

Minutes of the conference call on 4<sup>th</sup> September 2002 of the *Aspergillus fumigatus* (*Af*) genome sequencing group.

Participating in the call:

University of Manchester:

David Denning - chair

Michael Anderson (MA) - secretary

The Wellcome Trust Sanger Institute:

Bart Barrell (BB)

Neil Hall (NH)

David Harris (DH)

The Institute for Genomic Research:

William Nierman (WN)

Tamara Feldblyum (TF)

The University of Salamanca:

Miguel Sánchez-Pérez (MSP)

National Institute of Allergy and Infectious Diseases:

Dennis Dixon

Rory Duncan

1) Release of 10 x whole genome shotgun assembly:

BB and WN were happy for each other's data to be on each centre's website. NH recommended that the sequence be released as he had received requests for the data and it would be fairer for it to be available for everyone. He proposed that the data not be released to the public databases, but only to the two main centres' websites where the data would be covered by release policies. He would send the Sanger policy out for approval. BB made the point that if the project's intentions were clear regarding publication, then there would not be a vacuum which someone else could fill.

2) Pilot project:

DH reported that the sequencing of the 16 overlapping BAC clones was finished and that the sequence totalled 922 kb. NH stated that the annotation would be complete in a month or so. Writing the paper would take 1 or 2 months and the emphasis would be a comparison with the *Aspergillus nidulans* genetic map.

3) Whole genome shotgun assembly and closure:

TF stated that 13/14 Mb of the assembly had been sent to the Sanger. Half of the sequencing gaps in the TIGR scaffolds have been done and the next stage was to deal with the hard stops and the repeats. The sequence would almost certainly be misassembled in repetitive regions. For instance, all of the telomeric repeats bar one were in one contig. Another repeat class had a high A+T content consisting of retrotransposon-like sequence and probably originated from the centromeres.

DH stated that the Sanger assembly compared well with the scaffolds that they had received from TIGR. He projected that the sequencing gaps within these scaffolds (splits) would be done by the end of the year. He also stated that BAC end-pair sequence had almost been completed from another library (B26). Eighteen plates would be done.

The participants were informed that Owen White had set up an FTP site at TIGR which contained a list of the scaffolds and the reads associated with them. NH proposed that such a site be set up at Sanger.

4) Closure time line:

TF estimated that closure might be completed in the 2<sup>nd</sup> quarter of 2003, though it would depend on how difficult the repeats were. She suspected that one of the repeat classes might be problematic. NH stressed the importance of not involving external biologists too early as had occurred with the malarial project and suggested organising the meeting at the time when the annotation was started.

5) Spanish report:

MSP stated that they had received the cDNA library and should sequence the 1000 full-length clones in the next 3/4 months. They would send their traces to Sanger. They would therefore have money left over for 5 months of work. He stated that they were writing a new grant application for which they should have an answer in Jan. This project would be to analyse SNPs.

6) Additional sequencing:

cDNA: The question was raised if Sanger would have money remaining to carry out EST sequencing. Additional cDNA data would provide direct evidence for the presence of genes. NH stated that they could check the quality of the available library. The idea of specifically going after small RNAs was raised, but the problem of lots of incomplete mRNAs being present was pointed out.

Other genomes: MA proposed sequencing at least one additional genome to assist with the annotation of the *Af* genome. The question was whether to go for a closely related species such as *Neosartorya fischeri* or a more distantly related species such as a *Penicillium*. NH suggested that a suitable distance would be one where the average identity in coding regions was 70-80 %. It was suggested that 5x rather than 3x coverage would be preferable as the average contig size would increase and therefore provide more structural information. It was proposed that TIGR review their finances with 6 months to go in order to determine how much additional high-throughput sequencing could be done. TF stated that these costs were always decreasing and that it should take 2 months to make the libraries and sequence to 5x coverage. It was felt that this discussion needed to be continued and strategy finalised over the next few months. In the meantime MA would prepare genomic DNA from *Penicillium chrysogenum* (*Pc*). It was suggested that Spain could also be involved in sequencing *Pc*.

7) Fungal genetics conference in March 2003:

It was proposed that NH talk at this meeting where he could present the data from the pilot project.

Michael J. Anderson

September 2002