The rbcL Gene Sequence Variations among and within Prunus Species

S. Sarhan 1*, F. Hamed 2, and W. Al-Youssef 3

ABSTRACT

The objective of this study was to detect the level of SNP variations of rbcL gene sequences among and within Prunus species including 17 locally cultivated and wild relatives of Prunus, and two species of the subfamily Maloideae (Malus domestica and Pyrus communis), as out groups. The rbcL sequences were amplified, sequenced, and aligned to determine Single Nucleotide Polymorphisms (SNPs). The rbcL gene tree of the samples showed two main clusters. The first included the outgroup taxa (M. domestica and P. communis); and all Prunus samples in the second cluster including Prunus armeniaca, which separated in a subcluster. Our results indicate that rbcL gene sequence analysis provides a well-defined tool to study relationships within and among Prunus species, and can be successfully used in constructing reliable phylogenetic tree for Prunus accessions.

Keywords: Genetic relationship, Prunus, rbcL, Sequencing.

INTRODUCTION

The Rosaceae family contains about 100 genera and 3,000 species (Judd et al., 1999), Prunus is the largest genus in the subfamily Prunoideae (Amygdaloideae) including about 250 species (Lee and Wen, 2001), Peach (Prunus persica L.) and almond (Prunus dulcis Mill.; syn. P. amygdalus Batsch) are two commercially grown species that belong to the Prunus genus (Rehder, 1940). The two species have originated in Southeast and Central Asia, respectively (Watkins, 1979), while the wild relatives of Prunus were found growing from eastern China to the Mediterranean Sea (Browicz and Zohary, 1996). According to Rehder (1940), genus Prunus is divided into subgenera of Prunophora (Prunus), Amygdalus, Padus, Cerasus and Laurocerasus. Other authors consider them as a separate genus (McVaugh, 1951). The classification of Prunus species is complicated, since closely related taxa often differ by only a single morphological character (Shi et al., 2013). In addition, typical identification requires reproductively mature material that may be available for only a short period of the year (Julian et al., 2009).

Therefore, molecular markers became necessity to study phylogenetic relations of Prunus genus. The Single Nucleotide Polymorphism (SNP) and small Insertion and Deletion (InDels), are the most forms of genetic variations in natural populations. They reflect the results of evolution and adaptation (Yamanaka et al., 2004; Wright et al., 2005). They are frequently used in modern genetics for reverse genetics, linkage analysis, genome-wide association study, genotyping and markers assisted selection. In addition, SNP markers were developed to identify plant diseases resistance (Bakooie et al., 2015).
Several investigations were carried out to assess genetic variation and phylogenetic relationship among and within *Prunus* species using isozyme, RFLPs, RAPD's, SSRs and AFLP (Martínez-Goñi et al., 2003; Mowrey and Werner, 1990). The use of SNPs for genetic fingerprinting, parentage verification, and gene mapping was recommended to be applied in the study of genetic diversity of sweet cherry (*Prunus avium*) and was also proved to be useful in other related species within *Amygdaloideae* (Marti et al., 2012). At present, techniques for studying the molecular phylogeny and taxonomy of plants rely heavily on chloroplast genome sequence data. This is because the chloroplast genome is haploid with simple and stable genetic structure, where no or very rare recombinations take place, and universal primers can be used to amplify target sequences. In addition, the ease of PCR amplification and sequencing of chloroplast genes using universal primers facilitates phylogeny projects. The chloroplast DNA restriction sites were used to construct the phylogeny of eight cultivated members of *Prunus* (Badenes and Parfitt, 1995). On the other hand, sequences of chloroplast regions of *rbcL*, *matK*, *trnL/F*, *18S rDNA*, and *ITS* have been used in studies of *Rosaceae* and *Prunus* phylogeny (Morgan et al., 1994; Potter et al., 2001, 2003, 2007; Bortiri et al., 2001, 2006). Some of these regions were also shown to be informative within subfamilies such as *Amygdaleae* (Lee and Wen, 2001; Potter et al., 2002), and *Maloideae* (Potter et al., 2007).

The Ribulose-Bisphosphate Carboxylase/Oxygenase (RuBisCO) is the enzyme that facilitates the primary CO₂ fixation step in the Calvin cycle. The quaternary structure of the enzyme consists of 8 large and 8 small subunits. Sequences of the *rbcL* large subunit have been used to elucidate higher taxonomic relationships in the angiosperms (Olmstead et al., 1992; Chase et al., 1993; Qiu et al., 1993). The large size (more than 1,400 bp) of *rbcL* provides many characters that can be utilized in phylogenetic analysis. Additionally, the availability of conserved primers allow for rapid amplification and sequencing. The slow synonymous nucleotide substitution rate in chloroplast DNA (cpDNA) compared to nuclear genes, is another reason for the utility of *rbcL* in angiosperm phylogeny studies.

The first suggestions that *rbcL* gene sequence was appropriate in phylogenetic studies were from Ritland and Clegg (1987) and Zurawski and Clegg (1987), and small scale phylogenetic studies based on *rbcL* sequences were followed (Doebley, 1990; Kim et al., 1992). However, the first collaborative large-scale phylogenetic analysis using *rbcL* sequence data for a broad sampling of seed plants was conducted by Chase et al., (1993). The *rbcL* gene was also widely used in phylogeny studies of *Prunus* species (Morgan et al., 1994; Potter et al., 2007; Quan and Zhou, 2011), and was found to be useful for identifying variations among and within the genera and species.

The purpose of this study was to detect the level of *rbcL* sequence variations among *Prunus* species and to produce a phylogeny of genetic tree relationships among *Prunus* species based on *rbcL* gene sequence.

**MATERIALS AND METHODS**

**Plant Material**

Plant samples investigated in this study (Table 1) were provided by Cergaya and Hott Agricultural Stations which belong to the General Commission for Scientific Agricultural Researches in Syria. The samples comprised five local varieties of almond (*Prunus dulcis*), seven varieties of peach (*Prunus persica*), two genotypes of the wild relatives of *Prunus orientalis*, *Prunus korshinsky*, Hybrid GF-677 (*Prunus dulcis♂×Prunus persica♀*), one local sample of *Prunus armeniaca*, which belongs to the subgenus *Amygdalus*, one variety of each of *Malus domestica* and *Pyrus*.
Table 1. Investigated Prunus species and varieties and out group samples and their codes in this study.

<table>
<thead>
<tr>
<th>Species</th>
<th>Varieties name</th>
<th>Origin</th>
<th>Code</th>
</tr>
</thead>
<tbody>
<tr>
<td>Almonds</td>
<td>Shami Furk</td>
<td>Local</td>
<td>P.d_sh</td>
</tr>
<tr>
<td><em>Prunus dulcis</em></td>
<td>Hama34</td>
<td>Local</td>
<td>P.d_34</td>
</tr>
<tr>
<td></td>
<td>Hama47</td>
<td>Local</td>
<td>P.d_47</td>
</tr>
<tr>
<td></td>
<td>Babnis</td>
<td>Local</td>
<td>P.d_bb</td>
</tr>
<tr>
<td></td>
<td>Oja</td>
<td>Local</td>
<td>P.d_oja</td>
</tr>
<tr>
<td><em>P. persica</em> (L.) Batsch</td>
<td>Gioia</td>
<td>Italy</td>
<td>P.p_gi</td>
</tr>
<tr>
<td>Peach and Nectarine</td>
<td>Venus</td>
<td>Italy</td>
<td>P.p_ve</td>
</tr>
<tr>
<td></td>
<td>Caldesi2000</td>
<td>Italy</td>
<td>P.p_ca</td>
</tr>
<tr>
<td></td>
<td>Nectaross</td>
<td>Italy</td>
<td>P.p_ne</td>
</tr>
<tr>
<td></td>
<td>Laure</td>
<td>Italy</td>
<td>P.p_la</td>
</tr>
<tr>
<td></td>
<td>May Crest</td>
<td>Italy</td>
<td>P.p_ma</td>
</tr>
<tr>
<td></td>
<td>Pontina</td>
<td>Italy</td>
<td>P.p_po</td>
</tr>
<tr>
<td><em>Prunus armeniaca</em> L</td>
<td>Klabe</td>
<td>local</td>
<td>P.a_kl</td>
</tr>
<tr>
<td><em>Prunus korshinskyi</em></td>
<td>(Rootstock)</td>
<td>local</td>
<td>P.kor</td>
</tr>
<tr>
<td><em>Prunus orientalis</em> Mill</td>
<td>(Rootstock)</td>
<td>Local</td>
<td>P.ori1</td>
</tr>
<tr>
<td><em>Prunus orientalis</em></td>
<td>(Rootstock)</td>
<td>Local</td>
<td>P.ori2</td>
</tr>
<tr>
<td>Hybrid GF-677</td>
<td>(Rootstock)</td>
<td>Italy</td>
<td>GF-677</td>
</tr>
<tr>
<td><em>Malus domestica</em></td>
<td>Starking</td>
<td></td>
<td>M.d_st</td>
</tr>
<tr>
<td><em>Pyrus communis</em></td>
<td>Koshi</td>
<td></td>
<td>P.c_ko</td>
</tr>
</tbody>
</table>

*communis* (from the sub-family Maloideae) as out groups.

DNA Extraction, PCR Amplification, Electrophoresis and Sequencing

Young half folded leaves were used for total genomic DNA extraction based on the modified CTAB method (Doyle and Doyle, 1987). Primer pair used in PCR amplification were described by (Shokralla et al., 2010) to amplify about 0.6 Kbp of *rbcL* gene were as follow:

**rbcLaF:**
5'ATTGTCACCACAAAACAGAGACTAAAGC

**rbcLaR:**
5'GTAAAATCAAAGTCCACCCG

A total volume of 15 µL of PCR reaction mixture contained the following: 0.15 µL of 5 U µL⁻¹ Hot start *Taq* DNA Polymerase (Takara-bio, Japan), 0.25 µL of 10 mM dNTPs and 0.5 µL of 50 mM MgCl₂, 0.5 µL of each primer (10 pmol) (Vienna Biotech, Austria), 1.5 µL of 10X PCR buffer, 5 µL of 1 ng µL⁻¹ DNA sample, and sterile distilled water to adjust to final volume. PCR amplification was performed with a 96 well Veriti thermal cycler (Applied Biosystems, USA) as follows: 95°C for 5 minutes, followed by 40 cycles of 95°C for 1 minute, 60°C for 1 minute and 72°C for 1 minute, and a last elongation step at 72°C for 10 minutes. One microliter of each of the amplified PCR products was tested on 1.5% agarose gel stained with Ethidium Bromide, and visualized under UV for detecting the amplification efficiency.

PCR products were purified with the QIAquick PCR Purification Kit (Qiagen, Germany) and sequenced using the same primers by ABI Prism Big Dye Terminator Cycle Sequencing Ready Reaction Kit (Applied Biosystems, USA).

Data Analysis

The obtained sequences were subjected to BLASTn analysis by NCBI server (http://blast.ncbi.nlm.nih.gov/Blast.cgi) to detect homology of the target gene and species. Editing and assembling of
sequences were conducted using BioEdit Sequence Alignment editor (Hall, 1999). Phylogenetic and molecular evolutionary analyses were investigated using MEGA version 5 (Tamura et al., 2011). Bootstrap analyses were used to assess the robustness of the tree with 1,000 replicates (Felsenstein, 1985).

RESULTS

The primers were used on the DNAs of 19 samples, which showed complete amplifications with a clear band 600 bp which matched the amplified amplicon size, in all samples (Figure 1).

To confirm results and identify the SNPs, sequencing of \textit{rbcL} was conducted in all samples. The SNPs found in our work were both transversion, the substitution of a (two ring) purine for a (one ring) pyrimidine (A/T) and (T/G), and transition, a point mutation that changes a purine nucleotide to another purine (\(A \leftrightarrow G\)). Among \textit{P. armeniaca} and all other samples there was one transition (A/G), whereas \textit{P.dulcis} samples showed a transversion SNP (A/T).

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{image1.png}
\caption{An example for PCR amplification using primer pair of \textit{rbcL} F and \textit{rbcL} R. (M) Molecular weight markers; (a) Samples from \textit{P. dulcis} and \textit{P. persica}, and (b): \textit{Prunus} sp. and out group samples.}
\end{figure}
compared with the rest of the samples, and another clear transversion (T/G) appeared between all *Prunus* sp. and outgroup samples (Figure 2).

The computed nucleotide pair frequencies among all samples showed the presence of three SNPs: two transversions and one transition. This data analysis also showed the highest nucleotide changes in *P. armeniaca* when compared to the rest of *Prunus* input sequences. However, the analysis showed no nucleotide substitution within each of *P. persica* and *P. dulcis* (Table 2).

The average of the pairwise distance was 0.002, pairwise distance values within each of the group of *P. Persica* and the group of *P. dulcis* and the wild relatives including hybrid GF-677 was zero, whereas the value 0.002 was between *P. persica* and all other samples of *P. dulcis* and wild relatives, including hybrid GF-677. The pairwise distance between *P. armeniaca* and *P. persica* samples was 0.005. The

**Figure 2.** Partial samples of multiple sequence alignment of *rbcL* gene sequences of the samples sequenced (SNPs are indicated).

**Table 2.** Nucleotide pair frequencies among the studied sequences of the samples.  

<table>
<thead>
<tr>
<th>Samples</th>
<th>II</th>
<th>SI</th>
<th>SV</th>
<th>R</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Within all samples</td>
<td>8.00</td>
<td>1.00</td>
<td>1.00</td>
<td>0.61</td>
<td>10.00</td>
</tr>
<tr>
<td><em>P. armeniaca</em> and the rest of <em>Prunus</em> sp.</td>
<td>3.00</td>
<td>1.00</td>
<td>0.00</td>
<td>0.36</td>
<td>4.00</td>
</tr>
<tr>
<td>Wild relatives and <em>P. persica</em></td>
<td>0.00</td>
<td>0.00</td>
<td>1.00</td>
<td>0.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Wild relatives and <em>P. dulcis</em></td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>nc</td>
<td>0.00</td>
</tr>
<tr>
<td>Hybrid GF-677 and <em>P. dulcis</em></td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>nc</td>
<td>0.00</td>
</tr>
<tr>
<td>Hybrid GF-677 and <em>P. persica</em></td>
<td>1.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Within <em>P. persica</em></td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>nc</td>
<td>0.00</td>
</tr>
<tr>
<td>Within <em>P. dulcis</em></td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>nc</td>
<td>0.00</td>
</tr>
<tr>
<td>Within wild relatives</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>nc</td>
<td>0.00</td>
</tr>
<tr>
<td><em>P. dulcis</em> and <em>P. persica</em></td>
<td>0.00</td>
<td>0.00</td>
<td>1.00</td>
<td>0.00</td>
<td>1.00</td>
</tr>
</tbody>
</table>

*p II= Identical DNA nucleotide pairs; SI: Transitional Pairs; SV= Transversional Pairs, R= SI/SV.
All frequencies are averages (rounded) over all samples. Analysis conducted in MEGA5 (Tamura *et al.*, 2011).
corresponding value between \( P. \) \( \text{armeniaca} \) and all samples of \( P. \) \( \text{dulcis} \), and wild relatives, including the hybrid GF-677, was 0.007. The highest value of 0.017 was recorded between \( Malus \) \( \text{domestica} \) and the samples of \( P. \) \( \text{dulcis} \), and wild relatives, including the hybrid GF-677 (Table 3).

The aligned sequences were used to build a phylogenetic tree using the UPGMA method (Figure 3). Results indicated the presence of two main clusters. The smaller cluster contained plant species \( Malus \) \( \text{domestica} \) and \( Pyrus \) \( \text{communis} \), species of subfamily \( Pomoidae \), the other samples of \( Prunus \) genus (subfamily \( Prunoideae \)) were joined in the bigger cluster. The second main cluster was divided into two further subclusters of \( P. \) \( \text{armeniaca} \) (subgenus \( Prunus \)) alone, and all other samples of \( P. \) \( \text{persica} \), \( P. \) \( \text{dulcis} \). Wild relatives (subgenus \( Amygdalus \)) including Hybrid GF-677 joined together in another sub cluster at a high bootstrap value of 95. The subcluster of subgenus \( Amygdalus \), was divided into two branches, the first one contained all \( P. \) \( \text{dulcis} \) samples and wild relatives with hybrid GF-677, the second clad contained \( P. \) \( \text{persica} \) samples (Figure 3).

**DISCUSSION**

Our study took the advantage of using \( rbcL \) gene, taking into consideration that the coding \( rbcL \) gene is easily amplified and sequenced in most land plants and has an impact in phylogeny investigations by providing a reliable placement of a taxon into a plant family and genus (Kress and Erickson, 2007; Gyulai et al., 2012).

Sequences of \( rbcL \) marker showed several genetic differences among samples. The interspecific genetic diversity was lower than the intraspecific one, that was in contrary with Mattia et al. (2011) in their previous research on \( Lamiaceae \) member. Two SNPs in the \( rbcL \) region were detected between the three commercial oregano and the other analyzed samples. Whereas, the sequencing of the chloroplast gene \( rbcL \) of...
Eryngium spp. showed the presence of three SNPs between studied species (Jawdat et al., 2013). This result agrees, to some extent, with our results in the ability of rbcL to differentiate between species.

The results of UPGMA cluster analysis were close and agree with some other studies (Morgan et al., 1994; Shaw and Small, 2004; Potter et al., 2002, 2003, 2007; Lee and Wen, 2001) and showed that all samples gathered into groups which matched their taxonomic classification. This reflects the ability of rbcL gene to provide high discrimination of taxa at the level of subfamilies, genera, and species.

In our study, rbcL sequence variations were high between the outgroup Malus domestica and Pyrus communis (sub family Pomoideae) when compared with other taxa of Prunus (subfamily of Prunoideae), where the average pairwise distance was 0.002. The sample of P. armeniaca (sub genus Amygdalus) has been considerably differentiated by having four variation sites (Table 2) when compared with all samples of Prunus and with wild relatives (subgenus Amygdalus). Our results support Watkins (1979), Potter et al. (2003) and Mallikarjuna et al. (2004) studies.

Two groups of taxa including the wild relatives within subgenus Amygdalus formed two separate subclusters, the first one contained P. dulcis samples, the wild relatives and the hybrid GF-677. The second sub cluster contained only Prunus persica samples. Similar diversification was presented in the result of authors sharing some samples with our study (P. dulcis, P. armeniaca, P. persica and some other hybrids) using matk and SSRs, and by using CAPS (Cleavage Amplified Polymorphic Sequence) (Bouhadida et al., 2007).

The complete matching between the wild relatives P. korshenskyi and P. orientalis and the P. dulcis samples (domesticated almond), with the average of pairwise distance of 0.00, may be due to that Prunus wild relatives have been used for a long time as a source of genetic pool in almond breeding programs. However, rbcL sequence variations between Prunus dulcis samples and their wild relatives were not informative.
for their identification. Also, this result supports the results of Bortiri et al. (2001) and Shaw and Small (2004) in grouping Prunus dulcis samples with P. korshinskyi and P. orientalis.

The GF-677 rootstock is a vegetatively propagated hybrid between the Spanish almonds ‘Garfi’ [P. dulcis (Mill.) D. A., Webb] as the female parent and the North American peach ‘Nemared’ [P. persica (L.) Batsch] as the pollen donor. This hybrid has joined the group of P. dulcis and wild relatives in one major sub-cluster. This can be supported by the fact that rbcL gene is maternally inherited. Furthermore, a partial sequence may lead to a high match percentage that may not reflect an accurate identification of the query sequence (Schori and Showalter, 2011).

Reliable identification of variation below species level would provide valuable insight into subspecies ranges, and habitat differentiation, additionally, it would help in defining the important variation within species. (Kane et al., 2012).

Peach samples included in this study were vegetatively propagated. Therefore, it is not expected to obtain high variation level in a single gene locus such as rbcL. In addition, the fact is that rbcL gene is maternally inherited which makes it highly conserved within cultivars belonging to the same species and sharing the same origin of maternally chloroplast genome. The reason along the low evolutionary rate of this chloroplast gene suggests that cultivars from both species (P. persica and P. dulcis) included in this study seem to be one taxa in each group.

Finally, a good level of discrimination based on rbcL marker was observed between studied species and less or absence of variation within species. The analysis of genetic variations among groups of Prunus samples and wild relatives using rbcL gene allowed us to cluster successfully all samples supporting their morphological characteristics, and their botanical classification. Although rbcL gene sequence was not able to identify cultivars, which belonged to the same species, it is effectively variable at species level and much more detailed work is needed on the complete gene sequence. In addition, using other chloroplast genes such as matK or PCR based markers like microsatellites should also be investigated to further clarify genetic diversity of Prunus.

REFERENCES


تفییرات توالی $\text{rbcL}$ در درون و در بین گونه‌های هلوسای $\text{Prunus}$

س. سرهاش، ف. حامد، و. الیکس

چکیده

هدف این پژوهش بررسی سطح تغییرات SNP توالی $\text{rbcL}$ در درون و در بین گونه‌های هلوسای $\text{Prunus}$ سا بهبود و ۱۷ گونه کشت شده محلی و خویشاوندان وحشی هلوسای ها و دو گونه از زیرخانواده Maloideae (Malus domestica and Pyrus communis) به عنوان گروه پریت (outgroup) را در بر می‌گرفت. در انجام پژوهش، توالی یا $\text{rbcL}$ تکیه، توالی بندی و هم‌ریخت $\text{rbcL}$ (gene tree) شد تا چندیشکلی های توکانولید مجرد (SNP) تعیین شود. نتایج درخت زنی (M. domestica) مرتبه به نمونه‌ها، در خوشه اصلی را نشان می‌داد. اولین آنها شامل گروه پریت (Prunus armeniaca) می‌باشد و در خوشه دوم $\text{P. communis}$ قرار گرفتند که در یک زیرخوشه جدا شدند. نتایج چنین اشاره داشت که برای بررسی روابط درون گونه و روابط بین گونه‌ها، تجزیه توالی $\text{rbcL}$ ابزاری به دست می‌دهد که به خوبی مشخص است و می‌تواند با موافقت در ترسیم درخت نکاملی و تبار زایی برای نمونه‌های ثبت شده گونه هلوسای به کار رود.