

5th Annual Joint Bioinformatics Workshop

July 19, 2005
2229 Seamans Center
University of Iowa

by

Brent Kronmiller
Iowa State University

Sequencing a 1.3 Mb contig spanning the rf1 fertility restorer locus as a prototype to assess complex-genome coverage strategies

ABSTRACT

Brent Kronmiller 1,3, Karin Werner 2,3 and Roger Wise 1,2,3

1) Bioinformatics and Computational Biology, 2) USDA-ARS Corn Insects and Crop Genetics Research; 3) Plant Pathology, Iowa State University, Ames, Iowa, USA, 50011

In T-cytoplasm maize, cytoplasmic male sterility (CMS) is attributed to the presence of the unique mitochondrial gene, T-urf13. Full suppression of T-urf13-mediated CMS is directed by the combined action of dominant alleles of the nuclear (fertility restorer) genes, rf1 and rf2a. To facilitate a candidate approach towards identification of the rf1 gene, three B73 BAC libraries were used to create a physical map of 794 clones from the centromeric region of chromosome 3 anchored to the rf1 locus. A minimum-tiling path of 14 contiguous BACs covering 1.3 megabases were shotgun sequenced, assembled and finished to completion for annotation and display in the GBrowse viewer. Eighty-seven percent has been identified as repetitive sequences, with most transposable elements found in large nested clusters spanning up to 300 kb with insertion chronologies of -0.19 to -8.50 million years. GeneSeqer, Fgenesh, and GeneMark.hmm were used to predict consensus locations and structures for 53 genes. Thirty-seven of these are positioned in gene clusters with as many as 8 members. Two hundred fifty-four GSS assemblies (including MAGs, TIGR's AZM and PlantGDB's GSS) aligned to the 1.3 Mb contig, 36 of which aligned to predicted genes. Two hundred eighteen GSS assemblies aligned to regions not predicted as genes, revealing that only 15% of GSS contigs align to genes in this region. Seventeen predicted genes did not correspond to any GSS assembly indicating that at least in the centromeric region of chromosome 3, finished sequence can provide a significant number of previously undescribed gene predictions. Research funded by USDA-NRI 2002-35301-12064.