Date: Fri Aug 24 14:43:33 BST 2012
Author: F. Yalcin, Wellcome Trust Sanger Institute, Hinxton, U.K.
E-mail: fy2@sanger.ac.uk
Description: This document describes the delivered files and gives a description of the data analysis for your PAGe genome project.

The delivered files

Two files are delivered:

1. "name.mappings.txt" (text file)
   The Sanger Institute uses special IDs to store and track your isolates internally.This file maps the Sanger ID to your institute's isolate name. The file has the following four TAB delimited fields:

   1. Sanger ID for your isolate (For example "1234_1#23")
   2. Your institution's name for the isolate
   3. Sanger name for your isolate
   4. The collaborator name

2. “data.tar.gz" (a gzip compressed tar archive)
   This file contains the results of the analysis. You will need to decompress this file firstin order to see its contents. Please run this command in your unix terminal:

   ```
   tar -xzvf data.tar.gz
   ```

   The command above should create a subdirectory called "data" in your current working directory.The data directory contains one EMBL file per isolate. The file names have the Sanger IDsas prefix to indicate the isolates they represent. The EMBL files have the assembled genome sequence and the annotations for your isolates. The files can be visualised with the ARTEMIS genome browser or opened in a text editor.

A description of the data analysis

Each isolate was paired-end sequenced on the Illumina platform. The sequenced readswere assembled into contigs using the Sanger Pathogen Informatics assembly pipeline.
This pipeline assembled each isolate multiple times using various assembly settings results to create a multi fasta file with multiple contigs.

We employed Sanger institute’s automated annotation pipeline to annotate the contigsobtained from the assembly. This produced one EMBL file per annotated contig per isolate. We have merged these EMBL files and contigs into a single EMBL file with a single genome sequence. The coordinates of the original contigs (before the merging) can be seen in the EMBL file if desired.