

Initial Development of an Informatics Platform for Octoploid Strawberry

Problem:

The production of cultivated strawberry faces new and familiar challenges. These barriers may be understood or possibly overcome with the assistance of new technologies that may allow molecular characterization of physiological processes, development of molecular markers and establishment of transgenic research. Yet at the start of 2004 there was a negligible amount of gene sequence information available in public databases. To remedy this discrepancy my program proposed to prepare EST libraries from our own funding if NASGA would be willing to grant funds to sequence them. The proposal would finance sequencing of approximately 1000 expressed sequence tags (ESTs), substantially expanding the community database and providing tools to initiate molecular studies of traits of agricultural importance. \$2750 was generously allocated to my research program to expand the public database. The funding was matched by the Horticultural Sciences Department at the University of Florida.

Results:

Together, these funds financed sequencing of over 1800 (ESTs) from *Fragaria x ananassa* cv. 'Strawberry Festival'. Approximately 1300 unigenes were defined, as well as SSRs that would be used for molecular markers in subsequent mapping studies. The sequences were immediately deposited to GenBank and the Genome Database for Rosaceae (GDR), two publicly-accessible databases. Sequenced clones were made available to the public, along with the library itself.

Conclusions:

It is difficult to give a distilled synopsis of the results because they are far reaching and present at many levels. Perhaps the best conclusion is to present how the funding has shaped public databases and contributed to genetic, molecular and structural/functional genomics efforts. The following are the tangible results that would have not been possible without this seed funding:

- 1505 unique sequences have been delivered to GenBank and the GDR. A summary of the results is available at http://www.genome.clemson.edu/gdr/projects/fragaria/folta/FA_SEa/index.shtml
- 187 SSRs have been identified and have been used in mapping in the diploid species.
- The information gathered formed the foundation of the following publications:

Sargent, D., Stewart, PJ, Bassil, NV, Simpson, D., and KM Folta (2005) Toward merging mapping and gene function in *Fragaria* through EST-SSR development. *JASHS* (submitted).

Folta, KM, Staton, M, Main, D, Stewart, P, and Bies, DH (2005) Maximal mining of small data sets; an EST collection from *Fragaria x ananassa*. *Genome Research* (in preparation for December 2004 submission).

- The information obtained is the foundation of the following pending federal grant proposals studying structural and functional genomics questions in strawberry:

Davis, TM and Folta KM – Defining genome composition and dissecting a genetically-complex trait in octoploid strawberry (*Fragaria-Rosaceae*). NASGA funded sequencing has allowed us to identify the strawberry genes associated with flowering and photoperiod. The proposed work examines the molecular basis of this important trait in day-neutral and short-day cultivars and sequence variation between the A, A' and B genomes within the octoploid to study genome composition.

Davis, TM, Folta KM, SanMiguel, P – Gene pair haplotypes and sequence samples from strawberry (*Rosaceae*): multipurpose, transferable resources for genomics and variety improvement. This study uses NASGA-funded sequence information to test synteny and co-linearity between strawberry and *Arabidopsis* and use these relationships as a basis for mapping genes in the octoploid.

Sosinski, B, Folta, KM, Main, D, Dandekar, A, Silva H – fGDR, A standardized community platform for functional-genomics studies in *Rosaceae*. NASGA funds have allowed strawberry to be represented in the proposed *Rosaceae* microarray!