

cite this poster as:

AM Baldo, LD Robertson, SM Sheffer, WF Lamboy, and JA Labate. 2006. Evaluation of SNP markers across tomato landraces. Plant and Animal Genome XIV, San Diego, CA.

To improve SNP prediction methods from ESTs and to estimate ascertainment bias in EST markers with predicted SNPs we compared four marker types. Polymorphism was estimated across a genetically diverse panel of 30 *S. lycopersicum* (formerly *L. esculentum*) accessions held at our germplasm repository and breeding line TA496. We surveyed 11 carotenoid, 11 COSII, 10 EST markers where a predicted SNP had previously been verified, 11 EST markers where a predicted SNP had not been verified, and 5 EST markers that we hypothesized to be cryptic wild species alleles within *S. lycopersicum*. Improved SNP prediction methods and increased understanding of ascertainment bias will be useful for genetic diversity studies for all crops, especially those that are relatively low in genetic variation.

Evaluation Of SNP Markers Across Tomato Landraces

Angela M Baldo, Larry D Robertson, Susan M Sheffer, Warren F Lamboy, Joanne A Labate

USDA - ARS Plant Genetic Resources Unit, 630 W. North St., Geneva, NY 14456

http://www.ars.usda.gov/main/site_main.htm?modecode=19-07-15-00



PAG-XIV
San Diego, CA
January 2006

Abstract

To improve SNP prediction methods from ESTs and to estimate ascertainment bias in EST markers with predicted SNPs we compared four marker types. Polymorphism was estimated across a genetically diverse panel of 30 *L. esculentum* (syn *S. lycopersicum*) accessions held at our germplasm repository and breeding line TA496. We surveyed 10 EST markers where a SNP had previously been observed, 11 EST markers where no SNP had previously been observed, 11 COSII, and 11 arbitrary markers. EST markers included five that were hypothesized to be cryptic wild species alleles within *L. esculentum*. Improved SNP prediction methods and increased understanding of ascertainment bias will be useful for genetic diversity studies for all crops, especially those that are relatively low in genetic variation.

Results

There is a low amount of genetic diversity in cultivated tomato, necessitating creative methods for developing markers to measure genetic distances among accessions in our germplasm collection. We have resequenced gene fragments (markers) identified by a variety of means (Fig. 1). A set of public ESTs comprised of 3/4 TA496 plus other *esculentum* cultivars/lines was mined for predicted SNPs (Huntley et al. 2005). A number of these were validated in tomato lines TA496, E6203, Rio Grande, and Moneymaker (Labate and Baldo 2005). Wu and Tanksley (unpublished) have generated markers based on conservation of introns between Arabidopsis and tomato, available at SGN (Mueller et al. 2005). Finally, fragments of 11 arbitrary loci, 9 of which have been implicated in carotenoid biosynthesis or ripening, were resequenced.

A set of 30 tomato accessions plus the modern breeding line TA496 were used to quantify the amount of polymorphism in each of these markers (Fig. 2). Population diversity at each marker was estimated with $\theta = S / (\pi \Sigma 1/i)$ where S equals the number of segregating sites in the sample, π equals the size of the sequenced region in nucleotides, and $i = 1, \dots, (n-1)$ with n equal to sample size (Watterson 1975). Levels of θ ranged from 0.00704 to zero, with the highest found in COSII #10.

1. Markers Assayed

Marker	+TA496		-TA496		+TA496		-TA496	
	# SNPs	# indels	-# bp	# reads	theta	theta	theta	theta
EST-based	2488_1	5	1	210	71	0.00596	0.00119	
Possible	2534_1	10	0	600	64	0.00417	0.00083	
Wild Allele	432_2	9	4	690	115	0.00326	0.00000	
	220_1	2	0	165	68	0.00303	0.00000	
	2325_3	5	0	435	67	0.00288	0.00288	
EST-based	2875_4	9	1	590	64	0.00402	0.00402	
SNP	115_1	2	0	140	66	0.00358	0.00358	
Previously	241_2	2	1	305	61	0.00164	0.00164	
Observed	3155_3	5	0	770	68	0.00163	0.00163	
	1297_1	1	0	160	68	0.00156	0.00156	
	298_1	3	1	580	117	0.00129	0.00129	
	1909_2	1	0	200	65	0.00125	0.00125	
	3332_3	1	0	200	69	0.00125	0.00125	
	1260_2	1	0	350	68	0.00072	0.00000	
	3300_2	1	0	510	66	0.00049	0.00049	
EST-based	2819_5	16	3	665	68	0.00602	0.00602	
SNP Not	1724_1	3	0	265	65	0.00283	0.00283	
Previously	2189_1	3	0	270	67	0.00278	0.00278	
Observed	1863_3	5	1	700	64	0.00179	0.00179	
	2719_1	1	0	170	67	0.00147	0.00147	
	4301_3	1	0	420	69	0.00060	0.00060	
	1523_4	0	0	220	65	0.00000	0.00000	
	1589_1	0	0	160	65	0.00000	0.00000	
	1675_1	0	0	170	62	0.00000	0.00000	
	2280_1	0	0	160	66	0.00000	0.00000	
	2592_1	0	0	140	63	0.00000	0.00000	
COS II	COS10	9	2	320	66	0.00704	0.00704	
	COS7	6	1	700	72	0.00215	0.00215	
	COS4	6	1	840	68	0.00179	0.00179	
	COS3	3	0	700	68	0.00107	0.00107	
	COS12	3	2	900	70	0.00083	0.00083	
	COS11	2	1	750	68	0.00067	0.00067	
	COS5	1	0	450	65	0.00056	0.00056	
	COS2	1	0	700	68	0.00036	0.00036	
	COS1	0	0	600	68	0.00000	0.00000	
	COS13	0	2	200	69	0.00000	0.00000	
	COS9	0	1	1180	64	0.00000	0.00000	
Arbitrary	Fw22	8	1	500	68	0.00401	0.00401	
(Carotenoid and other)	Rin	3	1	370	85	0.00203	0.00203	
	TG11	4	0	510	64	0.00196	0.00196	
	CRTISO	3	0	400	63	0.00188	0.00188	
	Hp2 (exon2)	1	0	385	60	0.00065	0.00065	
	PTOX	2	1	800	62	0.00063	0.00063	
	Hp2 (3')	1	1	500	69	0.00050	0.00050	
	Ph5	1	0	560	60	0.00045	0.00045	
	Phy1	0	0	520	61	0.00000	0.00000	
	B	0	0	450	64	0.00000	0.00000	

Discussion

The pedigree of TA496 includes introgression from non-*esculentum* tomato species for disease resistance. A possible limitation of EST mining is ascertainment bias, in which polymorphisms identified may be at higher frequencies in the lines from which the ESTs are generated and not broadly representative.

In our verification of EST-based markers in TA496, E6203, Rio Grande, and Moneymaker, five showed an unusually high level of diversity (2486_1, 2534_1, 432_2, 220_1, 2325_3) that may indicate cryptic wild alleles in TA496 (Labate and Baldo 2005). In the current panel of only *esculentum* accessions, θ drops markedly for four of these markers (2486_1, 2534_1, 432_2, 220_1) if TA496 is removed from the dataset (Fig. 1). Only one other marker (1260_2) shows lower diversity without TA496. These results imply that TA496 alleles contributed disproportionately to diversity of these loci.

The markers tested in this study do include enough information to begin to differentiate landraces in our collection (Fig. 3). Both parsimony and distance methods distinguish a group of accessions from Panama, Ecuador, and Nicaragua. Additional resolution indicates closer relationships, including one between TA496 and PI155372 from Peru.

References

Huntley D, Baldo AM, Johri S, Sergot M. 2005. **SEAN: SNP prediction and display program utilizing EST sequence clusters**. Bioinformatics (doi:10.1093/bioinformatics/btk006)

Labate JA, and Baldo AM. 2005. **Tomato SNP discovery by EST mining and resequencing**. Mol Breed 16:343-349.

Mueller LA et al. 2005. **The SOL Genomics Network. A Comparative Resource for Solanaceae Biology and Beyond**. Plant Phys 138:1310-1317

Watterson G.A. 1975. **On the number of segregating sites in genetical models without recombination**. Theor Pop Biol 7:256-276.

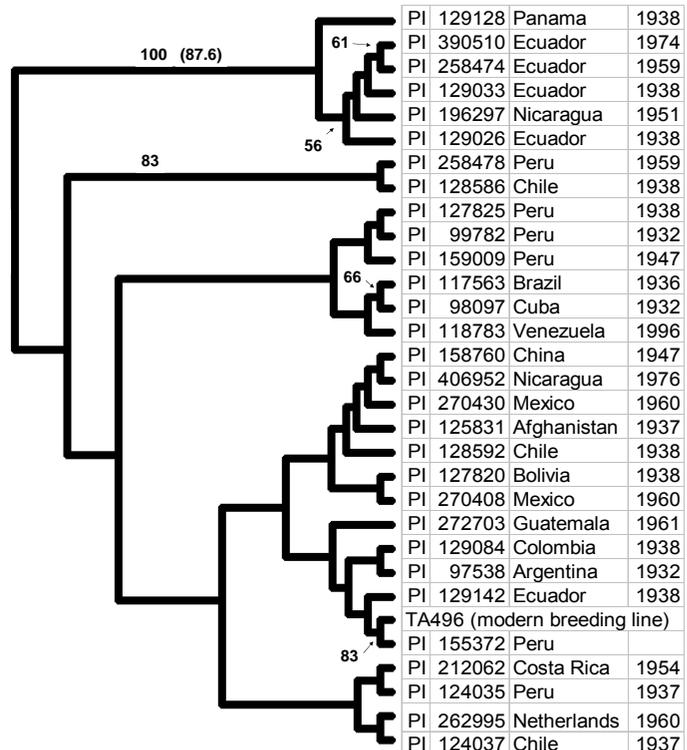
2. Tomato Accessions

Accession	Date deposited	Cultivar / taxon	Country	State	Locality	Narrative
PI 97538	2/29/1932	Cherry	Argentina		West of Tucuman	A small-fruited precocious variety of the ordinary tomato.
PI 88097	3/8/1932	Rinon	Cuba			A sm-med. sized tomato, flattened, oblatly rounded or curved towards the stem and wrinkled. Prof. Jack believes it to be a degenerate form of the improved cultivated tomato which has become a wild plant in Cuba. Can grow under very trying cond.
PI 99782	4/24/1932	Tomate	Peru		Market in Tarma	A small round red local variety, cultivated by the Indians.
PI 117563	7/23/1936	Sao Paulo	Brazil		Seeds collected at Rio de Janeiro	A common and popular red variety resembling a plum, irregularly globose and bilocular.
PI 118783	9/23/1996	Rinon	Venezuela		From Yarilagua, Yaracuy, Sept.23, 1936.	
PI 124035	3/11/1937		Peru		From Lima, Peru.	Fruit small-celled, thick-walled, very mealy, lobed but quite smooth; flesh red.
PI 124037	3/11/1937		Chile		From Valdivia, Chile.	
PI 125831	9/11/1937	Rumi Banjan	Afghanistan		From Khanabad.	Fruit yellow, irregular, flattened; 3 1/3 inches in diameter.
PI 127820	1/20/1938	Cultivar/subtax	Bolivia		From the market at Chulumani.	Fruit red, small, round; grown in this district for many years.
PI 127825	1/12/1938	Coolo-Chuma	Peru		Collected along a river near Pachacuccho.	a perennial vine with large woody stems, growing among shrubery. Used medicinally by the natives.
PI 128586	3/14/1938		Chile		From a garden in Limache.	Fruit large, fleshy, rose colored.
PI 128592	3/16/1938		Chile		From the market at Valparaiso.	Fruit large, smooth, pink.
PI 129026	4/10/1938		Ecuador		From the market at Guayaquil, Ecuador.	Fruit medium, flat, rough, very deep scarlet.
PI 129033	4/17/1938		Ecuador		From the market at Guayaquil, Ecuador.	Fruit small, very rough, flat, many celled.
PI 129084	5/4/1938		Colombia		From the market at Calarca, Columbia.	Fruit currant type, 2-celled, oval.

Accession	Date deposited	Cultivar / taxon	Country	State	Locality	Narrative
PI 129128	5/10/1938		Panama		From the market at Panama, Panama.	Fruit medium size, flat, variable.
PI 129142	4/12/1938		Ecuador		From Pasaje, Ecuador.	Fruit said to be yellow when ripe.
PI 155372			Peru			
PI 158760	4/8/1947	Chih-Mu-Tao-Se	China			
PI 159009	5/15/1947		Peru			
PI 196297	4/11/1951		Nicaragua	El Recreo	From El Recreo, Nicaragua	Apparently immune from wilt and nematodes.
PI 212062	1/5/1954	Turriaba	Costa Rica		From Turriaba, Costa Rica	
PI 258474	1/11/1959		Ecuador	Guayas	From Guayaquil, Guayas	From a market. Typical Guayas type; fruits very large, severely fasciated-distorted; skin yellow.
PI 258478	1/11/1959		Peru	Lambayeque	From Chiclayo, Lambayeque	From a market. Highly variable line. Fruits 3 cm diameter, fasciated, deformed.
PI 262995	1/11/1960	Alisa Craig	Netherlands	Noord-Holland	From Sluis en Groot, Enkhuzen	Greenhouse variety. Greenback 80 days, indet., 2 x 2 inches, above average yield, does very well in drought and heat.
PI 270408	22266	cerasiforme	Mexico	Puebla	From Huachirango, Puebla	Fruit 2-4 loculed, weighs to 30 grams; skin colorless, rough; flesh red.
PI 270430	12/16/196	cerasiforme	Mexico	Veracruz	From Coatzacoalcos, Veracruz	From at sea level. Fruit 8-9 loculed, weighs to 90 grams; skin yellow, rough; flesh red.
PI 272703	3/15/1961		Guatemala	Huehuetenango	From Huehuetenango, Guatemala	Fruit weight 68 grams.
PI 390510	1/11/1974	cerasiforme	Ecuador		Collected from the Instituto Nacional de Investigaciones Agropecuarias Experiment Station, Bolloche, Guayas	Fruit globose, 2.5 cm diameter, red. Wild seed
PI 406952	1/11/1976	Chicarojo	Nicaragua		From a public market in Managua.	Fruit somewhat lobed, red, to 10 cm in diameter.

3. Accession Relationships based on 48 SNP markers

Distance tree generated using the F84 matrix and Fitch algorithm in Phylip. Bootstrap values >50% are indicated on branches (100 replicates, each jumbled 3 times). Parsimony bootstrap values > 50% are indicated in parenthesis.



Acknowledgements

The authors gratefully thank Steve Tanksley and Feinan Wu for sharing their COSII primer sequences, David Spooner, David Francis, and Ed Buckler for stimulating discussion. We thank Paul Kisly for assistance with growing the plants.