THE *MEDICAGO TRUNCATUALA* GENOME SEQUENCING PROJECT: STRATEGY, STATUS AND RESULTS

Chris Town, The Institute for Genomic Research, 9712 Medical Center Drive, Rockville MD 20850.
cdtown@tigr.org; www.tigr.org/tdb/e2k1/mta1

Legumes are one of the world's most important crop families, unique in their ability to fix atmospheric nitrogen. Among legumes, *M. truncatula* with a genome size of ~ 500 Mb is considered an excellent model for genome research. Unlike most crop legumes, *M. truncatula* has a compact genome, simple genetics, short generation time, relatively high transformation, excellent mutants, a large collection of ecotypes and an active research community. Comparative genomics demonstrates that the *M. truncatula* genome is highly syntenic with the genomes of alfalfa and pea and substantially conserved with several other legumes including soybean. Genomic resources already developed include a large collection of ESTs, several BAC libraries and an FPC-based physical map that is linked in many places to the genetic map. Because of these desirable features, *M.truncatula* is now the subject of an international sequencing effort. Persuasive evidence based on fluorescent in situ hybridization (FISH) of sequenced BACs to pachytene chromosomes demonstrates that the *M. truncatula* genome is organized into pericentromeric heterochromatin, rich in repeats, and separate gene-rich euchromatic chromosome arms. Thus, the overwhelming majority of the gene space, estimated to be 200-250 Mb, can be sequenced in a highly efficient manner using a BAC-based approach. The international sequencing consortium consists of groups from the USA (University of Minnesota, Oklahoma University and The Institute for Genomic Research) and the EU (INRA-Toulouse with Genoscope and the John Innes Centre with the Sanger Centre). Chromosomes 1, 4, 6 and 8 will be sequenced by OU, 2 and 7 by TIGR, 3 by JIC/Sanger and 5 by INRA/Genoscope. As of July 1, 2004 there was approximately 50 Mb of finished sequence and 50 Mb of "sequencing in progress" in GenBank. The project aims to complete the sequencing of the euchromatic gene-rich space by the end of 2006. During this sequencing phase of the project, automated bioinformatics pipelines are being used by several groups to provide working annotation of the sequence. The US project coordinating web site is at www.medicago.org and provides links out to the various web sites providing annotation and other information in both the USA and the EU. In this talk, I will describe the sequencing strategy and the current status of the project and highlight some of the features of the *M. truncatula* genome organization and gene content that have been revealed to date.

The US effort is supported by the National Science Foundation, and previous sequencing at OU was supported by the Noble Foundation.