

Cite this abstract as:

A.M. Baldo, J. Labate, and L.D. Robertson. 2004. A search for molecular diversity in tomato. p. 147 *In* Final Abstracts Guide, Plant and Animal Genome XII, San Diego, CA.

P300

A SEARCH FOR MOLECULAR DIVERSITY IN TOMATO

Angela M. Baldo¹, Joanne Labate¹, Larry D. Robertson¹

¹ USDA,-ARS Plant Genetic Resources Unit, New York State Agricultural Experiment Station, Geneva, NY 14456-0462, USA

Cultivated tomato is known to be relatively low in genetic diversity. This is a result of microevolutionary processes such as founder events, genetic bottlenecks, and intense selection. For these reasons, a computational approach to predicting single nucleotide polymorphisms (SNPs) is valuable to direct laboratory efforts toward regions more likely to yield results. We have developed a method to screen an entire NCBI Unigene set for potential SNPs using SEAN (Huntley, 2003). Predictions are based on established criteria: A window on either side of the predicted SNP must be identical for all sequences in the alignment, at least two sequences agreeing on each of a minimum of two polymorphisms, etc. Polymorphisms were further examined in the context of the cultivars and clones in which they were identified. Using this method, we discovered 2,527 potential SNPs among 764 clusters from the unigene set. We are in the process of verifying these polymorphisms in the laboratory.

A search for molecular diversity in tomato

Angela M. Baldo, Joanne Labate, Larry Robertson

United States Department of Agriculture

Agricultural Research Service

Plant Genetic Resources Unit, Geneva, NY, USA

<http://www.ars-grin.gov/gen>

abaldo@pgru.ars.usda.gov



Abstract

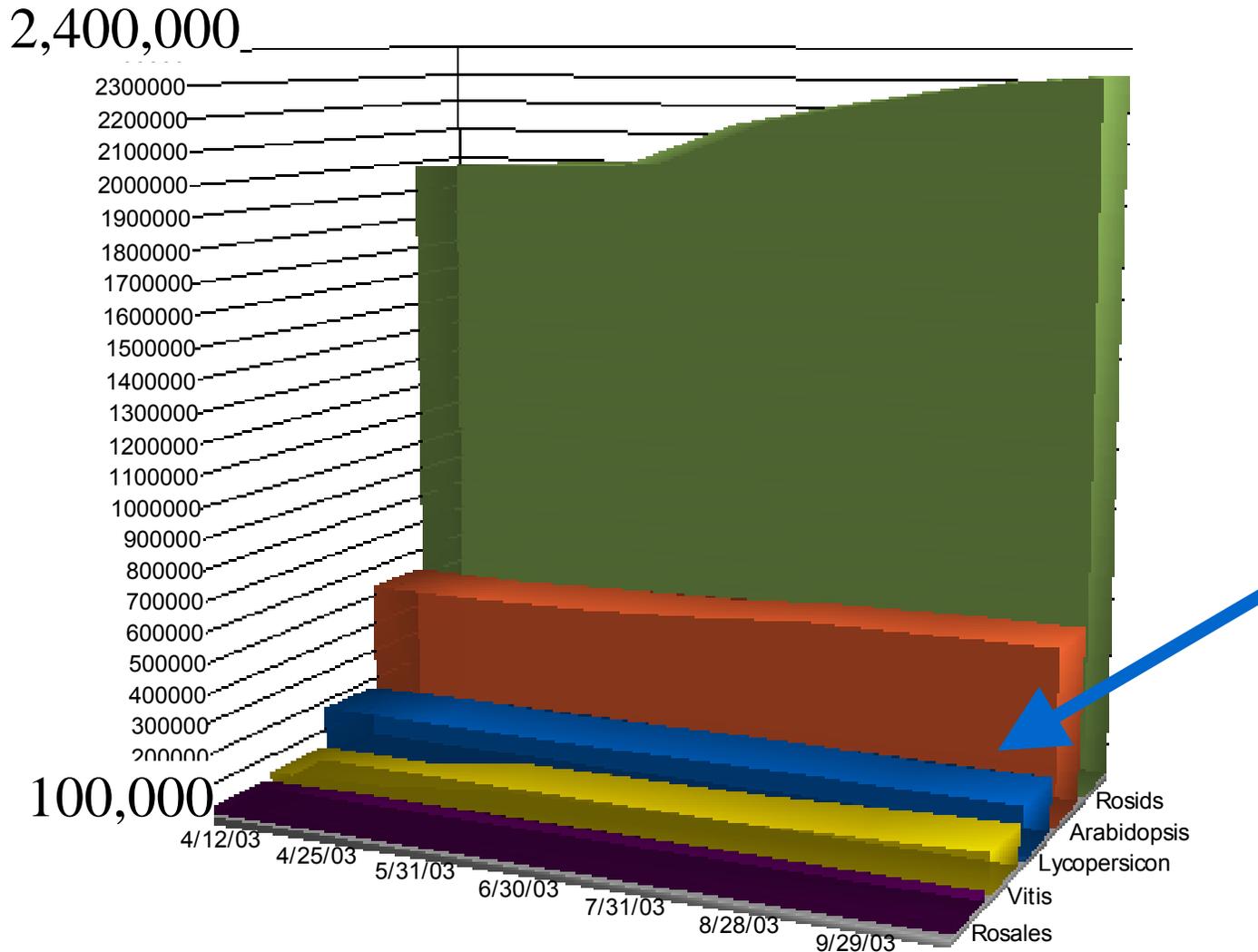
Cultivated tomato is known to be relatively low in genetic diversity. This is a result of microevolutionary processes such as founder events, genetic bottlenecks, and intense selection. For these reasons, a computational approach to predicting single nucleotide polymorphisms (SNPs) is valuable to direct laboratory efforts toward regions more likely to yield results. We have developed a method to screen an entire NCBI Unigene set for potential SNPs using the SEAN SNP Prediction Program (Huntley, 2003). Predictions are based on established criteria: A window on either side of the predicted SNP must be identical for all sequences in the alignment, at least two sequences agreeing on each of a minimum of two polymorphisms, etc. Polymorphisms were further examined in the context of the cultivars and clones in which they were identified. Using this method, we discovered 2,527 potential SNPs among 764 clusters from the unigene set. We are in the process of verifying these polymorphisms in the laboratory.

Why we need to find SNPs in Tomatoes

- Germplasm Collection Management
- Relationships among accessions
- Identifying duplicates
- Association Mapping for Traits
- Population Genetics
- * Low diversity between *Lycopersicon* varieties creates the need to predict SNPs from public sequences

Sequences available in GenBank

from 4/12/03 to 9/29/03



Rosids

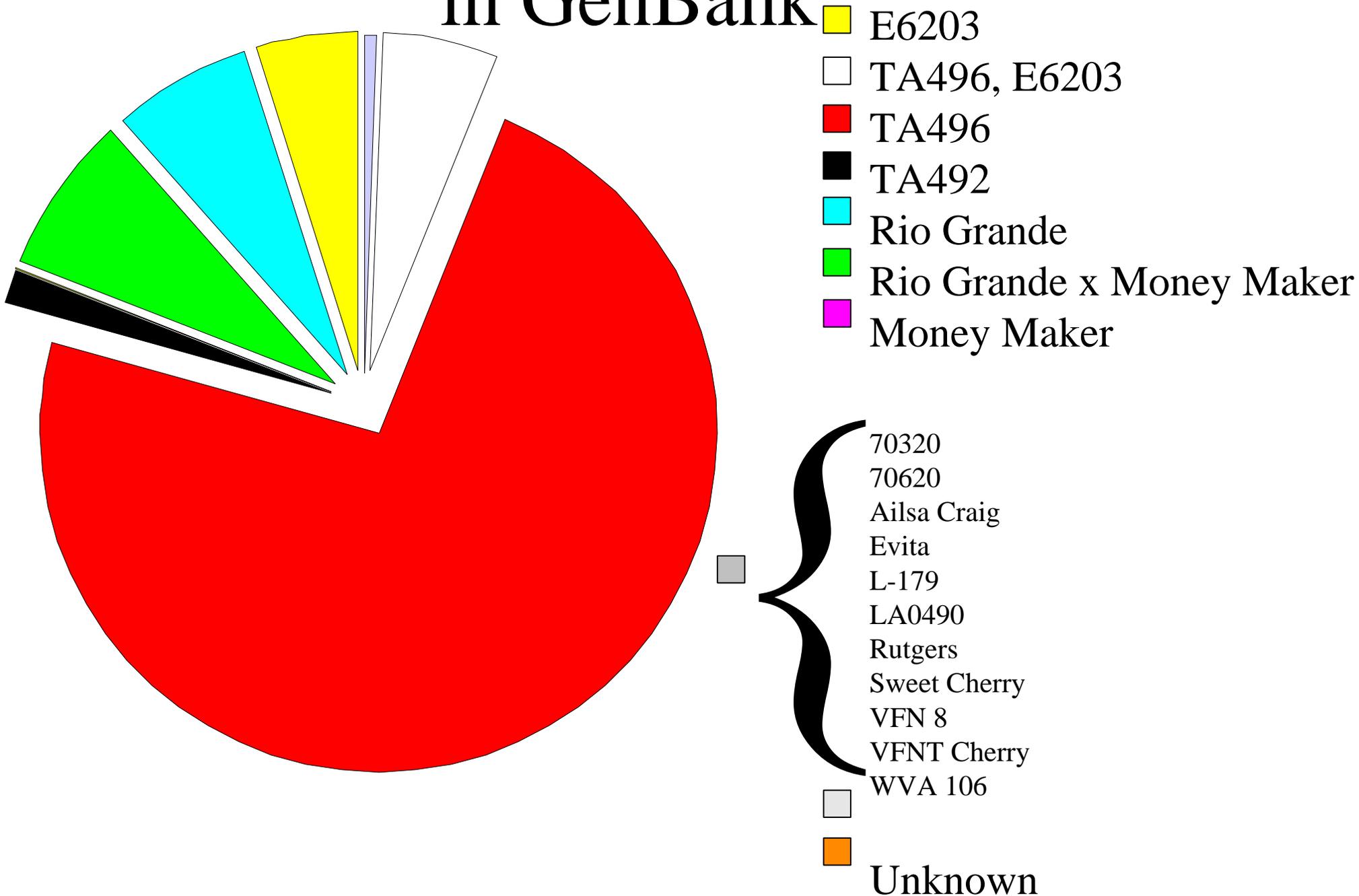
Arabidopsis

Lycopersicon

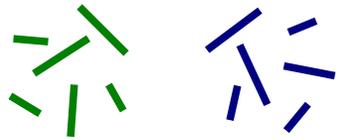
Vitis

Rosales

There are 148,000 Tomato ESTs in GenBank



SNP Discovery in Tomatoes



NCBI Unigene Set

- sequence similarity
- CDNA clone origin

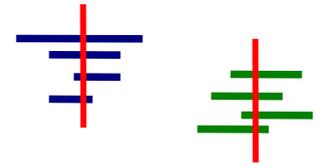
Other sets:

- TIGR Gene Index
- Sputnik
- SGN

Cluster
&
align

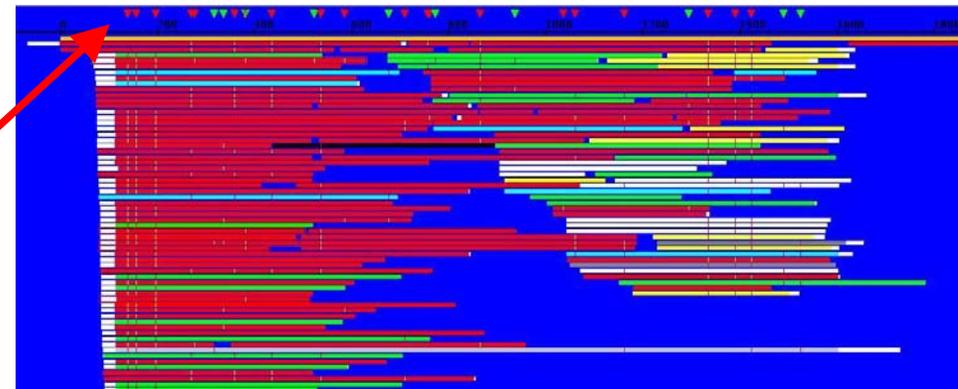
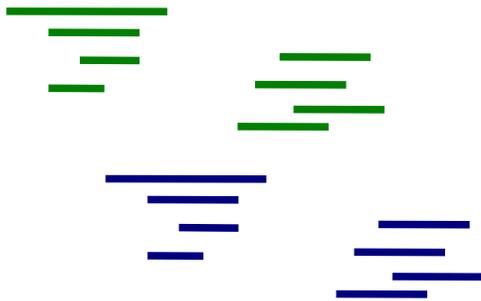
Identify potential SNPs

- 2 sequences each
- Perfect matches surrounding site
- 2 clones matching for a variety



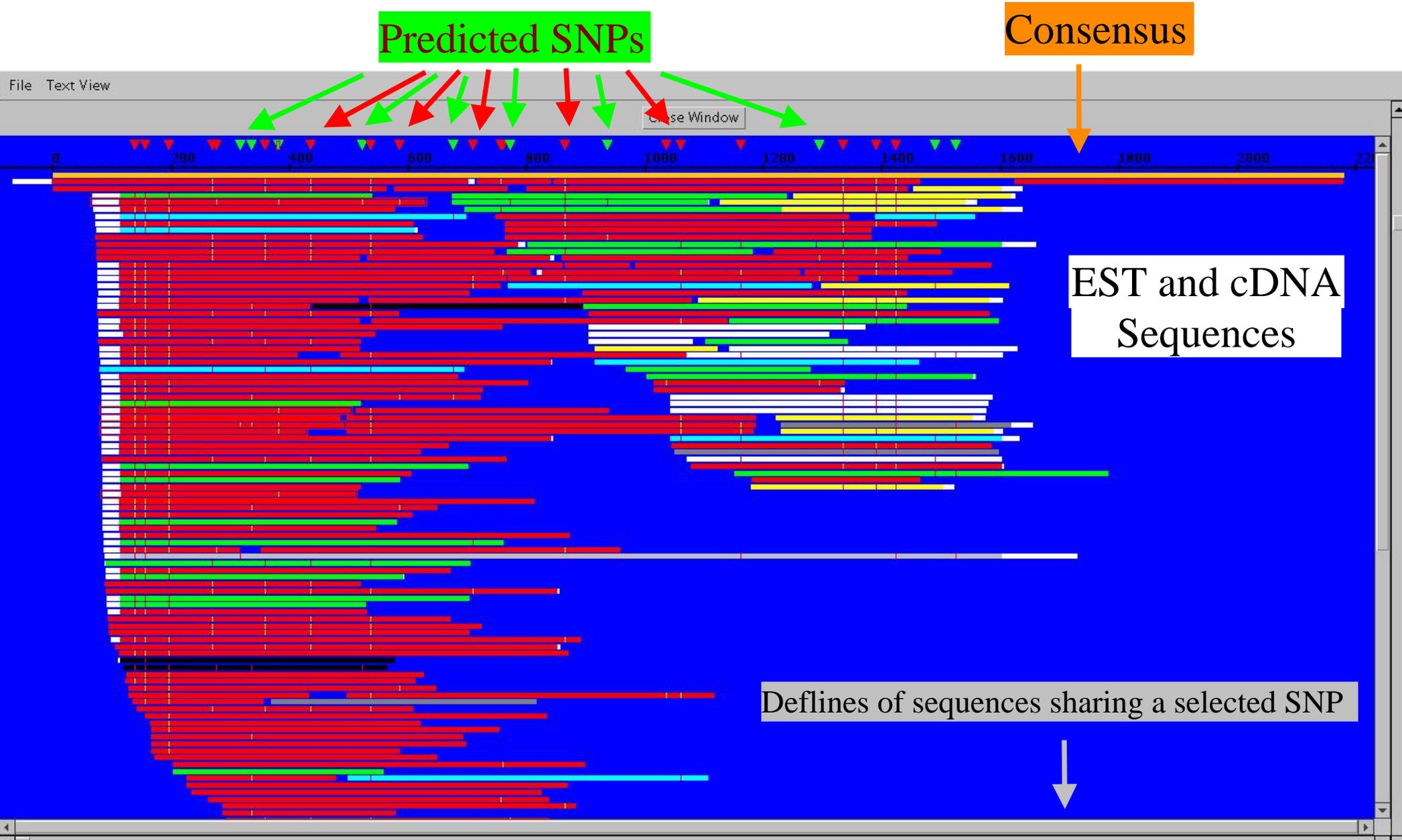
BLAST consensi against
Swissprot &
mapped tomato markers

Visually
Examine
Alignments



- ▼ SNP **confirmed** in more than one clone from the same cultivar
- ▼ SNP **not confirmed** in more than one clone from same cultivar

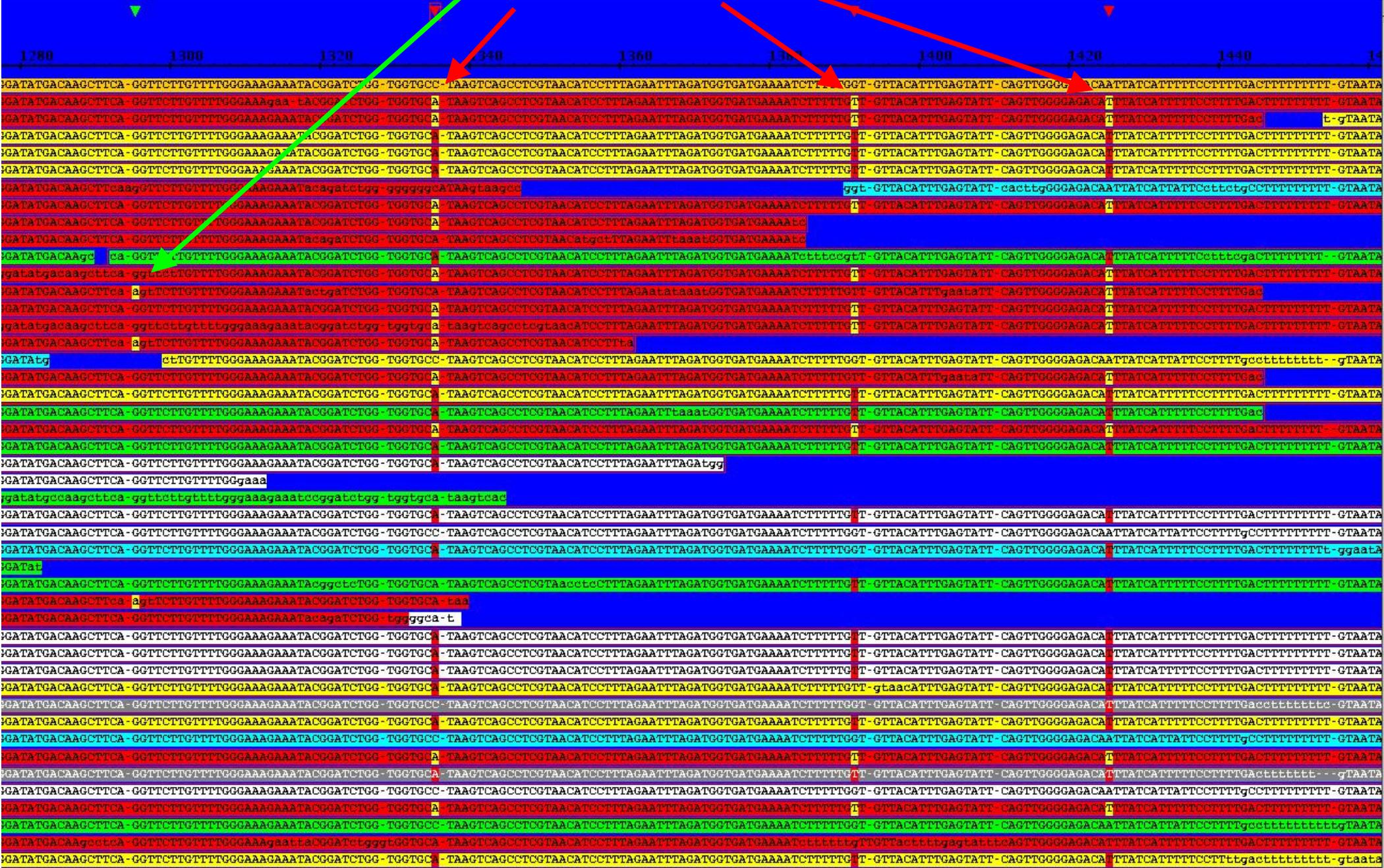
SEAN Viewer with Adenine Nucleotide Translocator Contig



+ gnl|UG|Les#S5219574 EST402149 tomato root, plants pre-anthesis, Cornell University Lycopersicon esculentum cDNA clone cLEY17D11, mRNA sequence /clone
+ gnl|UG|Les#S5178761 EST280416 tomato callus, TAMU Lycopersicon esculentum cDNA clone cLEC32P17 similar to ADP/ATP translocator, mRNA sequence /clor
+ gnl|UG|Les#S5179229 EST281689 tomato callus, TAMU Lycopersicon esculentum cDNA clone cLEC32D1 similar to ADP/ATP translocator, mRNA sequence /clon
+ gnl|UG|Les#S5178759 EST280414 tomato callus, TAMU Lycopersicon esculentum cDNA clone cLEC32P15 similar to ADP/ATP translocator, mRNA sequence /clor
+ gnl|UG|Les#S5235206 EST414675 tomato developing/immature green fruit Lycopersicon esculentum cDNA clone cLEM6L7, mRNA sequence /clone=cLEM6L7 /g

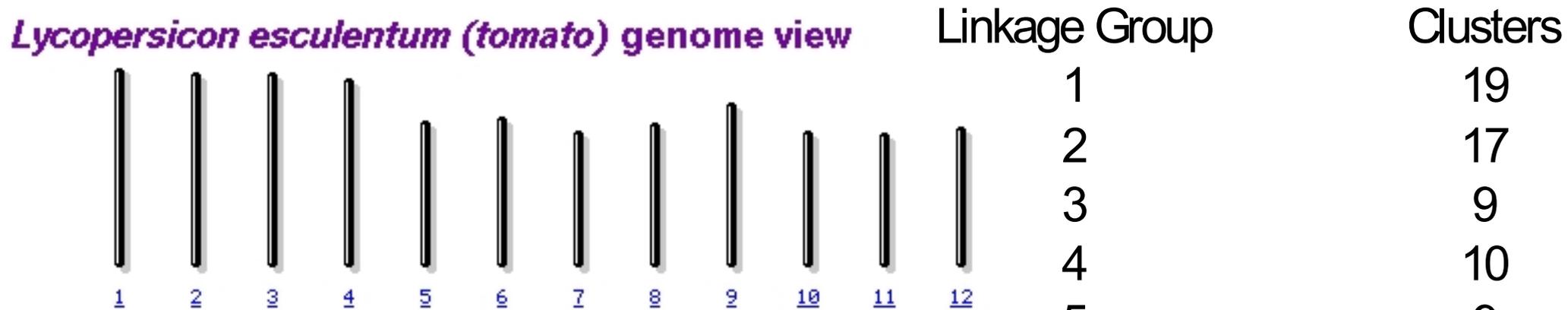
Zooming in on Adenine Nucleotide Translocator

Predicted SNPs



2,527 Potential Tomato SNPs in 764 Clusters

Frequency of predicted SNPs
 $\sim 3.9 \times 10^{-4}$ (~ 1 in every 2,500 BP examined)



Mapped markers matching clusters with SNPs:

23 RFLP

22 EST-by-clone

2 Protein?

81 COS (ESTs in Tomato & Arabidopsis)

3 originally thought COS

Clusters Matching Mapped Markers	131
----------------------------------	-----

Acknowledgements

- Derek Huntley (Imperial College, London)
- Bob Nearpass (PGRU, ARS, USDA)
- Robert Ahrens (SGN, Cornell University)
- USDA-ARS for Funding the Hardware

References

- Balasubramanian, S., Harrison, P., Hegyi, H., Bertone, P., Luscombe, N., Echols, N., McGarvey, P., Zhang, Z.L., Gerstein, M. 2002. SNPs on human chromosomes 21 and 22 – analysis in terms of protein features and pseudogenes. *Pharmacogenomics* 3(3):393-402
- Barker, G., Batley, J., O’Sullivan, H., Edwards, K.J., Edwards, D. 2003. Redundancy based detection of sequence polymorphisms in expressed sequence tag data using autoSNP. *Bioinformatics* 19(3):421-422
- Batley, J., Barker, G., O’Sullivan, H., Edwards, K.J., Edwards, D. 2003. Mining for Single Nucleotide Polymorphisms and Insertions/Deletions in Maize Expressed Sequence Tag Data. *Plant Physiology* 132:84-91
- Estivill, X., Cheung, J., Angel Pujana, M., Nakabayashi, K., Scherer, S.W., Tsui, L.C. 2002. Chromosomal regions containing high-density and ambiguously mapped putative single nucleotide polymorphisms (SNPs) correlate with segmental duplications in the human genome. *Hum. Mol. Genet.* 11(17):1987-1995
- Green, P. <http://www.phrap.org>
- Huntley, D. 2003. SEAN SNP Prediction and Display Programs <http://zebrafish.doc.ic.ac.uk/Sean>
- Jander, G; Norris, SR; Rounsley, SD; Bush, DF; Levin, IM; Last, RL 2002. *Arabidopsis* Map-Based Cloning in the Post-Genome Era *Plant Physiol.* 129:440-450
- Kuittinen, H., Salguero, D., Aguade, M. 2002. Parallel Patterns of Sequence Variation Within and Between Populations at Three Loci of *Arabidopsis thaliana*. *Mol. Biol. Evol.* 19(11):2030-2034
- Lazo, G.R., Lui, N., You, F., Hummel, D., Chao, S., and Anderson, O.D. 2003. Charting Contig-Component Relationships Within The Triticeae. 23rd Stadler Genetics Symposium, Genome Exploitation: Data Mining, J.P. Gustafson, ed., March 31-April 2, 2003, Columbia, MO. Kluwer Academic/Plenum Publishers, New York (in press).
- Lercher, M.J., Hurst, L.D. 2002. Human SNP variability and mutation rate are higher in regions of high recombination. *TIG* 18(7):337-340
- Majewski, J., Ott, J. 2003. Amino acid substitutions in the human genome: evolutionary implications of single nucleotide polymorphisms. *Gene* 305:167-173
- Morton, B.R., Oberholzer, V.M., Clegg, M.T. 1997. The Influence of Specific Neighboring Bases on Substitution Bias in Noncoding Regions of the Plant Chloroplast Genome. *J. Mol. Evol.* 45:227-231
- Nasu, S., Suzuki, J., Ohta, R., Hasegawa, K., Yui, R., Kitazawa, N., Monna, L., Minobe, Y. 2002. Search for and Analysis of Single Nucleotide Polymorphisms (SNPs) in Rice (*Oryza sativa*, *Oryza rufipogon*) and Establishment of SNP Markers. *DNA Research* 9:163-171
- Rafalski, A. 2002. Applications of single nucleotide polymorphisms in crop genetics. *Curr. Opin. Plant Bio.* 5:94-100
- Tomso, D.J., Bell, D.A. 2003. Sequence Context at Human Single Nucleotide Polymorphisms: Overrepresentation of CpG Dinucleotide at Polymorphic Sites and Suppression of Variation in CpG Islands. *JMB* 237:303-308
- Zhao, Z.M., Boerwinkle, E. 2002. Neighboring-Nucleotide Effects on Single Nucleotide Polymorphisms: A Study of 2.6 Million Polymorphisms Across the Human Genome. *Genome Research* 12:1679-1686
- Zhu, Y.L., Song, Q.J., Hyten, D.L., VanTassell, C.P., Matukumalli, L.K., Grimm, D. R., Hyatt, S. M., Fickus, E.W., Young, N.D., Cregan, P.B. 2003. Single-Nucleotide Polymorphisms in Soybean Genetics 163:1123-1134