JOURNAL OF PURE AND APPLIED MICROBIOLOGY, Nov. 2013. Vol. **7**(Spl. Edn.), p. 687-693

Phylogenetic Relatoinships of Varieties of Tomato (*Solanum lycopersicum*) using DNA Markers

Yan-Lin Sun[1]*, Jun Pill Baek[2], Mun-Haeng Lee[3], Hee-Kyoung Lee[3], Young-Sik Kim[4], Soon-Kwan Hong[5], She-Hwan Kim[6], Hyang-Hee Hwang[6], Mohammad Zahirul Islam[7] and Ho-Min Kang[7]*

[1]School of Life Sciences, Ludong University, Yantai, Shandong - 264-025, China
[2]Agriculture and Life Science Research Institute, Kangwon National University, Chuncheon - 200 701, Korea.
[3]Buyeo Tomato Experiment Station, CARES Buyeo, 323-814, Korea.
[4]Department of Plant & Food Science, Sangmyung University, Cheonan 330-720, Korea
[5]Department of Bio-Health Technology, and Institute of Bioscience and Biotechnology, Kangwon National University, Chuncheon, 200-701, Korea.
[6]Agricultural and Life Science Research Institute, Kangwon National University, Chuncheon 200-701, Korea.
[7]Department of Horticulture, Kangwon National University, Chuncheon 200-701, Korea.

(Received: 04 March 2013; accepted: 10 June 2013)

Tomato (*Solanum lycopersicum*) is one of the major vegetable crops in the world. To satisfy demands of people, increasing varieties have been bred out. However, a clear and accurate variety identification method is invisible. In this study, we collected 26 tomato varieties, and attempted to distinguish them based on the cpDNA rbcL, nrDNA ITS and rDNA 5S region. The sequence analysis of nrDNA ITS region suggested that four groups were divided at 87% similarity. One7 bp array deletion resulted in Abstract Saenggeurin, Rikopin 9, Yo-yo Captain and Rubiking forming one group. The rDNA 5S region also showed high nucleotide variation among these varieties. The cpDNA rbcL region showed high similarity among these varieties, suggested that this region was not suitable for variety identification. Combined with sequences of these three DNA markers, more efficient and accurate variety identification was found using software DNAMAN 6.0 version. In this combined phylogenetic tree, three groups were divided at 83% similarity, and more subgroups were divided at 98% similarity. This work would help and instruct variety authentication and grouping in molecular level.

Key words: Tomato, nrDNA ITS, cpDNA rbcL, rDNA 5S, variety identification.

This document provides instructions for style and layout, information on installing the Word template and how to submit the final version. The instructions are designed for the preparation of a camera-ready and accepted paper in MS Word and should be read carefully. Recently, molecular identification based on DNA markers is widely used in systematic of plants, fungi and even animals[1,2]. Among various DNA markers, the internal transcribed spacer (ITS) region of nuclear ribosomal DNA (nrDNA) is the most widely used DNA marker in plant systematic[3]. By 1990, investigations of fungus species identification based on the ITS region sequence have been done[4], in which four universal ITS primers, including ITS1, ITS2, ITS3 and ITS4 were designed and continued to use up to now. Due to high primer universality and efficient

* To whom all correspondence should be addressed.
Tel.: +86-535-6685003 (YL Sun);
+82-33-2506425 (HM Kang);
E-mail: laddiya@hotmail.com (YL Sun);
hominkang@kangwon.ac.kr (HM Kang)
Yan-Lin Sun and Jun-Pill Baek equally contributed to this work.

amplification, the ITS gene has been used as convenient target region for molecular identification of plants, fungi, and animals[5]. This region contains two variable non-coding regions that are nested within the nrDNA repeat between the highly conserved small subunit 18S, 5.8S, and large subunit 28S nrDNA genes. The entire ITS region is generally between 600 and 800 bp, with multicopy of the rDNA repeat[6,7]. All these characteristics of the ITS region described above resulted in frequent utilization in systematic study. Another rDNA marker, 5S including 5S non-transcribed region (5S-NTS) is also considered as a DNA marker for plant species identification[3,8]. This region has been found to have discrimination potential at both infrafamilial as well as infrageneric level in *Alibertia* group[8].

As development of DNA sequencing, chloroplast genome found to have abundant DNA and obvious sequence variation, could provide basic information of evolution rate[9]. The gene for the large subunit of ribulose-bisphosphate carboxylase (rbcL) located on the chloroplast DNA (cpDNA), is an important DNA marker for phylogenetic analysis[6]. Due to its slow evolutionary rate, rbcL gene is considered to be more useful method of species identification than some biochemical methods, however, narrow nucleotide variations usually reduce low distinguishment and limit at higher taxonomic levels[10].

Tomato (*Solanum lycopersicum*) is one of the major vegetable crops grown in many countries of the world. As it has not only good taste but high nutrient value, the tomato has been considered as an economically important agricultural crop[11, 12]. At present, tomato varieties which have been applied to commercial behaviors are mostly F1 hybrid or breeding-improved variety[13]. Many characteristics have been improved by breeding approach, such as fruit size, flavor, pigmentation, storage ability, and disease resistance. In decades, more and more new tomato varieties have been registered at some competent department for seed variety management. Only taken South Korea as example, there have been approximately 260 new varieties registered at the Korea Seed & Variety Service for commercial behavior[13]. How to define whether it is a new tomato variety and identify these varieties clearly and accurately becomes a new problem. In this study, we collected 26 tomato varieties registered at the Korea Seed & Variety Service, varying in fruit shape, color, and virus resistance (Table 1). We aimed to find an efficient, accurate, rapid approach to distinguish and identify these 26 tomato varieties. Because morphological characteristics do not always allow the quantification of genotypic difference, traditional identification methods by morphology show limitations. However, molecular identification methods based on DNA markers are considered as effective tool for variety identification. Thus, in this study, we analyzed the nrDNA ITS region, cpDNA rbcL and rDNA 5S gene sequences of 26 tomato varieties, attempted to group these varieties by nucleotide variation. Based on the results, we suggested that rDNA 5S showed the highest identification capability among these three DNA markers, while cpDNA rbcL showed the lowest; in the combined phylogenetic trees constructed by cpDNA rbcL + nrDNA ITS + rDNA 5S sequences, 26 varieties were classified into five groups.

MATERIALS AND METHODS

Plant materials

Twenty-six tomato varieties registered at the Korea Seed & Variety Service were investigated in the present study. Their morphological characteristics and other traits including fruit size, color and virus resistance were shown in Table 1. Fresh mature leaves were sampled from these tomato varieties and immediately stored in liquid nitrogen condition.

Isolation of DNA, PCR amplification and sequencing

DNA extractions were performed by using the modified cetyltrimethylammonium bromide (CTAB) [14]. The ITS1-5.8S-ITS2 region was amplified using universal primers ITS1 and ITS4 in 20 µl PCR reaction [4]. PCR amplifications of the *rbc*L gene were performed using the universal primers aF (5'-3': ATGTCACCACAAACAGAGACTAAAGC) and cR (5'-3': GCAGCAGCTAGTTCCGGGCTCCA) [15]. PCR amplifications of the rDNA 5S region were performed using the universal primers 5SF (5'-3': CGGTGCATTAATGCTGGTAT) and 5SR (5'-3': CCATCAGAACTCCGCAGTTA) [16]. PCR was performed using a Gene Amp 9700 PCR system

(Applied Biosystems Incorporate, Warrington, Cheshire, UK) in 20 µl volumes. The amplification products were checked by electrophoresis through 1.0% agarose gel, and then purified before DNA sequence analysis using a QIAquick PCR Purification Kit (QIAGEN, Korea) or Gel Purification Kit (QIAGEN, Korea) according to the manufacturer's instructions. Purified PCR products were then sequenced at MACROGENE Advancing through Genomics (Korea, http:// dna.macrogen.com/kor/).

Sequence editing and alignment

Sequencing results were edited and assembled by the software DNAMAN version 6.0 (Lynnon Biosoft Corporation, USA, www.lynon.com). Analogue of our sequences and nucleotide sequence comparisons were detected with Basic Local Alignment Search Tool (BLAST) network services against databases (http://www.ncbi.nlm.nih.gov/). The multiple sequence alignment of ITS1-5.8S-ITS2 region, cpDNA rbcL and rDNA 5S gene were performed using DNAMAN version 6.0 software, respectively, to detect single nucleotide polymorphisms. Assembled sequences were deposited in the National Centre for Biotechnology Information (NCBI, http://www.ncbi.nlm.nih.gov). The NCBI GenBank accession numbers of tomato varieties investigated in this study was shown in Table 2.

RESULTS AND DISCUSSION

Despite 26 DNAs of tomato varieties were investigated in this study, there were only 19 nrDNA ITS region sequences, 14 rDNA 5S gene sequences, and 24 cpDNA rbcL gene sequences obtained (Table 2). The reason of non-amplification of some DNA templates was come down to

Table 1. Detailed information of cultivars investigated in this study, and their fruit shape, color, and virus resistance situation

No.	Species	Variety	Fruit shape	Fruit color	Virus resistance
1	*Solanum lycopersicum*	Black kiss 20	Round	Black	
2	*Solanum lycopersicum*	Black kiss 20	Round	Black	
3	*Solanum lycopersicum*	Mini Chal	Chinese date-shapped	Red	
4	*Solanum lycopersicum*	Vitamini	Chinese date-shapped	Red	
5	*Solanum lycopersicum*	Rikopin 9[1]	Chinese date-shapped	Red	
6	*Solanum lycopersicum*	Rikopin 9[2]	Chinese date-shapped	Red	
7	*Solanum lycopersicum*	Sseommeoking	Chinese date-shapped	Red	
8	*Solanum lycopersicum*	Yellow Mountain View	Chinese date-shapped	Yellow	
9	*Solanum lycopersicum*	Acrn Gold	Chinese date-shapped	Yellow	
10	*Solanum lycopersicum*	Gold Sugar	Chinese date-shapped	Yellow	
11	*Solanum lycopersicum*	Saenggeurinbichwibol	Chinese date-shapped	Grass green	
12	*Solanum lycopersicum*	Abstract Saenggeurin	Round	Grass green	
13	*Solanum lycopersicum*	Seuwiteuking	Round	Red	
14	*Solanum lycopersicum*	Cutie	Round	Red	
15	*Solanum lycopersicum*	Rubiking	Round	Red	TY Resistance
16	*Solanum lycopersicum*	Ten Ten	Round	Red	
17	*Solanum lycopersicum*	Yo-Yo Captain	Round	Red	
18	*Solanum lycopersicum*	Unicorn	Round	Red	
19	*Solanum lycopersicum*	Hoyong		Peachblow	
20	*Solanum lycopersicum*	Lang Selection Procedure		Peachblow	
21	*Solanum lycopersicum*	Rafito		Peachblow	
22	*Solanum lycopersicum*	Nice Def	Round, or oval	Deep red	TY Resistance
23	*Solanum lycopersicum*	Max Thailang	Heart-shapped	Deep red	TY Resistance
24	*Solanum lycopersicum*	Madison	Quail egg-shapped	Deep red	
25	*Solanum lycopersicum*	Campari	Golf ball-shapped	Deep red	
26	*Solanum lycopersicum*	Temptation	Flat round to round	Deep red	

factitious handling factor or universal primer dismatch. There were only two tomato materials (both belonging to Black kiss 20) which were not amplified using universal primers of any DNA markers. The cpDNA rbcL region sequences showed very narrow variations after DNA alignment analysis (data now shown). Few nucleotide substitutions in relatively long sequence length (approximately 1300 bp) resulted in very high identity rate (data now shown, about 99%). However, the high similarity of rbcL sequences would help DNA alignment of combined DNA marker sequences more accurate and good matching in right nucleotide site easier. Although four groups were divided by alignment analysis of cpDNA rbcL sequences, it could not highly affect the variety identification.

Due to high nucleotide variation, the nrDNA ITS region showed the highest distinguishment capability among these three DNA markers used in this study. The total ITS length of sequencing results arranged between 402 bp (Saenggeurinbichwibol) and 926 bp (Madison). According to the Jukes and Cantor method analysis, all 19 nrDNA ITS sequences shared 58.16% identity rate, and among different regions in the total ITS region, the ITS2 region showed the highest variation rate, playing the most important role in variety identification. For instance, there was a deletion of 7 bp nucleotide array (-CGGCAAG-) only appearing in Abstract Saenggeurin, Rikopin 9[1], Yo-yo Captain and Rubiking, but not in other varieties, inducing these four varieties forming one group with 99% identity rate (Fig. 1). Same nucleotide substitutions in the ITS1 and ITS2 region were specific in Gold sugar, Max thailand and Madison (data now shown), resulting in these three varieties were relative near

Table 2. NCBI GenBank submission information of various DNA markers of varieties investigated in this study

No.	Variety	NCBI GenBank Accession Number		
		nrDNA ITS	cpDNA rbcL	rDNA 5S
1	Black kiss 20	-	-	-
2	Black kiss 20	-	-	-
3	Mini Chal	KC213731	KC436086	-
4	Vitamini	KC213732	KC436087	KF156909
5	Rikopin 9[1]	KC213733	-	KF156910
6	Rikopin 9[2]	-	KC436088	-
7	Sseommeoking	KC213734	KC436089	KF156911
8	Yellow Mountain View	KC213735	KC436090	-
9	Acrn Gold	KC213736	KC436091	-
10	Gold Sugar	KC213737	KC436092	KF156912
11	Saenggeurinbichwibol	KC213738	KC436093	KF156913
12	Abstract Saenggeurin	KC213739	KC436094	KF156914
13	Seuwiteuking	KC213740	KC436095	KF156915
14	Cutie	KC213741	-	KF156916
15	Rubiking	KC213742	KC436096	KF156917
16	Ten Ten	-	KC436097	-
17	Yo-Yo Captain	KC213743	KC436098	-
18	Unicorn	KC213744	KC436099	KF156918
19	Hoyong	KC213745	KC436100	KF156919
20	Lang Selection Procedure	KC213746	KC436101	KF156920
21	Rafito	-	KC436102	KF156921
22	Nice Def	-	KC436103	-
23	Max Thailang	KC213747	-	KF156922
24	Madison	KC213748	-	-
25	Campari	-	KC436104	-
26	Temptation	KC213749	KC436105	-

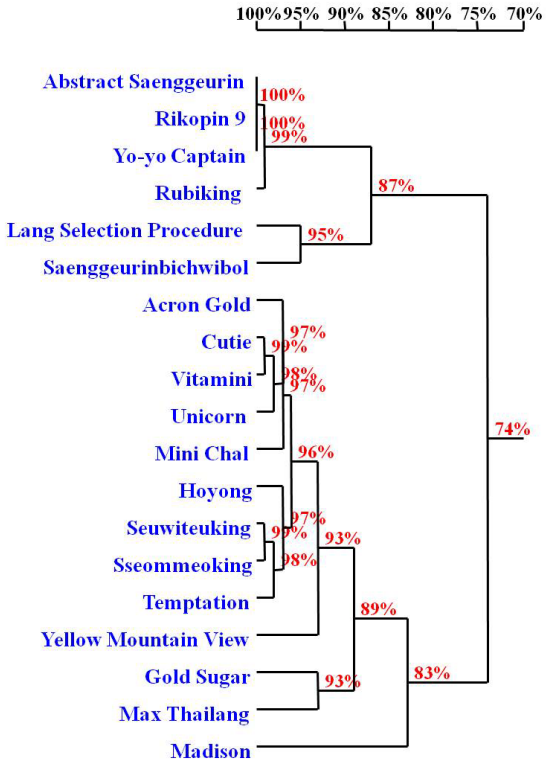

Fig. 1. Phylogenetic tree of 19 varieties investigated in this study based on the nrDNA ITS region sequence

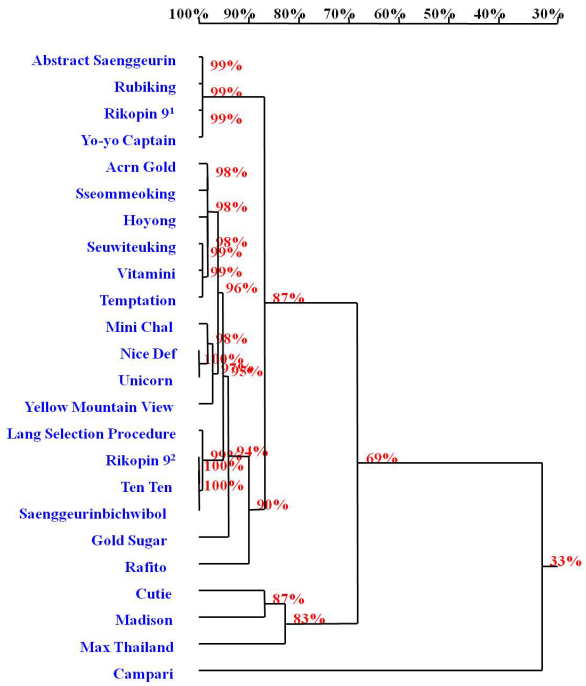

Fig. 2. Phylogenetic tree of 24 varieties investigated in this study based on combined DNA sequences (cpDNA rbcL+ nrDNA ITS + rDNA 5S). Rikopin 9^1 is combined sequences by nrDNA ITS (KC213733) and rDNA 5S (KF156910); Rikopin 9^2 is cpDNA rbcL sequence (KC436088).

in phylogenetic tree. Because of excessively long sequencing result of Madison, this variety showed respective group compared to other two varieties.

The rDNA 5S region showed relatively high conservatism, however, due to short sequencing result of this region (approximately 300 bp) its nucleotide variation rate looked high because of small denominator. According to DNA alignment result, multi-groups were divided in phylogenetic tree. And we noticed that the grouping was different from that according to the nrDNA ITS region sequence. To achieve more efficient and accurate variety identification, we combined all three DNA marker sequences investigated in this study, and constructed one combined phylogenetic tree (Fig. 2). Combined with affects on sequence length and nucleotide variation, three groups were divided, having one or more subgroups in each group. Differentiated at 83% similarity level, Campari formed one independent group alone, Cutie, Madison and Max Thailand formed another one, and the other tomato varieties formed the third one. Only cpDNA rbcL sequence was obtained from Campari DNA template, that this affected the monophyly in the phylogenetic tree (Table 2, Fig. 2). Within the biggest group, there were six subgroups divided at 98% similarity level, among which Abstract Saenggeurin, Rubiking, Rikopin 9, and Yo-yo Captain were still divided into one subgroup, also supported by analysis result of the nrDNA ITS sequence (Fig. 2). Both Rikopin varieties were divided into the same group, but not the same subgroups, suggested that other exterior factors might influence nucleotide sequence. However, in this study, both Rikopin combined sequences had no comparability, because No. 5 Rikopin sequence was combined by nrDNA ITS and rDNA 5S region, while No. 6 Rikopin sequence had only cpDNA rbcL region. Rafito was monophyletic in subgroup level, just like the result according to the rDNA 5S sequence analysis (Fig. 2).

Tomato is representatively self-pollination plant species, however, to improve its physiological and morphological characteristics, cross breeding is highly performed. The result of cross breeding is that not only some alterations of morphological traits are found, but some molecular nucleotide variations. In this study, we further identified 26 tomato varieties based on three DNA markers. Combined with their morphological characteristics, no obvious relation was found between morphological characteristic and molecular identification. Among these three DNA markers used in this study, the nrDNA ITS region played the preponderant role in variety identification, followed by the rDNA 5S and cpDNA rbcL region. The cpDNA rbcL region was considered not to adapt to species identification but not higher class identification, such as genus, family. More accurate variety identification needed further understanding their hybridization information, more variety samples and sequence analysis using more DNA markers.

In conclusion, this work advanced the method of variety identification, and provided ample theory evidences to distinguish tomato variety. This method would popularize to variety or even species identification of other crops. This work would help and instruct variety authentication and direction of breeding.

ACKNOWLEDGMENTS

This research was supported by Export Promotion Technology Development Program, Ministry of Agriculture, Food and Rural Affairs, Republic of Korea and Outstanding Young Scientist in Shandong Province Award Fund Project (No. BS2013SW021).

REFERENCES

1. Heubl, G. New aspects of DNA-based authentication of Chinese medicinal plants by molecular biological techniques. *Plant. Med.,* 2010; **76**: 1963-1974.
2. Li, M., Cao, H., But, P.P.H., Shaw, P.C. Identification of herbal medicinal materials using DNA barcodes. *J. Syst. Evol.,* 2011; **49**: 271-283.
3. Baldwin, B.G., Sanderson, M.J., Porter, J.M., Wojciechowsi, M.F., Campbell, C.S., Donoghue, M.J. The ITS region of nuclear ribosomal DNA: a valuable source of evidence on Angiosperm Phylogeney. *Ann. Mo. Bot. Gard.,* 1995; **82**: 247-277.
4. White, T.J., Bruns, T., Lee, S., Taylor, J.: Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. In: *PCR protocols-a guide to methods and applications* (Innis MA, Gelfand DH, Sninsky

JJ, White TJ, eds). San Diego, Calif: Academic Press, 1990; pp 315-322.

5. Álvarez, I., Wendel, J.F., Ribosomal ITS sequences and plant phylogenetic inference. *Mol. Phylogenet. Evol.,* 2003; **29**: 417-434.
6. Kress, W.J., Wurdack, K.J., Zimmer, E.A., Weigt, L.A., Janzen, D.H. Use of DNA barcodes to identify flowering plants. *Proc. Natl. Acad. Sci., USA,* 2005; **102**: 8369-8374.
7. Karehed, J., Groenincx, I., Dessein, S., Motley, T.J., Bremer, B. The phylogenetic utility of chloroplast and nuclear DNA markers and the phylogeny of the Rubiaceae tribe Spermacoceae. *Mol. Phylogenet. Evol.,* 2008; **49**: 843-866.
8. Persson, C. Phylogeny of the neotropical *Alibertia* group (Rubiaceae), with emphasis on the genus *Alibertia*, inferred from ITS and 5S ribosomal DNA sequences. *Am. J. Bot.,* 2000; 87: 1018-1028.
9. Clegg, M.T., Gaut, B.S., Learn, G.H.J., Morton, B.R. Rates and patterns of chloroplast DNA evolution. *Proc. Natl. Acad. Sci., USA,* 1994; **91**: 6795-6801.
10. CBOL Plant Working Group. A DNA barcode for land plants. *Proc. Natl. Acad. Sci., USA,* 2009; **106**: 12794-12797.
11. Schauer, N., Zamir, D., Fernie, A.R. Metabolic profiling of leaves and fruit of wild species tomato: a survey of the *Solanum lycopersicoides* complex. *J. Exper. Bot.,* 2005; **56**: 297-307.
12. Robertson, L.D., Labate, J.A. Genetic resources of tomato (*Lycopersicon esculentum* var. *esculentum*) and wild relatives. In: *Genetic improvement of Solanaceous crops, vol 2, tomato* (Razdan MK, Mattoo AK, eds). Enfield, NH, USA: Science Publish, 2007; pp 25-75.
13. Kwon, Y.S., Park, S.G., Yi, S.I. Assessment of genetic variation among commercial tomato (*Solanum lycopersicum* L.) varieties using SSR markers and morphological characteristics. *Genes Genom.,* 2009; **31**: 1-10.
14. Muhammad, I., Zhang, T.T., Wang, Y., Zhang, C.Y., Miao, Q., Zhang, L.J., Lin, F. Modification of CTAB protocol for maize genomic DNA extraction. *Res. J. Biotechnol.,* 2013; **8**: 41-45.
15. Hasebe, M., Omori, T., Nakazawa, M., Sano, T., Kato, M., Iwatsuki, K. *rbcL* gene sequences provide evidence for the evolutionary lineages of leptosporangiate ferns. *Proc. Natl. Acad. Sci., USA,* 1994; **91**: 5730-5734.
16. Hizume, M. Chromosomal localization of 5S rRNA genes in *Vicia faba* and *Crepis capillaries*. *Cytologia,* 1993; **58**: 417-421.