Detection of SNPs and validation of a SFP InDel (deletion) in inverted repeat region of the *Prunus* species chloroplast genome

B. Bielsa[1], D. Jiwan[2], A. Fernandez i Marti[3], A. Dhingra[2,] M.J. Rubio-Cabetas[1*]

[1]Unidad de Fruticultura CITA – Gobierno de Aragón, Avda. de Montañana 930, 50059 Zaragoza, SPAIN

[2]Department of Horticulture, Washington State University Pullman, Washington 99164, USA

[3]Lab. de Mejora Genética y Biología Molecular, Parque Científico Tecnológico Aula Dei (PCTAD), Av. de Montañana, 930, 50059 Zaragoza, SPAIN

[*]Corresponding author: mjrubioc@aragon.es

ABSTRACT

In order to control tree size, disease, precocity and stress most *Prunus* varieties are cultivated as composite plants grafted onto desirable rootstocks that impart all the afore-mentioned traits. Several *Prunus* rootstock breeding programs have been focused on the production of interspecific hybrids. The pedigree of most of these rootstocks remains unknown due to the lack of parental information necessitating the application of DNA-based knowledge in breeding programs. The amplification and sequencing of the chloroplast inverted repeat B (IRB) region spanning 25,960 bps from *P. cerasifera* (myrobalan plum) Ehrh., *P. amygdalus* (almond) and *P. persica* (peach) using the ASAP method revealed a single nucleotide polymorphisms (SNP) in the *rps*19-*rpl*2 IRB region in myrobalan when compared to almond and peach. In addition, a prominent and an easily identifiable single feature polymorphism (SFP-InDel (deletion)) of 18 nucleotides was discovered in reference to the peach chloroplast genome in the *ycf*1 gene in the IRB region. In this work, it has been developed a highly useful polymorphic molecular marker to characterize the maternal parent in interspecific hybrids of *Prunus* rootstocks as a first step towards developing pedigree information. The *ycf*1 SFP-InDel (deletion) has been successfully used in several 3-way hybrids generated in the stone fruit rootstock breeding program for the characterization of new interspecific plant material. This SFP is expected to be highly utile in characterizing the maternal lineage of *Prunus* hybrids in other breeding programs.

Keywords: ASAP, Rootstock, INDEL, Phylogenetic analysis, SFP, SNP.

1. **INTRODUCTION**

Prunus is a diverse genus including approximately 200 species with most of them growing in the temperate zone and some in the tropical and subtropical regions. This genus is economically important due to its diverse uses as fruit (plums, peaches, apricots, cherries, and almonds), oil, timber, and ornamentals (Lee and Wen, 2001). In addition, several species of *Prunus* (*P. amygdalus* Batsch; *P. persica* (L.) Batsch*; P. cerasifera* Ehrh.*; P. domestica* L.; *P. insititia* L.; and their hybrids, etc.) are used as rootstocks (Serrano et al., 2002; Lecouls et al., 2004; Felipe, 2009). Rootstocks are responsible for water and nutrient uptake, resistance to soil-borne pathogens, tolerance to environmental stresses, to name a few of the important traits (Layne, 1987). Currently, the aim of several stone fruit rootstock breeding programs is the production of interspecific hybrids. Commercial *Prunus* rootstocks that are a result of uncontrolled interspecific pollinations are available on the market. However, the pedigree of most of the clones is unknown due to the lack of parental information, and this can be a major constraint for their use in breeding programs. Thus, there is a need to develop pedigree information especially to draw upon novel genotypes for important traits. Prior to that, there is a need for classifying existing hybrid rootstocks that are currently available in the programs and in commercial use. One rapid method to categorize existing rootstocks is to establish maternal lineages using chloroplast-based polymorphisms since chloroplast genome is maternally inherited in *Prunus* (Panda et al., 2003). Chloroplasts are plant organelles with their own genome containing genes coding for transcription, translation machinery and components of the photosynthetic complex (Tangphatsornruang et al., 2001). Organelle genomes are typically non-recombinant, uniparentally inherited and effectively haploid. In angiosperms, the genome is circular with a quadripartite structure that includes two copies of an inverted repeat (IR) region (~25 kb), and separately, one large single copy (LSC) region (~90 kb) and one small single copy (SSC) region (~20 kb) (Provan et al., 2001). In *Prunus*, the IR region has a size of 26,381 bp, whereas the LSC and the SSC have a size of 85,969 and 19,060 pb respectively (Jansen et al., 2011). Furthermore, chloroplast genome is extremely well conserved in size, gene arrangement, and coding sequence, at least within major subgroups of the plant kingdom (Jansen et al., 2011; Odintsova and Iurina, 2003) This feature, exploited by the Amplification, Sequencing & Annotation of Plastomes (ASAP) method (Dhingra and Folta, 2005) allows for obtaining complete coverage of many higher plant plastid genome regions, and even substantial coverage from distant genera, rapidly and inexpensively, and enables identification of polymorphisms between different genotypes.

Genotypic differences within a germplasm collection can be tracked using single nucleotide polymorphisms (SNPs) and insertion/deletions (InDel). In most cases, the InDel occur as a consequence of slippage of a template or daughter strand at the replication fork (Lovett, 2004; Terakami et al., 2012). These polymorphisms represent the most frequent mutations found in

eukaryotic genomes (Galeano et al., 2009) SNPs have been widely used in a variety of research areas, such as association studies (Ohnishi et al., 2001), biodiversity assessment (van Tienderen et al., 2002) and genetic map construction (Batley and Edwards, 2007). The popularity of SNPs as valuable and efficient molecular markers has increased as these have been demonstrated to be the most abundant of all the classes of molecular markers (Gupta et al., 2001; Hayashi et al., 2004). A number of DNA marker systems have been developed in genetic and breeding studies. Molecular markers such RFLPs, AFLPs were used for phylogenetic relationships into *Prunus* genus. Furthermore, SSRs located in the LSC and SSC regions of the chloroplast genome (Ohta et al., 2005; Decroocq et al., 2004) have been developed. Finally plastid *ndhF* sequences have been used for phylogenetic studies within *Prunus* (Chin et al., 2010; Yazbek et al., 2013. The aim of this on-going work is to characterize the myrobalan plum genotypes and almond x peach hybrids using Inverted Repeat B (IRB) chloroplast region.

2. MATERIALS AND METHODS

Eight *Prunus* genotypes located at CITA (Zaragoza, Spain) were studied in this work: Almond (*P. amygdalus* Batsch), peach [*P. persica* (L.) Batsch], the plum 'Marianna 2624' (*Prunus cerasifera* Ehrh. x *P. munsoniana* W. Wight & Hedrick), the myrobalan plums (*P. cerasifera* Erhr.) Myrobalan '29C' and Myrobalan 'P.2175', and the almond x peach hybrids [*P. amygdalus* Batsch x *P. persica* (L.) Batsch] 'GF-677', 'Felinem', 'Garnem' and 'Hansen 536'. These genotypes are used as control rootstocks for several interspecific crossing for introgression of gene tolerance of abiotic stresses.

After DNA isolation, amplifications of cpDNA were done using ASAP-PCR method with 27 primer pairs (Dhingra and Folta, 2005). Two microliters of amplicons from the first PCR was used in a second amplification (Nested-PCR). The used primers were the same as those used in the previous amplification. After, the PCR products were treated with ExoSAP-IT®, and 30 ng of each amplified product were sequenced by ABI 3130 sequencer using Bigdye® terminator v3.1. For confirming the results, the PCR product of three different PCRs were used for sequencing. Sequences generated were aligned and assembled using SEQMAN II (Lasergene, DNAstar, Inc. Madison, WI, USA), and by comparing generated sequences against the chloroplast genome sequence from IRB region of *Prunus persica* (GenBank accession number DQ 768222.1). To verify the ycf1 InDel (deletion) polymorphism a pair of primers FIn1/RIn1 and a reverse primer 27RIn2 within the InDel (deletion) sequence, for use with the forward primer IRB27F (Dhingra and Folta, 2005) was designed with the PRIMER 3 software. The new primers were used to amplify DNA of genotypes analyzed by ASAP-PCR method using the same conditions as described earlier (Dhingra and Folta, 2005). The amplicons were verified by both agarose gel and by capillary electrophoresis, using the ABI 3130.

3. RESULTS

3.1. Identification of a SNP in the IRB region

The amplification and sequencing of the Inverted Repeat B region (IRB) using the ASAP approach and the analysis of the sequences with SEQMAN II Lasergene® software (DNAstar Inc., WI, USA) revealed single-nucleotide polymorphism (SNP) in the amplicons obtained using primers IRB1F and IRB1R (Dhingra and Folta, 2005) that amplify *rps*19-*rpl*2 intergenic region in the IRB region of peach chloroplast genome. Nucleotide position number 221 in the consensus sequence in almond and almond x peach hybrids 'GF-677', 'Felinem' and 'Garnem' was a guanine (G) while in myrobalan plum genotypes and peach an adenine (A) was present at that position.

3.2. Identification of a SFP-InDel in the IRB region

A SFP-InDel (deletion) spanning 18 nucleotides was identified in the *ycf*1 gene of the IRB region using the primer combination IRB27F/IRB27R (Fig. 1A), which was present in Myrobalan '29C' and Myrobalan 'P.2175' when compared to almond, peach and their hybrids 'GF-677', 'Garnem' and 'Felinem'. This was further confirmed by ASAP-PCR with IRB27F/27RIn2 primers (Fig. 1B) where an amplicon of 850 bp fragment was obtained in the almond x peach hybrids 'GF-677' and 'Hansen 536', whereas no amplicon was obtained in the myrobalan plum 'Myrobalan 29C' and 'Marianna 2624' plum since primer 27RIn2 anneals to the deleted region (Fig. 1A). Further, FIn1/RIn1 primers, that anneal in the vicinity of the deleted region, yielded an amplicon of 182 bp in almond, peach and their hybrids, while a 164 bp amplicon was obtained in myrobalan plum genotypes, confirming the absence of the InDel (Fig. 1C) in the *Amygdalus* subgenera but present in the *Prunophora* subgenera. Several additional hybrids of different origins were ASAP-PCR analyzed with IRB27F/27RIn2 primers. These results were further validated by capillary electrophoresis with FIn1/RIn1 and fluorescent primers, indicating a clear classification of the hybrid rootstocks whose maternal lineage had been derived theoretically from the *Amygdalus* subgenera, and from the *Armeniaca* section (Fig. 2). An amplicon indicating an absence of the InDel (Fig. 2) was found in all genotypes belonging to the *Euamygdalus* Schneid section, as well as to 10 genotypes belonging to the *Armeniaca* (Lam.) Koch section, and 3 genotypes from the *Euprunus* Koehne section, 'Apricot Plum', 'Japanese plum' and the *P. ursina* accesion. In contrast, most genotypes representing a plum maternal lineage did not show this delection, including the 22 accessions from the *Euprunus* Koehne section, the two accessions from the *Prunocerasus* Koehne section, and the two genotypes of *P. brigantiaca* from the *Armeniaca* (Lam.) Koch section (Fig. 2).

4. DISCUSSION

The presence of several SNPs in the *rps*19-*rpl*2 IRB region when using the primers IRB1F/IRB1R and the absence of InDel (deletion) in the *ycf*1 region of the chloroplast genome were consistent in almond, myrobalan peach and almond x peach hybrids. The deletion is present in plums confirming the cross direction in most of the hybridizations because of the maternal inheritance of these polymorphic features.

The results indicate the usefulness of the new primers in phylogenetic studies within *Prunus* genus as well as its use for a marker assisted selection (MAS) of the parents. The effectiveness of these new markers, validated by capillary electrophoresis and evaluation of different hybrids of varied origins and subgenera, have confirmed in which subgenera of *Prunus* represent the maternal lineage. Thus, we believe the InDel (deletion) reported in this work is expected to have important applications in systematic and evolutionary biology such as elucidating the maternal origin of domesticated *Prunus* species. Furthermore, these polymorphisms have been used to identify different interspecific hybrids and confirm the direction of crosses between species. It is worth investigating the hybridizations between plums and almond x peach hybrids, which belong to two different subgenera, and the hybrids between domesticated and wild species within *Prunus* genus.

This delection could be very useful as an intraspecific DNA marker in *Prunus*, especially in rootstock breeding programs, since it was only present in the apricot cultivars and in some accessions related to apricot, including the Afghan apricot, a suspected hybrid between apricot and myrobolan plum. Thus, the InDel (deletion) had to be inherited from the apricot parent, consequently the female parent, because not even a single myrobalan showed the presence of the InDel (deletion). The other genotypes also showing this InDel such as Apricot plum and Japanese plum, could also be hybrids between the two subgenera. Although they have been classified within *Euprunus*, they may be genetically closer to Apricot. Plastid inheritance has already been applied to change the classification of some *Prunus* species, as it happened when considering *P. manchurica* within *Prunus* (Chin et al 2010), and when confirming the classification of P. *amygdalus* and *P. persica* within the same phylogeny. Consequently, we have considering as valid Rehder´s classification for our broad group of species, taking into account the *Prunophora* subgenus and not the *Prunus* classification by Yazbek and Oz (2013).

It is noteworthy, that the chloroplast genome region where these polymorphisms have been identified is distinct from the regions where most of the chloroplast microsatellites (SSRs) have been reported previously (Provan et al., 2001; Ohta et al., 2005; Decroocq et al., 2004). Thus, the new designed markers are proposed to be used as candidate markers to identify false hybrids in progenies and determining the maternal inheritance of interspecific crosses and their offspring in addition to all new information becoming available on the genome of the whole *Prunus* genus (Koepke et al, 2013). Therefore, the presence of this InDel (deletion) may

indicate a point of divergence within the *Prunophora* subgenus between the *Armeniaca* and the *Euprunus* sections, as the large number of genotypes studied show.

ACKNOWLEDGEMENTS

This work was supported by RTA-011-00089 from the Instituto Nacional de Investigacion y Tecnologia Agraria (INIA) and by the foundation DGA (CONSI+D) for a fellowship to MJR to visit the Department of Horticulture at Washington State University, Pullman, WA, USA. Work in the Dhingra lab was supported by Washington State University Agriculture Center Research Hatch funds to AD.

We highly appreciate the comment of Dr R. Socias i Company.

REFERENCES

Batley, J., Edwards, D. 2007. SNP applications in plants. In: *Association mapping in plants*. Edited by Oraguzie N, Rikkerink E, Gardiner S, Nihal De Silva H. New York, Springer 95-102.

Chin, S.W., Wen, J., Johnson, G., Potter, D. 2010. Merging Maddenia with the morphologically diverse *Prunus* (Rosaceae). Botanical Journal of the Linnean Society, 164: 236–245. doi: 10.1111/j.1095- 8339.2010.01083.x

Decroocq, V., Hagen L.S., Favé, M.G., Eyquard, J.P. and Pierronnet, A. 2004. Microsatellite markers in the hexaploid *Prunus domestica* species and parentage lineage of three European plum cultivars using nuclear and chloroplast simple-sequence repeats. Molecular Breeding 13: 135-142.

Dhingra, A., Folta, K.M. 2005. ASAP: Amplification, sequencing & annotation of plastomes. BMC Genomics 6: 176.

Felipe, A.J. 2009. 'Felinem', 'Garnem', and 'Monegro' almond x peach hybrid rootstocks. HortScience 44 (1): 196-197.

Galeano, C.H., Fernández, A.C., Gómez, M., Blair, M.W. 2009. Single strand conformation polymorphism based SNP and Indel markers for genetic mapping and synteny analysis of common bean (*Phaseolus vulgaris* L.). BCM Genomics 10: 629.

Gupta, P.K., Roy, J.K., Prasad, M. 2001. Single nucleotide polymorphisms: a new paradigm for molecular marker technology and DNA polymorphism detection with emphasis on their use in plants. Curr. Sci. 80: 524-535.

Hayashi, K., Hashimoto, N., Daigen, M., Ashikawa, I. 2004. Development of PCR-based SNP markers for rice blast resistance genes at the *Piz* locus. Theor Appl Genet 108: 1212-1220.

213 Jansen, R.K., Saski, C., Lee, S., Hansen, A.K., Daniell, H. 2011. Complete Plastid Genome
214 Sequences of Three Rosids (*Castanea, Prunus, Theobroma*): Evidence for At Least
215 Two Independent transfers of *rpl22* to the Nucleus. Mol. Biol. Evol. 28 (1): 835-847.
216 Layne, R.E.C. 1987. Peach rootstocks. In: *Rootstocks for fruit crops*. Edited by Rom RC,
217 Carlson RF. New York, Wiley 185-216.
218 Lecouls, A.C., Bergougnoux, V., Rubio-Cabetas, M.J., Bosselut, N., Voisin, R., Poessel, J.C.,
219 Faurobert, M., Bonnet, A., Salesses, G., Dirlewanger, E., Esmenjaud, D. 2004. Marker-
220 assisted selection for the wide-spectrum resistance to root-knot nematodes conferred
221 by the *Ma* gene in Myrobalan plum (*Prunus cerasifera*) in interspecific *Prunus*
222 material. Molecular breeding 13: 113-124.
223 Lee, S., Wen, J. 2001. A phylogenetic analysis of *Prunus* and the *Amygdaloideae* (Rosaceae)
224 using ITS sequences of nuclear ribosomal DNA. Am. J. Bot. 88: 150-160.
225 Lovett, S.T. 2004. Encoded errors: mutations and rearrangements mediated by misalignment at
226 repetitive DNA sequences. Mol. Microbiol 52: 1243-1253.
227 Koepke,T., Schaeffer, S., Harper, A., Dicenta,F., Edwards, M., Henry, R. J., Møller, B. L.,
228 Meisel, L., Oraguzie, N., Silva, H., Sanchez-Perez, R., Dhingra, A. 2013. Comparative
229 genomics analysis in Prunoideae to identify biologically relevant polymorphisms. Plant
230 Biotechnology Journal 11: 883-893.
231 Odintsova, M.S., Iurina, N.P. 2003. Plastidic genome of higher plants and algae: structure and
232 function. Mol Biol 37:768-783.
233 Ohnishi, Y., Tanaka, T., Ozaki, K., Yamada, R., Suzuki, H., Nakamura, Y. 2001. A high-
234 throughput SNP typing system for genome-wide association studies. J Hum Genet 46:
235 471-477.
236 Ohta, S., Nishitani, C., Yamamoto, T. 2005. Chloroplast microsatellites in *Prunus*, Rosaceae.
237 Molecular Ecology Notes 5: 837-840.
238 Panda, S., Martín, J.P., Aguinagalde, I. 2003. Chloroplast DNA study in sweet cherry cultivars
239 (*Prunus avium* L.) using PCR-RFLP method. Genetic Resources and Crop Evolution.
240 50: 489-495.
241 Provan, J., Powell, W., Hollingsworth, P.M. 2001. Chloroplast microsatellites: new tools for
242 studies in plant ecology and evolution. Trends in Ecology and Evolution. 16 (3): 142-
243 147.
244 Serrano, B., Gómez-Aparisi, J., Hormaza, J.I. 2002. Molecular fingerprinting of *Prunus*
245 rootstocks using SSRs. J. Hort. Sci. Biotechnol. 77: 368-372.
246 Tangphatsornruang, S., Uthaipaisanwong, P., Sangsrakru, D., Chanprasert, J., Yoocha, T.,
247 Jomchai, N., Tragoonrung, S. 2011. Characterization of complete chloroplast genome
248 of *Hevea brasiliensis* reveals genome rearrangement, RNA editing sites and
249 phylogenetic relationships. Gene. 475: 104-112.

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

van Tienderen, P.H., de Haan, A.A., van der Linden, C.G., Vosman, B. 2002. Biodiversity assessment using markers for ecologically important traits. Trends Ecol Evol 17: 577-582.

Yazbek, M., Oh, S.H. 2013. Peaches and almonds: phylogeny of *Prunus* subg. *Amygdalus* (Rosaceae) based on DNA sequences and morphology. Plant Syst Evol 299: 1403-1418. doi: 10.1007/s00606-013-0802-1

FIGURES

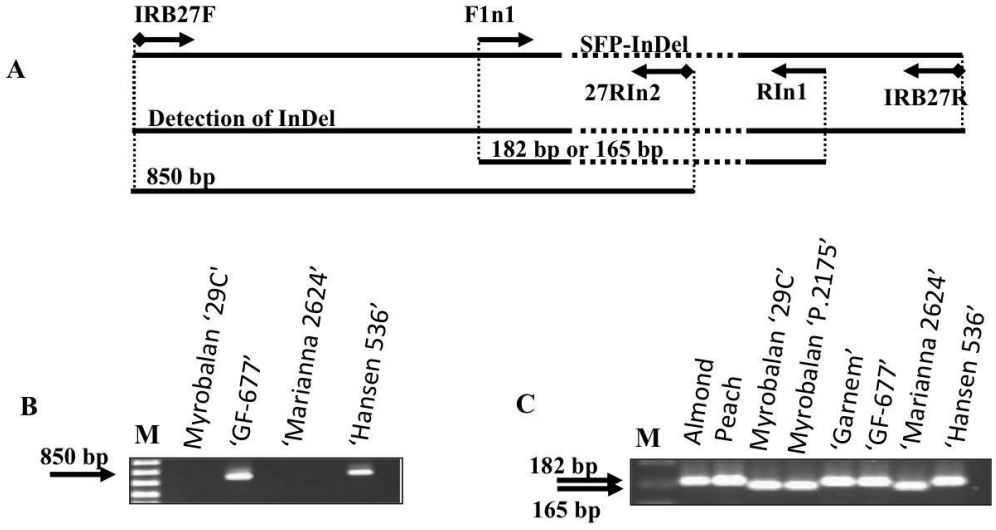

Fig. 1. Electrophoresis of ASAP-PCR products. (A) Representative scheme of the SPF-Indel in *Prunus* chloroplast genome. (B) PCR products obtained with primer pair IRB27F/27RIn2. (C) PCR products obtained with primer pair FIn1/RIn1. (M: 100 bp DNA Ladder).

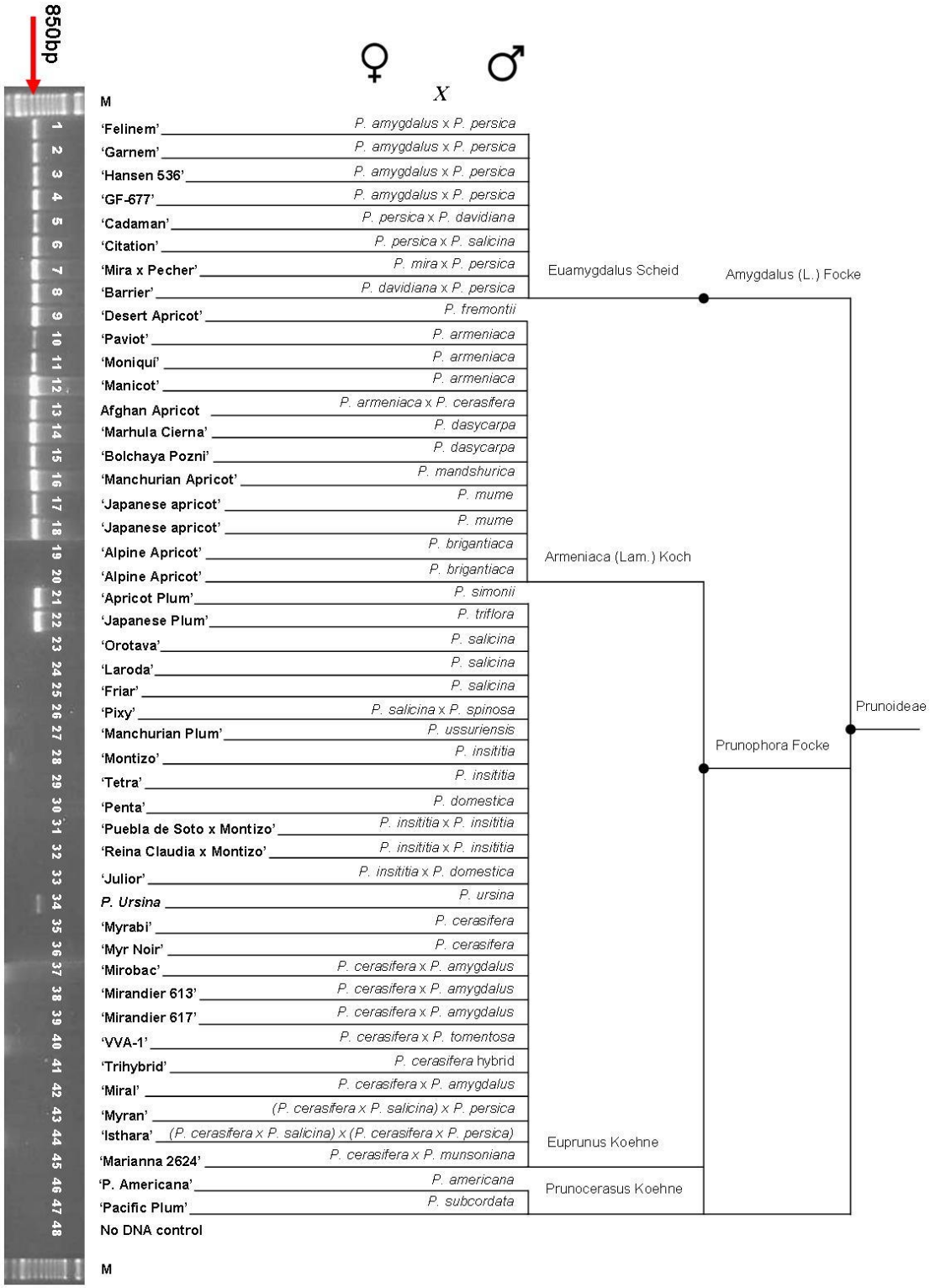

Fig. 2. Electrophoresis of ASAP-PCR products amplified with IRB27F/27RIn2 primers using ASAP-PCR. Based on the presence of the 850 bp amplicon the genotypes are classified into four distinct maternal lineages. (M: 100 bp DNA Ladder).