Cite this poster as:

Sheffer, S.M., L.D. Robertson, W.F. Lamboy, A.M. Baldo, and J.A. Labate. 2006. DNA polymorphism in tomato, a crop with a history of selection, bottlenecks and introgression. Annual Meeting of the Society for the Study of Evolution, Stony Brook, NY.

DNA polymorphism in tomato: a crop with a history of selection, bottlenecks, and introgression

Susan M. Sheffer, Larry D. Robertson, Angela M. Baldo, Warren F. Lamboy and Joanne A. Labate, USDA, ARS Plant Genetic Resources Unit, Geneva, NY 14456

Introduction

❖Bottlenecks have recurred in *L. esculentum* during migration away from centers of diversity in Chile, Ecuador and Peru (Jenkins, 1948).

❖This cultivated species has also undergone strong directional selection for crop improvement, and natural selection for adaptation to temperate regions.

❖During the past 70 years, wild relatives within the genus have been used extensively in breeding for disease resistance (Stevens and Rick, 1986).

❖The natural history of this crop has led to the hypothesis that genetic variation is lacking within the species, with the exception of known introgressed genes (Nesbitt and Tanksley, 2002).

Results

❖ Within *L. esculentum*, we verified 150 SNPs in 22.3 kb (about 1 SNP per 150 bp). Three classes of markers gave similar estimates of polymorphism (Table 1).

❖Five loci rejected the null model in the HKA test (Hudson, 1987) (Table 2).

❖Three loci gave significant results in Fu and Li's test using an outgroup (Fig. 1). In each case, one allele was more diverged from all other *L. esculentum* alleles than it was from the *L. peruvianum* allele (LR24 at LeSNP34, TA496 at LeSNP1, and LR29 at psy1, not shown).

❖Seven hypothetical introgressions from wild species into *L. esculentum* were identified using allele trees (not shown). EST-based marker LeSNP9b is an example (Fig. 2).

References

Colosi, J.C. and B.A. Schaal. 1993. Tissue grinding with ball bearing and vortex mixer for DNA extraction. Nuc. Acids.Res.21: 1051-1052.

Fu, X.Y. and W.H. Li. 1993. Statistical tests of neutrality of mutations. Genetics 133: 693-709.

Green P. 2004. Phred, Phrap, Consed. [Online] http://www.phrap.org/phredphrapconsed.html

Hamblin, M.T., et. al. 2004. Comparative population genetics of the Panicoid grasses: sequence polymorphism, linkage disequilibrium and selection in a diverse sample of Sorghum bicolor. Genetics 167: 471-483.

Hudson, R.R., M. Kreitman and M. Aguade. 1987. A test of neutral molecular evolution based on nucleotide data. Genetics 116: 153-159.

Jenkins, J.A. 1948. The origin of cultivated tomato. Econ. Bot. 2: 379-392.

Kumar S., K. Tamura, and M. Nei. 2004. MEGA3: Integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment. Briefings in Bioinfomatics 5: 150-163.

Labate, J.A., and A.M. Baldo. 2005. Tomato SNP discovery by EST mining and resequencing. Mol. Breeding 16: 343-349.

Roselius, K, W. Stephan and T. Stadler. 2005. The relationship of nucleotide polymorphism, recombination rate and selection in wild tomato species. Genetics 171:753-763.

Rozas J., Sánchez-DelBarrio J.C., et. al. R. 2003. DnaSP, DNA Polymorphism analyses by the coalescent and other methods. Bioinfomatics 19: 2496-24-97.

Stevens, M.A. and C.M. Rick. 1986. Chapter 2. Genetics and breeding, p. 35-109. in: J. Atherton and J. Rudich (eds.). The tomato crop. Chapman and Hall, NY, NY.

Nesbitt, T.C. and S.D. Tanksley. 2002. Comparative sequencing in the genus *Lycopersicon*: Implications for the evolution of fruit size in the domestication of cultivated tomatoes. Genetics. 162: 365-379.

Wu, et al. 2006. Combining bioinformatics and phylogenetics to identify large sets of single copy, orthologous genes (COSII) for comparative, evolutionary and systematic studies: A test case in Euasterid plant species. (in prep).

Table 1. Polymorphism of SNP marker types, 31 tomato DNAs by 48 loci.

Marker type	no. markers	mean no. sequences	no. SNPS	nt	mean π*	mean θ*	mean no. haplotypes
EST-based	26	33	86	9,183	1.66	2.07	2
COSII	11	34	33	7,587	1.20	1.40	3
Arbitrary	11	32	31	5,564	1.20	1.40	3

*nucleotide diversity x 1000

Table 2. Significant HKA tests for 48 tomato loci.

Marker	χ^2	p-value	observation*
LeSNP21	4.003	0.0454	1
COS1	4.348	0.0371	2
COS9	4.506	0.0338	2
COS10	5.356	0.0207	1
B	4.515	0.0336	2

* 1 = *L. esculentum* highly polymorphic, no fixed interspecific differences
2 = *L. esculentum* monomorphic, high interspecific divergence

Fig. 1 Theta versus pi estimates for 48 tomato loci
(significant Fu and Li's tests are indicated)

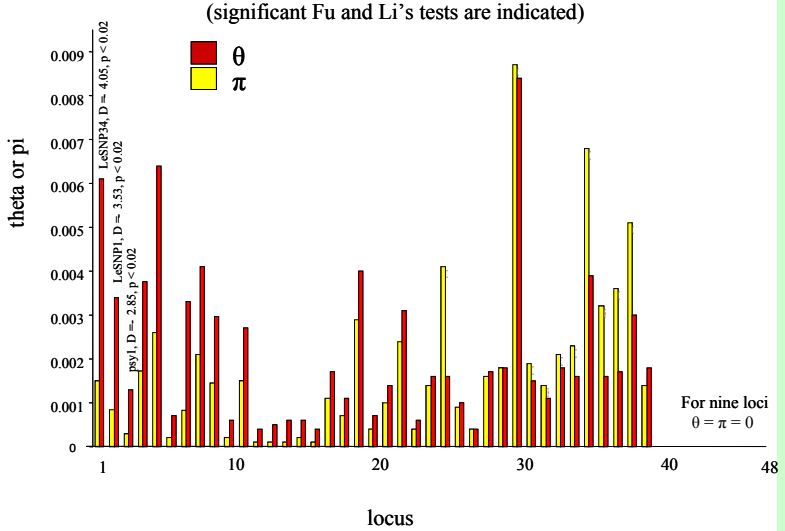
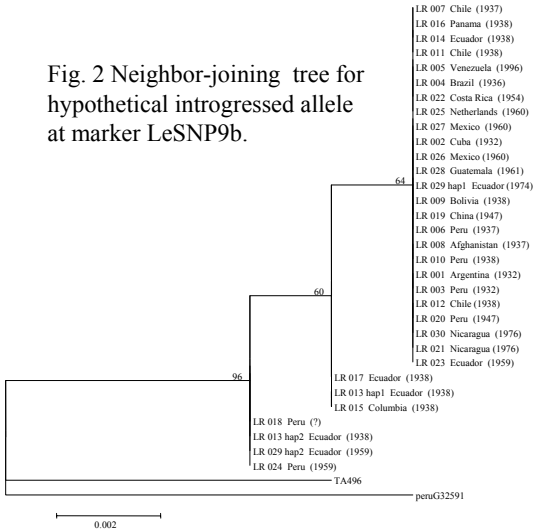


Fig. 2 Neighbor-joining tree for hypothetical introgressed allele at marker LeSNP9b.



Discussion

❖Our panel of 31 *L. esculentum* lines was approximately as diverse as one population of *L. pimpinellifolium* (mean theta = 1.6 x 10^{-3}), the closest wild relative. Nine of the 48 loci in our study were monomorphic, versus 2 of 15 loci in *L. pimpinellifolium* (Roselius et al., 2005).

❖*L. esculentum* polymorphism estimates were about 50% of those observed in another selfing crop, *Sorghum bicolor* (sorghum) (Hamblin et. al, 2004).

❖A neutral model of evolution was rejected for 8 loci. Introgression may explain these results for LeSNP1, COS10, LeSNP34, LeSNP21, and psy1. Directional selection is implicated at COS1, COS9 and B. Enzymes psy1 and B are key regulators of fruit color in tomato.

❖At least 7 loci showed evidence of introgressed alleles from wild species into *L. esculentum*. This can result from linkage drag during breeding, e.g., LeSNP9b maps within 1 cM of a disease resistance allele that was introgressed into TA496 (Labate and Baldo, 2005).

❖Genetic variation within cultivated tomato reflects its complex natural history. Sequences of wild species alleles aid in interpreting evidence of non-neutral evolution.

Materials and Methods

❖DNA from breeding line TA496, one *L. peruvianum*, and 30 geographically diverse *L. esculentum* accessions (inbred lines), was extracted using a CTAB protocol (Colosi and Schaal, 1993). One plant per accession was PCR amplified across 26 EST-based (Labate and Baldo, 2005), 11 Conserved Ortholog Set II (COSII) (Wu et al., 2006), and 11 arbitrary (mostly noncoding regions of genes) markers.

❖Primers used for PCR amplification were also used for DNA sequencing in separate forward and reverse reactions.

❖Sequences were assembled into contigs for each locus and polymorphisms were scored using Phred/Phrap and Consed software (Green, 2004). DnaSP (Rozas J., Sánchez-DelBarrio, 2003) was used for estimating sequence polymorphism and tests of neutrality. Bootstrapped neighbor-joining trees were generated by MEGA (Kumar, 2004) to explore relationships among alleles within a locus.