DNA BARCODE OF *TRNH-PSBA*, A PROMISING CANDIDATE GENE FOR EFFICIENT IDENTIFICATION OF BITTER AND SWEET ALMOND AND RELATED SPECIES

Hassan, Ahmed H.M.
Genetics and Cytology Unit, Department of Genetic Resources, Desert Research Center, El-Matareya, Cairo, Egypt
E-mail: ahmedhassantic2020@yahoo.com

Chloroplast genome sequencing is becoming useful for creating various DNA barcodes. Currently, the utilization of non-coding genes is vital for the systematics and development of flowering plants' plastid DNA barcode. The current work aimed to evaluate the molecular evolution in the almond genus for taxonomic levels and applicability of the two-hotspot barcode plastid coding genes, *trnH-* and *psbA*. In this instance, 14 of subg. *Amygdalus* species of chloroplast genomes were chosen for sequence annotation to look for the highly variable coding DNA barcode sections. Twelve genera were retrieved from the database, while two genera—including the bitter and sweet haplotype—were based on the data of the present study. Results indicated two plastid genes *trnH- psbA*, with an average length of 1058 bp, has been anticipated to have the higher nucleotide diversity, was the most variable area within the coding genes in the 14 specimens tested. The phylogenetic tree was accurately drawn the monophyletic annotations, providing clear identification without overlap across species among the bitter and sweet almond. The outcomes of data suggest that the sweet almond underwent a domestication happening that originated from the wild almond, peach, and cherry. Whereas the bitter almond was integrated with the wild and cultivated peach indicated its compatibility as the main standard rootstock of peach germplasm. The results obtained here, put forward the domain of the domain of *trnH- psbA* genes as the most promising coding plastid DNA barcode among the bitter and sweet almond at low taxonomic levels.

Keywords: *Prunus dulcis*, chloroplast, DNA barcode, *trnH- psbA* gene

INTRODUCTION

Prunus is a great genus in the family Rosaceae with almost 200 species, including multiple domesticated crops such as almond, apricot, cherry, peach, and plum (Rehder, 1940 and Potter, 2012). Almonds were derived from the wild forms of *Prunus dulcis*, abundant in the Levant countries and Iran and eastern Turkey with a few in southeastern Europe, the Mediterranean region, and east into Mongolia and western China. Almond [*Prunus dulcis* (L.) Batsch] and Peach [*Prunus persica* (Mill.) D. A. Webb] are two of the three most economically main domesticates in Prunus worldwide, and portion a few associations, including precocity, perennially and genome size and organization (Baird et al., 1994 and Arús et al., 2012).

Almond is a member of subg. Amygdalus are deciduous shrubs or small trees, typically arising in mountainous areas, regularly between 1,000 and 2,500 m in elevation (Yazbek and Oh, 2013). They occur in a diversity of environments, but mostly in relatively dry habitats, and occasionally in deserts. Almonds are a familiar type of nut. It is divided into two types in the term of taste: sweet and bitter. Sweet almond is produced from one type of almond tree (*Prunus amygdalus* var. dulcis) and does not contain poisonous chemicals. Bitter almond comes from a different type of almond tree (*Prunus amygdalus* var. amara) and does contain toxic chemicals. Diamond (1997) proposed almond as an example of simple domestication, where a dominant mutation at a single gene conferring a sweet taste to the otherwise bitter and toxic kernel would result in an edible and cultivable crop. Recent progress in genomic re-sequencing found that in wild almond trees, bHLH2 binds to two genes, instigating the production of amygdalin. In sweet domestic varieties, there is a mutated version of bHLH2 that is not able to bind with the genes, thus, production of amygdalin does not occur (Sánchez-Pérez et al., 2019).
Worldwide, almond tree is a tiny tree that can grow to a height of between 4 and 10 meters. The blossoms of sweet almond trees are typically white, while the blossoms of bitter almond trees are frequently pink in color. These two types of almonds have an appearance that is comparable to one another, with the primary distinction being that bitter almonds are significantly smaller in size compared to sweet almonds. Because of its agricultural significance, almonds receive an abundant deal of attention, particularly from horticulturists and plant breeders, commonly to evaluate the potential for their use in cultivar improvement (López-Moral et al., 2020).

Due to the high degree of resemblance and monophyletic clades notion within *Prunus dulcis* and *Prunus persica*, there may be significant complications in classifying the species. A new era for the advancement of comparative plastome genome research and the finding of DNA barcode genes in almonds opened up as a result of the recent advances towards the entire genome sequence and the assembly of Chloroplast DNA (cpDNA) genome data. Fortunately, the low cost of high-throughput sequencing of the

cp genome provides chances to acquire additional genome sequences and identify specific DNA barcodes of coding and noncoding plastid genes in almond species.

The chloroplasts, which are the photosynthetic organelles in plant cells that give green plants energy, are well recognized to serve a crucial part of sustaining life (Douglas, 1990). After the nuclear genome and mitochondria, the chloroplast genome is the third-largest genome, and it encodes a significant number of crucial proteins that are essential for photosynthesis and other metabolic functions (Daniell et al., 2016). The chloroplast genomes exhibit a variety of characteristics, including maternal inheritance in the majority of angiosperms and high levels of conservation in the genome's structure and gene content (Yu et al., 2018). The quadripartite structure of the highly conserved cpDNA genome, which consists of a self-replicating circular molecule with two inverted repeats (IRs) separated by a large single-copy region (LSC) and a short single copy region (SSC), is common (Jeon and Kim, 2019). About 110–133 genes, including protein-coding genes (CDS), ribosomal RNA genes, and transfer RNA genes, make up the majority of the cpDNA. The chloroplast genome's non-coding and coding sections were suitable for systematic and evolution research since they have a diversity signature at both high and low taxonomic levels (Li et al., 2018). While the coding area is highly conserved and only appropriate for higher taxonomic levels, the noncoding region is less functionally constrained than the coding region and offers greater levels of evolutionary rate for phylogenetic and barcoding research at the subspecies level (Pervaiz et al., 2015).

According to molecular phylogenetic research, chloroplast intergenic spacers are characterized as changeable universal sequences (Taberlet et al., 1991; Shaw et al., 2007 and Dong et al., 2012**)**. In plant molecular phylogenetics, the *trnH-psbA* intergenic spacer is the most often used non-coding universal sequence (CBOL, 2009 and Hollingsworth et al., 2011). The structural changes in the barcode sequence, particularly the insertions and deletions, have accumulated and they are useful molecular genetic tools for differentiating plant species (Kress et al., 2005; Liu et al., 2012 and Mahadani and Ghosh, 2014). Universality in amplification across several plant taxa *trnH* and *psbA* which are exceptionally, preserved primers resulting in the barcode sequence being the perfect target for genotyping mixed samples (Sang et al., 1997; Tate and Simpson, 2003; Kress et al., 2005; Shaw et al., 2005; 2007 and Bruni et al., 2010). The present work describes the genotyping of barcode fragment sizes in DNA-based identification of Egyptian sweet and bitter almonds. As a diagnostic barcode, the plastid *trnH-psbA* intergenic spacer was used to distinguish between sweet and bitter almonds and its most prevalent adulterant.

MATERIALS AND METHODS

Two Egyptian edible cultivars of almond were employed in this study: a sweet cultivar and a bitter cultivar (Table 1). These two samples were collected for the natural habitats in South Sinai.

Using a plant genomic DNA extraction kit (DP305-03, Tiangen Biotech, Beijing, China), total genomic DNA was isolated from 100 mg of fresh leaves in accordance with the manufacturer's instructions. A spectrophotometer (Nanodrop2000, Thermo Fischer, USA) was used to measure the amount of DNA. Prior to usage, both the stock and the diluted parts were kept at -80 °C.

1. DNA Sequencing, Genome Assembly, and Validation

The entire DNA of the two examined samples was sequenced using the Illumina HiSeq 2500 platform. After sequencing, low-quality areas were first filtered out of the raw data to ensure that the analysis would produce the desired clean results. The almond cv. *Prunus dulcis* plastid, public full chloroplast genome (GenBank accession NC_034696) was used to map the cpDNA genome (Jansen et al., 2010). Genome assembly and alignment analyses were carried out using the Geneious R10 programme (http://www.geneious.com; Biomatters Ltd., Auckland, New Zealand).

2. Genome Annotation and Analysis

In the present study, 14 cpDNA genomes were used for annotation, including the two specimens from our materials, in addition to the 12 cpDNA genomes that were downloaded from NCBI GenBank database. However, *Pyrus spinosa* and *Malus prunifolia* were used as outgroup. Using the online tool Dual Organellar GenoMe Annotator (DOGMA), gene annotation of the 14 cpDNA genomes was carried out (Wyman et al., 2004). Initial annotation, putative starts, stops, and intron positions were determined, and then the draft annotation was inspected and corrected manually by comparison with a homologous gene with the chloroplast genome of *Prunus dulcis* (NC_034696) from the NCBI database.

3. Identification of *trnH-psbA* Gene, Sequence Editing and Alignment

The *trnH-psbA* genes were annotated from the 14 cpDNA genomes using DOGMA analysis to compare the structural sequence, and the multiple sequence alignment was carried out using MUSCLE v3.70+ fix1-2 (Edgar, 2004) and manually modified, as needed. The MEGA X programme was used to calculate the nucleotide diversity, estimated values of transition/transversion bias (R), and nucleotide substitutions (r) for each sequence (Kumar et al., 2018).

4. Phylogenetic Inference

The analysis of the consensus phylogenetic tree was achieved using 14 nucleotide sequences of *trnH-psbA* gene, including 12 species of *Prunus*

in addition to the two species for *Pyrus* and *Malus* as an outgroup (*Malus prunifolia* and *Pyrus spinosa*). Using the ClustVis online application for visualizing clustering of multivariate data, a graphic demonstration of partition among specimens is shown to provide correct perspectives on genetic diversity (Metsalu and Vilo, 2015). Using the Tamura-Nei model as a foundation, the maximum likelihood approach (ML) was used to estimate the evolutionary history (Tamura and Nei, 1993). Using MEGA X software, the maximum likelihood (ML) tree was generated. To characterize the evolutionary history of the 14 examined specimens, the bootstrap consensus tree inferred from 1000 replicates occupied and searched for the best-scoring ML tree simultaneously to represent the evolutionary history of the 14 specimens tested.

RESULTS

1. Performance of *trnH-psbA* Genes Identifications

At first, to empirically test the regions recognized as the most suitable for barcoding in the plastid coding genes of almond, automatic genome annotations were performed among the 14 cpDNA genomes of *Prunus* (Fig. 1). Along with original gene annotations analysis and our earlier data on this material (data not published), numerous uniform loci in the analyses were ignored because of insufficient identification, e.g., *rbcL*, *matK*, *ndhA*, *ycf2*, *ycf3*, *ropC1*, *rpoC2*, *rpoB*, *rps16*, *clpP*, *psbB*, *atpF*, *atpA*, *trnKUUU*, and *trnH-psbA* (data not shown). To avoid the challenges associated to a single-locus approach, this study takes on a two-locus analysis with its overlapping or intergenic spacer (IGS) and insertion/deletion as a beneficial choice in explaining closely related almond sequence differences based on the grouping of complete two protein-coding genes *trnH-psbA* of the chloroplast genome. The position of *trnH-psbA* in coding sequence CDS regions is 1058 bp and the full-length regions were varied from 1315 in *Pyrus spinosa* to 1356 bp in *Prunus yedoensi* (Table 1 and Fig. 1).

Results indicated that the full length of *trnH-psbA* gene in sweet almond and *Prunus dulicus* were 1345 and 1346 bp, respectively. While bitter almond, *Prunus mira* and *Prunus persica* showed similar length with 1329 bp. It is worth noting that the species so-called *Prunus davidiana* and *Prunus mongolica* gave similar length with 1025 bp, while the remaining four species *Prunus takesimensis*, *Prunus mume*, *Prunus kansuensis* and *Malus prunifolia* recorded a slight diver length being 1348, 1333, 1324 and 1320 bp, respectively, as shown in Table (1) and Fig. (2).

In contrast, the intergenic region was identified within the 14 specimens which harbored IGS sequence ranged from 298 to 257 bp (Table 1). Among all, *Prunus yedoensi* and *Prunus subhirtella* showed the highest IGS sequence variation harbored 298 bp, followed by *Prunus takesimensis*, sweet almond and *Prunus mume* with 290, 287 and 275 bp, respectively.

While the three specimens bitter almond, *Prunus mira* and *Prunus persica* contained the similar IGS with 271 bp in length.

In contrast, the rest four specimens of *Prunus davidiana*, *Prunus mongolica*, *Prunus kansuensis* and *Malus prunifolia* showed a slight lower length being 267, 267, 266 and 262 bp, respectively, as shown in Table (1) and Fig. (2). However, overall, the two-loci *trnH-psbA*, showed abundant sequence variation than the single locus approach due to the variant overlaps of gene annotation and intergenic regions.

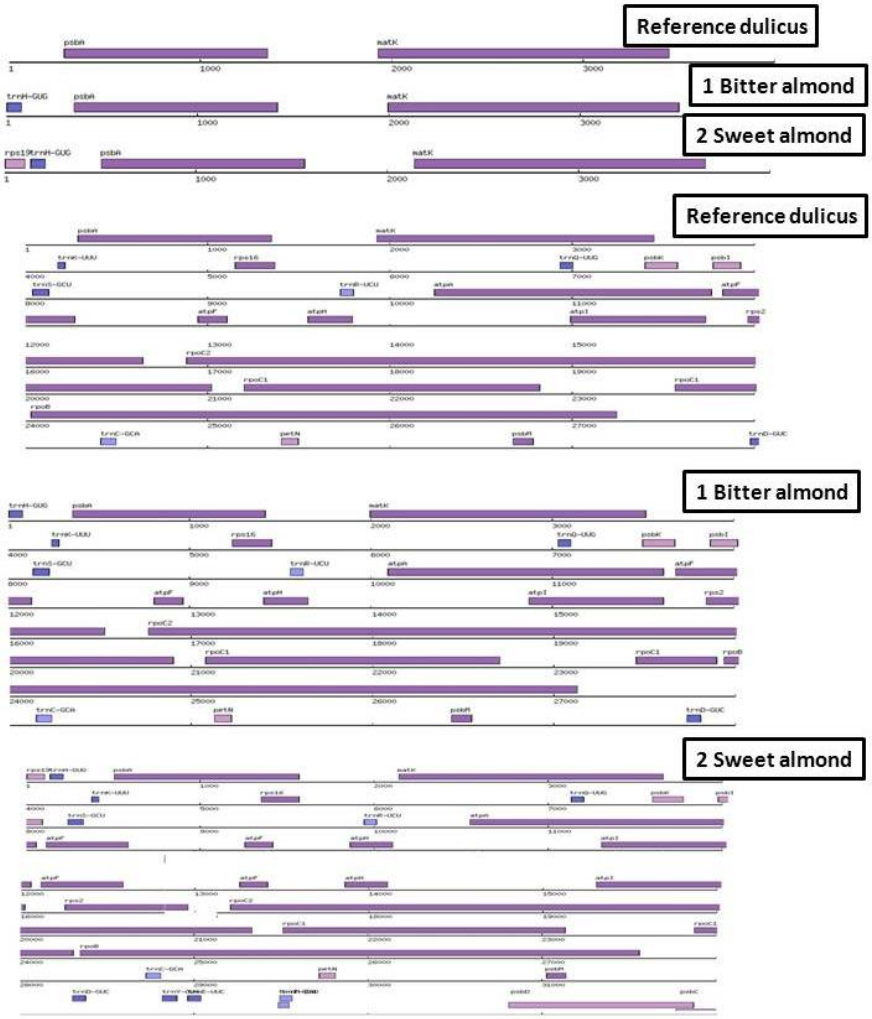

Fig. (1). Visualization of *trnH-psbA* genes within 3 species of bitter, sweet and reference using Dual Organellar GenoMe Annotator (DOGMA) where the putative starts, stops, and intron positions were annotated.

Table (1). List of sampled taxa with the *trnH-psbA* gene length and intergenic regions.

Sample code	Species name	IGS gene *trnH-psbA*	CDS gene trnH-psbA			IGS-CDS full length (bp)
			Start (bp)	End (bp)	Gene length (bp)	
1	Bitter almond	271	348	1406	1058	1329
2	Sweet almond	287	489	1547	1058	1345
3	*Prunus dulicus*	288	288	1346	1058	1346
4	*Mira*	271	348	1406	1058	1329
5	*Prunus davidiana*	267	377	1435	1058	1325
6	*Prunus mongolica*	267	367	1425	1058	1325
7	*Prunus kansuensis*	266	420	1478	1058	1324
8	*Prunus takesimensis*	290	380	1438	1058	1348
9	*Prunus mume*	275	352	1410	1058	1333
10	*Prunus yedoensi*	298	419	1477	1058	1356
11	*Prunus persica*	271	348	1406	1058	1329
12	*Prunus subhirtella*	298	419	1477	1058	1356
13	*Malus prunifolia*	262	378	1436	1058	1320
14	*Pyrus spinosa*	257	426	1484	1058	1315

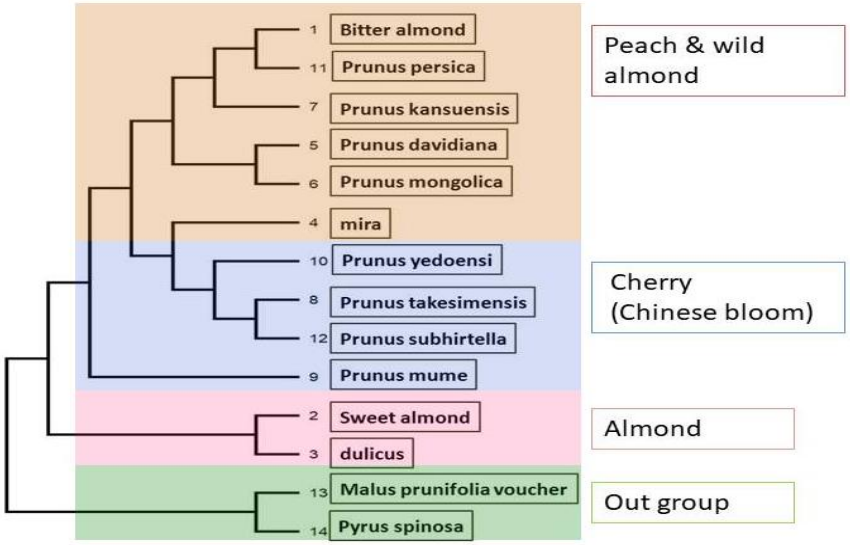

Fig. (2). Phylogenetic trees of 14 species within almond *Prunus* sp. The entire sequence dataset analyzed using maximum likelihood (ML), species group and out-group illustrates the relative genetic variability

2. Sequence Divergence of *trnH-psbA* Genes

To obtain a broad information on the *trn*H and *psb*A sequence divergence among taxa, the averages of nucleotide frequencies contents were A (28.10%), T/U (34.41%), C (18.77%), and G (18.72%) (Table 2).

To calculate the transition/transversion bias (R), the nucleotide substitution form was estimated to define the greater substitution pattern using Kimura 2- parameter analysis with five models named T92+G+I, HKY+ G+I, GTR+G+I, TN93+G+I, and K2+G+I. The rates of different transitional substitutions are shown in bold and those of transversionsal substitutions are shown in italics as presented in Table (2). The highest rate of substitutions values (r) for each nucleotide pair was discovered in *r* (GA ± 0.14) and *r* (CT ± 0.2), revealing great ranks of substitutions. In contrast, the minor values of substitution were detected within r (AC; GC; CG; TG ± 0.04), respectively (Table 3).

Table (2). Maximum composite likelihood estimates of the pattern of nucleotide substitution among 14 different nucleotide sequences.

	A	**T**	**C**	**G**
A	-	*7.34*	*4.01*	**7.85**
T	*6.00*	-	**13.30**	*4.00*
C	*6.00*	**24.39**	-	*4.00*
G	**11.78**	*7.34*	*4.01*	-

Note: Each entry shows the probability of substitution (r) from one base (row) to another base (column) (Arús et al., 2012). For simplicity, the sum of r values is made equal to 100. Rates of different transitional substitutions are shown in bold and those of transversionsal substitutions are shown in *italics*. The nucleotide frequencies are 28.10% (A), 34.41% (T/U), 18.77% (C), and 18.72% (G). This analysis involved 13 nucleotide sequences. Codon positions included were 1st+2nd+3rd+Noncoding. All ambiguous positions were removed for each sequence pair (pairwise deletion option). There were a total of 1351 positions in the final dataset

3. Phylogenetic Inference

Based on the entire sequence dataset data of *trnH-psb*A divergence, the entire species group and out-group illustrates the relative genetic variability (Fig. 2). However, the certain relationship between the 14 specimens have been tested, and the phylogenetic tree was assembled using Maximum Likelihood (ML) to define whether there is a variance dependent on the method of choice. Actually, some changes in the positioning among the 14 cultivars of various ranges were detected in the position of some specimens, and the phylogenetic based on the ML tree was the most compatible with the sequence dataset was analyzed using ML, and the

Table (3): Maximum Likelihood fits using the Kimura 2-Parameter model among 14 different nucleotide sequence for the combined locus of *trnH-psbA* genes.

Model	Invariant	R	Freq. A	Freq. T	Freq. C	Freq. G	r(AT)	r(AC)	r(AG)	r(TA)	r(TC)	r(TG)	r(CA)	r(CT)	r(CG)	r(GA)	r(GT)	r(GC)
T92+G+I	0.54	1.34	0.31	0.313	0.187	0.187	0.06	0.04	0.11	0.06	0.11	0.04	0.06	0.18	0.04	0.18	0.06	0.04
HKY+G+I	0.54	1.33	0.28	0.344	0.187	0.187	0.07	0.04	0.11	0.06	0.11	0.04	0.06	0.20	0.04	0.17	0.07	0.04
TN93+G+I	0.54	1.39	0.28	0.344	0.187	0.187	0.07	0.04	0.07	0.06	0.14	0.04	0.06	0.27	0.04	0.11	0.07	0.04
GTR+G+I	0.52	1.11	0.28	0.344	0.187	0.187	0.10	0.04	0.08	0.08	0.13	0.03	0.06	0.23	0.04	0.11	0.06	0.04
K2+G+I	0.56	1.34	0.25	0.250	0.250	0.250	0.05	0.05	0.14	0.05	0.14	0.05	0.05	0.14	0.05	0.14	0.05	0.05

previous taxonomy of *Prunus* (Fig. 2). All the 14 *Prunus* specimens were classified into four major clades with highly bootstrap value within the almond, peach and cherry. In the current revise, a graphic demonstration of the phylogenetic analysis revealed several well supported clades with strong bootstrap values. In general, results indicated three strongly haplotypes clades, supported that resolved the species into distinct branches among the *Prunus* specimens. In details, the first clade classify bitter almond was grouped with *Prunus persica* cv. Lovell formed a monophyletic clade and gathered into a common clade with *Prunus kansuensi*, *Prunus davidiana* and *Prunus mongolica* confirming a close genetic relationship to peach. The cherry group was comprised in a unifying clade, clade two involved *Prunus mira* (the oldest wild relative of peach) with the roots of three members of cherry species combining *Prunus yedoensis*, *Prunus takesimensis* and *Prunus subhirtella*, while the Chinese plum *Prunus mume* was placed independently in the basal position of the cherry clade. However, the third clade assemble sweet almond and *Prunus dulcis* with a high proportion of close relationships in the basal portion of the third clade, confirming a close genetic relationship among them. By contrast, *Malus prunifolia* and *Pyrus spinosa*, were shared individually as an out-group of the ML tree. Herein, our results suggest that the sweet almond underwent a domestication happening that originated from the wild almond, peach and cherry. Whereas the bitter almond was integrated with the wild and cultivated peach indicated its compatibility as the main standard rootstock of peach germplasm.

DISCUSSION

High-throughput sequencing has recently gained popularity, which has attracted attention. Through the comparison of whole cpDNA genomes, this method presents chances to develop a flowering plant's plastid DNA barcode that is more appropriate (Daniell et al., 2016). Although most chloroplast coding area genes lack sufficient sequence diversity to determine inter- and intraspecific connections at lower taxonomic levels of flowering plants, this issue does not arise from the non-coding regions (Dong et al., (2012). As a result, this study focused on loci that code for the *trnH-psbA* genes since they would also be the most useful coding hotspot loci in the almond *Prunus* sp. Because of their considerable diversity, *trnH-psbA* genes have been suggested in several prior papers to provide helpful information for DNA barcodes (Potter et al., 2007; Dong et al., 2012; Wang et al., 2013 and Song et al., 2015). Considering the analysis of the latest plastid data (Jeon and Kim, 2019), based on the combination of entire two protein-coding genes *trnH-psbA* of the chloroplast genome, *trnH-psbA* has two-locus analysis with its overlapping or IGS and insertion/deletion as a beneficial alternative in distinguishing closely related almond sequence variants. Contrarily, a frame shift mutation was discovered with a greater substitution rate (R), leading to multiple conservative and missense mutations. Since the replacement rate

within the genus has decreased, there is primarily very little genetic variation (Rohwer et al., 2009). It is generally known that this disparity arises because it is more likely to substitute a single ring for another single ring than it is to substitute a double ring for a single ring (Amar et al., 2014). Thses findings indicated that the DNA sequence variation had a greater transition/transversion (R) rate, with transitions arising more frequently than transversion. For investigations of molecular phylogeny, this variation in the rate of transition and transversion serves as a vital premise (Guo et al., 2017). Another remarkable characteristic is the overlapped phenomenon between the almond, peach, wild types and cherry genotypes; this is due to an unequal size difference or lack of overlapping in the expansion and contraction of the IR region (Choi et al., 2016), which point to fast-evolving events. Besides, a positive association in the IGS region was detected within this region, owing to sequence divergence in the cpDNA. Since the border of the IR region of genomes occasionally harbor insertion or deletion with a broad trend of IGS sequence, this might have led to higher IGS sequence divergence (Meng et al., 2019). This phenomenon was also reported in *Pyrus* and *Malus* (Terakami et al., 2012). Earlier study in Rosaceae (Wang et al., 2013) indicated that the non-coding regions are frequently responsible for the cpDNA genome size difference (Pervaiz et al., 2015), thus, given that higher levels of diversity and could be valuable for evolving DNA barcodes to assessment phylogeny at low level taxonomy (Daniell et al., 2016 and Wang et al., 2018).

Almond taxonomy and phylogeny are regularly the topics of controversy and the most important complication in almond breeding. Domestication is mainly of importance in subg. *Amygdalus*, but the origins of cultivation of almonds and peaches are still controversial (Yazbek and Oh, 2013). Despite its domesticated status, almond retains more genetic diversity than any of the peach species studied thus far, suggesting that mating system explains more of the differences in diversity among species than domestication. The advance of DNA barcoding in almond breeding is allowing the unmistakable molecular identification of cultivars, which is revealing cases in which a given cultivar was identified by different names or, conflicting, the use of the same name for different cultivars.

Almond and peach phylogeny are often the subjects of controversy and the major obstacle in subg. *Amygdalus* breeding. To verify the sensitivity of the phylogenetic tree, this study compared the results with the recent molecular phylogeny and evolution study (Yu et al., 2018). Earlier studies recognized *Amygdalus* as a genus that include peaches (Wen et al., 2008). Several recent phylogenetic studies of *Prunus* based the DNA barcode studies indicated that *Prunus persica* is nested within subg. *Amygdalus*, supporting the attachment of peaches in subg. *Amygdalus* (Yazbek and Oh, 2013). Herein, these results suggest that the bitter almond was nested with *Prunus persica* and their wild relatives, supports the previous finding that most of the peach cultivars originated directly from the wild relatives of *Prunus persica* and wild

almond (Amar, 2020). However, bitter almond is also still widely used as a rootstock for almond and peach breeding programs, because of its deep root system and tolerance to drought (Zrig et al., 2016). Additionally, sweet almond is widely cultivated and consumed, and it is closely related to several other important *Prunus* species cherry (Alioto et al., 2020). The presented data assume sweet almond as the most closely related to *Prunus dulcis*, and supports the hypothesis that sweet almond is of hybrid origin between different species of *Prunus* including *Prunus dulcis*, peach and cherry (Mirrahimi et al., 2011). Furthermore, the current analyses strongly support the monophyletic of *Prunus mume* as the rootstock for cherry species (Feng et al., 2018). This suggests that the introgression of the chloroplast genome from plum to cherry might have occurred (Amar et al., 2019). During the evolutionary history of a certain lineage, it is believed that one of the controversial issues raised in subg. *Amygdalus* species is the relationship between cultivated and wild taxa.

However, to this end, there are still wild populations of almond that are genetically distinct from its domesticated counterparts. These wild populations may have important genetic diversity that could be useful for breeding new varieties with improved traits, such as disease resistance or tolerance to environmental stress. As such, efforts are being made to conserve wild almond and peach populations and to study their genetic diversity. This revision was tentatively put forward that might draw the attention of other scientists who have been working on evaluating the evolutionary relationships among almond species and their relation to the *Amygdalus* species.

CONCLUSION

With the rapid progress of NGS technologies, many cpDNA genome sequences have been developed during the last two decades, which is beneficial for genome evolution and developing several DNA barcodes in plants. In the present work, targeting the chloroplast *trnH-psbA* intergenic spacer proved successful to discriminate almond from closely related species to identify minor quantities in almond. The described DNA-based assay in this study highlighted to check the resolution and sensitivity of two DNA barcoding hotspot locus *trnH-psbA* genes, which can offer a new approach to resolve the phylogeny and systematics for closely associated species in *Prunus* sp. Noteworthy, these results revealed that the -locus *trnH-psbA* was varied in length. The results of the phylogenetic tree analysis indicate that the first clade the sweet almond joined with *Prunus dulicus* and on the other hand the bitter almond separated on the other out-group.

REFERENCE

Alioto, T., K.G. Alexiou, A. Bardil, F. Barteri, R. Castanera, F. Cruz, A. Dhingra, H. Duval, Á. Fernándezi Martí, L. Frias and B. Galán

(2020). Transposons played a major role in the diversification between the closely related almond and peach genomes: results from the almond genome sequence. The Plant Journal, 101 (2): 455-472.

Amar, M.H. (2020). *ycf1-ndhF* genes, the most promising plastid genomic barcode, sheds light on phylogeny at low taxonomic levels in *Prunus persica*. Journal of Genetic Engineering and Biotechnology, 18 (1): 1-10.

Amar, M.H., A.H. Hassan, M.K. Biswas, E. Dulloo, Z.Z. Xie and W.W. Guo (2014). Maximum parsimony-based resolution of inter-species phylogenetic relationships in *Citrus* L. (Rutaceae) using ITS of rDNA. Biotechnol. Biotechnol. Equip., 28: 61–67.

Amar, M.H., M. Magdy, L. Wang, H. Zhou, B. Zheng, X. Jiang and Y. Han (2019). Peach chloroplast genome variation architecture and phylogenomic signatures of cpDNA introgression in *Prunus*. Canadian Journal of Plant Science, 99 (6): 885-896.

Arús, P., I. Verde, B. Sosinski, T. Zhebentyayeva and A.G. Abbott (2012). The peach genome. Tree Genetics and Genomes, 8: 531-547.

Baird, W.V., A.S. Estager and J.K. Wells (1994). Estimating nuclear DNA content in peach and related diploid species using laser flow cytometry and DNA hybridization. J. Amer. Soc. Hort. Sci., 119:1312–1316.

Bruni, I., F. De Mattia, A. Galimberti, G. Galasso, E. Banfi, M. Casiraghi and M. Labra (2010). Identification of poisonous plants by DNA barcoding approach. International Journal of Legal Medicine, 124: 595–603.

CBOL, Plant Working Group (2009). A DNA barcode for land plants. Proceedings of the National Academy of Sciences, 106: 12794–12797.

Choi, K.S., M.G. Chung and S. Park (2016). The complete chloroplast genome sequences of three Veronaceae species (Plantaginaceae): comparative analysis and highly divergent regions. Front. Plant Sci., 7: 355.

Daniell, H., C.S. Lin, M. Yu and W.J. Chang (2016). Chloroplast genomes: diversity, evolution, and applications in genetic engineering. Genome Biol., 17: 134.

Dong, W., J. Liu, J. Yu, L. Wang and S. Zhou (2012). Highly variable chloroplast markers for evaluating plant phylogeny at low taxonomic levels and for DNA barcoding. PloS One, 7: e35071.

Douglas, S.E. (1990). Plastid evolution: origins, diversity, trends. Curr. Opin. Genet. Dev., 8: 655–661.

Diamond, J. (1997). In: 'Guns, Germs, and Steel: the Fates of Human Societies'. New York: W.W. Norton and Company, 14 p.

Edgar, R.C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res., 32:1792–1797.

Feng, Y., T. Liu, X.Y. Wang, B.B. Li, C.L. Liang and Y.L. Cai (2018). Characterization of the complete chloroplast genome of the Chinese cherry *Prunus pseudocerasus* (Rosaceae). Conserv. Genet. Resour., 10: 85–88.

Guo, C., I.C. Mcdowell, M. Nodzenski, D.M. Scholtens, A.S. Allen, W.L. Lowe and T.E. Reddy (2017). Transversions have larger regulatory effects than transitions. BMC Genomics, 18: 394.

Hollingsworth, P.M., G.S. Wraham and D.P. Little (2011). Choosing and using a plant DNA barcode. PLoS One, 6: e19254.

Jansen, R.K., C.A. Saski, S. Lee, A.K. Hansen and H. Daniell (2010). Complete plastid genome sequences of three rosids (*Castanea*, *Prunus*, *Theobroma*): evidence for at least two independent transfers of rpl22 to the nucleus. Mol. Biol. Evol., 28: 835–847.

Jeon, J.H. and S.C. Kim (2019). Comparative analysis of the complete chloroplast genome sequences of three closely related East-Asian wild roses (Rosa sect. Synstylae; Rosaceae). Genes, 10: 23.

Kress, W. J., K.J.Wurdack, E.A. Zimmer, L.A. Weigt and D.H. Janzen (2005). Use of DNA barcodes to identify flowering plants. Proceedings of the National Academy of Sciences, 102: 8369–8374.

Kumar, S., G. Stecher, M. Li, C. Knyaz and K. Tamura (2018). MEGA X: molecular evolutionary genetics analysis across computing platforms. Mol. Biol. Evol., 35 (6): 1547–1549.

Li, Y., J. Zhang, L. Li., L. Gao., J. Xu., and M. Yang (2018). Structural and comparative analysis of the complete chloroplast genome of *Pyrus hopeiensis* "wild plants with a tiny population" and three other *Pyrus* species. Int. J. Mol. Sci., 19 (10): 3262.

Liu, J., J. Provan, L.M. Gao and D.Z. Li (2012). Sampling strategy and potential utility of indels for DNA barcoding of closely related plant species: A case study in Taxus. International Journal Molecular Sciences, 13: 8740–8751.

López-Moral, A., C. Agustí-Brisach, M. Lovera, O. Arquero and A. Trapero (2020). Almond anthracnose: current knowledge and future perspectives. Plants, 9 (8): 945.

Mahadani, P. and S.K. Ghosh (2014). Utility of indels for species-level identification of a biologically complex plant group: A study with intergenic spacer in Citrus. Molecular Biology Reports, 41: 7217–7222.

Meng, D., Z. Xiaomei, K. Wenzhen and Z. Xu (2019). Detecting useful genetic markers and reconstructing the phylogeny of an important medicinal resource plant, *Artemisia selengensis*, based on chloroplast genomics. PLoS One, 14 (2): e0211340.

Metsalu, T. and J. Vilo (2015). ClustVis: A web tool for visualizing clustering of multivariate data using principal component analysis and heatmap. Nucleic Acids Res., 43: W566–W570.

Mirrahimi, A., K. Srichaikul, A. Esfahani, M.S. Banach, J.L. Sievenpiper, C.W. Kendall and D.J. Jenkins (2011). Almond (*Prunus dulcis*) Seeds and Oxidative Stress. In: 'Nuts and Seeds in Health and Disease Prevention'. Academic Press, pp. 161-166.

Pervaiz, T., X. Sun, Y. Zhang, R. Tao, J. Zhang and J. Fang (2015). Association between chloroplast and mitochondrial DNA sequences in Chinese *Prunus* genotypes (*Prunus persica*, *Prunus domestica*, and *Prunus avium*). BMC Plant Biol., 15: 1–10.

Potter, D. (2012). Basic information on the stone fruit crops. In: 'Kole, C. and A.G. Abbott (Eds.)'. Genetics, Genomics and Breeding of Stone Fruits. CRC, New York, pp. 1–21.

Potter, D., T. Eriksson, R.C. Evans, S. Oh, J.E.E. Smedmark, D.R. Morgan, M. Kerr, K.R. Robertson, M. Arsenault and T.A. Dickinson (2007). Phylogeny and classification of Rosaceae. Plant Syst. Evol., 266: 5–43.

Rehder, A. (1940). In: 'Manual of Cultivated Trees and Shrubs Hardy in North America Exclusive of the Subtropical and Warmer Temperate Regions'. MacMillan, New York.

Rohwer, J.G., J. Li, B. Rudolph, S.A. Schmidt, H. van der Werff and H.W. Li (2009). Is *Persea* (Lauraceae) monophyletic. Evidence from nuclear ribosomal ITS sequences. Taxon, 58: 1153–1167.

Sánchez-Pérez, R., S. Pavan, R. Mazzeo, C. Moldovan, R.A. Cigliano, J. Del Cueto and B.L. Møller (2019). Mutation of a bHLH transcription factor allowed almond domestication. Science, 364 (6445): 1095-1098.

Sang, T., D. Crawford and T. Stuessy (1997). Chloroplast DNA phylogeny, reticulate evolution, and biogeography of *Paeonia* (Paeoniaceae). American Journal of Botany, 84: 1120–1136.

Shaw, J., E.B. Lickey, J.T. Beck, S.B. Farmer, W. Liu., J. Miller, R.L. Small (2005). The tortoise and the hare II: Relative utility of 21 noncoding chloroplast DNA sequences for phylogenetic analysis. American Journal of Botany, 92: 142–166.

Shaw, J., E.B. Lickey, E.E. Schilling and R.L. Small (2007). Comparison of whole chloroplast genome sequences to choose noncoding regions for phylogenetic studies in angiosperms: The tortoise and the hare III. American Journal of Botany, 94: 275–288.

Song, Y., W. Dong, B. Liu, C. Xu, X. Yao, J. Gao and R.T. Corlett (2015). Comparative analysis of complete chloroplast genome sequences of two tropical trees *Machilus yunnanensis* and *Machilus balansae* in the family Lauraceae. Front. Plant Sci., 6: 662–662.

Taberlet, P., L. Gielly, G. Pautou and J. Bouvet (1991). Universal primers for amplification of three non-coding regions of chloroplast DNA. Plant Molecular Biology, 17: 1105–1109.

Tamura, K. and M. Nei (1993). Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. Mol. Biol. Evol., 10: 512–526.

Tate, J.A. and B.B. Simpson (2003). Paraphyly of Tarasa (Malvaceae) and diverse origins of the polyploid species. Systematic Botany, 28: 723–737.

Terakami, S., Y. Matsumura, K. Kurita, H. Kanamori, Y. Katayose, T. Yamamoto and H. Katayama (2012). Complete sequence of the chloroplast genome from pear (*Pyrus pyrifolia*): genome structure and comparative analysis. Tree Genet. Genomes, 8 (4): 841–854.

Wang, S., C. Shi and L.Z. Gao (2013). Plastid genome sequence of a wild woody oil species, *Prinsepia utilis*, provides insights into evolutionary and mutational patterns of Rosaceae chloroplast genomes. PLoS One, 8: e73946.

Wang, J., C. Li, C.Yan, X. Zhao and S. Shan (2018). A comparative analysis of the complete chloroplast genome sequences of four peanut botanical varieties. PeerJ., 6: e5349.

Wen, J., S.T. Berggren, C.H. Lee, S. Ickert-Bond, T.S. Yi, K.O. Yoo, L. Xie, J. Shaw, D. Potter (2008). Phylogenetic inferences in *Prunus* (Rosaceae) using chloroplast *ndhF* and nuclear ribosomal ITS sequences. J. Syst. Evol., 46: 322–332.

Wyman, S.K., R.K. Jansen and J.L. Boore (2004). Automatic annotation of organellar genomes with DOGMA. Bioinformatics, 20: 3252–3255.

Yazbek, M. and S.H. Oh (2013). Peaches and almonds: phylogeny of *Prunus* subg. *Amygdalus* (Rosaceae) based on DNA sequences and morphology. Plant Systematics and Evolution, 299: 1403-1418.

Yu, Y., J. Fu, Y. Xu, J. Zhang, F. Ren, H. Zhao, S. Tian, W. Guo, X. Tu, J. Zhao and D. Jiang (2018). Genome re-sequencing reveals the evolutionary history of peach fruit edibility. Nat. Commun., 9: 5404.

Zrig, A., T. Tounekti, H. BenMohamed, H. Abdelgawad, A.M. Vadel, D. Valero and H. Khemira (2016). Differential response of two almond rootstocks to chloride salt mixtures in the growing medium. Russian Journal of Plant Physiology, 63: 143-151.

باركود الحامض النووي لجين *trn H- psb A* الواعد لتعريف كل من اللوز الحلو والمر والأنواع ذات الصلة

أحمد حسن محمد حسن
وحدة الوراثة، قسم الأصول الوراثية، مركز بحوث الصحراء، القاهرة، مصر

أصبحت تقنية تتابعات جينوم الكلوروبلاست لها أهمية في إنشاء العديد من ترميز الحامض النووي DNA. يعد حاليًا استخدام الجينات غير المشفرة أمرًا هامًا جدًا في إعادة تقسيم النباتات على أساس جزيئي بالإضافة إلى إعادة تطوير تقنيات ترميز الحامض النووي. ولذا فإن الغرض من العمل الحالي هو إضافة طرق تصنيفية جديدة للتطور الجزيئي في جنس اللوز باستخدام جينات الترميز المتواجدة في جينوم الكلوروبلاست مثل *trnH-* و *psbA*. تم عرض نتائج استخدام 14 تتابع لأنواع مختلفة لجينوم الكلوروبلاست من جنس البرونوس والبحث عن أنواع الباركود ذات الترميز العالي للحمض النووي. تم تجميع ١٢ جنس من قاعدة البيانات NCBI بالإضافة إلى الطرازان الوراثيان اللوز الحلو والمر، إلى أن اثنين من جينات الكلوروبلاست *trnH- psbA*، بمتوسط طول حوالى ١٠٥٨ قاعدة نيكليوتيدية، من المتوقع أن يكون لهما أكبر تنوع نيوكليوتيدات، مما يمثل المنطقة الأكثر تنوعًا ضمن جينات الترميز في جنس البرونوس. كان أثر الانتقال / والانقلاب يمثل (R = 1.3). نظرًا لتغير الجين من حيث الإضافة أو الحذف الكامل لتسلسل IGS، أظهر نطاق شكل *trnH- psbA* اتجاهًا إيجابيًا يمكن ملاحظته في التباين الهيكلي بين العينات الـ 14 التي تم تقييمها. تم استخدام الأنواع الأحادية من اللوز بدقة واستخدام بيانات شجرة ML، مما يوفر تحديدًا واضحًا دون وجود تداخل بين أنواع اللوز المر والحلو. النتائج التي تم الحصول عليها أشارت إلى أن جينات *trnH- psbA* يعتبران ذات أهمية للDNA المشفر في الكلوروبلاست ويمكن التمييز بين اللوز المر والحلو عند مستويات تصنيفية منخفضة.