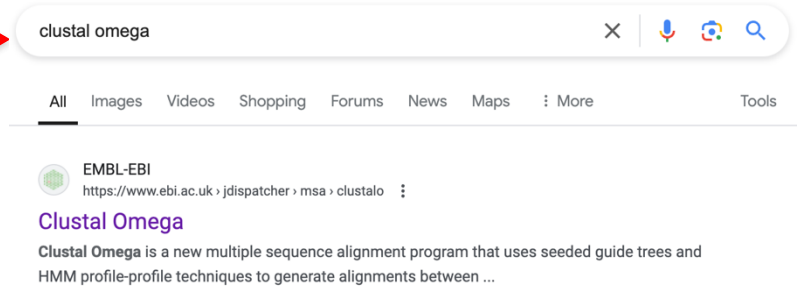
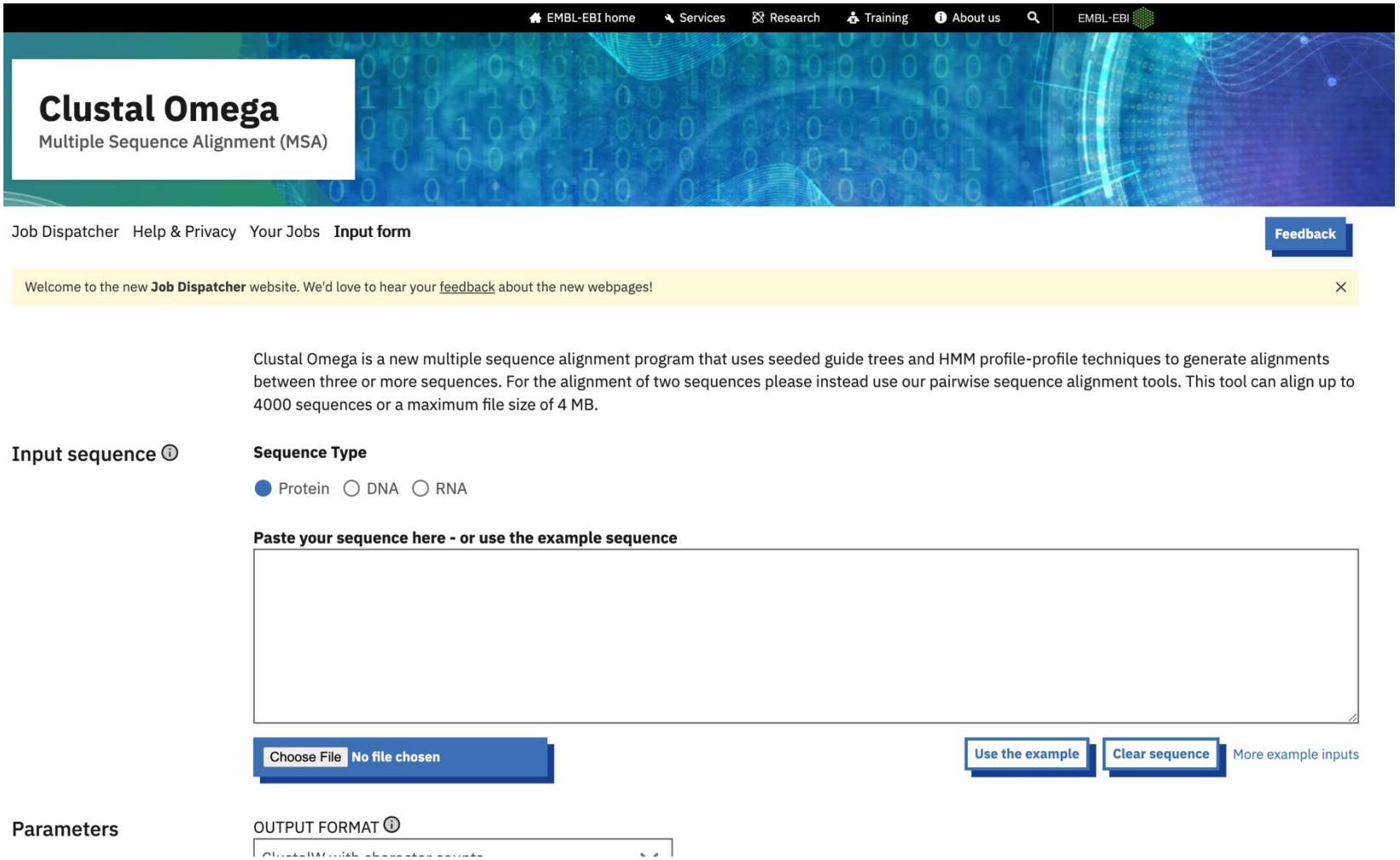
Clustal Omega: Align two or more DNA or Protein Sequences

Step 1. Internet search “clustal omega”

Step 2. Or click this link:
https://www.ebi.ac.uk/jdispatcher/msa/clustalo

Step 3. You’ll get this page:



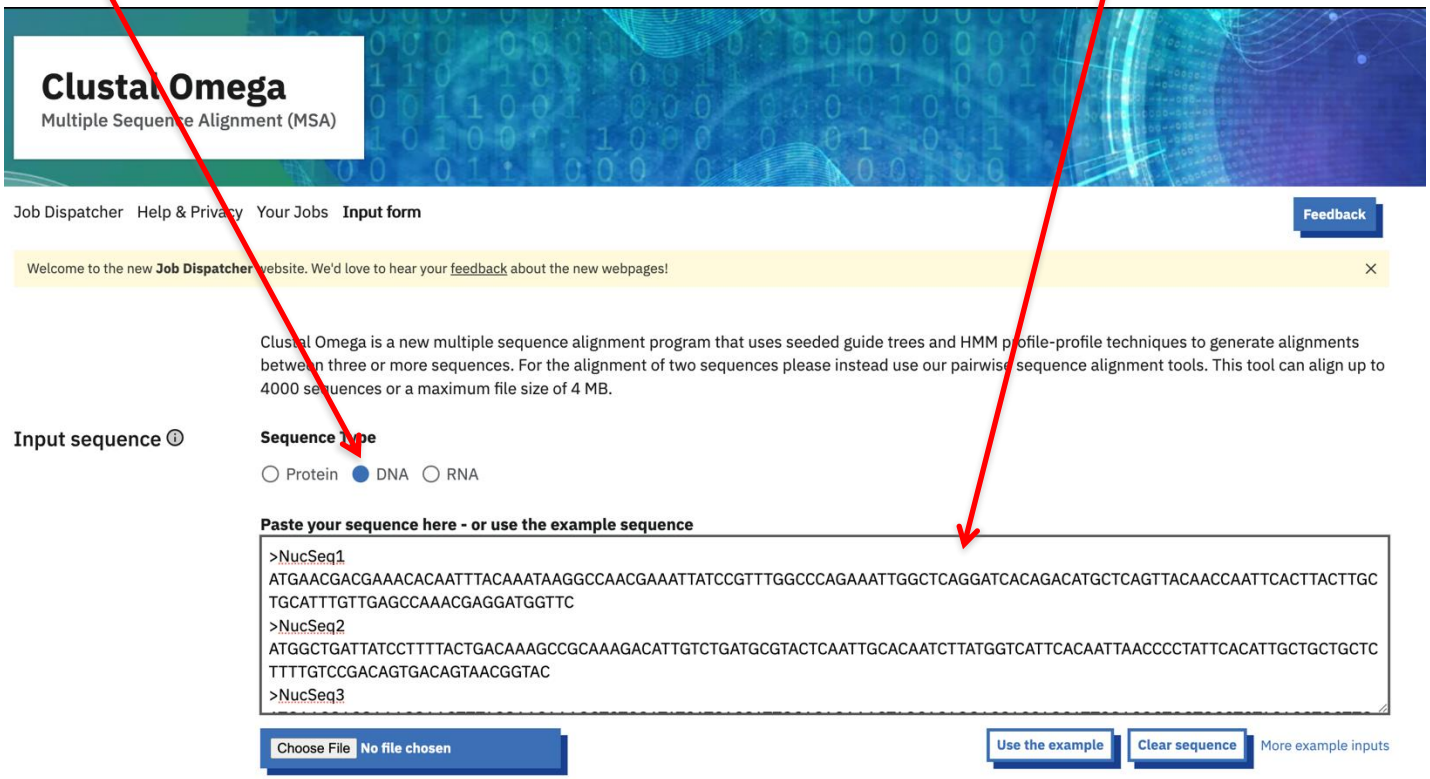


Step 4. Find some sequence data to align.

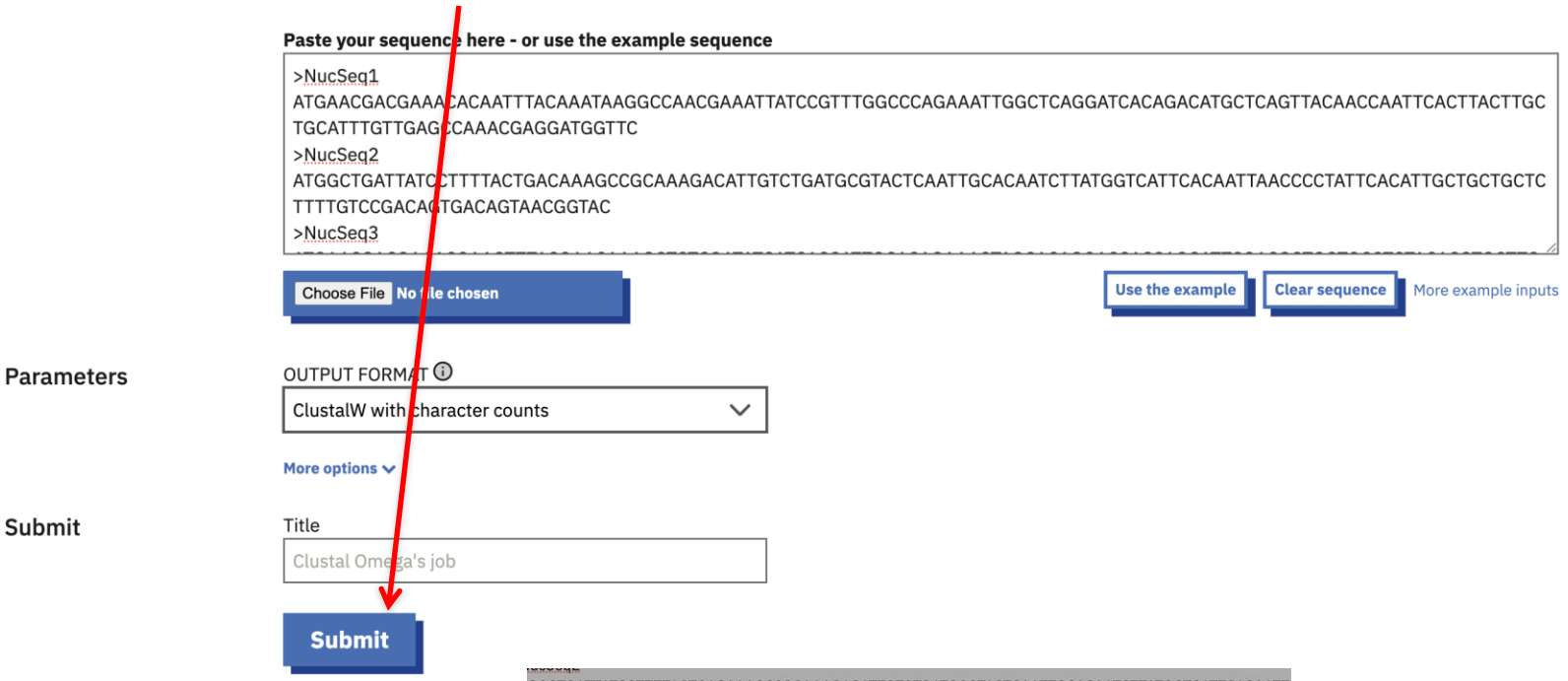
Here are 3 DNA sequences in FASTA format from http://kelleybioinfo.org/algorithms/data/DAli2.txt (The "Sequence files" link on the Basics page has more information on FASTA.)

```
>NucSeq1
ATGAACGACGAAACACAATTTACAAATAAGGCCAACGAAATTATCCGTTTGGCCCAGAAATTGGCTCAGGATCACAGACATGCTCAGTTACAACCAATT
CACTTACTTGCTGCATTTGTTGAGCCAAACGAGGATGGTTC
>NucSeq2
ATGGCTGATTATCCTTTTACTGACAAAGCCGCAAAGACATTGTCTGATGCGTACTCAATTGCACAATCTTATGGTCATTCACAATTAACCCCTATTCAC
ATTGCTGCTGCTCTTTTGTCCGACAGTGACAGTAACGGTAC
>NucSeq3
ATGAACGACGAAACGAAGTTTACGAACAAAGCTCTCGATATCATCACCATTGCACAGAAACTAGCACAGGACCACCAGCATTCGACGCTGGTGCCTCTA
CACGTGCTTGCAGCGTTCGTAGAGACACCTGCTGATGGTAG
```

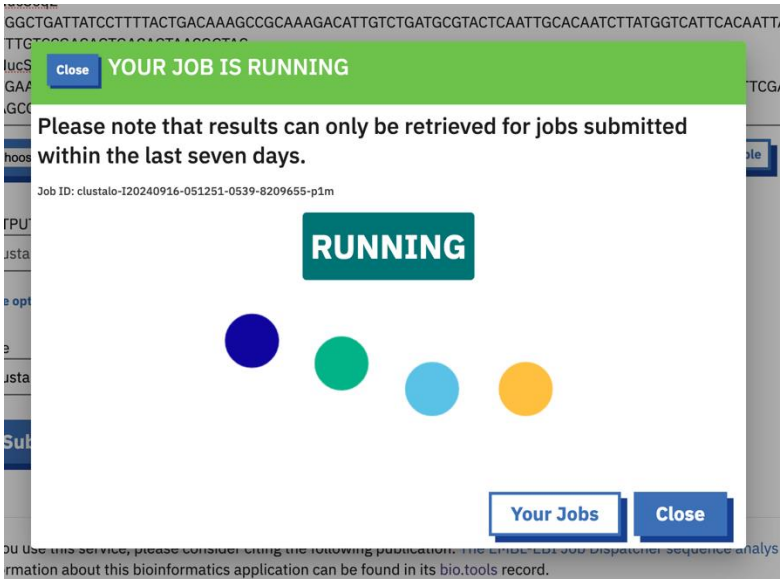
Step 5. Select DNA because these are DNA sequences, then Copy and Paste the data into the input window.



Step 6. Scroll Down to the bottom of the page and click the Submit button.



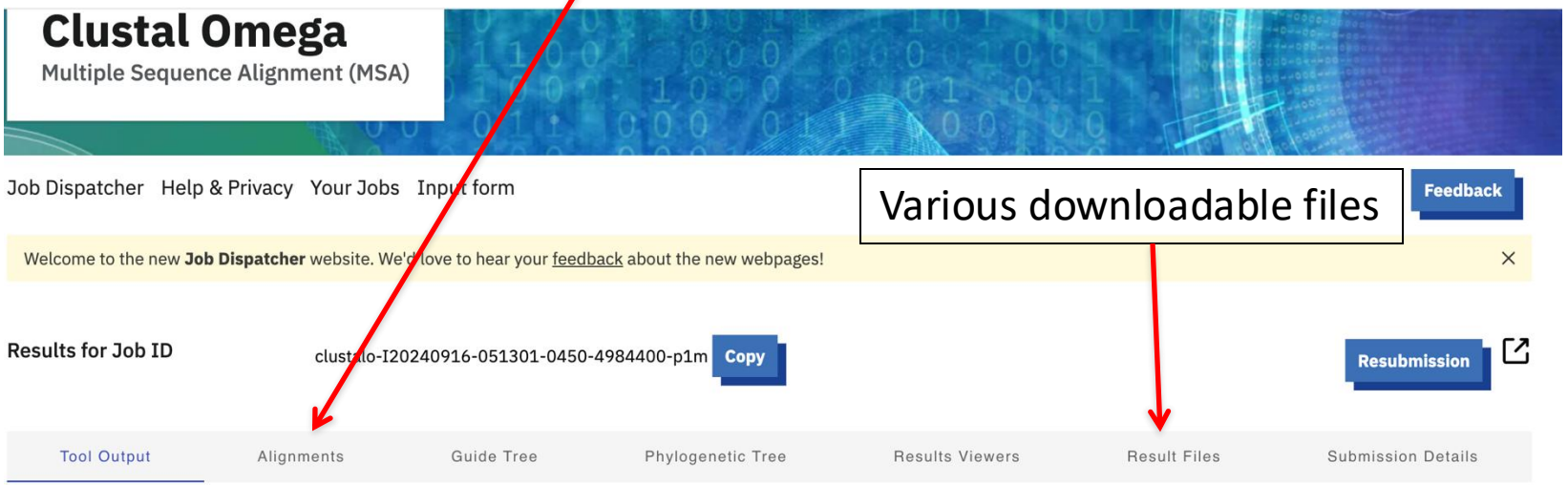
Step 7. Then wait….



Step 8. Interpreting the output.

After waiting a bit, the output will appear in the browser like this:

View alignments with different color labels



Various downloadable files

The sequence alignment:

Tool output

Download

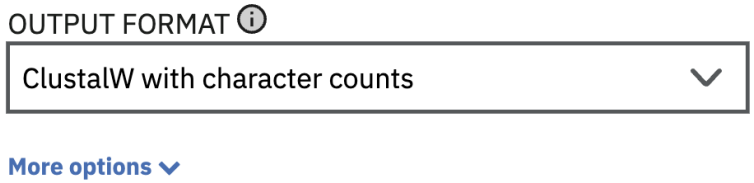
```
CLUSTAL O(1.2.4) multiple sequence alignment


NucSeq2      ---ATGGCTGATTATCCTTTTACTGACAAAGCCGCAAAGACATTGTCTGATGCGTACTCA      57
NucSeq1      ATGAACGACGAAACACAATTTACAAATAAGGCCAACGAAATTATCCGTTTGGCCCAGAAA      60
NucSeq3      ATGAACGACGAAACGAAGTTTACGAACAAAGCTCTCGATATCATCACCATTGCACAGAAA      60
                *  *  **       *****  * ** **     * *   *       **  *   *

NucSeq2      ATTGCACAATCTTATGGTCATTCACAATTAACCCCTATTCACATTGCTGCTGCTCTTTTG      117
NucSeq1      TTGGCTCAGGATCACAGACATGCTCAGTTACAACCAATTCACTTACTTGCTGCATTTGTT      120
NucSeq3      CTAGCACAGGACCACCAGCATTCGACGCTGGTGCCTCTACACGTGCTTGCAGCGTTCGTA      120
              * ** **     *    *** *     *    **  * *** *   *** **  *  *


NucSeq2      TCCGACAGTGACAGTAACGGTAC    140
NucSeq1      GAGCCAAACGAGGATG---GTTC    140
NucSeq3      GAGACACCTGCTGATG---GTAG    140
                      *    *    **
```

The Clustal format in “interleaved”: these **alignments show 60 alignment positions for all sequences**, then go to the next 60 until there is no more alignment.

A asterisk “*” indicates all the sequences have the same nucleotide.
Fully conserved.

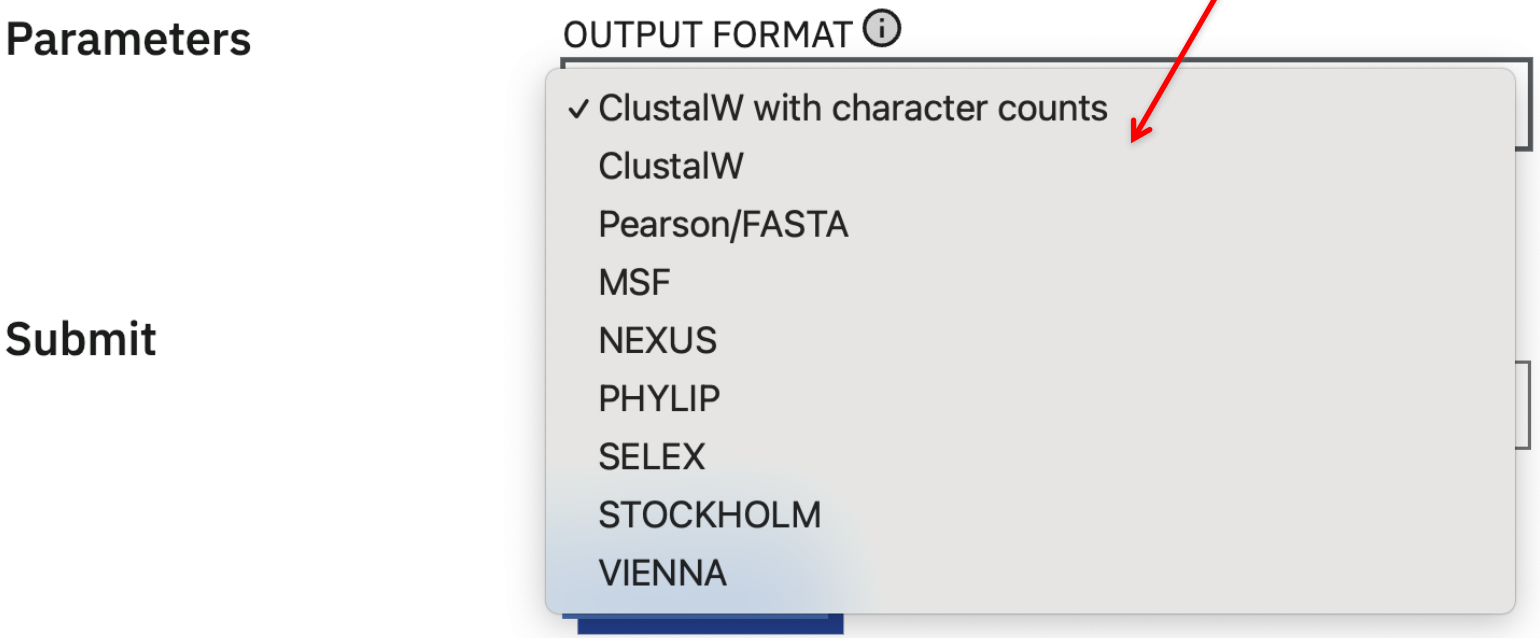
Clustal Omega Options

Clicking here…



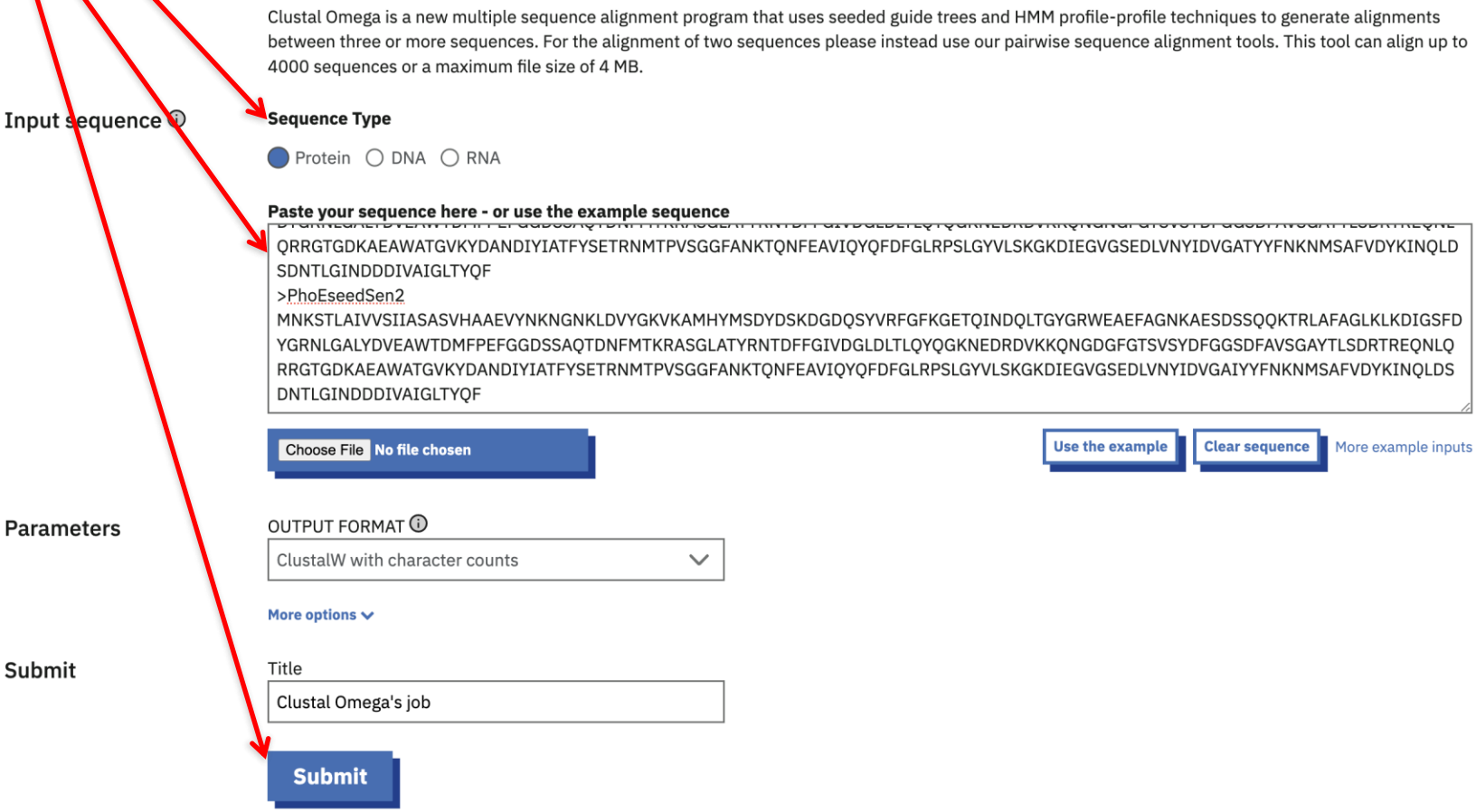
…will allow you to change alignment settings.

Most important for most people is how to change output formats.



Aligning Protein Sequences

Step 1. Set to Protein (default).
Step 2. Paste in Data (or choose file).
Step 3. Submit.



Aligning Protein Sequences

Step. 4 Wait for alignment

```
LCseedSfl        MKKLTVAISAVAASVLMAMSAQAAEIYNKDSNKLDLYGKVNAKHYFSSNDADDGDTTYVR  60
PhoEseedEco2     MKMKKSTLALVVMGIVASVSVQAAEIYNKDGNKLDVYGKVKAMHYMSDNDSKDGDQSYIR  60
PhoEseedEco1     MKMKKSTLALVVMGIVASASVQAAEIYNKDGNKLDVYGKVKAMHYMSDNDSKDGDQSYIR  60
PhoEseedEco4     --MKKSTLALVVMGIVASASVQAAEIYNKDGNKLDVYGKVKAMHYMSDNDSKDGDQSYIR  58
PhoEseedSen1     --MNKSTLAI-VVSIIASASVHAAEVYNKNGNKLDVYGKVKAMHYMSDYDSKDGDQSYVR  57
PhoEseedSen2     --MNKSTLAI-VVSIIASASVHAAEVYNKNGNKLDVYGKVKAMHYMSDYDSKDGDQSYVR  57
                     . :::  . .:: : *.:***:***:.****:****:* **:*. *:.*** :*:*

LCseedSfl        LGFKGETQINDQLTGFGQWEYEFKGNRAESQGSSKDKTRLAFAGLKFGDYGSIDYGRNYG  120
PhoEseedEco2     FGFKGETQINDQLTGYGRWEAEFAGNKAESDT-AQQKTRLAFAGLKYKDLGSFDYGRNLG  119
PhoEseedEco1     FGFKGETQINDQLTGYGRWEAEFAGNKAESDT-AQQKTRLAFAGLKYKDLGSFDYGRNLG  119
PhoEseedEco4     FGFKGETQINDQLTGYGRWEAEFAGNKAESDT-AQQKTRLAFAGLKYKDLGSFDYGRNLG  117
PhoEseedSen1     FGFKGETQINDQLTGYGRWEAEFAGNKAESDS-SQQKNRLAFAGLKLKDIGSFDYGRNLG  116
PhoEseedSen2     FGFKGETQINDQLTGYGRWEAEFAGNKAESDS-SQQKTRLAFAGLKLKDIGSFDYGRNLG  116
                 :***************:*:** ** **:***:  :::*.********  * **:***** *
```

Meaning of symbols in protein alignment:

*** (Asterix)** positions with a single, fully conserved residue.
: (colon) positions with conservation between amino acid groups of similar properties.
. (period) positions with conservation between amino acid groups of weakly similar properties.