Research in Biotechnology, 3(6): 33-41, 2012 ISSN: 2229-791X
www.researchinbiotechnology.com

Regular Article

Phylogeny of some *Solanum* species (*Solanaceae*) based on complete chloroplast genomes (cpDNA) and individual chloroplast genes

Ertugrul Filiz*

Department of Crop and Animal Production, Cilimli Vocational School, Duzce University, Duzce, Turkey
*Corresponding author E-mail: ertugrulfiliz@gmail.com

Chloroplast DNA sequences can be used to estimate phylogenetic relationships among plant taxa. There are many chloroplast sequences have been made available in public databases and these data have been used for phlyogenetic relationship and evolutionary processes. We used three *Solanum* species (*Solanum bulbocastanum*, *Solanum tuberosum* and *Solanum lycopersicum*) and *Nicotiana tabacum* as outgroup. We performed phylognetic analysis four species based on complete chloroplast DNA (cpDNA), matK coding region and trnL-trnF noncoding region by using Maximum Likelihood (ML), Neighbor-joining (NJ) and unweighted pair group method with arithmetic mean (UPGMA) models. The all tree resulting from complete cpDNA clearly separated *Solanum* species with high bootstrap values and all the chloroplast segments clearly differentiated the genus *Nicotiana tabacum* from *Solanum* species. Tajima's neutrality test revealed that the highest nucleotide diversity (0.01667) was observed in trnL-trnF non-coding region. UPGMA phylogenetic inference has been shown to be more robust to ML and NJ due to bootstrap values. Particularly, ML inference has lower phylogenetic resolution than UPGMA and NJ tree.

Key words: cpDNA, phylogenetics, *Solanaceae*, *Solanum*.

Solanaceae contains between 3,000 and 4,000 species in about 90 genera, the largest of which is *Solanum* genus (Knapp et al., 2004). *Solanum* is one such giant genus, has approximately 1,250 to 1,700 species, it is the largest genus in *Solanaceae* (Frodin, 2004). *Solanum* is the most economically important genus of plants, containing important crop species such as the tomato (*S. lycopersicum*), potato (*S. tuberosum* L.), eggplant (*S. melongena*) etc. (Weese and Bohs, 2007). Despite the great wealth of data collected, the taxonomy of this section has been controversial (Zuriaga et al., 2009). Taxonomy studies based on molecular markers (Isshiki et al., 2008; Knapp et al., 2004; Stedje and Bukenya-Ziraba, 2003) and many molecular phylogenetic studies have been helped to solve phylogenetic relationships between the genera (Bohs and Olmstead, 1997; Marshall et al., 2001; Peralta and Spooner, 2001; Bohs, 2005; Bohs and Olmstead, 2007; Stern and Bohs, 2012).
Chloroplast genomes exhibit a much more stable structure than mitochondria, and also higher substitution rates can be observed (Wolfe et al., 1987). Chloroplast genes have been extensively used to reconstruct the phylogeny of related species. In order to gain more information for phylogenetic

Received: 5.10.2012; Revised: 28.11.2012; Accepted: 29.11.2012

reconstruction many chloroplast coding and non-coding regions have been analyzed (Melotto-Passarin et al. 2008). Variations of non-coding DNA regions was considered that its better phylogenetic performance when compared to coding DNA regions (Borsch and Quandt, 2009). Melotto-Passarin et al. (2008) studied the phylogenetic relationships in 16 *Solanaceae* species inferred from the plastid DNA regions trnE-trnT. Weese and Bohs (2007) reconstructed phylogenies of genus *Solanum* (*Solanaceae*) based on sequence variation of chloroplast ndhF and trnT-F, and nuclear waxy gene. In similar study, Garcia and Olmstead (2003) reconstructed phylogenies of tribe *Anthocercideae* (*Solanaceae*) based on sequence variation of two cpDNA markers ndhF and trnL-F. Bohs and Olmstead (1997) analyzed the phylogenetic relationship in *Solanum* (*Solanaceae*) based on ndhF sequences which encodes a subunit of the NADH dehydrogenase complex. Data from the study of cpDNA and nrDNA have contributed to resolve some controversial problems in *Solanaceae* systematics. In order to infer phylogenetic relationships within section *Gonatotrichum,* DNA sequence data from two nuclear regions (ITS and the granule bound starch synthase gene) and the chloroplast region trnT-F were used (Stern and Bohs, 2012). The phylogenetic relationships of *Solanum* section *Lycopersicum*, were investigated the internal transcribed spacer (ITS) region of nuclear ribosomal DNA (Marshall et al., 2001). The classification and phylogeny of the *Solanum* section *Lycopersicon* was used to characterize based on two nuclear-gene sequences, CT179 and CT66 (Zuriaga et al., 2009).

The *matK* gene encodes maturase in the land plant chloroplast genome and has high nucleotide and amino acid substitution rate provides high phylogenetic potential. Thus, *matK* is frequently employed in phylogenetic analyses at various taxonomic levels of angiosperms (Hilu et al., 2008; Hilu et al., 2003; Döring et al., 2007). The trnL-F region includes the trnL gene, an intergenic spacer and the trnF exon. The *trnL–trnF* intergenic spacer and trnL intron have been used widely for phylogenetic analysis (Cech, 1988; Drabkova et al., 2004). In this study, we investigated the question that the individual chloroplast genes (matK and trnF-trnL) provide the same phylogenetic information as compared to complete chloroplast DNA (cpDNA).

Materials and Methods

The sequences of matK and trnF-trnL of three *Solanum* species including *Solanum bulbocastanum*, *Solanum tuberosum* and *Solanum lycopersicum* were obtained from Genbank. The respective sequences of tobacco (*Nicotiana tabacum*) were used as outgroup due to its close relationship to *Solanum* genera. The details of these sequences are given in Table 1.

Table 1. Sequences of the *Solanum* species used in the comparative phylogeny

	Accession no.		
Species	**cpDNA**	**matK**	**trnL-trnF**
Solanum bulbocastanum	NC_007943	EF439049	DQ180444
Solanum tuberosum	NC_008096	HQ619841	HM006842
Solanum lycopersicum	NC_007898	HQ593449	HM006841
Nicotiana tabacum	NC_001879	HQ619839	FJ490822

The sequences were aligned by using CLUSTAL-X software (Thompson et al., 1997). The evolutionary history was investigated using the maximum likelihood (Tamura and Nei, 1993), unweighted pair group method with arithmetic mean (Sneath and Sokal, 1973) and neighbor-joining methods (Saitou and Nei, 1987). All the phylogenetic analyses were performed in MEGA software (Tamura et al., 2011). The percentage of replicate trees in which the associated taxa clustered together was determined by the bootstrap test (100 replicates) and bootstrap values of 95% or

greater be considered statistically significant and indicate "support" for a clade; alternative nodes can be rejected if they occur in less than 5% of the bootstrap estimates (Felsenstein, 1985). The sequence data were subjected to three different methods of phylogenetic reconstruction: Maximum Likelihood (ML), Neighbor-joining (NJ) and unweighted pair group method with arithmetic mean (UPGMA). For ML method, the phylogenetic analyses were performed using Tamura-Nei model and tree inference option as Nearest-Neighbor-Interchange (NNI) was used. For NJ and UPGMA method, the phylogenetic analyses were performed using Maximum Composite Likelihood model (Tamura et al., 2004). Also, Tajima's test of neutrality (Tajima, 1989), Maximum Likelihood Estimate of Transition/Transversion Bias (Kimura, 1980), overall sequence pairwise distance estimation and Maximum Likelihood Estimate of Substitution Matrix were computed (Tamura and Nei, 1993; Tajima, 1989).

Results

The complete cpDNA genome, matK and trnL-F sequences of some *Solanum* taxa were analyzed. The estimated transition /transversion bias (*R*) was found 0.85 for cpDNA, 0.56 for matK and 0.56 for trnL-trnF region. The overall pairwise distance was found 0.11 for cpDNA, 0.10 for matK and 0.11 for trnL-F region. The results of Tajima's neutrality are given in Table 2. Although, complete chloroplast genomes have 3365 segregating sites (S), trnL-trnF region has the highest nucleotide diversity (π) even containing small number segregating sites with 14. Tajima's neutrality test showed that segregating sites of matK and trnL-trnF regions were equal.

Table 2. Tajima's neutrality test for 4 taxa in *Solanaceae*.

	Number of sites (m)	Number of segregating sites (S)	Ps = S/m	Nucleotide diversity (π)	Tajima test statistics (D)
cpDNA	4	3365	0.021981	0.011990	-0.673840
matK	4	14	0.024648	0.012324	-0.845323
trnL-trnF	4	14	0.033333	0.016667	-0.845323

Table 3. The test of homogeneity of substitution patterns for different chloroplast regions

	Solanum bulbocastanum	*Solanum lycopersicum*	*Solanum tuberosum*	*Nicotiana tabacum*
Complete cpDNA				
Solanum bulbocastanum	-	0.005	0.000	0.012
Solanum lycopersicum	0.114	-	0.000	0.000
Solanum tuberosum	0.386	1.000	-	0.002
Nicotiana tabacum	0.190	1.000	0.354	-
matK				
Solanum bulbocastanum	-	0.000	0.019	0.000
Solanum lycopersicum	1.000	-	0.011	0.000
Solanum tuberosum	0.156	0.210	-	0.011
Nicotiana tabacum	1.000	1.000	0.236	-
trnL-trnF				
Solanum bulbocastanum	-	0.000	0.000	0.000
Solanum lycopersicum	1.000	-	0.000	0.000
Solanum tuberosum	1.000	1.000	-	0.000
Nicotiana tabacum	1.000	1.000	1.000	-

The test of homogeneity of substitution patterns revealed certain identities and variations in disparity index as well as Monte Carlo probability for different chloroplast markers (Table 3).

The estimates of the disparity index per site are shown for each sequence pair above the diagonal. The P values based on Monte Carlo test (100 replicates) are shown below the diagonal. *P ≤ 0.05, statistically significant. Complete cpDNA phylogenetic analysis revealed that all *Solanum* species clustered one *Solanum* monophyletic clade and have high bootstrap supports except *N. tabacum*. The NJ and ML trees showed that *Solanum lycopersicum* and *Solanum tuberosum* were grouped together generally based on matK and trnL-trnF regions except that UPGMA analysis only indicates different cluster based on trnL-trnF region with *Solanum bulbocastanum* and *Solanum tuberosum* were grouped together. All the tree analyses also showed close relationship between *Solanum* species except outgroup genus *Nicotiana tabacum*.

Discussion

The chloroplast is an essential plant organelle which makes photosynthesis and to be the basis of various metabolic reactions. Gene duplication, gene loss and gene transfer between chloroplast and nuclear genomes were a strong resource of evolution (Xiong et al., 2009). In our study, we used complete cpDNA, coding matK gene and non-coding trnL-trnF intergenic spacer region sequences and compared phylogenetic data based on different tree methods as ML, NJ and UPGMA methods. The entire complete chloroplast genome trees showed a very strong monophyletic *Solanum* clade with high bootstrap values (100%) including all tree topologies except outgroup as *Nicotiana tabacum* (Fig. 1.).

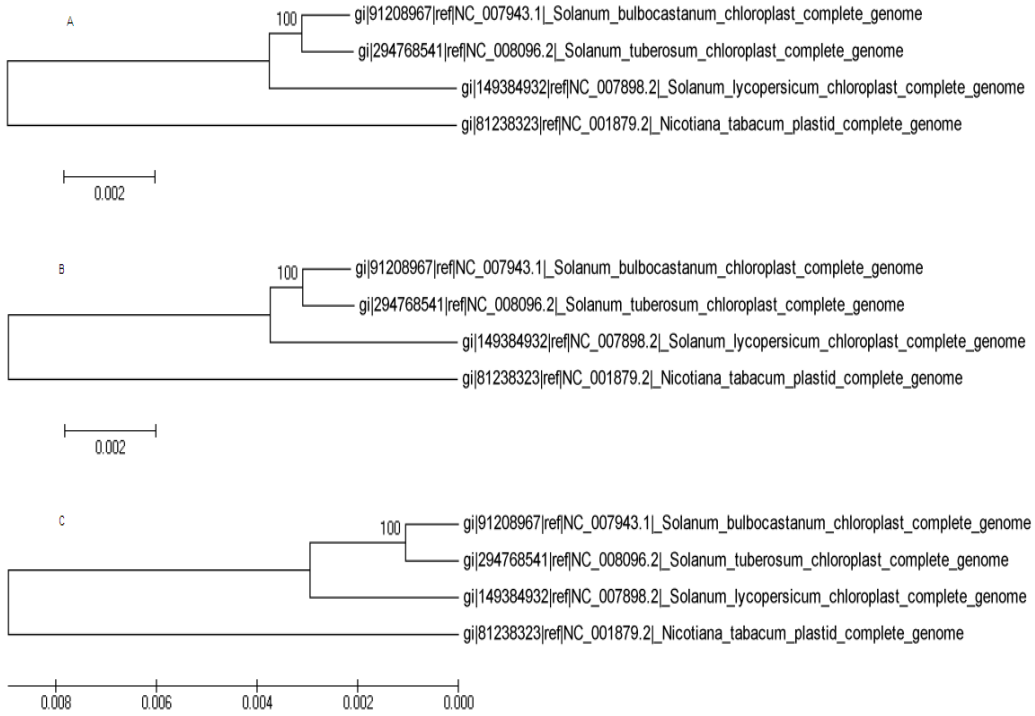

Figure 1. Phylogenetic relationship among *Solanaceae* species based on complete cpDNA Maximum-likelihood dendrogram (A), Neighbor-Joining dendrogram (B), Unweighted Pair Group Method with Arithmetic Mean dendrogram (C). Notes: Bootstrap values (expressed as percentages of 100 replications; ≥50%) are shown at branching points and GenBank gi and accession numbers are written.

Solanum tuberosum and *Solanum bulbocastanum* are same section in *Petota* while *Solanum lycopersicum* is in *Lycopersicon* section (Peralta and Spooner, 2001). Phylogenetic analysis of complete chloroplast genome corroborates this data that *Solanum tuberosum* and *Solanum bulbocastanum* were grouped in a dichotomy clade in all types of tree with high bootstrap supports (100%).

The matK gene has strong potential in providing insight into evolutionary and systematic problems at various levels (Johnson and Soltis, 1994). The results of matK gene phylogenetic analysis indicate similar tree topologies and have lower bootstrap values than complete cpDNA results (Fig. 2.). In all types of tree topologies, *Solanum lycopersicum* and *Solanum tuberosum* were grouped in a dichotomy clade and this finding is contrast to complete cpDNA inference. Bohs and Olmstead (1997) studied phylogenetic relationship in *Solanum* and revealed that *Solanum lycopersicum* and *Solanum tuberosum* were grouped in a dichotomy clade based on ndhF sequences. This finding supports our results. UPGMA phylogenetic inference has been shown to be more robust to ML and NJ due to bootstrap values. Particularly, ML inference has lower phylogenetic resolution than UPGMA and NJ tree.

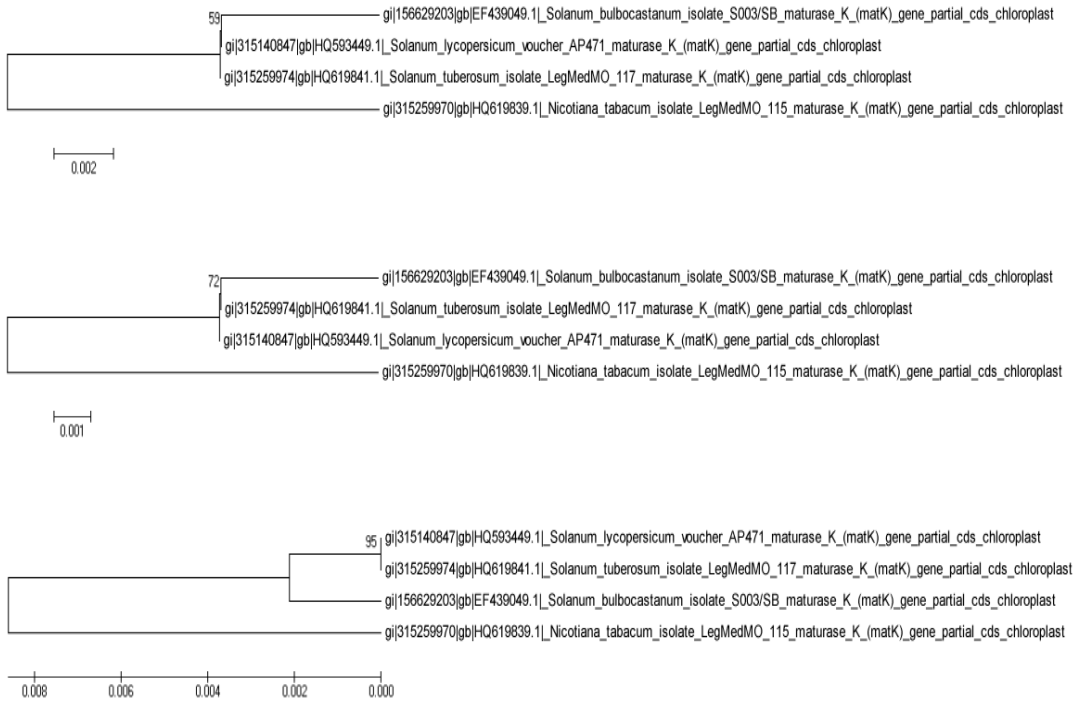

Figure 2. Phylogenetic relationship among *Solanaceae* species based on matK gene Maximum-likelihood dendrogram (A), Neighbor-Joining dendrogram (B), Unweighted Pair Group Method with Arithmetic Mean dendrogram (C). Notes: Bootstrap values (expressed as percentages of 100 replications; ≥50%) are shown at branching points and GenBank gi and accession numbers are written.

Non-coding regions in DNA are efficient source to analyze phylogenetic relationships of subspecies, varieties in modern phylogenetic studies. In cpDNA regions, trnL intron and the trnL–trnF intergenic spacer often were used in combination (Koch et al., 2006). The trnL–trnF intergenic spacer has highly variable potential in the chloroplast genome (Hilu and Liang, 1997). The phylogenetic analysis of trnL-trnF

intergenic spacer revealed that ML and UPGMA tree showed low bootstrap values 40% and 50% respectively (Fig. 3.). UPGMA tree model resulted different phylogenetic inferences that *Solanum tuberosum* and *Solanum bulbocastanum* were grouped in a dichotomy clade while NJ and ML trees were contrast to this finding that *Solanum lycopersicum* and *Solanum tuberosum* were grouped together. Melotto-Passarin et al. (2008) carried out phylogenetic study in *Solanaceae,* cpDNA trnE-trnT intergenic spacer sequences were used and *Solanum* clade indicates that *Solanum lycopersicum* and *Solanum tuberosum* were grouped together. This finding is in agreement with our results. It was expected that same section species (*Solanum tuberosum* and *Solanum bulbocastanum* in *Petota* section) would cluster together but only UPGMA tree supported this clade in spite of low bootstrap value (50%).

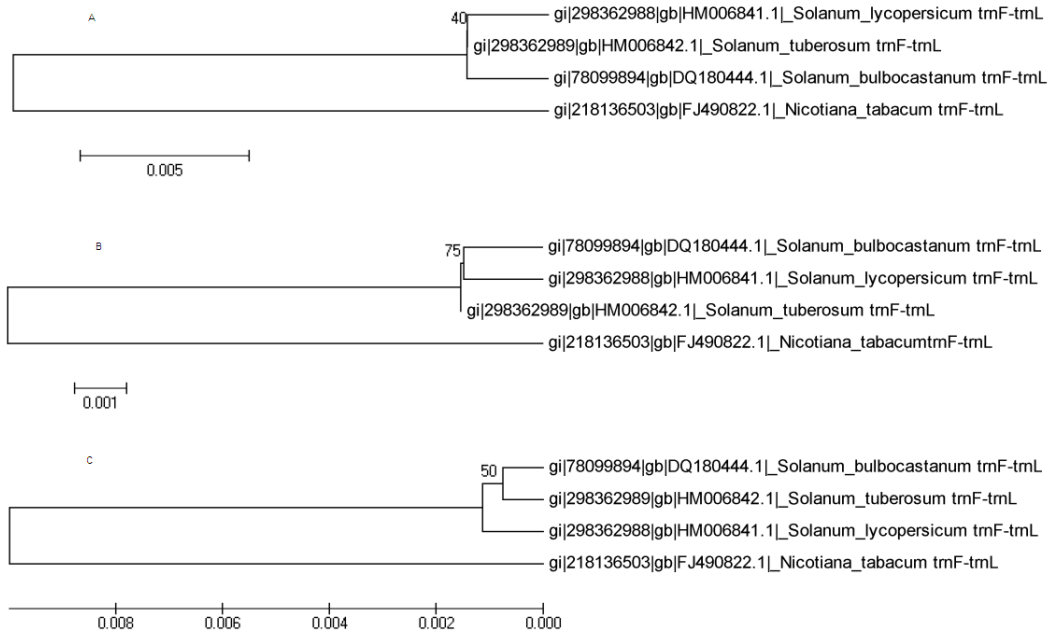

Figure 3. Phylogenetic relationship among *Solanaceae* species based on trnL-trnF intergenic spacer Maximum-likelihood dendrogram (A), Neighbor-Joining dendrogram (B), Unweighted Pair Group Method with Arithmetic Mean dendrogram (C). Notes: Bootstrap values (expressed as percentages of 100 replications; ≥40%) are shown at branching points and GenBank gi and accession numbers are written.

The study of *Solanum* genus phylogeny was performed based on chloroplast trnT-F intergenic spacer region, chloroplast ndhF and nuclear waxy gene. Phylogenetic analysis implies that *Solanum bulbocastanum* and *Solanum lycopersicum* were grouped together in all trees (Weese and Bohs, 2007). The complete chloroplast genome sequences of *Solanum* species were studied to compare with *Solanaceae* species (Chung et al., 2006; Daniell et al., 2006). *S. tuberosum* and related species (9 different species) inferred from 13 open reading frames (ORFs) in chloroplast genome sequences were phylogenetically analyzed and obtained one *Solanum* clade that *Solanum bulbocastanum* and *Solanum tuberosum* were grouped together (Chung et al., 2006). This finding is consistent with our results related to complete cpDNA phylogeny. All phylogenetic analysis support that *Nicotiana tabacum* as outgroup separated the *Solanum* clades. The lowest bootstrap values were observed in trnL-trnF intergenic spacer region (40%) while the highest bootstrap values were observed in complete cpDNA genome analysis (100%). The trees obtained from complete cpDNA sequences exhibit

greater resolution with high bootstrap support as compared to phylogeny inferred from chloroplast genes (matK and trnL-trnF). According to our results, ML trees have the lowest bootstrap values (Fig. 2 and Fig. 3.). Comparisons of sequence divergence of *Solanum tuberosum* and *Solanum lycopersicum* showed that the between noncoding and coding regions indicate that the noncoding regions are more divergent than coding regions (Daniel et al., 2006). In our analysis revealed that trnL-trnF noncodig region exhibits higher nucleotide diversity (0.01667) than coding region and complete cpDNA genome.

In conclusion, the results of our study indicated that complete chloroplast genome provides a greater resolution in phylogenetic analysis of related taxonomic groups. Complete cpDNA sequence efficiently separates four taxa of *Solanacea* including *Solanum* species while the phylogeny of *Solanum* species based on chloroplast genes (matK and trnL-trnF) failed to get identical tree topology with complete cpDNA. Cluster analysis dendrogram based on the complete cpDNA genome clearly distinguished the examined *Solanum* species. Therefore, complete chloroplast DNA sequences should be used to revealed complex phylogenetic relationship among *Solanum* taxa than individual chloroplast genes.

References

Bohs L. 2005. Major clades in Solanum based on ndhF sequence data. In: Keating RC, Hollowell VC, Croat TB (Eds) A Festschrift for William G. D'Arcy: the legacy of a taxonomist. Missouri Botanical Garden Press, St. Louis, pp. 27–49.

Borsch T, Quandt D. 2009. Mutational dynamics and phylogenetic utility of noncoding chloroplast DNA. Plant Syst Evol 282: 169–199.

Bohs L, Olmstead RG. 1997. Phylogenetic relationship in *Solanum* (*Solanaceae*) based on ndhF sequences. Systematic Botany 22(1): 5-17.

Cech TR. 1988. Conserved sequences and structures of group I introns: building an active site for RNA catalysis—a review. Gene 73: 259–271.

Chung HJ, Jung JD, Park HW et al. 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.

Daniell H, Lee SB, Grevich J. et al. 2006. Complete chloroplast genome sequences of *Solanum bulbocastanum*, *Solanum lycopersicum* and comparative analyses with other *Solanaceae* genomes. Theor Appl Genet 112: 1503–1518.

Döring E, Schneider J, Hilu KW, Röser M. 2007. Phylogenetic relationships in the *Aveneae/Poeae* complex. Kew Bull. 62: 407–424.

Drabkova L, Kirschner J, Vlcek C, Paces V. 2004. rnL–trnF Intergenic Spacer and trnL Intron Define Major Clades Within Luzula and Juncus (*Juncaceae*): Importance of Structural Mutations. J Mol Evol 59: 1–10.

Felsenstein J. 1985. Confidence limits on phylogenies: An approach using the bootstrap. Evolution 39: 783–91.

Frodin DG. 2004. History and concepts of big plant genera. Taxon 53: 753–776.

Garcia VF and Olmstead RG. 2003 Phylogenetics of Tribe *Anthocercideae* (*Solanaceae*) based on ndhF and trnL/F sequence data. Systematic Botany 28: 609–615.

Hilu KW, Liang H, 1997. The matK gene: sequence variation and application in plant systematics. Am. J. Bot. 84: 830–839.

Hilu KW, Black C, Diouf D, Burleigh JG. 2008. Phylogenetic signal in matK vs.

trnK: a case study in early diverging eudicots (angiosperms). Mol. Phylogenet. Evol. 48: 1120–1130.

Hilu KW, Borsch T, Muller K, Soltis DE, Soltis PS, Savolainen V, et al. 2003. Angiosperm phylogeny based on matK sequence information. Am. J. Bot. 90: 1758–1776.

Isshiki S, Iwata N, Khan MR. 2008. ISSR variations in eggplant (*Solanum melongena* L.) and related *Solanum* species. Scientia Horticulturae 117: 186–190.

Johnson LA, and Soltis DE. 1994. matK DNA sequences and phylogenetic reconstruction in Saxifragaceae s. str. Systematic Botany 19: 143-156.

Kimura M. 1980. A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. Journal of Molecular Evolution 16: 111-120.

Knapp S, Bohs L, Nee M, Spooner DM. 2004. *Solanaceae*- a model for linking genomics with biodiversity. Comparative and Functional Genomics 5: 285–291. doi: 10.1002/cfg.393.

Koch MA, Kiefer C, Ehrich D, Vogel J, Brochmann C, Mummenhoff K. 2006. Three times out of Asia Minor: the phylogeography of *Arabis alpina* L. (*Brassicaceae*). Molec Ecol 15: 825–839.

Marshall JA, Knapp S, Davey MR, Power JB, Cocking EC, Bennett MD, Cox AV. 2001. Molecular systematics of *Solanum* section *Lycopersicum* (*Lycopersicon*) using the nuclear ITS rDNA region. Theor Appl Genet 103: 1216–1222.

Melotto-Passarin DM, Berger IJ, Dressano K, De Martin VF, Oliveira GCX, Bock R, Carrer H. 2008. Phylogenetic relationships in *Solanaceae* and related species based on cpDNA sequence from plastid species based on cpDNA sequence from plastid trnE-trnT region. Crop Breeding and Applied Biotechnology 8: 85-95.

Peralta IE and Spooner DM 2001. Granule-bound Starch Synthase (GBSSI) Gene Phylogeny of Wild Tomatoes (*Solanum* L. Section *Lycopersicon* [Mill.] Wettst. Subsection *Lycopersicon*). American Journal of Botany 88(10): 1888–1902.

Saitou N, Nei M. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Molecular Biology and Evolution. 4: 406–25.

Sneath PHA and Sokal RR. 1973. Numerical Taxonomy. Freeman, San Francisco.

Stedje B and Bukenya-Ziraba R. 2003. RAPD variation in *Solanum anguivi* Lam. and S. aethiopicum L. (*Solanaceae*) in Uganda. Euphytica 131: 293–297.

Stern S, Bohs L. 2012. An explosive innovation: Phylogenetic relationships of *Solanum* section *Gonatotrichum* (*Solanaceae*). PhytoKeys 8: 89–98. doi: 10.3897/phytokeys.8.2199

Tajima F. 1989. Statistical methods to test for nucleotide mutation hypothesis by DNA polymorphism. Genetics 123: 585-595.

Tamura K, Nei M. 1993. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. Molecular Biology and Evolution. 10: 512–26.

Tamura K, Nei M. and Kumar S. 2004. Prospects for inferring very large phylogenies by using the neighbor-joining method. Proc. Natl. Acad. Sci. U.S.A. 101: 11030–5.

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, and Kumar S. 2011. MEGA5: Molecular Evolutionary Genetics Analysis using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. Molecular Biology and Evolution 28: 2731-2739.

Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG. 1997. The CLUSTAL_X windows interface:

flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Research. 24: 4876–82.

Weese TL, Bohs L. 2007. A Three-Gene Phylogeny of the Genus *Solanum* (*Solanaceae*). Systematic Botany 32(2): 445–463.

Wolfe KH, Li W-H, Sharp PM. 1987. Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. Proc Natl Acad Sci USA 84: 9054–9058.

Xiong AS, Peng RH, Zhuang J. et al. 2009. Gene duplication, transfer, and evolution in the chloroplast genome. Biotechnology Advances 27: 340–347.

Zuriaga E, Blanca J, Nuez F. 2009. Classification and phylogenetic relationships in *Solanum* section *Lycopersicon* based on AFLP and two nuclear gene sequences. Genet Resour Crop Evol 56: 663–678.