

**Tomato Germplasm Conservation at the
Plant Genetic Resources Unit – Geneva, NY**

May 10, 2006

Currently, there are 5964 accessions of 10 *Lycopersicon* species conserved at the Northeast Regional Plant Introduction Station located at the PGRU, Geneva, New York (Table 1). Approximately 11% of these are maintained with local G numbers, most of which are *L. esculentum* accessions. We are in the process of reviewing the G number accessions to determine which should have PI numbers, which should be maintained as G numbers (and for how long), and which to discard. Additionally, we have started the process of transferring unique tomato accessions housed at the National Center for Genetic Resources Preservation (NCGRP) located at Ft. Collins, Colorado to PGRU in Geneva, NY.

Table 1. Tomato germplasm conserved at the Plant Genetic Resources Unit located at Geneva, NY

Species	Accessions	Backup
<i>cheesmaniae</i>	12	11
<i>chilense</i>	1	1
<i>chmielewskii</i>	1	1
<i>esculentum</i>	5330	4769
<i>glandulosum</i>	12	11
<i>hirsutum</i>	60	40
<i>parviflorum</i>	6	6
<i>pennellii</i>	10	4
<i>peruvianum</i>	128	101
<i>pimpinellifolium</i>	231	228
sp.	15	3
Hybrids	158	158
TOTAL	5964	5332

Characterization for the minimal descriptor list and digital imaging of the germplasm accessions has been incorporated as part of the regeneration process. We now have 174 images of fruit and 227 leaf images available on GRIN.

All cultivated tomato accessions are now available for distribution and we are in the process of backing up remaining accessions that are not backed up at NCGRP. Seed is being pulled for these for backup and necessary regenerations are being completed where necessary. Additionally, those backups that do not meet seed number and germination standards are being replaced by samples that meet the requirements. As mentioned previously, those that have sufficient seed were transferred to PGRU and assigned PI accession numbers. The file with duplicate tomato accessions was reviewed and the most representative accession was identified. The other accessions will be inactivated as duplicates. Once this is completed, the remaining NCGRP tomato accessions will be transferred to Geneva and assigned PI accession numbers and the majority of the G numbered accessions will be assigned PI

numbers.

Distribution of the tomato germplasm collection at PGRU for the past five years is given in Tables 2 and 3 (accessions and samples). There has been a significant increase in distribution of *Lycopersicon* the past three years. This year there was a major distribution to a core collection project in the framework of the SOL project. This project was started to develop and explore a cultivated tomato core-collection that includes a few thousand varieties representing traditional and modern types from around the world, including Cerasiforme from Mexico and some wild species. The objective of this project is through the union of association analysis with genetic mapping of simple and complex traits to create a common platform for linking natural tomato phenotypes with the underlying DNA polymorphism. This core collection and the accompanying database will be freely available to the SOL community for research and gene discovery. The majority of the *Lycopersicon* accessions distributed have been *Lycopersicon esculentum*. Additionally, 93 DNA samples taken from the germplasm collection were distributed in the past year.

Table 2. Distribution of tomato germplasm accessions at PGRU for 2001 through 2006*

Genus/Species	2001	2002	2003	2004	2005	2006	Total
<i>cheesmaniae</i>	8	6	3	8	5	0	30
<i>chilense</i>	1	1	1	1	1	0	5
<i>chmielewskii</i>	1	1	1	1	1	0	5
<i>esculentum</i>	404	774	246	632	659	2889	5604
<i>glandulosum</i>	6	6	4	4	8	0	28
<i>hirsutum</i>	17	35	20	21	42	4	139
<i>hybrid</i>	9	66	1	45	29	13	163
<i>parviflorum</i>	4	6	3	2	5	0	20
<i>pennellii</i>	2	5	3	3	3	1	17
<i>peruvianum</i>	44	58	117	41	77	91	428
<i>pimpinellifolium</i>	33	223	41	17	65	3	382
<i>sp.</i>	1	2	1	2	3	6	15
Total	530	1183	441	777	898	3007	6836

Table 3. Distribution of tomato germplasm samples at PGRU for 2001 through 2006*

Genus/Species	2001	2002	2003	2004	2005	2006	Total
<i>cheesmaniae</i>	11	6	3	19	7	0	46
<i>chilense</i>	2	1	2	6	2	0	13
<i>chmielewskii</i>	4	1	1	5	1	0	12
<i>esculentum</i>	486	887	258	1025	886	3204	6746
<i>glandulosum</i>	7	10	4	4	13	0	38
<i>hirsutum</i>	30	51	28	29	97	7	242
<i>hybrid</i>	9	80	1	50	36	16	192
<i>parviflorum</i>	5	8	3	3	5	0	24
<i>pennellii</i>	3	16	5	7	4	2	37
<i>peruvianum</i>	59	66	142	60	137	91	555
<i>pimpinellifolium</i>	39	315	48	17	70	5	494
<i>sp.</i>	1	2	1	2	3	6	15
Total	656	1443	496	1227	1261	3331	8414

*2006 distributions are through April 15, 2006.

Single nucleotide polymorphisms (SNPs) are rarely applied to germplasm management because DNA sequence data that is necessary for developing SNP markers has been lacking for many plant species. This situation is rapidly changing. According to NCBI, there are at least 6 large-scale genome and 102 expressed sequence tag (EST) sequencing projects in plants as of April 2006, the majority of which are crop species. We anticipate that SNPs will be increasingly applied to crop germplasm conservation and management.

A possible limitation of EST mining is ascertainment bias, in which discovered polymorphisms are at higher frequencies in the lines from which the ESTs were generated and not broadly representative. 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* accessions held at our germplasm repository and breeding line TA496. We surveyed 15 EST markers where a predicted SNP had previously been verified, 11 EST markers where a predicted SNP had not been verified, 11 arbitrary and 11 Conserved Ortholog Set II (COSII) markers (Table 4). Five of the EST markers were hypothesized to be wild species alleles within *L. esculentum* originating from linkage drag. Despite the fact that 75% of public tomato ESTs originate from line TA496, we found a majority of SNPs (73%) and indels (64%) in the 26 EST-based markers were represented among the other 30 tomato lines in our panel. This can be compared to SNPs (94 to 100%) and indels (92 to 100%) for the COSII and arbitrary markers, respectively. Yang et al. (2004) also found the majority of SNPs (53%) they EST-mined between TA496 and Rio Grande were present in 19 additional tomato varieties. TA496 alleles contributed disproportionately to diversity of 4 of the 5 hypothesized wild species alleles (Table 4). Increased understanding of ascertainment bias and linkage drag will be useful for genetic diversity studies for all crops, especially those such as tomato that are relatively low in genetic variation.

Table 4. Comparison of SNP marker types, 31 tomato DNAs by 48 loci

Marker type	no. markers	with TA496			without TA496			bp
		no. SNPs	theta (SNPs/kb)	no. indels	no. SNPs	theta (SNPs/kb)	no. indels	
EST-based, confirmed ^a , hypothetical introgression	5	31	3.67	6	8	0.95	1	2100
EST-based, confirmed ^a	10	26	1.71	4	26	1.72	4	3775
EST-based, unconfirmed	11	29	2.16	4	29	2.17	4	3340
(all EST-based)	26	86	2.32	14	63	1.71	9	9215
COSII	11	33	1.11	12	31	1.05	11	7370
arbitrary ^b	11	31	1.31	6	31	1.32	6	5890
total	48	150	1.66	32	125	1.39	26	22475

^a SNP previously confirmed using 2 or 3 lines (TA496, Rio Grande, Moneymaker, TA209)

^b includes genes related to fruit color, weight, and one anonymous marker

total, proportion w/o TA496	SNPs: 83.33%	indels: 81.25%
EST-based, proportion w/o TA496	SNPs: 73.26%	indels: 64.29%
COS II, proportion w/o TA496	SNPs: 93.94%	indels: 91.67%
arbitrary, proportion w/o TA496	SNPs: 100.00%	indels: 100.00%