Systematic Development of Tomato BioResources in Japan

Tohru Ariizumi¹, Koh Aoki² and Hiroshi Ezura¹*

¹Graduate School of Life and Environmental Sciences, University of Tsukuba, Tsukuba 305-8572, Japan
²Kazusa DNA Research Institute, 2-6-7 Kazusa-Kamatari, Kisarazu 292-0818, Japan

Subject areas: Resources

Author contribution: T.A., K.A. and H.E. wrote this article.

*Correspondence and requests for materials should be addressed to H.E. (ezura@gene.tsukuba.ac.jp).

Reviewer: Yasutaka Kubo, Okayama University, Japan; Masatomo Kobayashi, RIKEN BioResources Center, Japan

Editor: Yongpyo Lim, Chungnam National University, Republic of Korea

Received December 22, 2010; Accepted January 05, 2011; Published January 17, 2011


Funding: This work is supported by Ministry of Education, Culture, Sports, Science and Technology (MEXT), Japan.

Competing interest: All authors declare no financial or personal conflict that could inappropriately bias their experiments or writing.

Copyright: This article is licensed under a Creative Commons Attribution License, which freely allows to download, reuse, reprint, modify, distribute, and/or copy articles as long as a proper citation is given to the original authors and sources.

Keywords: tomato, National BioResource Project, micro-tom, mutants, full length cDNA

SYNOPSIS

Recently, with the progress of genome sequencing, materials and information for research on tomato (Solanum lycopersicum) have been systematically organized. Tomato genomics tools including mutant collections, genome sequence information, full-length cDNA and metabolomic datasets have become available to the research community. In Japan, the National BioResource Project Tomato (NBRP Tomato) was launched in 2007, with aims to collect, propagate, maintain and distribute tomato bioresources to promote functional genomics studies in tomato. To this end, the dwarf variety Micro-Tom was chosen as a core genetic background, due to its many advantages as a model organism. In this project, a total of 12,000 mutagenized lines, consisting of 6000 EMS-mutagenized and 6000 gamma-ray irradiated M2 seeds, were produced, and the M3 offspring seeds derived from 2236 EMS-mutagenized M2 lines and 2700 gamma-ray irradiated M2 lines have been produced. Micro-Tom mutagenized lines in the M3 generation and monogenic Micro-Tom mutants are provided from NBRP tomato. Moreover, tomato cultivated varieties and its wild relatives, both of these are widely used for experimental study, are available. In addition to these bioresources, NBRP Tomato also provides 13,227 clones of full-length cDNA which represent individual transcripts non-redundantly. In this paper, we report the current status of NBRP Tomato and its future prospects.
Introduction

Tomato (Solanum lycopersicum) is suggested to be native to South America, perhaps in the mountainous regions in Peru. It is also believed that tomatoes were transported to Central America, Mexico and Europe during the Age of Geographical Discovery. Although the exact date of tomato domestication is not known, humanity has produced approximately 8000 tomato varieties since then. Tomato is one of the most popular vegetable crops in the world, and is sufficiently loved to give rise to several figurative expressions due to high contents of various nutrients, such as ‘Golden apple’, ‘Love apple’, and ‘When tomatoes turn red, doctor’s faces turn green’. In addition to its nutritional value, tomato is economically important, with one of the highest productions of edible crops in the world. According to the statistics of the Food and Agricultural Organization tomato production is about 1.3 million tons per year and ranks the 12th largest among all crops.

To improve the quality and yield of tomato, it is crucial to understand the molecular bases underlying these complex traits. Recent advances in plant genome sequencing have enabled us to link phenotype to DNA sequences. For tomato, an international genome sequencing project started in 2004, and the first draft genome assembly was placed in the public domain in 2009 (http://solgenomics.net/genomes/index.pl). To fully exploit the wealth of genome sequence information, we need genomic tools that support the functional analysis of genes, such as mutagenized lines, monogenic mutants, expressed sequence tags (EST) and full-length cDNAs. To this end, several tomato resource centers have been constructed internationally.

In Japan, development of tomato bioresources is performed in a framework of the National BioResource Project (NBRP). The NBRP started in 1998 with the aim of developing systematic bioresources for the comprehensive promotion of life science. NBRP currently covers 28 core organisms including animal species, microorganisms, embryonic stem (ES) cells and plants. Tomato was added to this project in 2007. The main purpose of NBRP Tomato is to be a resource center at the hub of the tomato research community, as well as to establish bases of tomato genomics research. In this paper, we introduce the main activities of NBRP Tomato and its future prospects.

Why Study Tomato?

From the phylogenetic point of view, tomato is regarded as a model plant of the Solanaceae family. The Solanaceae comprises 1000–2000 species, and includes a number of vegetable crops and ornamental plants of economic importance, including potato (Solanum tuberosum L.), eggplant (S. melongena), petunia (Petunia × hybrida), tobacco (Nicotiana tabacum) and pepper (Capsicum annum L.). Many species of the Solanaceae share a basic set of 12 chromosomes and similar gene content. Nevertheless, they show remarkably different phenotypic outcomes, which make the Solanaceae an excellent system to study plant diversity and adaptation to environment.

The development of new tomato varieties with novel properties such as high sugar content, high nutritional value, high yield and high degree of disease resistance are the main targets of commercial breeding companies. Such new varieties promise high prices due to higher quality.

From the scientific point of view, tomatoes have three major unique features compared to the other model plants, including Arabidopsis and rice. First, tomato is a day-neutral plant that produces a floral meristem regardless of day length, whereas Arabidopsis is a long-day and rice is a short-day plant. Secondly, tomato is a fruit-bearing plant. Plant species showing fruit formation are taxonomically diverse, including taxa in the Solanaceae, Cucurbitaceae, Ribaceae, Rutaceae, Rosaceae and Vitaceae, and tomato serves as a model for the study of fruit development. Thirdly, tomato fruits contain numbers of functional metabolites such as carotenoids, flavonoids and triterpenoids, most of which are suggested to play roles in promoting human health. Thus, study of tomato provides an excellent opportunity to determine mechanisms controlling primary and secondary metabolisms by which the functional metabolites are synthesized.

An extensive EST analysis suggested that approximately 30% of tomato genes encode proteins showing limited sequence similarity to proteins in rice and Arabidopsis. Developing an infrastructure for tomato genomics would lay a foundation to explore biological functions of these novel proteins, thus revealing novel mechanisms underlying development, yield and quality of tomato fruit.

Micro-Tom, a Suitable Variety for Experimental Research of Tomato

There are several general requirements for model organisms for molecular genomics studies: small size, short generation time, ease of proliferation and high efficiency in genetic transformation. Micro-Tom, a dwarf variety of tomato, fulfills these requirements. The height of a mature Micro-Tom plant is about 20-30 cm, and its shape is characterized by short internodes and clear determinate phenotype, which makes it possible to grow at high density (1357 plants/m²). Micro-Tom has a short generation time of 3-4 months, and can grow under normal fluorescent lamps, making it possible to obtain seeds even in indoor conditions (as for Arabidopsis). The transformation efficiency of Micro-Tom is high, so that transgenic plants can be produced routinely, which has enabled production of several tagging lines. An additional advantage is that Micro-Tom can cross with other tomato varieties and many related Solanum spp., allowing transfer of the mutations found in Micro-Tom to other varieties. The capacity for interspecific crossing allows generation of F2 mapping populations between Micro-Tom mutants and relatives such as S. pennelli and S. pimpin nellifolium to perform positional cloning of candidate genes.

Micro-Tom reportedly has several mutations: dwarf (d), self-pruning (sp), miniature (mn) and two additional mutations affecting susceptibility to Cladosporium fulvum. Despite these mutations, since mutation appeared to be confined within these genes, we chose Micro-Tom as a core genetic background for our bioresource project due to the advantages previously described.

Activity of NBRP Tomato: Plant Resources

NBRP Tomato consists of two institutes, the University of Tsukuba as a core facility, and Kazusa DNA Research Institute (KDRI) as a sub-facility. The main activities of the University of Tsukuba are taking charge of developing, maintaining, propagating and providing tomato plant materials, and of KDRI similarly for tomato DNA materials (Figure 1).

Tomato plant materials available from NBRP Tomato include seeds of S. lycopersicum varieties, wild relatives, introgression lines, M3 generations of Micro-Tom mutagenized lines, and Micro-Tom individual mutants. All of these have been distributed by the University of Tsukuba. The Micro-Tom mutagenized lines were generated by ethylmethanesulphonate (EMS) treatment and gamma-ray irradiation, and a population of 12,000 lines consisting of 6000 EMS-mutagenized M2 lines and 6000 gamma-irradiated M2 lines were obtained. To date, a total of 2236 EMS-mutagenized M2 lines and 2700 gamma-ray irradiated M2 lines have been grown,
and their M3 offspring seeds harvested. Consequently, a total of 4936 M3 seed lots have been produced (Figure 2A). In the process of propagating M3 offspring seeds, individual M2 plants were initially inspected for visual phenotypic changes. A number of phenotypes were seen: ranging from dwarfs to giants, from early flowering to late flowering. In addition, phenotypes with diverse leaf structure and fruit morphology were recovered\(^{1,14}\). These mutant phenotypes were recorded and registered into the database called Tomato Mutants Archive (TOMATOMA; http://tomatoma.nbrp.jp/) (Figure 2B). The mutagenized M3 seeds and seeds of individual mutants are available upon request in exchange for a material transfer agreement (MTA). The primary goal of NBRP is to develop the M3 seeds from all of M2 populations.

Along with Micro-Tom bioresources, the University of Tsukuba also provides tomato varieties widely used for experimental study, such as cvs Rutgers, Alisa Craig, Money Maker and M82. Also provided are tomato wild relatives, such as S. pennelli, S. peruvianum, S. chilense and S. pimpinellifolium. This supports experimental research in the Japanese tomato research community.

As a technology related to the NBRP Tomato’s plant resources, a high-throughput platform for the Targeting Local Lesions IN Genomes (TILLING)\(^{12}\), was developed by the University of Tsukuba using DNA extracted from the M2 generation of Micro-Tom plants to accelerate the identification of mutated genes. TILLING has great potential to screen for mutated alleles of genes in mutant populations. The University of Tsukuba has prepared DNA from approximately 1800 different mutagenized lines, and is in the process of testing the quality of the mutant populations (Okabe et al., unpublished data). Recently, several other Solanaceae research groups have also developed high-throughput TILLING platforms\(^{11,21}\). Establishment of TILLING platforms will strengthen the value of mutagenized lines as a bioresource for key-gene identification that governs the important traits of tomato.

### Activity of NBRP Tomato: DNA Resources

Tomato DNA materials provided by NBRP Tomato include full-length cDNA, tomato promoter clones and tomato BAC clones; and all have been collected, maintained and distributed by KDRI. KDRI has collected approximately 100,000 full-length cDNA clones from various tissues of Micro-Tom, and among them, full-length sequences of 13,227 full-length cDNA clones have been determined\(^{22}\). The 13,227 clones represent individual transcripts non-redundantly. Sequences of these full-length cDNA clones are available at the Kazusa full-length tomato cDNA database (KatTom; http://www.pgb.kazusa.or.jp/katfom/)\(^{22}\). In addition to the full-length sequences, KatTom provides BLAST annotations, predicted protein sequences, predicted protein domains and results of mapping full-length cDNAs onto Solanaceae Genomics Network (SGN) tomato BAC sequences. Based on mapping onto BAC sequences, we also provide the prediction of exons and introns (Figure 3A). Interestingly, the frequency of nucleotide mismatches between Micro-Tom full-length cDNA and tomato genome sequences of cv Heinz 1706 was estimated to be 0.06%, suggesting that nucleotide identity in the exon region is so high that Micro-Tom can be a good
In this section, we describe a comprehensive catalog of Micro-Tom and more than 90 tomato varieties were identified, and are also attached to the reference for tomato genome annotation.

Additionally, sequences of the 89,872 ESTs derived from 5'-end of the full-length cDNA clones are available in the Micro-Tom transcriptome database MiBASE (http://www.pgb.kazusa.or.jp/mibase). MiBASE also provides a set of tomato unigenes assembled using 280,000 tomato ESTs that are publicly available. According to the EST alignment, single nucleotide polymorphisms (SNPs) between tomato varieties were identified, and are also available in MiBASE. In addition to sequence information, MiBASE contains two forms of gene-expression data representation. First, users can estimate tissue-dependent gene expression using EST counts in tissue-specific libraries. Second, using publicly available Affymetrix Tomato Genechip datasets, MiBASE provides results of gene-to-gene co-expression analysis (Figure 3B).

In addition to cDNA sequences, NBRP Tomato has been accumulating genome sequence information of Micro-Tom. NBRP Tomato has determined approximately 93,682 end-sequences of Micro-Tom BAC clones, and the sequence information is available in the Tomato SBM & Marker Database (http://www.kazusa.or.jp/tomato). The Micro-Tom genome sequencing project is underway (November 2010) in collaboration with National Institute of Genetics, and the Micro-Tom genome information will be in the public domain by 2012.

As a DNA information resource, KDRI also provides a dataset of tomato fruit metabolomes. Metabolite composition of maturing fruit was determined using liquid chromatography–Fourier transform ion cyclotron resonance mass spectrometry. Information on approximately 900 metabolites, including >60 flavonoids and >90 triterpenoids, are available in the database KOMICS (http://webs2.kazusa.or.jp/komics). The comprehensive annotation of metabolites will provide a basis for improving the nutritional quality of tomato fruit.

**Features of NBRP Tomato Compared with Other Tomato BioResources Internationally**

As described in the previous sections, NBRP Tomato focuses on organizing materials and information concerning Micro-Tom-centered genomic bioresources. In this section, we describe a comparative overview of other tomato resource projects in the world, and see how they complement each other (Table 1).

The SGN focuses on organizing information on genomic and transcriptome data for Solanaceae species, including tomato. The Solanaceae International Consortium (SIC) provides a dataset of approximately 900 Solanaceae species, including tomato. The SIC focuses on the comprehensive annotation of metabolites will provide a basis for improving the nutritional quality of tomato fruit.

**Table 1. List of databases providing tomato resources and genome information**

<table>
<thead>
<tr>
<th>Available materials and sequence information</th>
<th>Database</th>
<th>URL</th>
</tr>
</thead>
<tbody>
<tr>
<td>NBRP Information center</td>
<td>MiBASE</td>
<td><a href="http://www.nbrp.jp/">http://www.nbrp.jp/</a></td>
</tr>
<tr>
<td>NBRP Tomato</td>
<td>MiBASE</td>
<td><a href="http://tomato.nbrp.jp/">http://tomato.nbrp.jp/</a></td>
</tr>
<tr>
<td>TOMATOMA</td>
<td>MiBASE</td>
<td><a href="http://tomatoma.nbrp.jp/">http://tomatoma.nbrp.jp/</a></td>
</tr>
<tr>
<td>TOBOMATOMA</td>
<td>MiBASE</td>
<td><a href="http://tomatoma.nbrp.jp/">http://tomatoma.nbrp.jp/</a></td>
</tr>
<tr>
<td>Tomato SBM &amp; Marker Database</td>
<td>MiBASE</td>
<td><a href="http://www.kazusa.or.jp/tomato/">http://www.kazusa.or.jp/tomato/</a></td>
</tr>
<tr>
<td>Tomato metabolites</td>
<td>MiBASE</td>
<td><a href="http://www.pgb.kazusa.or.jp/katom/">http://www.pgb.kazusa.or.jp/katom/</a></td>
</tr>
<tr>
<td>Solanaceae international consortium</td>
<td>KOMICS</td>
<td><a href="http://webs2.kazusa.or.jp/komics/">http://webs2.kazusa.or.jp/komics/</a></td>
</tr>
<tr>
<td>Genomic DNA information of tomato (cv. Heinz 1706) and wild relatives</td>
<td>SGN</td>
<td><a href="http://solgenomics.net/">http://solgenomics.net/</a></td>
</tr>
<tr>
<td>Seeds of cultivated tomato and wild relatives</td>
<td>TGRC</td>
<td><a href="http://tgrc.ucdavis.edu/">http://tgrc.ucdavis.edu/</a></td>
</tr>
<tr>
<td>Seeds of individual M82 mutants</td>
<td>The Genes That Make Tomatoes</td>
<td><a href="http://zamir.sgn.cornell.edu/mutants/">http://zamir.sgn.cornell.edu/mutants/</a></td>
</tr>
<tr>
<td>Seeds of individual Red Setter mutants</td>
<td>LycoTILL</td>
<td><a href="http://www.agrobios.it/tilling/index.html">http://www.agrobios.it/tilling/index.html</a></td>
</tr>
<tr>
<td>Seeds of cultivated tomato and wild relatives</td>
<td>EU-SOL germplasm and phenotype database</td>
<td><a href="http://www.eu-sol.net/">http://www.eu-sol.net/</a></td>
</tr>
</tbody>
</table>

**NBRP** - National BioResource Project, **TGRC** - Tomato Genetics Resource Center
genetic resources24,25. Sequence information of tomato reference genome (S. lycopersicum cv Heinz 1706) and genomes of two wild relatives (S. pimpinellifolium and S. pennellii) is available at the SGN website (http://solgenomics.net/). SGN also provides > 160,000 tomato EST sequences (November 2010), and DNA markers including RFLP (Restriction Fragment Length Polymorphism), SSR (Simple Sequence Repeats), CAPS (Cleavage Amplified Polymorphic Sequence), AFLP (Amplicon Fragment Length Polymorphism), RAPD (Random Amplified Polymorphic DNA), SCAR (Sequence Characterized Amplified Region) and SNP. High-density molecular maps of tomato, generated based on progenies of S. lycopersicum × S. pennellii, S. lycopersicum × S. pimpinellifolium and other crosses, are available at SGN.

Tomato Genetics Resource Center (TGRC; University of California, Davis) has a large seed stock collection (> 3700 lines). TGRC has 1186 lines of wild species, 1023 lines of monogenic mutants and 1560 lines of miscellaneous genetic lines. These seed materials are available upon request in exchange with an MTA.

The philosophy of the EU-SOL project in Europe is similar to NBRP Tomato in that they focus both on materials and information. EU-SOL has developed the ‘EU-SOL germplasm and phenotype database that provides an interface to data of diverse germplasm and populations consisting of > 7000 lines of tomato varieties and related species. EU-SOL has developed EMS mutagenesis lines consisting of 6677 M2 and 5508 M3 families, and phenotypes are categorized. Mutagenized seeds are available for distribution through the database called ‘LycoTILL’ upon the signing of a MTA. Tomato TILLING platforms have also been developed in the EU-SOL framework26.

The Hebrew University group in Israel focuses more on the material resources and phenotyping. They have generated and maintain 13,000 families of M2 populations using tomato cv M82 mutagenized by EMS and fast-neutron bombardment27. Such comprehensive mutant populations have been visually phenotyped and can be browsed in the phenotypic database ‘The Genes That Make Tomatoes’ (http://zamir.sgn.cornell.edu/mutants/), and mutant seeds are available upon request.

Future Prospects of Tomato BioResources

Tomato research is stepping into the post-genome phase. The reference genome sequence of tomato cv Heinz 1706 has been released, and the draft genome sequences of wild relatives, S. pimpinellifolium (LA1589) and S. pennellii (LA716), can be consulted. In this post-genome situation, the first expectation for tomato bioresources is to reinforce the link between genome sequence information and phenotypic information. Induced-mutant lines play an important role as the core material in establishing this link. The incredible progress in DNA sequencing technology will enable re-sequencing of genomes of mutant lines, which will allow the identification of mutations associated with the mutant phenotype. In parallel, the phenotypes of the mutant lines will be described in more detail. For instance, in addition to the visible phenotypes, typing of metabolite- and hormone-profiles will provide valuable information for tomato research. Mutant lines to which DNA sequence information and phenotypic information are attached will surely contribute materials for applied research programs as well as basic research. Since tomato is preferentially used when scientists attempt to translate their knowledge into industrial and agricultural applications, we expect that tomato bioresources will serve as a plant scientist’s primary box of tools in tackling the challenges of improving nutrition that contributes for promoting human health, and creation of a sustainable society.

Acknowledgements

We apologize to all of our colleagues whose work was not mentioned due to space limitations. We thank all members at the University of Tsukuba and KDIR for helpful discussions concerning this project. ‘National BioResource Project, Core Facility Upgrading Program, Tomato’ is supported by Ministry of Education, Culture, Sports, Science and Technology (MEXT), Japan.

References

15. Matsukura, C., Aoki, K., Fukuda, N., Mizoguchi, T., Asamizu, E.,


