



A snapshot of the Chinese SOL Project

Changbao Li^{a, c}, Jiuhai Zhao^{a, b}, Hongling Jiang^a, Yu Geng^a, Yuanyuan Dai^a, Huajie Fan^{a, b},
Dongfen Zhang^{a, b}, Jinfeng Chen^{a, b}, Fei Lu^{a, b}, Jinfeng Shi^a, Shouhong Sun^a, Jianjun Chen^d,
Xiaohua Yang^d, Chen Lu^d, Mingsheng Chen^a, Zhukuan Cheng^a, Hongqing Ling^a,
Ying Wang^d, Yongbiao Xue^a, Chuanyou Li^{a, *}

^a Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing 100101, China

^b Graduate School of Chinese Academy of Sciences, Beijing 100039, China

^c State Key Laboratory of Crop Biology, Agronomy College, Shandong Agricultural University, Taian 271018, China

^d Wuhan Botanical Garden, Chinese Academy of Sciences, Wuhan 430074, China

Received for publication 2 June 2008; revised 4 June 2008; accepted 5 June 2008

Abstract

In 2003, the International Solanaceae Project (SOL) was initiated by an international consortium of ten countries including Korea, China, the United Kingdom, India, the Netherlands, France, Japan, Spain, Italy and the United States. The first major effort of the SOL aimed to produce a DNA sequence map for euchromatin regions of 12 chromosomes of tomato (*Solanum lycopersicum*) before 2010. Here we present an update on Chinese effort for sequencing the euchromatin region of chromosome 3.

Keywords: tomato; *Solanum lycopersicum*; Solanaceae; SOL

Introduction

The Solanaceae is the third most important plant taxon economically and the most valuable in terms of vegetable crops. It encompasses the most variable of crop species in terms of their agricultural utility, as it includes the tuber-bearing potato (a food staple over much of the world), a number of fruit-bearing vegetables (e.g. tomato, eggplant, peppers, husk tomato), ornamental flowers (petunias, *Nicotiana*), edible leaves (*Solanum aethiopicum*, *S. macrocarpon*), and medicinal plants (e.g. *Datura*, *Capsicum*). Fruits and tubers are major contributors of vitamins, fibers, carbohydrates, and phyto-nutrient compounds in our diet (Knapp, 2002; Chung et al., 2006; Stupar et al., 2006; Chen et al., 2007). The Solanaceae is unique in that multiple crop species in this family are major contributors to fruit and vegetable consumption and thus to our quality of life. In addition to being important food crops, the Solana-

ceae also has roles as scientific model plants, for the study of fruit development, tuber development, biosynthesis of anthocyanin pigments and for plant defenses (Palmer and Zamir, 1982; Borisjuk et al., 1994; Parry et al., 1998; Lorenz-Lemke et al., 2006; Foolad, 2007).

The International Solanaceae Project (SOL)

The Solanaceae is remarkable in that the gene content of the different species is similar despite the markedly different phenotypic outcomes, making the Solanaceae an excellent model for the study of adaptation to natural and agricultural environments (Knapp et al., 2004). Many of the Solanaceae are diploid and share a basic set of 12 chromosomes, indicating an absence of polyploidizations during the evolutionary history of the family. The International Solanaceae Project (SOL) was initiated in November 2003 for the next decade (<http://sgn.cornell.edu/solanaceae-project/>) to create a network of map based resources and information to address two of the most sig-

* Corresponding author. Fax: +86-10-6487 3428
E-mail address: cyli@genetics.ac.cn

nificant questions in plant biology and agriculture (Mueller et al., 2005): 1) How can a common set of genes/proteins give rise to a wide range of morphologically and ecologically distinct organisms that occupy our planet? 2) How can a deeper understanding of the genetic basis of diversity be harnessed to better meet the needs of society in an environmentally-friendly way? The first goal of SOL is to determine the nucleotide sequence of tomato genome with great precision and link it to Solanaceae map.

Tomato is an ideal model system for the study of genome evolution, plant development and plant responses to the environment due to its simple diploid genetics, short generation time, routine transformation technology, availability of rich genetic and genomic resources, and an advanced BAC based physical map (Carland and Staskawicz, 1993; Budiman et al., 2000; Li et al., 2001; Ryan and Pearce, 2003; Liu et al., 2007). In addition, tomato genome (950 Mb) structure is made up of three-quarters (730 Mb) heterochromatin regions and one-quarter (220 Mb) euchromatin regions. More than 90% of the whole 35,000 genes are corresponding to the euchromatic regions (220 Mb). Therefore, the strategy is to sequence only the euchromatic portion of the genome to cover most of the gene space.

The International Tomato Sequencing Project is performed by several international consortium centers to sequence the gene-rich euchromatic portions of the 12 tomato chromosomes (http://www.sgn.cornell.edu/about/tomato_sequencing.pl). Currently, 10 countries are involved in sequencing the tomato genome. The 12 chromosomes have been split up into the countries as follows: USA (chromosome 1 and 10), Korea (2), China (3 and 11), UK(4), India (5), the Netherlands (6), France (7), Japan (8), Spain (9) and Italy (12). Fig. 1 shows an overview of the current status of the International Tomato Genome Sequencing Project (http://www.sgn.cornell.edu/about/tomato_sequencing.pl).

ing.pl).

The Chinese SOL Project (CSOL)

The Solanaceae is one of the most economically important crop families in China. There is a large scientific community including breeders, geneticists and molecular biologists with research focused on tomato and potato, two major Solanaceae species in China. The Chinese SOL Project (CSOL) was established in May 2004 and the centers of Solanaceae research in China included the Institute of Genetics and Developmental Biology (IGDB), Chinese Academy of Sciences; Wuhan Botanical Garden (WBG), Chinese Academy of Sciences; National Center for Gene Research (NCGR), Chinese Academy of Sciences; and Shandong Agricultural University.

Key activities of CSOL include construction of tomato physical map, FISH (Fluorescence *in situ* hybridization) confirmation, sequencing of the entire euchromatic portion of the tomato chromosome 3, sequence annotation and bioinformatics, building databases for genetics and genomics of tomato, establishing a platform for comprehensive genomics analysis, and providing technical workshop for Solanaceae community. Recently, they have made a good progress on these projects.

Construct a fine physical map for the tomato genome

To facilitate BAC-by-BAC based tomato genome sequencing, we are constructing a tomato genome-wide physical map (Tanksley et al., 1992; Wang et al., 2005). First, we have made extensive manual editing of FPC (Fingerprinted contigs) physical map based on fingerprinting of the *Hind* III BAC library of Heinz 1,706 developed by Arizona Genomics Institute. The contig num-

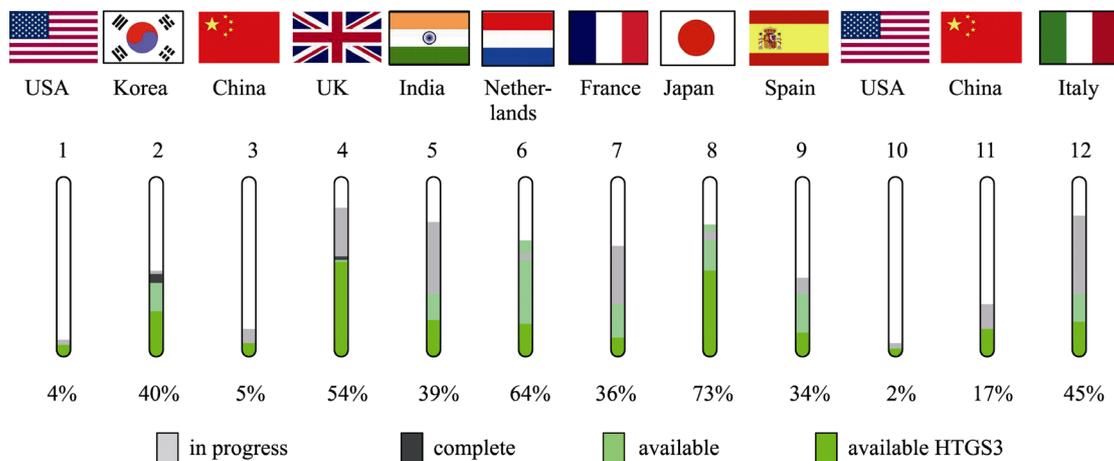


Fig. 1. Current status of the SOL project. The 12 tomato chromosomes are sequenced by an international consortium of 10 countries including Korea, China, the United Kingdom, India, the Netherlands, France, Japan, Spain, Italy and the United States. This overview is an update on May 30, 2008 from the SGN website (http://sgn.cornell.edu/help/about/tomato_sequencing.pl) and will be continuously updated as sequencing progresses.

ber has been reduced from about 7,000 to 3,000. The current FPC contains 88,650 BAC clones, representing a $10 \times$ coverage of the tomato genome and covering about 788 Mb of physical regions. After integrating the overgo hybridization, BAC end sequence, genomic sequence, and partial PCR-based anchoring data, 457 contigs were anchored to the euchromatic regions in the F2.2000 genetic map, which cover about 200 Mb of the tomato genome. The FPC data can be accessed through WebFPC (<http://tomato.genetics.ac.cn>). Second, we have made *in silico* digestion of sequenced BACs and integrated with the FPC. The sequenced BACs downloaded from GenBank were named with the GenBank accession number plus suffix—sd1, such as AF411806sd1. The sequenced BACs downloaded from SGN were named with the BAC name plus suffix—sd1, such as a0016A12sd1. The *in silico* digested BACs were color-coded and easily recognized. These efforts can be used to ensure the integrity of the physical map, to make additional anchoring, and to provide a framework for the tomato genome sequencing. Third, a total of 950 genetic markers with no overgo hybridization data were selected for PCR screening of BACs containing the markers. The thereafter remaining unanchored physical contigs will be mapped to F2.2000 using the BAC end sequences.

Sequence the euchromatic region of tomato chromosome 3

We FISHed 47 BACs onto chromosome 3, with each BAC connected to one genetic marker and identified 36 contigs anchored to the chromosome based on FPC (Peterson et al., 1999). By merging contigs with FPC, our

36 contigs extend to about 8 Mb. The length of each contig ranges between 100 kb and 1.4 Mb. Almost all contigs are distributed on paracentromeric regions and the long arm of chromosome 3 including the so-called “heterochromatic” regions. No contig has been identified on the short arm because no overgo hybridization data is available for genetic markers located on this arm. The next phase of our effort will be to FISH additional BACs to confirm the identity of merged contigs, to do more overgo hybridization to anchor additional contigs on the short arm, and to fill gaps between contigs.

Based on FISH analysis of 47 seed BACs, 20 BACs were selected for sequencing, and sequence information of 15 BACs was submitted to the NCBI and SGN network (Table 1).

Future directions

In order to further promote CSOL, several workshops were held in Beijing and Shanghai. Through discussions, we raised the following directions for the next step of the chromosome 3 sequencing efforts:

In order to build a strong SOL-community in China, we will invite more scientists, especially bioinformatics experts to participate in the chromosome 3 sequencing efforts and continue to seek funding support through both governmental and private channels.

Based on the current progress of the physical map, we plan to identify 200 BAC clones for sequencing. The identity of these BAC clones on chromosome 3 will be confirmed with FISH.

Table 1
Information of the 15 sequenced BACs on tomato chromosome 3

GenBank accession no.	BAC name	Size (kb)	Marker	Map position (cM)	Chromosome location
EU124730	C03HBa0001E24	119.104	T1235	160.0	Euchromatic region
EU124731	C03HBa0012D06	113.681	T0761	133.0	Euchromatic region
EU124732	C03HBa0030O03	119.955	TG244	171.0	Euchromatic region
EU124733	C03HBa0033A22	123.316	T1064	138.0	Euchromatic region
EU124734	C03HBa0034B23	126.477	T0482	137.0	Euchromatic region
EU124735	C03HBa0040F22	82.956	T1511	92.0	Euchromatic region
EU124737	C03HBa0054O21	165.852	T0196	96.0	Euchromatic region
EU124738	C03HBa0127K18	126.965	T1424	74.8	Euchromatic region
EU124736	C03HBa0143N09	190.913	SSR27	169.0	Euchromatic region
EU124739	C03HBa0233O20	117.727	T1607	105.0	Euchromatic region
EU124740	C03HBa0295I12	31.852	SSR11	164.0	Euchromatic region
EU124741	C03HBa0318C22	114.286	TG134	110.0	Euchromatic region
EU124742	C03HBa0323D22	102.153	TG284	123.0	Euchromatic region
EU124743	C03HBa0007J09	134.937	TG247	72.0	Heterochromatic region
EU124744	C03HBa0030F10	120.843	T0772	87.0	Heterochromatic region

We will continue to identify cutting edge opportunities in Solanaceae research to better meet the needs of biology and sustainable agriculture in China.

Acknowledgements

The Chinese SOL Project is financially supported by Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, and Wuhan Botanical Garden, Chinese Academy of Sciences. Changbao Li is supported by a grant from the Ministry of Science and Technology of China (No. 2006AA10A116).

References

- Borisjuk, N., Borisjuk, L., Petjuch, G., and Hemleben, V.** (1994). Comparison of nuclear ribosomal RNA genes among *Solanum* species and other Solanaceae. *Genome* **37**: 271–279.
- Budiman, M.A., Mao, L., Wood, T.C., and Wing, R.A.** (2000). A deep-coverage tomato BAC library and prospects toward development of an STC framework for genome sequencing. *Genome Res.* **10**: 129–136.
- Carland, F.M., and Staskawicz, B.J.** (1993). Genetic characterization of the *Pto* locus of tomato: Semi-dominance and cosegregation of resistance to *Pseudomonas syringae* pathovar tomato and sensitivity to the insecticide Fenthion. *Mol. Gen. Genet.* **239**: 17–27.
- Chen, S., Matsubara, K., Omori, T., Kokubun, H., Kodama, H., Watanabe, H., Hashimoto, G., Marchesi, E., Bullrich, L., and Ando, T.** (2007). Phylogenetic analysis of the genus *Petunia* (Solanaceae) based on the sequence of the *Hfl* gene. *J. Plant Res.* **120**: 385–397.
- Chung, H.J., Jung, J.D., Park, H.W., Kim, J.H., Cha, H.W., Min, S.R., Jeong, W.J., and Liu, J.R.** (2006). The complete chloroplast genome sequences of *Solanum tuberosum* and comparative analysis with Solanaceae species identified the presence of a 241-bp deletion in cultivated potato chloroplast DNA sequence. *Plant Cell Rep.* **25**: 1369–1379.
- Foolad, M.R.** (2007). Genome mapping and molecular breeding of tomato. *Int. J. Plant Genomics* **2007**: 64358–64409.
- Knapp, S.** (2002). Tobacco to tomatoes: A phylogenetic perspective on fruit diversity in the Solanaceae. *J. Exp. Bot.* **53**: 2001–2022.
- Li, L., Li, C., and Howe, G.A.** (2001). Genetic analysis of wound signaling in tomato. Evidence for a dual role of jasmonic acid in defense and female fertility. *Plant Physiol.* **127**: 1414–1417.
- Liu, H.F., Genard, M., Guichard, S., and Bertin, N.** (2007). Model-assisted analysis of tomato fruit growth in relation to carbon and water fluxes. *J. Exp. Bot.* **58**: 3567–3580.
- Lorenz-Lemke, A.P., Mader, G., Muschner, V.C., Stehmann, J.R., Bonatto, S.L., Salzano, F.M., and Freitas, L.B.** (2006). Diversity and natural hybridization in a highly endemic species of *Petunia* (Solanaceae): A molecular and ecological analysis. *Mol. Ecol.* **15**: 4487–4497.
- Mueller, L.A., Solow, T.H., Taylor, N., Skwarecki, B., Buels, R., Binns, J., Lin, C., Wright, M.H., Ahrens, R., Wang, Y., Herbst, E.V., Keyder, E.R., Menda, N., Zamir, D., and Tanksley, S.D.** (2005). The SOL Genomics Network: A comparative resource for Solanaceae biology and beyond. *Plant Physiol.* **138**: 1310–1317.
- Palmer, J.D., and Zamir, D.** (1982). Chloroplast DNA evolution and phylogenetic relationships in Lycopersicon. *Proc. Natl. Acad. Sci. USA* **79**: 5006–5010.
- Parry, S., Newbigin, E., Craik, D., Nakamura, K.T., Bacic, A., and Oxley, D.** (1998). Structural analysis and molecular model of a self-incompatibility RNase from wild tomato. *Plant Physiol.* **116**: 463–469.
- Peterson, D.G., Lapitan, N.L., and Stack, S.M.** (1999). Localization of single- and low-copy sequences on tomato synaptonemal complex spreads using fluorescence *in situ* hybridization (FISH). *Genetics* **152**: 427–439.
- Ryan, C.A., and Pearce, G.** (2003). Systemins: A functionally defined family of peptide signals that regulate defensive genes in Solanaceae species. *Proc. Natl. Acad. Sci. USA* **100** (Suppl. 2): 14577–14580.
- Stupar, R.M., Beaubien, K.A., Jin, W., Song, J., Lee, M.K., Wu, C., Zhang, H.B., Han, B., and Jiang, J.** (2006). Structural diversity and differential transcription of the patatin multicopy gene family during potato tuber development. *Genetics* **172**: 1263–1275.
- Tanksley, S.D., Ganai, M.W., Prince, J.P., de Vicente, M.C., Bonierbale, M.W., Broun, P., Fulton, T.M., Giovannoni, J.J., Grandillo, S., Martin, G.B., Messeguer, R., Miller, J.C., Miller, L., Paterson, A.H., Pineda, O., Riider, M.S., Wing, R.A., Wu, W., and Young, N.D.** (1992). High density molecular linkage maps of the tomato and potato genomes. *Genetics* **132**: 1141–1160.
- Wang, Y., van der Hoeven, R.S., Nielsen, R., Mueller, L.A., and Tanksley, S.D.** (2005). Characteristics of the tomato nuclear genome as determined by sequencing undermethylated *EcoR* I digested fragments. *Theor. Appl. Genet.* **112**: 72–84.