

PLSC 611 Genomics

Assignment: Genome sequence, annotation, and gene analysis.

Points: 300 points (toward homework portion of grade)

Due: December 3, 2004

The actual sequence and initial description of the *Arabidopsis thaliana* genome was completed and published in December 2000. As we discussed in class, the annotation of the genome is an on-going process. One of the annotation groups is centered at TIGR (The Institute for Genomic Research). In addition, other researchers have taken that sequence data and are performing individual analyses of genes and gene families of interest to themselves. The review you will write will touch on all of these topics.

You will use four research articles as the basis of your review. If you wish to use other papers to support your writing, that will be fine. As long as your review is complete and addresses all of the narrative requirements (see below), these four basic references should be sufficient. The four references are listed below and you can download them from the class homework site.

TAGI. 2000. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 408:796-815.

Wortman et al. 2003. Annotation of the Arabidopsis genome. *Plant Physiology* 132:461-468.

Heim et al. 2003. The basic helix-loop-helix transcription factor family in plants: a genome-wide study of protein structure and functional diversity. *Molecular Biology and Evolution* 20:735-747.

Parenicova et al. 2003. Molecular and phylogenetic analyses of the complete MADS-box transcription factor family in Arabidopsis: New openings to the MADS world. *The Plant Cell* 15:1538-1551.

Format (Note: if you fail to follow these instructions you will lose up to a full letter grade)

Page length: Minimum of five pages. The first page is the cover page that will include the title of your paper, your name, and the name of the course. The remaining three (or more pages) will contain the narrative of your review. The last page will contain your list of references in the same format used in Wortman et al (2003) article.

Spacing: All paragraphs will be single-spaced and begin with a 0.5 inch indent. Paragraphs will be separated by a double space.

Font: 12 point, Times Roman only

Margins: One-inch margins (top, bottom, left, right) for all pages except the title page.

Narrative requirements

1. Describe the approach used to generate the actual DNA sequence. Describe the basic conclusions regarding the organization of the genome. You will need to discuss: the coding region of the genome and how it compares to other species; general genome organization; distribution of transposable elements; and the organization of at least two functional families. (1 page minimum)
2. What approach was initially use to define gene models and annotate the genes? What approaches did TIGR use and how did they differ from that initial approach? (1 page minimum)
3. Compare and contrast the approaches used to define the basic helix-loop-helix and MADS-box gene families. Compare the conclusions regarding the evolution of the two different gene families. Describe the expression studies used to analyze these gene families. (1 page minimum)