

2007 STEP EUR Report
Guoqing Lu and Mark Pauley
Summer Research Course in Bioinformatics Programming

The names of the students who took the course and a brief description of their projects follow:

Faisal Ahmed – Faisal wrote a sequence (contig) assembly program. The program takes input in the form of a simple FASTA-formatted input file of sequences and generates contig(s).

Adam Cornish – Adam's project was entitled *Automatic Peptide Mass Fingerprint (PMF) File Submission*. The program that he wrote helps in the analysis of proteomic data from mass spectrometry experiments. In particular, it submits a list of masses to the Mascot search engine, which then performs database searches and attempts to identify the unknown protein using a probabilistic approach. It is important to note that his program is in active use by researchers at the Nebraska Center for Virology (UNMC; <http://www.unl.edu/virologycenter>).

Auxence Meminisima – Auxence was initially interested in doing a project that involved the generation and analysis of primers for PCR (polymerase chain reaction) experiments. After several weeks of work, it was determined that what he wanted to do was overly complicated and a more straightforward project, one that involved the creation of a program to manipulate genetic sequences, was developed. Auxence completed his project during Fall 2007.

Kossi Noameshie – Kossi's project involved the writing of a program to automatically perform BLAST (Basic Local Alignment Search Tool) searches. In particular, the program appropriately formats the target database, BLASTs a series of query sequences against this database and parses the results into a simple file. The program uses elements from BioPerl (<http://www.bioperl.org>).

Thaine Rowley – Thaine's project involved the development of *ChangeSeq*, a program that converts genetic sequences from one format to another. In particular, *ChangeSeq* converts between FASTA, Phylip and Genbank formats. In addition, *ChangeSeq* includes an advanced file upload script that checks for acceptable file extensions, illegal characters, reverse path inclusion, and data contents format. It is important to note that *ChangeSeq* is being used in the FluGenome project centered at UNO and the National Institutes of Health (<http://h5n1.flugenome.org/index.php>).

Pictures of Faisal (Figure 1), Adam (Figure 2) and Thaine (Figure 3) giving their final reports are included below.



Figure 1 – Faisal Ahmed **Figure 2** – Adam Cornish



Figure 3 – Thaine Rowley