# Tutorial: How to generate a publication-quality multiple sequence alignment

(Thomas Weimbs, University of California Santa Barbara, 11/2012)

**1) Get your sequences in FASTA format:**

- Go to the NCBI website; find your sequences and display them in FASTA format. Each sequence should look like this (http://www.ncbi.nlm.nih.gov/protein/6678177?report=fasta):

```
>gi|6678177|ref|NP_033320.1| syntaxin-4 [Mus musculus]
MRDRTHELRQGDNISDDEDEVRVALVVHSGAARLGSPDDEFFQKVQTIRQTMAKLESKVRELEKQQVTIL
ATPLPEESMKQGLQNLREEIKQLGREVRAQLKAIEPQKEEADENYNSVNTRMKKTQHGVLSQQFVELINK
CNSMQSEYREKNVERIRRQLKITNAGMVSDEELEQMLDSGQSEVFVSNILKDTQVTRQALNEISARHSEI
QQLERSIRELHEIFTFLATEVEMQGEMINRIEKNILSSADYVERGQEHVKIALENQKKARKKKVMIAICV
SVTVLILAVIIGITITVG
```

2) **In a text editor, paste all your sequences together** (in the order that you would like them to appear in the end). It should look like this:

```
>gi|6678177|ref|NP_033320.1| syntaxin-4 [Mus musculus]
MRDRTHELRQGDNISDDEDEVRVALVVHSGAARLGSPDDEFFQKVQTIRQTMAKLESKVRELEKQQVTIL
ATPLPEESMKQGLQNLREEIKQLGREVRAQLKAIEPQKEEADENYNSVNTRMKKTQHGVLSQQFVELINK
CNSMQSEYREKNVERIRRQLKITNAGMVSDEELEQMLDSGQSEVFVSNILKDTQVTRQALNEISARHSEI
QQLERSIRELHEIFTFLATEVEMQGEMINRIEKNILSSADYVERGQEHVKIALENQKKARKKKVMIAICV
SVTVLILAVIIGITITVG
>gi|151554658|gb|AAI47965.1| STX3 protein [Bos taurus]
MKDRLEQLKAKQLTQDDDTDEVEIAVDNTAFMDEFFSEIEETRVNIDKISEHVEEAKRLYSVILSAPIPE
PKTKDDLEQLTTEIKKRANNVRNKLKSMERHIEEDEVQSSADLRIRKSQHSVLSRKFVEVMTKYNEAQVD
FRERSKGRIQRQLEITGKKTTDEELEEMLESGNPAIFTSGIIDSQISKQALSEIEGRHKDIVRLESSIKE
LHDMFMDIAMLVENQGEMLDNIELNVMHTVDHVEKAREETKRAVKYQGQARKKLVIIIVIVVVLLGILAL
IIGLSVGLK
>gi|37577287|ref|NP_001971.2| syntaxin-2 isoform 1 [Homo sapiens]
MRDRLPDLTACRKNDDGDTVVVVEKDHFMDDFFHQVEEIRNSIDKITQYVEEVKKNHSIILSAPNPEGKI
KEELEDLNKEIKKTANKIRAKLKAIEQSFDQDESGNRTSVDLRIRRTQHSVLSRKFVEAMAEYNEAQTLF
RERSKGRIQRQLEITGRTTTDDELEEMLESGKPSIFTSDIISDSQITRQALNEIESRHKDIMKLETSIRE
LHEMFMDMAMFVETQGEMINNIERNVMNATDYVEHAKEETKKAIKYQSKARRKLMFIIICVIVLLVILGI
ILATTLS
```

3) **In your text editor, change the sequence names** to something that will look good in the end (make sure you keep the “>” symbol). For example:

```
>Stx4
MRDRTHELRQGDNISDDEDEVRVALVVHSGAARLGSPDDEFFQKVQTIRQTMAKLESKVRELEKQQVTIL
ATPLPEESMKQGLQNLREEIKQLGREVRAQLKAIEPQKEEADENYNSVNTRMKKTQHGVLSQQFVELINK
CNSMQSEYREKNVERIRRQLKITNAGMVSDEELEQMLDSGQSEVFVSNILKDTQVTRQALNEISARHSEI
QQLERSIRELHEIFTFLATEVEMQGEMINRIEKNILSSADYVERGQEHVKIALENQKKARKKKVMIAICV
SVTVLILAVIIGITITVG
>Stx3
MKDRLEQLKAKQLTQDDDTDEVEIAVDNTAFMDEFFSEIEETRVNIDKISEHVEEAKRLYSVILSAPIPE
PKTKDDLEQLTTEIKKRANNVRNKLKSMERHIEEDEVQSSADLRIRKSQHSVLSRKFVEVMTKYNEAQVD
FRERSKGRIQRQLEITGKKTTDEELEEMLESGNPAIFTSGIIDSQISKQALSEIEGRHKDIVRLESSIKE
LHDMFMDIAMLVENQGEMLDNIELNVMHTVDHVEKAREETKRAVKYQGQARKKLVIIIVIVVVLLGILAL
IIGLSVGLK
>Stx2
```

```
MRDRLPDLTACRKNDDGDTVVVVEKDHFMDDFFHQVEEIRNSIDKITQYVEEVKKNHSIILSAPNPEGKI
KEELEDLNKEIKKTANKIRAKLKAIEQSFDQDESGNRTSVDLRIRRTQHSVLSRKFVEAMAEYNEAQTLF
RERSKGRIQRQLEITGRTTTDDELEEMLESGKPSIFTSDIISDSQITRQALNEIESRHKDIMKLETSIRE
LHEMFMDMAMFVETQGEMINNIERNVMNATDYVEHAKEETKKAIKYQSKARRKLMFIIICVIVLLVILGI
ILATTLS
```

**4) Generate the multiple sequence alignment using T-Coffee:**

- Go to T-Coffee: http://tcoffee.crg.cat/apps/tcoffee/do:regular
- Paste your edited FASTA sequences into the input window.
- Hit “Submit” and wait for the result to come up.

**5) Generate a publication-quality output:**

- Because the colored output of T-Coffee is not suitable for publications, you need to format the alignment using another program called Boxshade.
- Download the “fasta_aln” result file from T-Coffee which should look like this:

```
>Stx4
MRDRTHELRQGDN-ISDDEDEVRVALVVHSGAARLGSPDDEFFQKVQTIRQTMAKLESKVRELEKQQVTILATPLPEESM
KQGLQNLREEIKQLGREVRAQLKAIEPQKEEADE--NYNSVNTRMKKTQHGVLSQQFVELINKCNSMQSEYREKNVERIR
RQLKITNAGMVSDEELEQMLDSGQSEVFVSNILKDTQVTRQALNEISARHSEIQQLERSIRELHEIFTFLATEVEMQGEM
INRIEKNILSSADYVERGQEHVKIALENQKKARKKKVMIAICVSVTVLILAVIIGITITVG
>Stx3
MKDRLEQLKAKQLTQDDDTDEVEIAVD-------NTAFMDEFFSEIEETRVNIDKISEHVEEAKRLYSVILSAPIPEPKT
KDDLEQLTTEIKKRANNVRNKLKSMERHIEE-DE--VQSSADLRIRKSQHSVLSRKFVEVMTKYNEAQVDFRERSKGRIQ
RQLEITGK-KTTDEELEEMLESGNPAIFTSGII-DSQISKQALSEIEGRHKDIVRLESSIKELHDMFMDIAMLVENQGEM
LDNIELNVMHTVDHVEKAREETKRAVKYQGQARKKLVIIIVIVVVLLGILALIIGLSVGLK
>Stx2
MRDRLPDLTACR--KNDDGDTVVV-VE-------KDHFMDDFFHQVEEIRNSIDKITQYVEEVKKNHSIILSAPNPEGKI
KEELEDLNKEIKKTANKIRAKLKAIEQSFDQ-DESGNRTSVDLRIRRTQHSVLSRKFVEAMAEYNEAQTLFRERSKGRIQ
RQLEITGR-TTTDDELEEMLESGKPSIFTSDIISDSQITRQALNEIESRHKDIMKLETSIRELHEMFMDMAMFVETQGEM
INNIERNVMNATDYVEHAKEETKKAIKYQSKARRKLMFIIICVIVLLVILGIILA--TTLS
```

- Go to Boxshade: http://www.ch.embnet.org/software/BOX_form.html
- Paste the “fasta_aln” result file from T-Coffee into the Boxshade input window.
- Select “other” as the Input sequence format.
- Select “RTF_new” as the output format.
- Hit “Run Boxshade...”, wait for the result and then click “here is your output…” to download your output file.
- Open your output file in MS-Word (most other text editors do not seem to be able to handle this file format).
- The final result should look like this (hopefully):


Stx4 1 MRDRTHELRQGDN-ISDDEDEVRVALVVHSGAARLGSPDDEFFQKVQTIRQTMAKLESKV
Stx3 1 MKDRLEQLKAKQLTQDDDTDEVEIAVD-------NTAFMDEFFSEIEETRVNIDKISEHV
Stx2 1 MRDRLPDLTACR--KNDDGDTVVV-VE-------KDHFMDDFFHQVEEIRNSIDKITQYV
Stx4 60 RELEKQQVTILATPLPEESMKQGLQNLREEIKQLGREVRAQLKAIEPQKEEADE--NYNS
Stx3 54 EEAKRLYSVILSAPIPEPKTKDDLEQLTTEIKKRANNVRNKLKSMERHIEE-DE--VQSS
Stx2 51 EEVKKNHSIILSAPNPEGKIKEELEDLNKEIKKTANKIRAKLKAIEQSFDQ-DESGNRTS
Stx4 118 VNTRMKKTQHGVLSQQFVELINKCNSMQSEYREKNVERIRRQLKITNAGMVSDEELEQML
Stx3 111 ADLRIRKSQHSVLSRKFVEVMTKYNEAQVDFRERSKGRIQRQLEITGK-KTTDEELEEML
Stx2 110 VDLRIRRTQHSVLSRKFVEAMAEYNEAQTLFRERSKGRIQRQLEITGR-TTTDDELEEML
Stx4 178 DSGQSEVFVSNILKDTQVTRQALNEISARHSEIQQLERSIRELHEIFTFLATEVEMQGEM
Stx3 170 ESGNPAIFTSGII-DSQISKQALSEIEGRHKDIVRLESSIKELHDMFMDIAMLVENQGEM
Stx2 169 ESGKPSIFTSDII-SDSQITRQALNEIESRHKDIMKLETSIRELHEMFMDMAMFVETQGEM
Stx4 238 INRIEKNILSSADYVERGQEHVKIALENQKKARKKKVMIAICVSVTVLILAVIIGITITV
Stx3 229 LDNIELNVMHTVDHVEKAREETKRAVKYQGQARKKLVIIIVIVVVLLGILALIIGLSVGL
Stx2 229 INNIERNVMNATDYVEHAKEETKKAIKYQSKARRKLMFIIICVIVLLVILGIILA--TTL
Stx4 298 G
Stx3 289 K
Stx2 287 S