Available online at www.sciencedirect.com

SCIENCE DIRECT

Gene 356 (2005) 127 – 134

GENE
SECTION
FUNCTIONAL GENOMICS

www.elsevier.com/locate/gene

Expressed sequence tags from the laboratory-grown miniature tomato (*Lycopersicon esculentum*) cultivar Micro-Tom and mining for single nucleotide polymorphisms and insertions/deletions in tomato cultivars

Naoki Yamamoto[a], Taneaki Tsugane[b], Manabu Watanabe[c], Kentaro Yano[c], Fumi Maeda[c], Chikara Kuwata[d], Moez Torki[e,1], Yusuke Ban[e], Shigeo Nishimura[e], Daisuke Shibata[a,c,*]

[a]*The Graduate School of Life Sciences, Tohoku University, Kazusa-Kamatari 2-6-7, Kisarazu, Chiba 292-0818, Japan*
[b]*Chiba Prefectural Agriculture Research Center, Chiba 266-0006, Japan*
[c]*Kazusa DNA Research Institute, Kazusa-Kamatari 2-6-7, Kisarazu, Chiba 292-0818, Japan*
[d]*Chiba Agricultural Research Center, Crop Breeding Institute, Mobara, Chiba 299-4335, Japan*
[e]*Institute of Agriculture and Forestry, University of Tsukuba, Tsukuba, Ibaraki 305-8572, Japan*

Received 21 September 2004; received in revised form 8 March 2005; accepted 7 April 2005
Available online 21 June 2005
Received by B. Hohn

Abstract

Laboratory-grown miniature tomato (*Lycopersicon esculentum*) cultivar Micro-Tom has attracted attention as a host for functional genomics research. In this study, we generated 35,824 expressed sequence tags (ESTs) from leaves and fruits of Micro-Tom. The ESTs comprised 10,287 unigenes (5007 contigs and 5280 singletons), including 1858 novel tomato unigenes. Of the 18 unigenes that shared strong homology with tobacco chloroplast genome sequences, one unigene was likely derived from polyadenylated transcripts of the *atpH* gene. Interestingly, ESTs for vacuolar invertase, pectate lyase and alcohol acyl transferase were underrepresented in the Micro-Tom data set. From all of the ESTs, we mined 2039 candidate single nucleotide polymorphisms (SNPs) and 121 candidate insertions and deletions (indels) based on homology with four tomato inbred lines, E6203, R11-13, Rio Grande PtoR and R11-12, and a wild relative, *L. pennellii* TA56, for which sequence data was publicly available with more than 5000 entries. Direct genome sequencing of several SNP or indel sites in Micro-Tom and *L. esculentum* E6203 suggested that more than 69% of the candidate sites were truly polymorphic, making them useful for the preparation of DNA markers.

Keywords: cDNA; DNA marker; unigene; *Lycopersicon pennellii*

1. Introduction

Tomato (*Lycopersicon esculentum*) has a moderately sized genome of 950 Mb (Arumuganathan and Earle, 1991), which is estimated to encode ~35,000 genes (Van der Hoeven et al., 2002); the majority of the genes are located in contiguous stretches of gene-rich euchromatin which accounts for less than 25% (220 Mb) of the DNA (Peterson et al., 1998). Interestingly, 30% of the genes have no significant correspondence to *Arabidopsis* genes, and the function of the majority of these genes remains unknown (Van der Hoeven et al., 2002). Research on the tomato, a

Abbreviations: EST, expressed sequence tag; SNP, single nucleotide polymorphism; indel, insertion and deletion.

* Corresponding author. Kazusa DNA Research Institute, Kazusa-Kamatari 2-6-7, Kisarazu, Chiba 292-0818, Japan. Tel.: +81 438 52 3947; fax: +81 438 52 3948.
E-mail address: shibata@kazusa.or.jp (D. Shibata).
[1] Present address: National Agronomical Institute of Tunisia, University 7th Novenber of Carthago, 43 Avenue Charles Nicolle-1082, Tunis Mahrajene, Tunisia.

0378-1119/$ - see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.gene.2005.04.026

nodal species in the Solanaceae family, complements biodiversity research using the most informative model plant, *Arabidopsis thaliana*, because these two plants diverged from their common ancestor early in the radiation of dicots. Recently, tomato genome sequencing programs have been launched as part of the internationally coordinated International Solanaceae Genome Project (SOL) consortium (http://www.sgn.cornell.edu/solanaceae-project/). With the wealth of the information generated by tomato genome sequencing, various resources for genetics and genomics research are expected to facilitate the understanding of gene function. However, the necessity of greenhouses or fields for growing tomatoes might impede their utilization because most molecular biology laboratories have limited access to such facilities.

Recently a miniature tomato cultivar, Micro-Tom, which was originally bred for home gardening (Scott and Harbaugh, 1989), has attracted attention as a host for functional genomics. With its small size of 10–20 cm in height, ability to grow well at densities as high as 1357 plants per 1 m^2 and short life cycle of 70–90 days, Micro-Tom is suitable for cultivation and experimentation in most plant biology laboratories (Meissner et al., 1997). The potential of Micro-Tom as a forward and reverse genetics tool has been demonstrated by the generation of mutant populations by ethyl methane sulfonate, Ac/Ds transposon and T-DNA with activation tag (reviewed in Emmanuel and Levy, 2002). Over 1400 mutants exhibiting various phenotypic mutations of leaves, fruits and flower shape and color among the mutagenized populations have been found. Gene identification from these mutants will lead to a better understanding of the biological processes unique to tomato. Although to isolate mutated genes by map-based cloning, DNA markers are necessary to identify genomic clones such as BAC (bacterial artificial chromosome) clones that contain the target locus, no DNA marker specific for Micro-Tom is available at present. As the tomato genome sequencing uses BAC clones derived from the cultivar Heinz 1706, the resulting sequences might not be identical to that of Micro-Tom, as exemplified in the single nucleotide polymorphisms (SNPs) identified in this study. Alternatively, EST sequencing of Micro-Tom will provide transcribed sequence information. Further development of resources such as DNA markers and ESTs will fully expand the potential usefulness of the cultivar for use in genetic and functional genomics approaches.

ESTs are a fundamental resource for genomic research. ESTs generated from various genotypes, such as cultivars and inbred lines, within a species provide information on nucleotide differences of allelic genes among the genetic variations is a source of DNA markers. Several algorithms have been developed for mining candidate SNPs in large EST collections produced from human (Buetow et al., 1999; Marth et al., 1999; Picoult-Newberg et al., 1999). In plants, large-scale mining of SNPs from EST databases have been conducted for maize (Useche et al., 2001; Batley et al., 2003), barley (Kota et al., 2003) and *A. thaliana* (Schmid et al., 2003). Further, protein sequences derived from the ESTs of various genotypes within a species will serve as a proteome resource for the functional analysis of genes. EST databases also can be used for global gene expression analysis if the cDNA libraries from which the EST information was generated were not normalized during preparation. A large collection of ESTs from 20 tomato cultivars or lines and a few wild relatives have been deposited in the public NCBI database (dbEST, 158,865 ESTs, February 20, 2004). However, there is no report on either Micro-Tom ESTs or SNP mining at present.

In this study, we generated 35,824 ESTs comprising 10,287 unigenes (5007 contigs and 5280 singletons) from leaves and maturing fruits of Micro-Tom. Comparisons with previously identified ESTs from a public database that were derived from other tomato cultivars and a few wild relatives showed that 1858 unigenes in the present unigene set were novel. We mined 2039 candidate SNPs and 121 candidate insertions and deletions (indels) from comparisons with four *L. esculentum* inbred lines and a wild relative, *L. pennellii* and evaluated the reliability of the candidate polymorphisms by direct genome sequencing. Other merits of having an abundance of ESTs were discussed. Information of EST sequences, SNPs and indels is available at the Micro-Tom database MiBASE (http://www.kazusa.or.jp/microtom/).

2. Materials and methods

2.1. Plant materials

Germinated seeds of Micro-Tom were grown under natural conditions in a greenhouse. Maturing fruits and fully expanded leaves were collected, immediately frozen in liquid nitrogen, and stored at −80 °C until RNA was extracted. Maturing fruits were collected at the green stage (45–60 days after germination), the breaker stage (55–65 days), the turning stage (60–70 days) and the red ripe stage (65–75 days), as defined according to color changes described in “United States Standards for Grades of Fresh Tomatoes” (United States Department of Agriculture, http://www.ams.usda.gov/standards/).

2.2. Preparation of cDNA libraries

RNA was extracted from fruits collected at each of the four stages and fully expanded leaves. Briefly, the entire fruit (12 g) were ground to powder in liquid nitrogen by mortar and pestle, mixed with RNA extraction buffer (100 mM Tris–HCl pH 7.5, 100 mM LiCl, 25 mM EDTA, 2% cetyl trimethyl ammonium bromide, 0.5% Sarkosyl, 0.5% SDS and 0.5% β-mercaptoethanol) and incubated with acid phenol at 80 °C for 10 min. After mixing with chloroform, phenol/chloroform extraction was performed three times as previously described (Sambrook et al., 1989). Following chloroform extraction, total RNA was precipitated by 2 M