

Valley Oak Data Overview

Colin Farrell / Zijun Zhang

MCDB 187

4/13/2018

BLAST

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

Input

TGATTCAGTGTAGTGTGACATATGATAGGAAGGCCATTATC
ATAAATGGGCAAAGGAGAATTCTCTTCTCTGGCTCTATACA
TTACCCCAGAAGCACTCCTGAGATGTGGGAGGATCTGATAC
AGAAAGCAAAAGATGGAGGTTTAGATGTGGTTGAGACCTAT
ATTTTTTTGGAATGTTTCATGAGCCTTCTCCTGGCAATTACAA

Target

ATGGAAATCAACTCAGTTTCCAAGTTGTTCTTTTTTATTTGGT
TTAGCGTGGTTCATTGGTTTCTTCCAGCTGATTCAGTGTAGT
GTGACATATGATAGGAAGGCCATTATCATAAATGGGCAAAGG
AGAATTCTCTTCTCTGGCTCTATACATTACCCCAGAAGCACT
CCTGAGATGTGGGAGGATCTGATACAGAAAGCAAAAGATGGA
GGTTTAGATGTGGTTGAGACCTATATT

BLAST

Output

100% Match

```
TGATTCAGTGTAGTGTGACATATGATAGGAAGGCCATTATC  
ATAAATGGGCAAAGGAGAATTCTCTTCTCTGGCTCTATACA  
TTACCCCAGAAGCACTCCTGAGATGTGGGAGGATCTGATAC  
AGAAAGCAAAAGATGGAGGTTTAGATGTGGTTGAGACCTAT  
ATTTTTTTGGAATGTTTCATGAGCCTTCTCCTGGCAATTACAA
```

Problem?

Multiple Sequence Alignment

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

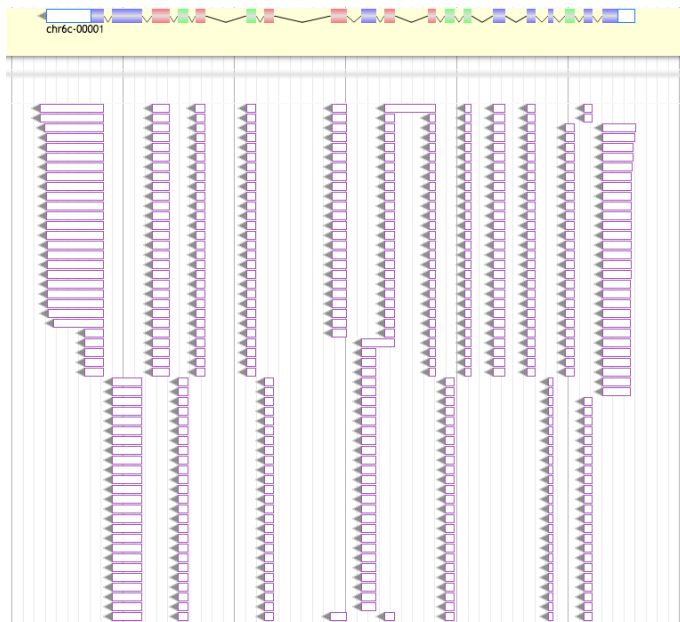
>sequence_label_1

```
TGATTCAGTGTAGTGTGACATATGATAGGAAGGCCATTATC
ATAAATGGGCAAAGGAGAATTCTCTTCTCTGGCTCTATACA
TTACCCCAGAAGCACTCCTGAGATGTGGGAGGATCTGATAC
AGAAAGCAAAAGATGGAGGTTTAGATGTGGTTGAGACCTAT
ATTTTTTTGGAATGTTTCATGAGCCTTCTCCTGGCAATTACAA
```

>sequence_label_2

```
TGATTCAGTGTAGTGTGACATATGATAGGAAGGCCATTATC
ATAAATGGGCAAAGGAGAATTCTCTTCTCTGGCTCTATACA
TTACCCCAGAAGCACTCCTGAGATGTGGGAGGATCTGATAC
AGAAAGCAAAAGATGGAGGTTTAGATGTGGTTGAGACCTAT
ATTTTTTTGGAATGTTTCATGAGCCTTCTCCTGGCAATTACAA
```

Clustal Omega Example



Clustal Omega Example

>Quercus_lobata

ATGGAAATCAACTCAGTTTCCAAGTTGTTCTTTTTATTTGGTTTAGCGT
TGATTCAGTGTAGTGTGACAT...CAACCATTAAGAACTGGCA

>Quercus_suber_beta-galactosidase_3

GACCTTTGAGTGTGGAGTCCAAAGCACAAATGGACCCTGTAGTAGTCTAG
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:

```
AGTGCATAGTGTGTCAGATGGTTAGTACCTCTGGCCAAAGAAGTGTTTGCAGCCGTAAGGCAATTGGT
TCTTAATTTAAGTTGGGTAATGGACTAATGTAATATAGTGATTGTTAGGTAGATAAGTTTGTGAATCCC
GGGGTCCTATAATTTGGGGCCCTTAGGTGAATTCGAGCCCCAGTCCCGACCTTTAACCAGGAAATGGGC
ACTGGGAAAAGAGGTGATTCTGTTAGTGTGATTATTATATTTATGTTTGTATCATTGGGGAAAAATC
AAACCCAGTTTGAGTTAGGTTTCATGTTGGGAGCGGATCCTACTATGACCATTGTTTATGTGCAGTTCCT
CAATTTGAAGTCCATTTTACATCCAAGTGGCATCGTTTTTTGTC
```

Or, upload a 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.

(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	GACCTTTGAGTGTGGAGTCCAAAGCACAAATGGACCCTGTAGTAGCTAGTCTTGTTCG	60
Quercus_lobata	-----	0
Quercus_suber_beta-galactosidase_3	TCTGCACTAGTAGACCAAAACCTCCATTAAAGACCTCTCTCTCTCTCTCTCTCTCTCT	120
Quercus_lobata	-----ATGG	4
Quercus_suber_beta-galactosidase_3	CTCTCTCTTTGCATTTCCAAAAAGATTGTAACCTTTAGTGTGTACACAGAGAGCAATGG	180

Quercus_lobata	AAATCAACTCAGTTTCCAAGTTGTTCTTTTATTTGGTTTAGCGTGGTTCATTGGTTTCT	64
Quercus_suber_beta-galactosidase_3	AAATCAACTCAGTTTCCAAGTCGTTCTTTTATTTGGTTTAGTGTGGTTCATTGGTTTCT	240

Quercus_lobata	TCCAGCTGATTCACTGTAGTGTGACATATGATAGGAAGGCCATTATCATAAATGGGCAAA	124
Quercus_suber_beta-galactosidase_3	TCCAGCTGATTCACTGTAGTGTGACATATGATAGGAAGGCCATTATCATAAATGGGCAAA	300

Quercus_lobata	GGAGAATTCTCTTCTCTGGCTCTATACATTACCCAGAAAGCACTCCTGAGATGTGGGAGG	184
Quercus_suber_beta-galactosidase_3	GGAGAATTCTCTTCTCTGGCTCTATACATTACCCAGAAAGCACTCCTGAGATGTGGGAGG	360

Quercus_lobata	ATCTGATACAGAAAGCAAAAGATGGAGGTTTAGATGTGGTTGAGACCTATATTTTTTGA	244
Quercus_suber_beta-galactosidase_3	ATCTGATACAGAAAGCAAAAGATGGAGGTTTGGATGTGGTTGAGACCTATATTTTTTGA	420



Multiple Sequence Alignment

Solanum tuberosum(potato)_beta-galactosidase_3	GCAGCTTTCTGTGCAACAAT-----GACTGGAAATCTGCTGC-TAGAGTAATGTCAAT	2610
Arabidopsis_thaliana_beta-galactosidase_3	AGCCCTTGGATATGGACAGAG-----GCATGGAGCGGGT-----GGTAGAGATCTCATTT	1833
Quercus lobata	-----ATGGAAATCAACTCAGTTTCCAAGTGTGTTCTT	32
Quercus_suber_beta-galactosidase_3	A-ACCTTAGTGTGTACACAGAGAGCAATGGAAATCAACTCAGTTTCCAAGTGTGTTCTT	208

Solanum tuberosum(potato)_beta-galactosidase_3	AACATGCACATATAATTGCGCTCCTTGGCTCATTAGCATCTCTTCGTATGGCAAAATGTA	2670
Arabidopsis_thaliana_beta-galactosidase_3	TGCTTTCCCTTTATATCTGTACATGGATTTGTCTCTCTCAATCGG-----TAACCTTTT	1888
Quercus lobata	TTTATTGGTTTAGCGTGGTTCATTGGTTCTTCCAGCTCATTCAG-----TGATGTGTG	87
Quercus_suber_beta-galactosidase_3	TTTATTGGTTTAGTGTGGTTCATTGGTTCTTCCAGCTGATTTCAG-----TGATGTGTG	263
	* ** * * * *	
Solanum tuberosum(potato)_beta-galactosidase_3	GTCTTCAACACGGCCAAAGTACGTAATACACCTGTGTAGGAACCTCTGTACATTCAATAG	2730
Arabidopsis_thaliana_beta-galactosidase_3	TTATCTCTGAAATTTGGAATGCATTACCG-----GTTCACTGAGTTCGGTGGGCCAATG	1943
Quercus lobata	ACATA-----TGATAGGAAGGCCATTATCATAAATGGGCAAGGAGAATCTCTCTCTCGG	143
Quercus_suber_beta-galactosidase_3	ACATA-----TGATAGGAAGGCCATTATCATAAATGGGCAAGGAGAATCTCTCTCTCGG	319
	* ** * * * *	
Solanum tuberosum(potato)_beta-galactosidase_3	TACTCATGAGCTGCTTAGCATCTTCCGAAGATATAGTTCATTTATGCTCATCGCTAACCA	2790
Arabidopsis_thaliana_beta-galactosidase_3	CACCATGACACAGTTCAGGATCTGGCGTTGCTGTGCGAGCT-----TTCATACA	1993
Quercus lobata	CTCTATACATTAGCCCGAAGCACTCTCGAGATGTGGGAGGA-----TCTGATACA	194
Quercus_suber_beta-galactosidase_3	CTCTATACATTAGCCCGAAGCACTCTCGAGATGTGGGAGGA-----TCTGATACA	370
	* * * * * *	
Solanum tuberosum(potato)_beta-galactosidase_3	ACTCTGAA-AGCA--AGGATTTAAAGCTTTTAAATACCATATGTAGGTTGGA-----	2840
Arabidopsis_thaliana_beta-galactosidase_3	AAAAG-----GAGGATCTTTGTTAACTACTACATGATGGTATGATGATTTTGG	2047
Quercus lobata	GAAAGCAAAAGATGGAGGTTTAGATGTGG-----TTGAGACCTATATTTTTGGAATGTCA	251
Quercus_suber_beta-galactosidase_3	GAAAGCAAAAGATGGAGGTTTAGATGTGG-----TTGAGACCTATATTTTTGGAATGTCA	427
	* ** * * * *	
Solanum tuberosum(potato)_beta-galactosidase_3	-----GTTCAAA-----CATCAAAAGATGGAATGCTACCGACTAATCAGAGA	2883
Arabidopsis_thaliana_beta-galactosidase_3	TTAATCATCTAGCTCCTTACACACTGTAAAGCTTTGGTTATCATCGAGCATCTTAAATA	2107
Quercus lobata	TGAGCCTTCTCTCGGCAATTACAATTTTGAAGGGAGATAT-----	291
Quercus_suber_beta-galactosidase_3	TGAGCCTTCTCTCGGCAATTACAATTTTGAAGGGAGATAT-----	467
	** * * *	
Solanum tuberosum(potato)_beta-galactosidase_3	TGCTATCTTGGGAGACTTACAGTGAAGATATGCTGCAATTTAGATGACAGCTCATCAATTA	2943
Arabidopsis_thaliana_beta-galactosidase_3	TAAGATGCCATTGTCTTTCTTGAGAAGCATATCATGGAGGAACATCT--TTGGAAGAAC	2165
Quercus lobata	---GATTTGGTGAGATTCATTAGAACACATACGAAGCTGGGGCTTATGCTCATATTCCG	348
Quercus_suber_beta-galactosidase_3	---GATTTGGTGAGATTCATTAGAACACATACGAAGCTGGGGCTTATGCTCATATTCCG	524
	* * * * *	
Solanum tuberosum(potato)_beta-galactosidase_3	GGTCTTTTGGTCTCTTGGAGCAATAAATGTAACCGA---GATACAAGTGATTACTGTG	3000
Arabidopsis_thaliana_beta-galactosidase_3	AG-----CCGGAGGTGCCATTGTGACACACAGCTATGATTTATGATGCTCCGA	2212
Quercus lobata	AT-----TGGACCTTATGTTTGTGCGAGAGTGAATTTGGAGGATTTCTGCTGT	395
Quercus_suber_beta-galactosidase_3	AT-----TGGACCTTATGTTTGTGCGAGAGTGAATTTGGAGGATTTCTGCTGT	571
	* * * * *	
Solanum tuberosum(potato)_beta-galactosidase_3	GGTACATAACCCAGGTGAGTTTGGTTTATATTAACCTGTCTTCTCGAATGCTCTCTC	3060
Arabidopsis_thaliana_beta-galactosidase_3	TCGAC---GAATATGTT-AGAGCA-----AAGACGACAGACAA---ACCTTTCTA	2254
Quercus lobata	TGGGCTCAGTATGTCCG-AGGCATCAGT---TTCAGAACAGACATAG---CCTTTCAA	449
Quercus_suber_beta-galactosidase_3	TGGGCTCAGTATGTGCC-AGGAATCAGT---TTCAGAACAGACATAG---CCTTTCAA	625
	* * * * *	
Solanum tuberosum(potato)_beta-galactosidase_3	ACGAGGAATGGATTTGTTGGCAGTGCAGAAATGTTTATGTGCTA---TATCTTTT-ATT	3116
Arabidopsis_thaliana_beta-galactosidase_3	CACATTCATTGAAATATAAGTTGATGCAAGCTTTTATGTACTAAGTGAGTGGTGTGTA	2314
Quercus lobata	GAGGCGCAATGCAGAAAGTTCTAGTGA---AGATTGTGGAGCTGA-----TGAAGAG	497
Quercus_suber_beta-galactosidase_3	GAGGCGCAATGCAGAAAGTTCTAGTGA---AGATTGTGGAGCTGA-----TGAAGAG	673

Phylogenetic Tree

Results for job clustalo-l20180412-222345-0717-10122482-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


Structural Modeling

- <https://swissmodel.expasy.org/>
- Input peptide sequence
- <https://swissmodel.expasy.org/interactive/fGErQZ/models/>

MEINSVSKLFFLFGLA...CASNPTTNWGG

Structural Modeling

Start a New Modelling Project

Target  Target

Sequence:

(Format must be FASTA, Clustal, plain string, or a valid UniProtKB AC)

Target

Target

Project Title:

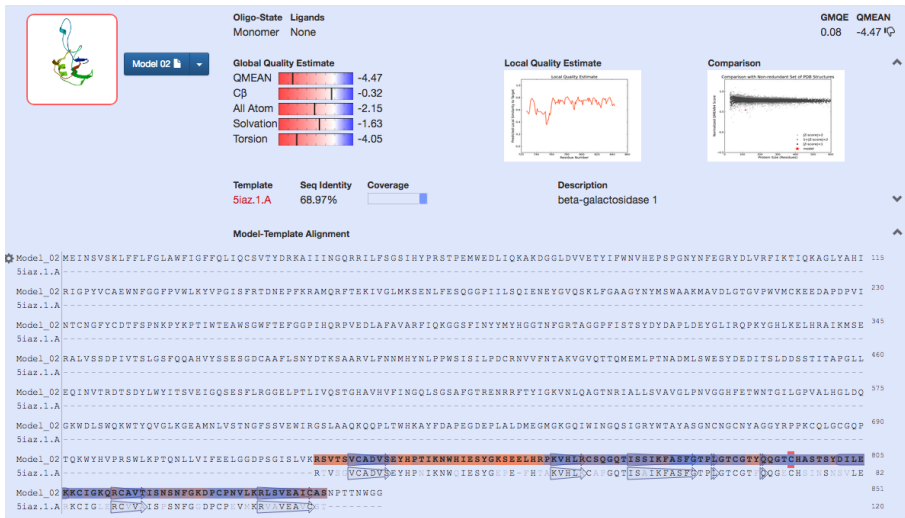
Email:

By using the SWISS-MODEL server, you agree to comply with the following [terms of use](#) and to cite the corresponding [articles](#).

Supported Inputs

Sequence(s)	▼
Target-Template Alignment	▼
User Template	▼
DeepView Project	▼

Structural Modeling



Additional Resources

- PDB
- AMIGO - Gene Ontology
- Pathway Analysis