



# Staden Course Exercises

*version 3.2 (public)*



## Licence

This manual is © 2007-8, Simon Andrews.

This manual is distributed under the creative commons Attribution-Non-Commercial-Share Alike 2.0 licence. This means that you are free:

- to copy, distribute, display, and perform the work
- to make derivative works

Under the following conditions:

- Attribution. You must give the original author credit.
- Non-Commercial. You may not use this work for commercial purposes.
- Share Alike. If you alter, transform, or build upon this work, you may distribute the resulting work only under a licence identical to this one.

Please note that:

- For any reuse or distribution, you must make clear to others the licence terms of this work.
- Any of these conditions can be waived if you get permission from the copyright holder.
- Nothing in this license impairs or restricts the author's moral rights.

Full details of this licence can be found at

<http://creativecommons.org/licenses/by-nc-sa/2.0/uk/legalcode>



## Exercise 1: Running Pregap

- ◆ You should have a directory containing a batch of example sequencing files
- ◆ Start Pregap by opening the Software repository and double clicking on the Pregap4 icon
- ◆ Use the Add Files button, and the file selector to add in all the SCF files in the Staden directory
- ◆ Make the following changes to the default module configuration
  - ◆ In Quality Clip increase the “Average Confidence” cutoff to 20
- ◆ Switch to the Textual Output window and run Pregap
- ◆ Look at the log which is generated to see if everything worked OK

## Exercise 2: Importing Sequences into Gap

- ◆ Start Gap by selecting double clicking the Gap4 icon in the Molbio Repository
- ◆ Create a new database called “assembly” in the Staden directory
- ◆ Use the “Normal Shotgun Assembly” to add in all the files listed in the pregap.passed file into this new database. Use the default assembly parameters
- ◆ Check how many contigs are created by looking at the Contig Selector
- ◆ Visualise how the traces are distributed across the contigs by generating a template display for all contigs

## Exercise 3: Using the Contig Editor

- ◆ Open the longest contig in the contig editor
- ◆ Adjust the settings so that any differences are highlighted with a different background colour. Do not make difference highlighting case sensitive
- ◆ Save your view options so they will be restored in future
- ◆ Scroll through the contig and identify where there are problems. Look at the underlying traces in the problem regions to see how they might be resolved



## Exercise 4: Correcting Mistakes

- ◆ Use the editing techniques you are shown to correct any mistakes in the assembly. Always use the lowest edit mode you can
- ◆ Save the changes you have made to your contig

## Exercise 5: Joining Contigs

- ◆ Use the Find Internal Joins tool to look for potential overlaps between your contigs
- ◆ Use the join editor to join contigs together where the overlaps are solid. Edit the individual contigs as necessary to get the joins to work
- ◆ You should be able to get your assembly down to two contigs. You may find a small overlap between them but do not use this to make a final join.

## Exercise 6: Adding More Data

- ◆ Add the “rawseq.txt” file as a single file to your existing assembly
- ◆ Use the find internal joins tool and the join editor to assemble your sequences to a single contig.
- ◆ Correct the contig so you are happy with the consensus it provides and save these changes.

## Exercise 7: Exporting Data

- ◆ Export the consensus sequence for your contig into a FastA format file
- ◆ Dump the whole alignment of your assembly project into a text file and import it into a Word document.