead use our pairwise sequence alignment tools.

Important note: This tool can align up to 4000 sequences or a maximum file size of 4 MB.

STEP 1 - Enter your input sequences

Enter or paste a set of

DNA

sequences in any supported format:

```
AGTGCATAGTTTTTGTCAGATGGTTAGTACCTCTGGCCAAAGAAGTGTTTGCAGCCGTAAGGCAATTGGT
TCTTAATTTAAGTTGGGTAATGGACTAATGTAATATAGTGATTGTTAGGTAGATAAGTTTGTGAATTCCC
GGGGTCCTATAATTTGGGGCCTTAGGTGAATTTCGAGCCCCAGTCCCCGACCTTTTAACCGGAAATGGGC
ACTGGGAAAAGAGGTGATTTCTGTTAGTGTGATTTATTTATATTTATGTTTGTATCATTGGGGAAAAATC
AAACCCAGTTTGAGTTAGGTTCATGTTGGGAGCGGATCCTACTATGACCATTGTTTTATGTGCAGTTCCT
CAATTTGAAGTCCATTTCATTTTACATCCAAGTGGCATCGTTTTTTGTC
```

Or, upload a file: Choose File No file chosen

STEP 2 - Set your parameters

OUTPUT FORMAT

ClustalW with character counts

The default settings will fulfill the needs of most users.

More options... *(Click here, if you want to view or change the default settings.)*

STEP 3 - Submit your job

Be notified by email *(Tick this box if you want to be notified by email when the results are available)*

Clustal Omega Example

Alignments | Result Summary | Phylogenetic Tree | Submission Details

Download Alignment File | Send to Simple Phylogeny | Send to MView

```
CLUSTAL O(1.2.4) multiple sequence alignment


Quercus_lobata                          ------------------------------------------------------------    0
Quercus_suber_beta-galactosidase_3      GACCTTTGAGTGTGGAGTCCAAAGCACAAATGGACCCTGTAGTAGTCTAGTTCTTGTTGC    60

Quercus_lobata                          ------------------------------------------------------------    0
Quercus_suber_beta-galactosidase_3      TCTGCACTAGTAGACCAAAACCTCCATTAAAGACCTCTCTCTCTCTCTCTCTCTCTCTCT    120

Quercus_lobata                          ------------------------------------------------------ATGG    4
Quercus_suber_beta-galactosidase_3      CTCTCTCTTTGCATTTCACAAAAGATTGTAACTTTAGTGTGTACACAGAGAGAGCAATGG    180
                                                                                                ****

Quercus_lobata                          AAATCAACTCAGTTTCCAAGTTGTTCTTTTTATTTGGTTTAGCGTGGTTCATTGGTTTCT    64
Quercus_suber_beta-galactosidase_3      AAATCAACTCAGTTTCCAAGTCGTTCTTTTTATTTGGTTTAGTGTGGTTCATTGGTTTCT    240
                                        ********************* ******************** ******************

Quercus_lobata                          TCCAGCTGATTCAGTGTAGTGTGACATATGATAGGAAGGCCATTATCATAAATGGGCAAA    124
Quercus_suber_beta-galactosidase_3      TCCAGCTGATTCAGTGTAGTGTGACATATGATAGGAAGGCCATTATCATAAATGGGCAAA    300
                                        ************************************************************

Quercus_lobata                          GGAGAATTCTCTTCTCTGGCTCTATACATTACCCCAGAAGCACTCCTGAGATGTGGGAGG    184
Quercus_suber_beta-galactosidase_3      GGAGAATTCTCTTCTCTGGCTCTATACATTACCCCAGAAGCACTCCTGAGATGTGGGAGG    360
                                        ************************************************************

Quercus_lobata                          ATCTGATACAGAAAGCAAAAGATGGAGGTTTAGATGTGGTTGAGACCTATATTTTTTGGA    244
Quercus_suber_beta-galactosidase_3      ATCTGATACAGAAAGCAAAAGATGGAGGTTTGGATGTGGTTGAGACCTATATTTTTTGGA    420
                                        ******************************** ****************************
```

Multiple Squence Alignment

```
Solanum_tuberosum(potato)_beta-galactosidase_3      GCAGCTTTCTTGTCAAACAAT-----GACTGGAAATCTGCTGC-TAGAGTAATGTTCAAT      2610
Arabidopsis_thaliana_beta-galactosidase_3           AGCCTTTGATATGGACAGAG------GCATGGAGCGGCT-----GGTAGAGATTCTCATT      1833
Quercus_lobata                                      ----------------------------ATGGAAATCAACTCAGTTTCCAAGTTGTTCTT      32
Quercus_suber_beta-galactosidase_3                  A-ACTTTAGTGTGTACACAGAGAGAGCAATGGAAATCAACTCAGTTTCCAAGTCGTTCTT      208
                                                                                ****                  *  *   *

Solanum_tuberosum(potato)_beta-galactosidase_3      AACATGCACTATAATTTGCCTCCTTGGTCCATTAGCATCCTTCCTGATTGCAGAAATGTA      2670
Arabidopsis_thaliana_beta-galactosidase_3           TGCTTTCCCTTTATATCTGTACATGGATTTGTTCTCTTCAAATCGG-----TAACCTTTT      1888
Quercus_lobata                                      TTTATTTGGTTTAGCGTGGTTCATTGGTTTCTTCCAGCTGATTCAG-----TGTAGTGTG      87
Quercus_suber_beta-galactosidase_3                  TTTATTTGGTTTAGTGTGGTTCATTGGTTTCTTCCAGCTGATTCAG-----TGTAGTGTG      263
                                                        *    * **        * * * *   **          * *          * *

Solanum_tuberosum(potato)_beta-galactosidase_3      GTCTTCAACACGGCCAAAGTACGTAATACACCTGTGTAGGAACTTCTGTACATTCAATAG      2730
Arabidopsis_thaliana_beta-galactosidase_3           TTATTCTGTAAATTGGAAATGCATTACCAG-----GTTCACTGAGTTCGGTGGGCCAATG      1943
Quercus_lobata                                      ACATA----TGATAGGAAGGCCATTATCATAAATGGGCAAAGGAGAATTCTCTTCTCTGG      143
Quercus_suber_beta-galactosidase_3                  ACATA----TGATAGGAAGGCCATTATCATAAATGGGCAAAGGAGAATTCTCTTCTCTGG      319
                                                       *           **   * * *        *                  *    *

Solanum_tuberosum(potato)_beta-galactosidase_3      TACTCATGAGCTGCTTAGCATCTTCCGAAAGATATAGTTCATTTATGCTACTCGCTAACA      2790
Arabidopsis_thaliana_beta-galactosidase_3           CACCATAGACCAGTTCAGGATCTGGCGTTTGGTGTTGCACGT----------TTCATACA      1993
Quercus_lobata                                      CTCTATACATTACCCCAGAAGCACTCCTGAGATGTGGGAGGA---------TCTGATACA      194
Quercus_suber_beta-galactosidase_3                  CTCTATACATTACCCCAGAAGCACTCCTGAGATGTGGGAGGA---------TCTGATACA      370
                                                      *     *       ** * *   *    * * * *                     ***

Solanum_tuberosum(potato)_beta-galactosidase_3      ACTCTGAA-AGCA--AGGATTTAAAAGCTTTTAAATACCATATGTAGGTTGGA-------      2840
Arabidopsis_thaliana_beta-galactosidase_3           AAAAG------GAGGATCTTTTGTTAACTACTACATGGTATGTACACTTTAGCATTTTGG      2047
Quercus_lobata                                      GAAAGCAAAAGATGGAGGTTTAGATGTGG---TTGAGACCTATATTTTTTGGAATGTTCA      251
Quercus_suber_beta-galactosidase_3                  GAAAGCAAAAGATGGAGGTTTGGATGTGG---TTGAGACCTATATTTTTTGGAATGTTCA      427
                                                                  *   **                  * *     ** *

Solanum_tuberosum(potato)_beta-galactosidase_3      -----GTTCAAA------------CATCAAAGATGGAAATGCTACCGACTAACTCAGAGA      2883
Arabidopsis_thaliana_beta-galactosidase_3           TTAATCATCTAGCTCCTTACACACTGTAAGCTTTGGTTATCATCGAGCATCTTTAAAATA      2107
Quercus_lobata                                      TGAGCCTTCTCCTGGCAATTACAATTTTGAAGGGAGATAT--------------------      291
Quercus_suber_beta-galactosidase_3                  TGAGCCTTCTCCTGGCAATTACAATTTTGAAGGGAGATAT--------------------      467
                                                           **                 *       *  **

Solanum_tuberosum(potato)_beta-galactosidase_3      TGCTATCTTGGGAGACTTACAGTGAAGATATGTCTGCATTAGATGACAGCTCATCAATTA      2943
Arabidopsis_thaliana_beta-galactosidase_3           TAAGATCCATTGTCTTTTCTTGAGAACAGTATCATGGAGGAACTAACT--TTGGAAGAAC      2165
Quercus_lobata                                      ---GATTTGGTGAGATTCATTAAGACCATACAGAAAGCTGGGCTTTATGCTCATATTCGC      348
Quercus_suber_beta-galactosidase_3                  ---GATTTGGTGAGATTCATTAAGACCATACAGAAAGCTGGGCTTTATGCTCATATCCGC      524
                                                       **     *    *      **  *                *      *

Solanum_tuberosum(potato)_beta-galactosidase_3      GGTCTTTTGGTCTCTTGGAGCAAATAAATGTAACCAGA---GATACAAGTGATTACTTGT      3000
Arabidopsis_thaliana_beta-galactosidase_3           AG-------------CCGGAGGTCCATTTGTCACCACCAGCTATGATTATGATGCTCCGA      2212
Quercus_lobata                                      AT-------------TGGACCTTATGTTTGTGCAGAGTGGAATTTTGGAGGATTTCCTGT      395
Quercus_suber_beta-galactosidase_3                  AT-------------TGGACCTTATGTTTGTGCAGAGTGGAATTTTGGAGGATTTCCTGT      571
                                                                  *          ***    *       *      ***     *

Solanum_tuberosum(potato)_beta-galactosidase_3      GGTACATAACCAGGTGAGTTTGGTTTAGTATATTAAACCTGTTTCCTCGATTTGCTTCTC      3060
Arabidopsis_thaliana_beta-galactosidase_3           TCGAC----GAATATGGT-AAGGAC----------AAGCAGCGAACAAA---ACTTTCTA      2254
Quercus_lobata                                      TTGGCTCAAGTATGTCCC-AGGCATCAGT--TTCAGAACAGACAATGAG---CCTTTCAA      449
Quercus_suber_beta-galactosidase_3                  TTGGCTCAAGTATGTCCC-AGGAATCAGT--TTCAGAACAGACAATGAG---CCTTTCAA      625
                                                       *      *  *      *             * * *             ***

Solanum_tuberosum(potato)_beta-galactosidase_3      ACGAGGAATTGGATTGTTTGCCAGTGCAAATGTTTTATTGCTTA---TATCTTTTT-ACT      3116
Arabidopsis_thaliana_beta-galactosidase_3           CACATTCATTGAAATATAAGTTGATGCAATGCTTTATGTACTAAGTGAGTCGTTGTTAAA      2314
Quercus_lobata                                      GAGGGCAATGCAAAGGTTCACTGAAA---AGATTGTTGGACTGA---------TGAAGAG      497
Quercus_suber_beta-galactosidase_3                  GAGGGCAATGCAAAGGTTCACTGAAA---AGATTGTTGGACTGA---------TGAAGAG      673
```

Phylogenetic Tree

Results for job clustalo-I20180412-222345-0717-10122482-pg

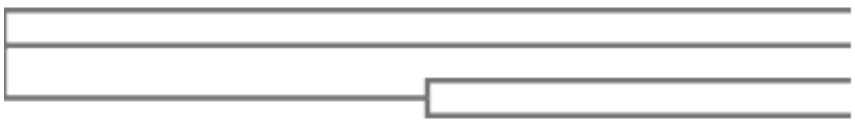
Alignments | Result Summary | Phylogenetic Tree | Submission Details

Phylogenetic Tree

This is a Neighbour-joining tree without distance corrections.

Download Phylogenetic Tree Data

Branch length: Cladogram Real



Solanum_tuberosum_potato__beta-galactosidase_3 0.28303
Arabidopsis_thaliana_beta-galactosidase_3 0.24119
Quercus_lobata 0.00441
Quercus_suber_beta-galactosidase_3 0.0085

Questions?