LETTER

Sequencing of cDNA Clones from the Genetic Map of Tomato (*Lycopersicon esculentum*)

Martin W. Ganai,¹ Rosemarie Czihal, Ulrich Hannappel,
Dorothee-U. Kloos, Andreas Polley, and Hong-Qing Ling

Institute for Plant Genetics and Crop Plant Research, D-06466 Gatersleben, Germany

The dense RFLP linkage map of tomato (*Lycopersicon esculentum*) contains >300 anonymous cDNA clones. Of those clones, 272 were partially or completely sequenced. The sequences were compared at the DNA and protein level to known genes in databases. For 57% of the clones, a significant match to previously described genes was found. The information will permit the conversion of those markers to STS markers and allow their use in PCR-based mapping experiments. Furthermore, it will facilitate the comparative mapping of genes across distantly related plant species by direct comparison of DNA sequences and map positions.

[cDNA sequence data reported in this paper have been submitted to the EMBL database under accession nos. AA824695–AA825005 and the dbEST Id database under accession nos. 1546519–1546862.]

Molecular markers used in mapping experiments should ideally have a known function. This requirement is fulfilled best by the use of known genes in the form of genomic clones or cDNA clones. Many genetic maps based on molecular markers, such as RFLPs (restriction fragment length polymorphisms) include a number of known genes. However, the number of known genes available for a given organism is usually very limited. To circumvent this problem, many maps contain a large number of randomly selected cDNA clones. cDNAs have the additional advantage that they are frequently single or low copy and thus ideal for genetic mapping (Bernatzky and Tanksley 1986). Furthermore, they show a higher level of conservation between species than genomic clones (Zamir and Tanksley 1988) and allow more efficient cross-mapping in related species for the determination of synteny (Ahn and Tanksley 1993).

In recent years random cDNA clones have attracted considerable interest, as the sequencing of such clones provides a means to obtain a fast catalog of expressed genes from a given organism without sequencing the entire genome. In plants, major efforts are now under way to sequence random cDNAs in *Arabidopsis* and rice (Sasaki et al. 1994; Delseny et al. 1997). Many thousands of such fragments have been deposited in the respective databases. For ~30%–40% of the cDNA sequences a possible function can be deduced based on homologies

to known genes from bacteria, animals, or plants (Delseny et al. 1997).

For all expressed genes, not only the DNA sequence and the possible function should be known but also their map position on the chromosomes. In the long term this will allow merging of the classical genetic map based on mutations with potential candidate genes for these mutants. A common repertoire of mapped cDNA clones will, in the future, enable us to study synteny even between distantly related species (Paterson et al. 1996) for which studies by cross-hybridization are very difficult. The availability of sequence information for mapped cDNA clones should make conclusions from hybridization studies more firm and testable.

Tomato (*Lycopersicon esculentum*) has one of the most densely populated genetic maps among plants. Currently, >1000 RFLP markers have been published by Tanksley et al. (1992), of which >300 are random cDNA clones. An additional 500 RFLP markers have been localized in reference to this map through the fact that all markers from potato can be readily transferred to tomato on the basis of extensive synteny (Gebhardt et al. 1991; Jacobs et al. 1995). We report here the results of the sequencing of ~90% of the mapped cDNA clones from the high-density tomato map.

RESULTS

cDNA Sequencing and Analysis

A total of >300 cDNA clones have been mapped

¹Corresponding author.
E-MAIL ganai@ipk-gatersleben.de; FAX 49-39482-5137.

onto the high-density map of tomato by Tanksley et al. (1992). These clones were derived from two different libraries. CD clones were generated from leaf tissue and CT clones from epidermal tissue. The clones were in four different vectors. For sequencing, only those clones were used that were generated in vectors with M13 or pBluescript primer sites. This excluded some of the early cDNA clones (Bernatzky and Tanksley 1986) because they were generated in pBR322. A total of 272 cDNA clones could be sequenced in this way from at least one direction. Sequences from another 15 clones were not included in the analysis because they contained obvious cloning artifacts that could not be resolved. From both sides 145 clones were sequenced, and for 73 clones the entire insert was sequenced. The largest clone completely sequenced was 896 bp.

Because these clones had already been placed onto the genetic map of tomato, it was anticipated that most should represent different genes. Nevertheless, for some cDNA clones duplicates could be identified (clones CT71 and CT210, CT72 and CT242, CT75 and CT166, CT88 and CT257, CT115 and CT218, CT154 and CT214, CT223 and CD61, CD18 and CD34). In these cases, the cDNA clones show a complex hybridization pattern upon which it was not possible to state previously that they are derived from the same gene.

Sequence Homologies

For 156 of the 275 analyzed clones (57%), significant nucleic acid and/or protein homologies were found with the respective databases. Of those, 125 clones showed matches with known DNA sequences and 141 clones showed matches with protein sequences with a BLAST score of $<10^{-10}$. Some sequenced clones showed only homology to other plant sequences at the DNA level and not at the protein level because the cDNA clones were not integrated directionally into the cloning vector and for these clones sequence information was obtained only from the untranslated 3' region with the poly(A) tail. If matches on the DNA and protein level were found, only those were considered significant that matched to the same protein type or corresponding genes. Table 1 shows a summary of the data for the clones with significant homologies. For 25 tomato genes already described we found a direct match with the database.

DISCUSSION

The availability of 272 sequenced cDNA clones from

the tomato map, together with other previously mapped genes of known function (Pillen et al. 1996b), creates a framework of >350 markers for the tomato genome for which at least part of the DNA sequence is known. This information will be sufficient for the generation of PCR-based markers for most regions of the tomato genome. Such sequence-tagged sites (STSs) will function as genetic anchor markers (Inoue et al. 1994) and permit fast and high throughput analysis of loci in large populations (Schumacher et al. 1995). In plant breeding and genetic experiments, STS markers, for example, in the form of CAPs (cleaved amplified polymorphisms), are very useful to follow linked genes in an economical manner through generations (Konieczny and Ausubel 1993). Such markers are especially useful for preselecting recombinants in specific regions of the tomato genome for the purpose of high-resolution mapping of genes targeted for map-based cloning (Alpert and Tanksley 1996). At the same time, they provide starting points for the rapid isolation of large insert clones from tomato DNA libraries, such as yeast artificial chromosomes (YACs) (Martin et al. 1992; Bonnema et al. 1996; Pillen et al. 1996a) or bacterial artificial chromosomes (BACs) (Shizuya et al. 1992) and binary BAC (BiBAC) clones (Hamilton et al. 1996).

Considerable effort is currently spent on the comparative mapping of highly conserved cDNA clones across a wide range of plant taxa. This has revealed that in limited regions, synteny exists even between distantly related plant species (Paterson et al. 1996). Sequence data from mapped cDNA clones will eventually help to reveal synteny between such plant species. For example, using the information from these mapped tomato clones, it will be possible to identify genes or expressed sequence tags (ESTs) with very high sequence homology on the DNA and/or protein level to the model plant organism *Arabidopsis thaliana*. Such probes are very likely cross-hybridizing between *Arabidopsis* and tomato. If they are single copy in hybridization on both genomes they can be mapped comparatively in these species and it can be determined whether there is conservation of linkage in certain areas of their genomes. Similarly, such experiments can be expanded to additional plant species for which large numbers of ESTs will be generated in the future. A carefully chosen approach involving sequence comparisons in combination with genetic mapping by hybridization is absolutely necessary for the study of synteny between distantly related plant species to discriminate between orthologous and paralogous genes (Tatusov et al. 1997). Large numbers of cDNA

Table 1. Nucleic Acid and Protein Homologies of Mapped cDNA Clones from Tomato

Number	Homology DNA	Homology protein	Match to organism
CT			
1 RBCS 3a	2.7x10E-33 LERBCS3A	1.5x10E-7 A05118	*Tomato
8 seven in absentia isolog	8.7x10E-6 ATU90439	4.6x10E-10 U90439	Arabidopsis
11 lipid transfer protein	1.4x10E-46 TOBLTP	8.9x10E-49 Q03461	Tobacco
17 chloroplast protein IEP110	2.7x10E-44 PSIEP110	-	Pea
19 aconitase	1.3x10E-49 QUCACO	4.2x10E-42 P49608	Cucurbita
21 mandelonitrile lyase	5.8x10E-19 PAMDL1	7.4x10E-31 P52706	Prunus
25 glycolate oxidase	1.9x10E-45 MCU80071	8.7x10E-34 U80071	Mesembryanthemum
27 alpha-tubulin	4.2x10E-17 ZMALPTUB	-	Maize
28 osmotin-like protein	7.0x10E-222 TOMNP24	2.5x10E139 P12670	*Tomato
29 phospholipase D	3.9x10E-31 NTZ84822	2.4x10E-28 Z84822	Tobacco
30 polygalacturonase inhibitor	1.9x10E-81 TOMPIP	8.2x10E-44 S47965	*Tomato
31 shaggy-like kinase	8.8x10E-72 ATASKRNA	2.5x10E-55 S71269	Arabidopsis
52 calmodulin binding protein	2.4x10E-142 NTU58971	1.0x10E-39 U58971	Tobacco
54 thiazole monophosphate biosynthesis protein	-	1.3x10E-24 U34923	E. coli
55 ADP/ATP translocator	2.3x10E-191 LEU89839	2.1x10E-50 U89839	*Tomato
59 arabinopolysaccharide protein	9.9x10E-124 LEAGP	1.0x10E-7 S55925	*Tomato
61 patched protein homolog	-	5.1x10E-16 U46155	Mouse
63 golgi-associated protein	5.4x10E-152 PSU31565	1.7x10E-130 U89897	Pea
66 arginine decarboxylase	1.2x10E-72 TOMARGDECA	9.4x10E-42 U35367	Tomato
68 no definition line found protein	-	2.9x10E-10 AF013294	Arabidopsis
70 endo beta 1-4 glucanase	2.2x10E-192 LEU78526	6.0x10E-106 U78526	*Tomato
71 glycine rich cell wall protein	8.4x10E-168 LECEWAPR	8.5x10E-20 X77373	*Tomato
72 protein kinase	-	2.7x10E-22 S42868	Spinach
75 ferredoxin reductase	1.8x10E-157 NTY14032	7.8x10E-93 Y14032	Tobacco
77 lysophospholipase	-	4.7x10E-14 U95973	Arabidopsis
79 profilin	8.6x10E-74 PVPROFILIN	2.0x10E-48 S35798	Phaseolus
80 serinehydroxymethyltransferase	2.7x10E-75 ATFCAD	5.0x10E-72 Z97335	Arabidopsis
82 DNA binding protein	3.9x10E-49 TOBRY4D	3.3x10E-33 U89255	Tobacco/Tomato
84 tryptophan decarboxylase	1.4x10E-126 CAU73657	1.0x10E-126 U73657	Camptotheca
90 S3A ribosomal protein	2.1x10E-44 RICT151	9.5x10E-48 P49397	Rice
91 S13 ribosomal protein	4.4x10E-34 PSRBPS13A	5.5x10E-26 P49203	Pea
93 adenosylmethionine decarboxylase	2.2x10E-44 STISTPB	-	Potato
100 xyloglucan endotransglycosylase	2.1x10E-22 LETXETB1	1.1x10E-30 X93175	Tomato
103 glutaredoxin-like protein	-	1.7x10E-29 L46863	Legionella
106 alpha tubulin	7.6x10E-128 HVU40042	2.2x10E-92 P14641	Barley
107 L12 ribosomal protein	4.7x10E-114 AB005246	4.0x10E-80 P50884	Arabidopsis/Clamydomonas
108 uridine kinase	5.4x10E-39 AC002304	4.3x10E-48 Z69635	Arabidopsis/Caenorhabditis
110 S18 ribosomal protein	1.0x10E-17 ATS18RP1	1.9x10E-15 Z17755	Arabidopsis
113 ubiquitin	9.0x10E-170 AMUBIQMRP	5.1x10E-117 S34285	Snapdragon
114 phosphoglycerat kinase	1.7x10E-191 NTPGKCYT	1.5x10E-119 Z48976	Tobacco
116 proteasome subunit P40	2.2x10E-82 ATU54560	1.3x10E-83 U54560	Arabidopsis
118 EREBP-4	1.6x10E-11 ATFCAB	4.2x10E-27 D38125	Arabidopsis
120 mitogen activated protein kinase	2.0x10E-123 NTP43NTF6	3.6x10E-40 Q40531	Tobacco
122 EIF-5A	7.5x10E-82 AB004825	2.8x10E-40 P24922	Potato
125 glycolate oxidase	1.9x10E-45 MCU80071	8.7x10E-34 U80071	Mesembryanthemum
126 vacuolar H ⁺ -ATPase	2.8x10E-127 AF010228	3.8x10E-69 Q40585	*Tomato
127 calcium dependent protein kinase	5.8x10E-72 VRU08140	1.1x10E-55 U08140	Vigna
128 CAB 1B	2.9x10E-259 TOMCPBA	1.5x10E-133 P07370	*Tomato
130 CAB 6A	1.6x10E-144 TOMPSI	3.7x10E-87 S00443	*Tomato
131 subtilisin like protein	2.7x10E-79 AGAG12	4.4x10E-65 S52769	Alnus
132 alpha tubulin 2	9.2x10E-55 HVU40042	2.0x10E-41 U40042	Barley
136 calcium binding pollen allergen	8.3x10E-12 SORBPY	2.6x10E-18 S54819	Spinach
137 propionyl CoA carboxylase	-	4.3x10E-68 Y07660	Mycobacterium
138 ankyrin homolog	8.6x10E-26 ATFCAB	7.4x10E-45 Z97336	Arabidopsis
141 remorin	2.0x10E-141 STU72489	2.0x10E-35 U72489	Potato
143 lamin receptor sterol reductase	1.7x10E-7 YSPC14SR	5.1x10E-26 A53616	Human/Yeast
146 salt tolerance protein	-	3.9x10E-13 X95572	Arabidopsis
148 superoxide dismutase	-	4.9x10E-13 S50633	Human
149 hypothetical yeast protein	2.2x10E-16 IG005110	1.1x10E-15 AC000103	Arabidopsis
150 secreted glycoprotein	6.5x10E-14 ITU51740	1.5x10E-19 U51740	Arabidopsis/Ipomea
155 UFD1-protein	-	2.1x10E-23 Z97338	Arabidopsis
156 S6 ribosomal protein	8.0x10E-41 NTRPS6	-	Tobacco
161 14-3-3 ribosyltransferase activator	1.2x10E-51 S76737	3.5x10E-34 S76737	Tobacco
166 ferredoxin reductase	7.3x10E-125 NTY14032	1.9x10E-64 Y14032	Tobacco
168 thioredoxin II	-	4.2x10E-11 P22217	Yeast

Table 1. (Continued)

Number	Homology DNA	Homology protein	Match to organism
171	TER ATPase	1.6x10E-18 CELME1SP	1.9x10E-53 AC000375 <i>Caenorhabditis/Arabidopsis</i>
172	SF3 transcription factor	1.7x10E-10 ATRNAATL2	6.8x10E-9 AF002109 <i>Arabidopsis</i>
173	pre-pro-cystein proteinase	3.1x10E-127 LECYSPRO	4.4x10E-25 S24988 *Tomato
174	hypothetical Arabidopsis protein	-	2.0x10E-28 AC000348 <i>Arabidopsis</i>
176	transporter protein	-	8.2x10E-12 U49082 Human
178	abscisic acid stress protein	5.7x10E-23 SCU12439	- Potato
179	aquaporin	1.7x10E-60 HAAP	4.7x10E-50 U62778 Cotton/Sunflower
182	UTP-glucose pyrophosphorylase	3.0x10E-80 POTUTPGP	1.8x10E-15 P19595 Potato
183	TER ATPase, MEI1-protein	7.1x10E-33 AC000375	8.2x10E-30 P39955 <i>Arabidopsis/Yeast</i>
184	clathrin coat assembly protein	4.3x10E-74 AF0009631	5.5x10E-64 AF0009631 <i>Arabidopsis</i>
187	GDP dissociation protein	2.1x10E-67 ATRABGD1	2.0x10E-63 D83531 <i>Arabidopsis</i>
188	annexin	2.7x10E-123 CAANN	1.4x10E-76 S66274 Capsicum
189	S19 ribosomal protein	1.4x10E-52 RICYK44	2.8x10E-62 P40978 Rice
193	RNA polymerase II	-	1.7x10E-11 U85510 Human
197	oleate desaturase	4.6x10E-158 SCD120LDS	5.0x10E-97 X92847 Potato
198	hypothetical Arabidopsis protein	3.2x10E-20 ATFCA1	8.3x10E-28 Z97336 <i>Arabidopsis</i>
199	preproplastocyanine	1.2x10E-75 LEPLACY	9.1x10E-5 P17346 *Tomato
201	NADP malic enzyme	2.2x10E-186 AF001270	1.7x10E-103 AF001270 *Tomato
202	CaMB channel protein	1.5x10E-35 NTU65390	1.5x10E-40 U65390 Tobacco
204	ascorbate oxidase	9.6x10E-17 D45066	1.8x10E-41 P45066 Cucurbita
206	protein kinase	-	6.1x10E-16 Z46636 <i>Phycomyces</i>
208	class III- ADH	2.3x10E-56 ZMFDH	1.8x10E-34 S71244 <i>Arabidopsis/Maize</i>
210	glycin rich cell wall protein	4.2x10E-98 LECEWAPR	3.6x10E-16 X77373 *Tomato
211	proline rich protein	1.1x10E-29 GMADR11A	1.7x10E-32 S33621 Soybean
212	MER1 5 gene	5.9x10E-9 ATMER15DA	8.2x10E-12 X82683 <i>Arabidopsis</i>
214	MIF-protein	-	1.5x10E-45 F14405 <i>Arabidopsis</i>
215	receptor kinase	1.4x10E-41 ATLEOGENE	9.3x10E-66 X95909 <i>Arabidopsis</i>
217	calcium binding protein	7.6x10E-81 POTCBP	2.5x10E-46 Q09011 Potato
220	rab11 ras-related protein	1.6x10E-44 TOBNTAE	4.4x10E-13 L29272 Tobacco
223	PSII 10kd protein	3.7x10E-193 LEPSIIPRE	2.4x10E-64 X95987 *Tomato
224	alpha amylase	2.4x10E-101 A21343	7.3x10E-60 A21434 Potato
225	nucleotide diphosphate kinase	4.2x10E-58 BTNBRA	1.3x10E-88 S60363 Spinach
226	cytochrom P450	5.2x10E-29 CTRP450B	4.2x10E-25 L19074 <i>Catharanthus</i>
228	vacuolar processing enzyme	-	5.0x10E-17 P49047 <i>Arabidopsis</i>
229	SAM2 dimethylmenaquinone methyltransferase	1.2x10E-37 AB005242	8.1x10E-14 Z83864 <i>Arabidopsis/Mycobacterium</i>
232	RAD23-protein	4.3x10E-22 DCRAD23I	7.7x10E-16 Y12013 Carrot
234	auxine responsive product	-	2.7x10E-21 S17433 Soybean
239	pectinesterase	1.9x10E-190 LEPE	2.8x10E-126 Z94058 *Tomato
240	L23A ribosomal protein	1.1x10E-119 TOBRPL25A	2.4x10E-52 Q07761 Tobacco
241	lipase	-	2.7x10E-13 S68410 <i>Arabidopsis</i>
242	protein kinase	-	5.2x10E-22 S42868 Spinach
243	calmodulin	9.2x10E-18 CICAM	4.9x10E-22 AC000103 Ciona/ <i>Arabidopsis</i>
245	prolin rich protein	2.4x10E-49 CR14KDPP	1.5x10E-29 L22305 <i>Catharanthus/Medicago</i>
247	CKC-protein	1.4x10E-12 ATU41339	1.8x10E-12 U44028 <i>Arabidopsis</i>
248	ribulose 3phosphate epimerase	6.0x10E-65 STPPEPIMR	4.6x10E-35 S58081 Potato
250	recessive suppressor of secretory effect	6.6x10E-11 CEC3487	1.7x10E-42 P42837 <i>Caenorhabditis/Yeast</i>
252	nucleoside diphosphate kinase	9.2x10E-30 SOU55019	1.3x10E-56 S60363 Saccharum/Spinach
253	expansin	5.4x10E-42 CSU30382	4.8x10E-69 U30460 Cucumis
255	Arabidopsis protein	3.0x10E-11 AC0002330	- <i>Arabidopsis</i>
259	apospory-associated protein	-	3.5x10E-42 U13149 Pennisetum
260	cell wall protein	2.5x10E-94 LECWAPR	- *Tomato
264	CUTA protein	-	3.1x10E-44 U78721 <i>Arabidopsis</i>
265	acetylornithine aminotransferase	2.8x10E-88 AGAG118	3.3x10E-75 Y08680 Alnus
267	xyloglucan endotransferase	6.8x10E-122 TOMEXT3	2.1x10E-103 D49539 *Tomato
268	leucine rich receptor kinase	1.2x10E-38 OSTMK	5.7x10E-47 Y07748 Rice
269	DNAJ-isolog	4.4x10E-36 CSCSDJ	4.6x10E-24 X94301 Cucumis/Potato
274	cytochrome c1	2.0x10E-32 STCYC1	- Potato
276	ubiquitin conjugating enzyme	-	7.0x10E-24 U23517 <i>Caenorhabditis</i>
277	pistil extensin	7.3x10E-8 NAPR44	2.1x10E-12 Q03211 Tobacco
278	tripeptidyl peptidase	-	6.1x10E-32 M55169 Human
280	glutathion peroxidase	8.4x10E-31 ATU94495	4.9x10E-42 U94495 <i>Arabidopsis</i>
283	ascorbate reductase	1.6x10E-95 TOMAFRR	6.6x10E-39 L41345 *Tomato
285	PSI subunit II	1.5x10E-166 TOMPSAD	4.9x10E-75 P12372 *Tomato
286	TFMP1-protein	6.6x10E-12 TOBTFHP1	1.1x10E-40 D29976 Tobacco
287	polyphenoloxidase	7.1x10E-88 LEPOOXE	9.2x10E-9 Q08296 *Tomato

GANAL ET AL.

Table 1. (Continued)

Number	Homology DNA	Homology protein	Match to organism
CD			
2 serine-glyoxylate aminotransferase	5.1x10E-13 D86125	1.9x10E-9 D86125	Hyphomicrobium
3 chloroplast mRNA binding protein	1.5x10E-34 SOU49442	7.9x10E-30 U49442	Spinach
5 fructose 1,6 biphosphatase	3.2x10E-25 BNAFBPCPA	1.6x10E-25 P22418	Spinach/Brassica
6 cellulose synthase	9.2x10E-30 AC002304	1.2x10E-28 U58284	Arabidopsis
15 cyclophilin	1.3x10E-45 ATHRCC4ARA	-	Arabidopsis
16 L3 ribosomal protein	1.2x10E-54 ATHRPCA	8.9x10E-36 P17094	Arabidopsis
18 L21 ribosomal protein	3.6x10E-41 AC000132	2.9x10E-36 F20014	Arabidopsis
19 Rieske FeS pecursor	2.0x10E-55 NTRFES2	2.3x10E-24 Q02585	Tobacco
25 adenylate kinase	-	5.6x10E-33 P43188	Maize
29 BIP binding protein, glucose regulated protein	3.6x10E-101 TOMBIPGRBC	1.9x10E-54 L08829	*Tomato
31 starch phosphorylase	1.3x10E-211 STSTP1	5.9x10E-128 P53535	Potato
34 L21 ribosomal protein	1.8x10E-53 AC000132	5.5x10E-45 AC000132	Arabidopsis
35 hypothetical yeast protein	-	2.0x10E-15 P40482	Yeast
38 transketolase	3.1x10E-75 STTKETMR	3.5x10E-29 S54300	Craterostigma
52 translationally controlled tumor protein	2.0x10E-79 SHP23H	1.8x10E-46 P35681	Potato/Rice
54 proton ATPase	1.5x10E-89 STDPHA2	1.6x10E-46 S50752	Potato
56 CAB 8	3.4x10E-88 LECAB8	2.1x10E-38 P27522	*Tomato
59 hypothetical Arabidopsis protein	-	9.3x10E-20 AC002337	Arabidopsis
61 PSII 10 KD precursor	7.2x10E-129 LEPSIIPRE	2.2x10E-70 X95987	*Tomato
64 dihydrolipoamide acetyltransferase	1.7x10E-53 D90915	2.5x10E-37 D90915	Synechocystis
65 PSI subunitV	4.2x10E-36 SOPSIA	1.2x10E-33 P12357	Spinach
74 suppressor like protein	4.9x10E-48 ATERF13	2.3x10E-47 S31445	Arabidopsis
75 global transcription factor	1.3x10E-32 SCYBR245C	3.2x10E-47 Z36114	Yeast
77 nucleolin	2.6x10E-26 AF013294	2.1x10E-53 AF013294	Arabidopsis
78 CAB 1b	1.0x10E-74 TOMCBPA	5.4x10E-37 U21115	*Tomato
79 glycolate oxidase	1.5x10E-126 LEGLYCOX	2.4x10E-70 X92888	*Tomato

Those cDNA clones listed show a homology score of $<10^{-10}$ at least on either the DNA or protein level. Homology scores of $>10^{-10}$ were included when the lower match agreed with the higher score ($>10^{-10}$) on the DNA or protein level, respectively. For homology searches on the DNA level, the GenBank DNA database was used. For homology searches on the protein level, the GenBank protein database was used. Accession numbers are listed in the corresponding columns for the respective databases. If the same match was found on the DNA and protein level, only a single organism is listed in the match to organism column. In some cases, two organisms are indicated—the most significant matches are to different organisms but the same protein type on the DNA and protein level, respectively. Clones that match directly to known tomato genes have been indicated by an asterisk (*).

clones are currently sequenced and mapped in some plant genomes. For rice, EST sequences were used for the construction of a high-density genetic map (Kurata et al. 1994). Large numbers of ESTs are mapped onto the genetic map or YAC contig map of *A. thaliana* (Agyare et al. 1997). With the sequencing of the entire *Arabidopsis* and rice genomes in the foreseeable future, it will be easier to study exclusively orthologous sequences from those two genomes in comparison to data from other plant species.

METHODS

cDNA Clones

All cDNA clones from the map of Tanksley et al. (1992) that have been cloned in pUC vectors, the pBluescript SK vector, or the pCR II (Invitrogen) vector were used for this study. Insert sizes were confirmed prior to sequencing by PCR using M13 forward and reverse sequencing primers.

Plasmid Preparation and Sequencing

Plasmids were prepared according to standard protocols using Qiagen columns (Qiagen, Hilden, Germany) from 5-ml cultures. Plasmid DNA was sequenced using commercially available sequencing kits and analyzed on ALF (automated laser fluorescence) sequencers (Pharmacia) and ABI sequencers (Perkin Elmer) using either M13 forward (pUC vectors) or SK primers (pBluescript). Most of the clones were also sequenced from the opposite side using M13 reverse (pUC vectors) or KS primers (pBluescript).

Sequence Analysis

Raw sequences were transferred into the sequence analysis program DNAsis (Hitachi) and edited for vector sequences, poly(A) tails, and other cloning artifacts. If sequencing was performed from both sides of a clone, it was determined whether the two sequences overlap and in such cases they were edited and merged into a single file. The edited sequences were analyzed. The DNA sequence and the translated protein sequences were compared to all available DNA and protein sequences using the NIH BLAST server (BLASTN and

BLASTX). Matches with a score of $<10^{-10}$ were considered to be significant. Accession numbers correspond to the respective entries in the GenBank Nucleotide Sequence and GenBank Protein databases, respectively.

ACKNOWLEDGMENTS

The technical support for sequencing of S. König and S. Gentz is acknowledged. The data will be also available through the SolGenes database. Part of this research has been supported by the Deutsche Forschungsgemeinschaft (Ga470/1-2).

The publication costs of this article were defrayed in part by payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 USC section 1734 solely to indicate this fact.

REFERENCES

- Agyare, F.D., D.A. Lashkari., A. Lagos, A.F. Namath, G. Lagos, R.W. Davis, and B. Lemieux. 1997. Mapping expressed sequence tag sites on yeast artificial chromosome clones of *Arabidopsis thaliana* DNA. *Genome Res.* 7: 1–9.
- Ahn, S. and S.D. Tanksley. 1993. Comparative linkage maps of the rice and maize genomes. *Proc. Natl. Acad. Sci.* 90: 7980–7984.
- Alpert, K.B. and S.D. Tanksley. 1996. High resolution mapping and isolation of a yeast artificial chromosome contig containing *fw2.2*: A major fruit weight quantitative trait locus in tomato. *Proc. Natl. Acad. Sci.* 93: 15503–15507.
- Bernatzky, R. and S.D. Tanksley. 1986. Majority of random cDNA clones correspond to single loci in the tomato genome. *Mol. & Gen. Genet.* 203: 8–14.
- Bonnema, G., J. Hontelez, R. Verkerk, Y.Q. Zhang, R. van Daelen, A. van Kamen, and P. Zabel. 1996. An improved method for partially digesting plant megabase DNA suitable for YAC cloning: Application to the construction of a 5.5 genome equivalent YAC library of tomato. *Plant J.* 9: 125–133.
- Delseny, M., R. Cooke, M. Raynal, and F. Grellet. 1997. The *Arabidopsis thaliana* cDNA sequencing projects. *FEBS Lett.* 403: 221–224.
- Gebhardt, C., E. Ritter, A. Barone, T. Debener, B. Walkemeier, U. Schachtschnabel, H. Kaufmann, R.D. Thompson, M.W. Bonierbale, M.W. Ganai, S.D. Tanksley, and F. Salamini. 1991. RFLP maps of potato and their alignment with the homeologous tomato genome. *Theor. Appl. Genet.* 83: 49–57.
- Hamilton, C.M., A. Frary, C. Lewis, and S.D. Tanksley. 1996. Stable transfer of intact high molecular weight DNA into plant chromosomes. *Proc. Natl. Acad. Sci.* 93: 9975–9979.
- Inoue, T., H.S. Zhong, A. Miyao, I. Ashikawa, L. Monna, S. Fukuoka, N. Miyadera, Y. Nagamura, N. Kurata, T. Sasaki, and Y. Minobe. 1994. Sequence-tagged sites (STSs) as standard landmarks in rice genome. *Theor. Appl. Genet.* 89: 728–734.
- Jacobs, J.M.E., H.J. van Eck, P. Arens, B. Verkerk-Bakker., B. te Lintel Hekkert, H.J.M. Bastiaanssen, A. El-Kharborty, A. Periera, E. Jacobsen, and W.J. Stiekema. 1995. A genetic map of potato (*Solanum tuberosum*) integrating molecular markers, including transposons and classical markers. *Theor. Appl. Genet.* 91: 289–300.
- Konieczny, A. and F.M. Ausubel. 1993. A procedure for mapping Arabidopsis mutations using co-dominant ecotype-specific PCR-based markers. *Plant J.* 4: 403–410.
- Kurata, N., Y. Nagamura, K. Yamamoto, Y. Harushima, N. Sue, J. Wu, B.A. Antonio, A. Shomura, T. Shimizu, S.-Y. Lin et al. 1994. A 300 kilobase interval genetic map of rice including 883 expressed sequences. *Nature Genet.* 8: 365–372.
- Martin, G.B., M.W. Ganai, and S.D. Tanksley. 1992. Construction of a yeast artificial chromosome library of tomato and identification of cloned segments linked to two disease resistance loci. *Mol. Gen. Genet.* 233: 25–32.
- Paterson, A.H., T.-H. Lan, K.P. Reischmann, C. Chang, Y.-R. Lin, S.-C. Liu, M.D. Burow, S.P. Kowalski, C.S. Katsar, T.A. DelMonte et al. 1996. Toward a unified genetic map of higher plants, transcending the monocot-dicot divergence. *Nature Genet.* 14: 380–382.
- Pillen, K., K.B. Alpert, J.J. Giovannoni, M.W. Ganai, and S.D. Tanksley. 1996a. Rapid and reliable screening of a tomato YAC library exclusively based on PCR. *Plant Mol. Biol. Rep.* 14: 58–67.
- Pillen, K., O. Pineda, C.B. Lewis, and S.D. Tanksley. 1996b. Status of genome mapping tools in the taxon Solanaceae. In *Genome mapping in plants* (ed. A.H. Paterson), pp. 281–308. R.G. Landes, Austin, TX.
- Sasaki, T., J. Song, Y. Koga-Ban, E. Matsui, F. Fang, H. Higo, H. Nagasaki, M. Hori, M. Miya, E. Murayama-Kayano et al. 1994. Toward cataloging all rice genes: Large scale sequencing of randomly chosen rice cDNAs from a callus cDNA library. *Plant J.* 6: 615–624.
- Schumacher, K., M. Ganai, and K. Theres. 1995. Genetic and physical mapping of the *lateral suppressor (ls)* locus in tomato. *Mol. Gen. Genet.* 246: 761–766.
- Shizuya, H., B. Birren, U.-J. Kim, V. Mancino, T. Slepak, Y. Tachiiri, and M. Simon. 1992. Cloning and stable maintenance of 300-kilobase-pair fragments of human DNA in *Escherichia coli* using an F-factor-based vector. *Proc. Natl. Acad. Sci.* 89: 8794–8797.
- Tanksley, S.D., M.W. Ganai, J.P. Prince, M.C. de Vicente, M.W. Bonierbale, P. Broun, T.M. Fulton, J.J. Giovannoni, S. Grandillo, G.B. Martin et al. 1992. High density molecular linkage maps of the tomato and potato genomes. *Genetics* 132: 1141–1160.
- Tatusov, R.L., E.V. Koonin, and D.J. Lipman. 1997. A genomic perspective on protein families. *Science* 278: 631–637.
- Zamir, D. and S.D. Tanksley. 1988. Tomato genome is comprised largely of fast-evolving low copy-number sequences. *Mol. & Gen. Genet.* 213: 254–261.

Received January 13, 1998; accepted in revised form July 9, 1998.

Sequencing of cDNA Clones from the Genetic Map of Tomato (*Lycopersicon esculentum*)

Martin W. Ganai, Rosemarie Czihal, Ulrich Hannappel, et al.

Genome Res. 1998 8: 842-847

Access the most recent version at doi:[10.1101/gr.8.8.842](https://doi.org/10.1101/gr.8.8.842)

References

This article cites 21 articles, 7 of which can be accessed free at:
<http://genome.cshlp.org/content/8/8/842.full.html#ref-list-1>

License

Email Alerting Service

Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#).

Affordable, Accurate
Sequencing.

To subscribe to *Genome Research* go to:
<https://genome.cshlp.org/subscriptions>