Research Article

DNA barcodes and microsatellites: How they complement for species identification in the complex genus *Tamarix* (Tamaricaceae)

Alejandro Terrones*1, Michelle van der Bank2, Joaquín Moreno3, and Ana Juan1

1Departamento de Ciencias Ambientales y Recursos Naturales, University of Alicante, Carretera San Vicente del Raspeig s/n, 03690 San Vicente del Raspeig, Alicante, Spain
2The African Centre for DNA Barcoding (ACDB)/Department of Botany and Plant Biotechnology, University of Johannesburg, PO Box 524, Auckland Park, 2006 Johannesburg, South Africa
3Departamento de Biología Aplicada, Miguel Hernández University of Elche, Avda. Universidad s/n, Edf. Torreblanca, 03202 Elche, Alicante, Spain

*Author for correspondence. E-mail: alejandro.terrones@ua.es

Received 20 July 2021; Accepted 7 January 2022; Article first published online 13 January 2022

Abstract DNA barcoding allows the identification of an organism by comparing the sequence of selected DNA regions (barcodes) with a previously compiled database, and it can be useful for taxonomic identification of species in complex genera, such as *Tamarix*. Many species of this genus show convergent morphology, which leads to frequent errors in their identification. Highly variable genetic markers, such as microsatellites or short sequence repeats (SSR), could be used to differentiate species where DNA barcodes fail. Here, we tested the ability of both, 5 different marker regions (*rbcL*, *matK*, ITS, *trnH-psbA*, and *ycf1*), and 14 microsatellites, to properly identify *Tamarix* species, especially those from the Mediterranean Basin, and compared the pros and cons of the different analytical methods for species identification. DNA barcoding allows the genetic identification of certain species in *Tamarix*. The two-locus barcodes *matK* + ITS and ITS + *ycf1* were the best-performing combinations, allowing up to 69% and 70%, respectively, correct identification. However, DNA barcoding failed in phylogenetically close groups, such as many Mediterranean species. The use of SSR can aid the identification of species, and the combination of both types of data (DNA barcoding and SSR) improved the success. The combination of data was especially relevant in detecting the presence of hybridization processes, which are common in the genus. However, caution must be exercised when choosing the clustering methods for the SSR data since different methods can lead to very different results.

Key words: DNA barcoding, identification methods, Mediterranean Basin, microsatellites, *Tamarix*.

1 Introduction

DNA barcoding is a technology that enables the identification of an organism by comparing the sequence of selected DNA regions (barcodes) with a previously compiled database (Hebert et al., 2003). DNA barcoding has proved useful in many fields, such as taxonomy, biosecurity, biodiversity, and conservation, among others (Pečnikar & Buzan, 2014). An ideal barcode should be easy to amplify with a single primer pair, require little manual editing of sequences, and provide sufficient sequence variability to allow to differentiate close species (CBOL Plant Working Group, 2009). In animals, the Cytochrome Oxidase subunit 1 (*CO1*) region meets these criteria and has become the standard barcode region (Hebert et al., 2003). However, in plants, no single region has been found to date that can be used as a standard barcode. The CBOL Plant Working Group (2009) proposed the combination of Ribulose-bisphosphate carboxylase (*rbcL*) and Maturase K (*matK*) as the core barcode for plants. However, these two markers do not always show sufficient genetic variability, and hence other markers have been proposed as additional barcodes, such as the nuclear ribosomal internal transcribed spacers (ITS) and the plastid *trnH-psbA* region (Hollingsworth et al., 2011), and some even as core barcodes, such as *ycf1* (Dong et al., 2015).

One of the biggest challenges of DNA barcoding is the existence of sequences of properly identified taxa in genetic databases (Ratnasingham & Herbert, 2007). An incomplete representation of taxa, or inaccuracy in the identification of taxa to which the sequences belong, might lead to incorrect identifications using those databases. Consequently, and especially in complex genera, it is of uppermost importance...
that the plant material is correctly identified and that it has a voucher specimen deposited in a registered collection that allows to double-check its identification in the future. In this sense, BOLD (http://www.boldsystems.org; Ratnasingham & Herbert, 2007) is a useful cloud-based platform designed to support DNA barcode data since it allows to include public data of the vouchers, sequences, trace files, and images.

Many methods have been described to date to identify species based on sequence data, including similarity matches, phylogenetic approaches, distance-based algorithms, and genealogical methods (Austerlitz et al., 2009). Up to date, no one method has proved to be universally applicable for all taxa, and success depends on the structure and amount of available data (Austerlitz et al., 2009). Consequently, when performing a DNA barcoding study on a specific group of species, it is convenient to test the efficacy of different methods, especially in taxonomically complex groups.

DNA barcoding might be a useful tool to aid the taxonomic identification of species of complex genera, such as Tamarix L. This genus comprises between 54 and 90 species of trees and shrubs that grow mainly in salt marshes, ravines and more or less saline rivers (Baum, 1978; Villar et al., 2019). The genus has a natural distribution from Asia to Europe and Africa (Baum, 1978), but it has been introduced in the Americas and Australia, where it is invasive due to high clonal dispersion ability (Brock, 1994; Gaskin & Schaal, 2003). The leaves of Tamarix trees are reduced to scales, and hence, many species show a convergent morphology, which leads to frequent errors in their identification. In fact, the majority of the species can only be differentiated by their floral characters, making it impossible to determine to species level outside the flowering season (Gaskin, 2003). Besides, in some species, floral characters have shown great variability between individuals and even at different times of the year (Baum, 1978; Villar et al., 2019). In addition, the recognition of the taxa is further complicated, especially for those Tamarix woodlands formed by more than one Tamarix species since the presence of certain hybrid forms is not uncommon (see Gaskin & Shafroth, 2005; Samadi et al., 2013; Mayonde et al., 2016). These hybridization processes can be especially challenging in invaded places, where the origin of the plants might be unknown. Consequently, the taxonomic species identification of this genus could be considered a challenging task.

Previous phylogenetic relationships have shown significant incongruencies between different gene trees in Tamarix (Villar et al., 2019), which might be caused by a rapid radiation of the genus. In this context, the expected presence of incomplete lineage sorting (ILS) and hybridization processes could hinder species identification with DNA barcodes, especially for closely related species. The presumably low genetic differentiation of certain species might be reflected in the absence of a barcode gap, which circumscribes the assumption that the amount of genetic variation within species is smaller than the amount of variation between species (Meyer & Paulay, 2005), and hence makes it impossible to differentiate between species with the genetic markers used. Nevertheless, highly variable genetic markers, such as microsatellites or short sequence repeats (SSR), could be used to differentiate species where DNA barcodes fail, though they lack the universality of DNA barcoding markers. However, SSR have proved to be useful for the identification of close species in other complex taxonomic groups, such as Indian bamboos (Sharma et al., 2008) or Rhododendron L. (Wang et al., 2019), and are also helpful for the identification of hybridization processes (Twyford & Ennos, 2012). Many methods have been described to date to analyze the information of the genetic clusters present in SSR data (Pritchard et al., 2000; Corander & Marttinen, 2006; Jombart et al., 2010; Beugn et al., 2018). Similar to the case of DNA barcoding, diverse approaches yielded strong differences in the genetic clusters retrieved (Pinho et al., 2019), and hence it is advisable to compare the results of various methods.

In this paper, we checked the ability of both, 5 different marker regions (rbcl, matK, ITS, trnH-psbA, and ycf1), and the recently described 14 microsatellites, for the widespread Mediterranean species Tamarix gallica L. (Terrones & Juan, 2020) to properly identify Tamarix species, especially those from the Mediterranean Basin. The objectives of the paper are: (i) to identify the best combination of DNA barcode markers for the identification at the species level in Tamarix, (ii) to check the utility of SSR markers for species identification in Tamarix, and (iii) to compare the pros and cons of the different methods and select the most suitable one for species identification in the genus.

2 Material and Methods

According to recent phylogenetic analyses in Tamarix (Villar et al., 2019), 67 individuals of 20 different species were selected in order to include a minimum of one specimen from all main phylogenetic clades and, at least, two to three specimens for each Mediterranean species (Table 1). Details of the project, including voucher information, GPS coordinates, images, and DNA barcodes are available on BOLD (http://www.boldsystems.org; Ratnasingham & Herbert, 2007) within the project “Barcoding of the genus Tamarix” (BTAM) (http://www.boldsystems.org/index.php/Public BIN?search?searchtype=records).

DNA extractions were carried out using silica-gel dried leaves and stems by a modified cetyltrimethylammonium bromide (CTAB) procedure (Doyle & Doyle, 1987). For each sample, four plastid regions (rbcl, matK, trnH-psbA, and ycf1) and one nuclear region (ITS) were amplified and sequenced using the primers as detailed in Table 2. The ITS region included ITS1, 5.8 S, and ITS2. Polymerase chain reactions (PCR) were performed in a final volume of 25 µL with DreamTaq PCR Master Mix (2x) (Thermo Scientific, Vilnius, Lithuania), 0.5 µg of bovine serum albumin (BSA), 5 pmol of each primer, 100 ng of template DNA, and, for matK and ITS, 0.5 µl of dimethyl sulfoxide (DMSO). Cycling conditions for rbcl, trnH-psbA, and ycf1 were as follows: an initial step of 5 min at 95 °C; followed by 35 cycles of 30 s at 95 °C, 40 s at 52 °C, and 1 min at 72 °C; and a final extension at 72 °C for 10 min. Cycling conditions for matK were as follows: an initial step of 4 min at 94 °C; followed by 10 cycles of 30 s at 94 °C, 30 s at 52 °C, and 1 min at 72 °C; and a final extension at 72 °C for 10 min. Cycling conditions for ITS were as follows: an initial step of 3 min at 94 °C; followed by 26 cycles of 1 min at 94 °C, 1 min at 48 °C, and 3 min at 72 °C; and a final extension at 72 °C for 7 min. PCR products were cleaned with ExoSAP-IT™ PCR Product Cleanup Reagent (Thermo
Table 1  Specimens used in the study with voucher information and NCBI GenBank numbers of the regions used in the article

| Species                  | Field ID | Herbarium number | Location                           | Collectors and no. | Coordinates (lat, lon) | rbCL GenBank | matK GenBank | ITS GenBank | trnH-psbA GenBank | ycf1 GenBank |
|--------------------------|----------|------------------|------------------------------------|--------------------|-----------------------|--------------|--------------|-------------|------------------|---------------|
| Tamarix africana Poir.   | Aguilas21| 70744 (ABH)      | Spain, Murcia, Águilas             | A. Terrones, J. Moreno s.n. | 37.433, -1.629         | MZ563047     | MZ563115     | MZ563183    | MZ573483         | MZ563225      |
|                          | Aguilas8 | 70742 (ABH)      | Spain, Murcia, Águilas             | A. Terrones, J. Moreno s.n. | 37.433, -1.629         | MZ563053     | MZ563121     | MZ563189    | MZ573482         | MZ563231      |
| T3.19Cas                 | T3.19Cas | 70790 (ABH)      | Spain, Valencian Community, Castellón | A. Terrones, J. Moreno s.n. | 39.879, -0.055         | MZ563050     | MZ563118     | MZ563186    | MZ573479         | MZ563228      |
| T3.6Cas                  | T3.6Cas  | 70789 (ABH)      | Spain, Valencian Community, Castellón | A. Terrones, J. Moreno s.n. | 39.879, -0.055         | MZ563049     | MZ563117     | MZ563185    | MZ573478         | MZ563227      |
| Ebro7                    | Ebro7    | 69754 (ABH)      | Spain, Catalonia, Tarragona        | A. Terrones, J. Moreno s.n. | 40.629, 0.753          | MZ563048     | MZ563116     | MZ563184    | MZ573484         | MZ563216      |
| FrCG1.19                 | FrCG1.19 | 69962 (ABH)      | France, Occitanie, Languedoc-Roussillon | A. Terrones, J. Moreno s.n. | 42.967, 2.999          | MZ563054     | MZ563122     | MZ563190    | MZ573485         | MZ563232      |
| Port2.22                 | Port2.22 | 73511-2 (ABH)    | Portugal, Alentejo, Baixo Alentejo | A. Terrones, J. Moreno s.n. | 38.129, -8.789         | MZ563051     | MZ563119     | MZ563187    | MZ573480         | MZ563229      |
| Port2.24                 | Port2.24 | 73511-3 (ABH)    | Portugal, Alentejo, Baixo Alentejo | A. Terrones, J. Moreno s.n. | 38.129, -8.789         | MZ563052     | MZ563120     | MZ563188    | MZ573481         | MZ563230      |
| Tamarix amplexicaulis Ehrenb. | T24AG   | 70685 (ABH)      | Algeria, Biskra, Biskra            | A. Terrones, M.A. Alonso, A. Juan s.n. | 34.719, 5.739         | MZ563058     | MZ563126     | MH626269    | MZ573496         | MZ563236      |
|                          | T25AG    | 70686 (ABH)      | Algeria, Biskra, Biskra            | A. Terrones, M.A. Alonso, A. Juan s.n. | 34.719, 5.739         | MZ563057     | MZ563125     | MH626270    | MZ573497         | MZ563235      |
|                          | T26AG    | 70687 (ABH)      | Algeria, Biskra, Biskra            | A. Terrones, M.A. Alonso, A. Juan s.n. | 34.719, 5.739         | MZ563056     | MZ563124     | MH626271    | MZ573498         | MZ563234      |
|                          | T27AG    | 70688 (ABH)      | Algeria, Biskra, Biskra            | A. Terrones, M.A. Alonso, A. Juan s.n. | 34.719, 5.739         | MZ563055     | MZ563123     | MH626272    | MZ573499         | MZ563233      |

Continued
| Species                     | Field ID | Herbarium number | Location                          | Collectors and no. | Coordinates (lat, lon) | rbcl GenBank | matK GenBank | ITS GenBank | trnH-psbA GenBank | ycf1 GenBank |
|----------------------------|----------|------------------|-----------------------------------|--------------------|------------------------|-------------|--------------|-------------|------------------|--------------|
| *Tamarix aphylla* (L.) H. Karst | Mar5     | 71909 (ABH)      | Morocco, Oriental Region, Nador    | A. Terrones, M.B. Crespo, M.A. Alonso s.n. | 35.103, −2.36 | MZ563059    | MZ563127    | MZ563191    | MZ573502         | MZ56337     |
|                            | TA1Mo    | 54208 (ABH)      | Morocco, Tangier-Tetouan Region, Nador | J.L. Villar, M.B. Crespo, A. Juan, M.A. Alonso, L. Sáez et al. s.n. | 34.972, −3.36 | MZ563060    | MZ563128    | MH626274    | MZ573501         | MZ56338     |
|                            | TACD     | 70064 (ABH)      | Italy, Sardinia, Oristano          | A. Terrones, J.L. Villar s.n. | 39.954, 8.403 | MZ563061    | MZ563129    | MH626275    | MZ573500         | MZ56339     |
| *Tamarix arceuthoides* Bunge | ARC782   | 5568719 (MO)     | Iran, Esfahan                      | J.F. Gaskin 782    | 33.0152, −52.5238  | MZ563062    | MZ563130    | MZ563192    | MZ573503         | MZ56340     |
|                            | ARC825   | 5568891 (MO)     | Iran, Qom                          | J.F. Gaskin 825    | 35.1705, −50.9777  | MZ563063    | MZ563131    | MZ563193    | MZ573504         | MZ56341     |
| *Tamarix boveana* Bunge     | Gata10   | 68315-3 (ABH)    | Spain, Andalusia, Almeria          | A. Juan, P. Vila s.n. | 36.773, −2.238 | MZ563068    | MZ563136    | MZ563198    | MZ573488         | MZ56346     |
|                            | Gata8    | 68315-1 (ABH)    | Spain, Andalusia, Almeria          | A. Juan, P. Vila s.n. | 36.773, −2.238 | MZ563066    | MZ563134    | MZ563196    | MZ573486         | MZ56344     |
|                            | Gata9    | 68315-2 (ABH)    | Spain, Andalusia, Almeria          | A. Juan, P. Vila s.n. | 36.773, −2.238 | MZ563067    | MZ563135    | MZ563197    | MZ573487         | MZ56345     |
|                            | StaP29   | 70782 (ABH)      | Spain, Valencian Community, Alicante | A. Terrones, J.L. Villar s.n. | 38.184, −0.602 | MZ563064    | MZ563132    | MZ563194    | MZ573489         | MZ56342     |
|                            | StaP33a  | 70783 (ABH)      | Spain, Valencian Community, Alicante | A. Terrones, J.L. Villar s.n. | 38.184, −0.602 | MZ563069    | MZ563137    | MZ563199    | MZ573490         | MZ56347     |
|                            | StaP35b  | 70785 (ABH)      | Spain, Valencian Community, Alicante | A. Terrones, J.L. Villar s.n. | 38.184, −0.602 | MZ563065    | MZ563133    | MZ563195    | MZ573491         | MZ56343     |
| *Tamarix canariensis* Willd. | TGC4.1   | 69607 (ABH)      | Spain, Canary Islands, Gran Canaria | J.L. Villar, E. Martinez s.n. | 27.996, −15.824 | MZ563070    | MZ563138    | MZ563200    | MZ573492         | MZ56348     |
|                            | TGC4.21  | 69609 (ABH)      | Spain, Canary Islands, Gran Canaria | J.L. Villar, E. Martinez s.n. | 27.996, −15.824 | MZ563071    | MZ563139    | MZ563201    | MZ573494         | MZ56349     |
|                            | TGC4.28  | 69644 (ABH)      | Spain, Canary Islands, Gran Canaria | J.L. Villar, E. Martinez s.n. | 27.996, −15.824 | MZ563072    | MZ563140    | MZ563202    | MZ573495         | MZ56350     |
| Species                  | Field ID    | Location                        | Collectors and no.            | Coordinates (lat, lon) | rbcL GenBank | matK GenBank | ITS GenBank | trnH-psBA GenBank | ycf1 GenBank |
|-------------------------|-------------|---------------------------------|------------------------------|------------------------|--------------|--------------|-------------|-------------------|-------------|
| Tamarix chinensis Lour.  | TGC4.3      | Spain, Canary Islands, Gran Canaria | J.L. Villar, E. Martinez s.n. | 27.996, -15.824        | MZ563074     | MZ563142     | MZ563204    | MZ573493          | MZ563252    |
| Tamarix dalmatica B. R. Baum | TGC4.30    | Spain, Canary Islands, Gran Canaria | J.L. Villar, E. Martinez s.n. | 27.996, -15.824        | MZ563073     | MZ563141     | MZ563203    | MZ573539          | MZ563251    |
| Tamarix gallica L.       | 202         | South Korea                     | J.F. Gaskin 202               |                        | MZ563075     | MZ563143     | MH626289    | MZ573505          | MZ563253    |
|                         | T1Alb       | Albania, Vlore, Sarande Montenegro, Bar | J.L. Villar, E. Martinez s.n. | 39.904, 20.084         | MZ563079     | MZ563147     | MH626290    | MZ573509          | MZ563257    |
|                         | T1MNE       | Albania, Vlore, Sarande Montenegro, Bar | J.L. Villar, E. Martinez s.n. | 42.093, 19.104         | MZ563078     | MZ563146     | MZ563207    | MZ573506          | MZ563256    |
|                         | T2Alb       | Albania, Vlore, Sarande Montenegro, Bar | J.L. Villar, E. Martinez s.n. | 40.047, 19.846         | MZ563077     | MZ563145     | MZ563206    | MZ573507          | MZ563255    |
|                         | T6Alb       | Albania, Shkoder                | J.L. Villar, E. Martinez s.n. | 41.968, 19.547         | MZ563076     | MZ563144     | MZ563205    | MZ573508          | MZ563254    |
|                         | A-El1       | Spain, Valencian Community, Alicante | A. Terrones, J.L. Villar s.n. | 38.313, -0.724         | MZ563082     | MZ563150     | MZ563210    | MZ573476          | MZ563260    |
|                         | Ebro49      | Spain, Catalonia, Tarragona     | J.L. Villar, A. Vicente, J.C. Agullo s.n. | 40.629, 0.753 | MZ563080     | MZ563148     | MZ563208    | MZ573472          | MZ563258    |
|                         | FrCG1.9     | France, Occitanie, Languedoc-Roussillon | A. Juan, A. Terrones, M.A. Alonso, J. Pérez s.n. | 42.967, 2.999 | MZ563085     | MZ563153     | MZ563213    | MZ573473          | MZ563263    |
|                         | MoKen15     | Morocco, Gharb-Chrarda-Beni Hssen Region, Kenitra | J.L. Villar, A. Vicente, M.A. Alonso, M.B. Crespo s.n. | 34.272, -6.556 | MZ563084     | MZ563152     | MZ563212    | MZ573474          | MZ563262    |
|                         | TD195       | Spain, Castilla-La Mancha, Ciudad Real | J.L. Villar, J. Moreno s.n. | 39.163, -3.684         | MZ563081     | MZ563149     | MZ563209    | MZ573477          | MZ563259    |
|                         | TMa2        | Spain, Andalusia, Málaga        | J.L. Villar, E. Martinez, A. Ortiz s.n. | 36.668, -4.466         | MZ563083     | MZ563151     | MZ563211    | MZ573475          | MZ563261    |
| Species          | Field ID | Herbarium number | Location                                  | Collectors and no.                                      | Coordinates (lat, lon) | rbcl. GenBank | matK GenBank | ITS GenBank | trnH-psbA GenBank | ycf1 GenBank |
|------------------|----------|------------------|------------------------------------------|---------------------------------------------------------|------------------------|---------------|--------------|-------------|------------------|-------------|
| *Tamarix hampeana* Boiss. & Heldr. | T3MNE    | 57891 (ABH)      | Montenegro, Ulcinj, Greece, Epirus, Igoumenitsa | J.L. Villar, E. Martínez s.n.                          | 41.87, 19.352          | MZ563088      | MZ563156     | MH626298   | MZ573511          | MZ563266    |
|                  | T6Gre    | 59025 (ABH)      | Greece, Central Greece, Molos-Agiós Konstantinos | J.L. Villar, E. Martínez s.n.                                | 39.525, 20.198          | MZ563086      | MZ563154     | MH626299   | MZ573510          | MZ563264    |
|                  | T72Gre   | 59877 (ABH)      | Greece, Central Greece, Molos-Agiós Konstantinos | J.L. Villar, E. Martínez s.n.                                | 38.834, 22.703          | MZ563087      | MZ563155     | MH626300   | MZ573512          | MZ563265    |
| *Tamarix hispida* Willd.         | 10164    | -                | China                                     | J.F. Gaskin 10164                                         | -                      | MZ563089      | MZ563157     | MH626302   | MZ573538          | MZ563267    |
| *Tamarix hohenackeri* Bunge       | HOH828   | 5568893 (MO)     | Iran, Gilan                               | J.F. Gaskin 828                                          | 36.8641, 49.4811       | MZ563091      | MZ563159     | MZ563215   | MZ573513          | MZ563269    |
|                  | HOH915   | 5568696 (MO)     | Iran, Semnan                              | J.F. Gaskin 915                                          | 36.7252, 55.2975       | MZ563090      | MZ563158     | MZ563214   | MZ573514          | MZ563268    |
| *Tamarix leptostachya* Bunge        | 10177    | -                | China                                     | J.F. Gaskin 10177                                        | -                      | MZ563092      | MZ563160     | MH626309   | MZ573537          | MZ563270    |
| *Tamarix minoa* J. L. Villar, Turland, Juan, Gaskin, M. A. Alonso & M. B. Crespo | T2CR     | 54194 (ABH)      | Greece, Crete, Chania                     | A. Vicente, M.A. Alonso, J.L. Villar s.n.                 | 35.365, 24.248          | MZ563095      | MZ563163     | MH626312   | MZ573515          | MZ563273    |
|                  | T4CR     | 54195 (ABH)      | Greece, Crete, Chania                     | A. Vicente, M.A. Alonso, J.L. Villar s.n.                 | 35.356, 24.26           | MZ563094      | MZ563162     | MH626313   | MZ573516          | MZ563272    |
|                  | MoCRTurland | 6207620 (MO) | Greece, Crete, Nomos Chanion              | P. Bareka, N.J. Turland sub N.J. Turland 178             | 35.359, 24.266         | MZ563093      | MZ563161     | MH626311   | MZ573517          | MZ563271    |
| *Tamarix nilotica* (Ehrenb.) Bunge  | TCR1     | 54314 (ABH)      | Greece, Crete, Heraklion                  | J.L. Villar, A. Vicente s.n.                               | 35.33, 25.327           | MZ563098      | MZ563166     | MZ563218   | MZ573518          | MZ563276    |
|                  | TCR11    | 54333 (ABH)      | Greece, Crete, Heraklion                  | J.L. Villar, A. Vicente s.n.                               | 34.979, 25.324          | MZ563097      | MZ563165     | MZ563217   | MZ573522          | MZ563275    |
|                  | TCR14    | 54320 (ABH)      | Greece, Crete, Heraklion                  | J.L. Villar, A. Vicente s.n.                               | 35.223, 23.67           | MZ563101      | MZ563169     | MH626314   | MZ573523          | MZ563279    |
|                  | TCR3     | 54316 (ABH)      | Greece, Crete, Lassithi                   | J.L. Villar, A. Vicente s.n.                               | 35.051, 26.232          | MZ563096      | MZ563164     | MZ563216   | MZ573519          | MZ563274    |
|                  | TCR5     | 54317 (ABH)      | Greece, Crete, Lassithi                   | J.L. Villar, A. Vicente s.n.                               | 35.014, 26.046          | MZ563099      | MZ563167     | MH626315   | MZ573520          | MZ563277    |
|                  | TCR9     | 54326 (ABH)      | Greece, Crete, Heraklion                  | J.L. Villar, A. Vicente s.n.                               | 34.979, 25.335          | MZ563100      | MZ563168     | MZ563219   | MZ573521          | MZ563278    |
| Species                     | Field ID | Herbarium number | Location                    | Collectors and no. | Coordinates (lat, lon) | rbCl GenBank   | matK GenBank | ITS GenBank | trnH-psBA GenBank | ycf1 GenBank |
|----------------------------|----------|------------------|-----------------------------|--------------------|-----------------------|----------------|-------------|-------------|------------------|--------------|
| *Tamarix parviflora* DC.   | T8CR     | 54197 (ABH)      | Greece, Crete, Heraklion    | M.A. Alonso, J.L.  | 35.321, 25.327         | MZ563104       | MZ563172    | MH626318    | MZ573526         | MZ563282     |
|                            | TCR10    | 54321 (ABH)      | Greece, Crete, Heraklion    | J.L. Villar, A.    | 34.979, 25.324         | MZ563103       | MZ563171    | MH626319    | MZ573525         | MZ563281     |
|                            | TparvBiar | 55398 (ABH)     | Spain, Valencian Community  | J.L. Villar, I.    | 38.629, −0.76          | MZ563102       | MZ563170    | MZ563220    | MZ573524         | MZ563280     |
| *Tamarix ramosissima* Lede. | W09-43   | 2009–19143 (W)   | Argentina, San Juan, Ullum  | J. Chiapella, E.   | −31.03, −68.75         | MZ563105       | MZ563173    | MH626320    | MZ573527         | MZ563283     |
|                            | 4690-06  | -                | Turkey                      | J.F. Gaskin s.n.   | -                     | MZ563106       | MZ563174    | MH626323    | MZ573529         | MZ563284     |
|                            | W03-43   | 2003–14043 (W)   | Armenia, Vayots'Dzor        | G. Fayvush et al. | 39.68, 45.22           | MZ563107       | MZ563175    | MH626325    | MZ573528         | MZ563285     |
| *Tamarix tetragyna* Ehrenb. | TTG07-28 | 2007–25728 (W)   | Egypt, South Sinai          | T. Eberl s.n.      | 28.4532, 34.4492       | MZ563109       | MZ563177    | MH626327    | MZ573530         | MZ563287     |
|                            | TTG07-48 | 2007–14048 (W)   | Egypt, New Valley           | J. Walter s.n.     | 25.667, 28.87          | MZ563108       | MZ563176    | MH626328    | MZ573533         | MZ563286     |
|                            | W07-64   | 2007–07364 (W)   | Jordan, Al Asimah           | J. Walter 7311     | 31.833, 35.676         | MZ563111       | MZ563179    | MZ563222    | MZ573532         | MZ563288     |
| *Tamarix usneoides* E. Mey. | T1NMB    | 58684 (ABH)      | Namibia, Erongo             | M. Martínez Azorín | −22.708, 14.961        | MZ563113       | MZ563181    | MH626329    | MZ573534         | MZ563290     |
|                            | T2NMB    | 58685 (ABH)      | Namibia, Erongo             | M. Martínez Azorín | −22.708, 14.961        | MZ563114       | MZ563182    | MZ563223    | MZ573536         | MZ563291     |
|                            | TsA7     | 58683 (ABH)      | South Africa, Western Cape  | M. Martínez Azorín | −33.085, 21.579        | MZ563112       | MZ563180    | MH626330    | MZ573535         | MZ563289     |

Herbarium acronyms: ABH, Universidad de Alicante (Spain); MO, Missouri Botanical Garden (USA); W, Naturhistorisches Museum Wien (Austria).
To analyze the different combinations of the five regions as DNA barcodes, 31 data matrices were constructed with all possible combinations: five with one region, 10 with two regions, 10 with three regions, five with four regions, and one with all five regions. Neighbor joining (NJ) trees for all combinations were constructed with the maximum composite likelihood model, uniform rates among sites, pairwise deletion in the gaps, and a bootstrap (BS) based on 1000 replications with MEGA v. 7.0.26 (Kumar et al., 2019). All trees were rooted using Tamarix aphylla (L.) H. Karst and Tamarix usneoides E. Mey. as outgroups, based on Villar et al. (2019).

The identification of each individual was considered as “correct”, “incorrect”, or “ambiguous” following Meier et al. (2006). Species with only one individual (singleton) were considered as “unidentifiable”. To test the barcoding efficacy of the 31 combinations of regions, different distance-based measures of identification accuracy implemented in the package “Spider” v. 1.5.0 (Brown et al., 2012) were calculated using software R v. 3.6.3 (R Core Team, 2020): nearest neighbor (NN), a threshold-based criterion (tID), and best close match (BCM). The nearest neighbor (NN) criterion finds the closest individual and returns if their names are the same (“True”) or different (“False”). If there is more than one individual closest to the target, it returns the species of the major component of the group. Hereafter, “True” results will be considered as “correct” identifications and “False” results as “incorrect” identifications. The threshold-based analysis (tID) is similar to the analysis conducted by the “Identify Specimen” tool provided by the Barcode of Life Database (http://www.boldsystems.org/index.php/IDS_OpenIdEngine). The analysis looks at all specimens within a threshold of the query, and the identification can be “correct”, “incorrect”, or “ambiguous”, if all, none, or a part, of the matches are the same species than the query, respectively. If no match is found below the threshold, the analysis retrieves “no identification”. Best close match (BCM) conducts the analysis described by Meier et al. (2006), considering the nearest neighbor only if it is closer than a given threshold. The identification can be “correct”, “incorrect”, “ambiguous”, or “no identification”, similar to the tID method. Hereafter, “incorrect” and “ambiguous” results will be considered as misidentifications. For tID and BCM, the threshold value can be optimized to increase the identification success of each combination of markers. Two different thresholds were used: the default 1% threshold used by BOLD, and an optimized threshold for each combination of markers calculated with the function “localMinima” (Appendix A1). This function looks for a dip in the density of genetic distances, which indicates the transition between intra- and inter-specific distances (Brown et al., 2012). In addition, the inter- and intra-specific pairwise distances were calculated with the function “sppDist” of the package “Spider” v. 1.5.0 (Brown et al., 2012).

We also performed the survey of nuclear microsatellites initially developed for T. gallica (Terrones & Juan, 2020; Appendix B). PCR amplifications were performed in 3 different mixes that included 14 microsatellites (further details in Terrones & Juan, 2020), and run on an ABI 3730xl genetic analyzer (Thermo Fisher Scientific, Carlsbad, CA, USA) with GeneScan 400HD Size Standard (Thermo Fisher Scientific) at Macrogen Spain (Madrid, Spain). Allele calling was done with Peak Scanner Software 2 (Applied Biosystems, Carlsbad, CA, USA), and allelic binning was done manually with the use of cumulative frequency plots of size distribution (Guichoux et al., 2011).

Microsatellite data were analyzed for the presence of private alleles and converted to different formats with

### Table 2 Primers used in the polymerase chain reactions (PCR) of the regions used in the article

| Region   | Primer     | Sequence                      | Reference                  |
|----------|------------|-------------------------------|----------------------------|
| rbcl     | rbclLaF    | ATGTCACCACAAACAGACTAAAGC      | Levin et al. (2003)        |
|          | rbclLaR    | GTAAACTAAGATCACCACCRG         | Kress et al. (2009)        |
| matK     | xf         | TAATTTAGATCAATTACTTC          | Ford et al. (2009)         |
|          | MALP_R1    | ACAAGAAAATTCGAAAT             | Dunning & Savolainen (2010) |
| ITS      | 1R         | ACCACGTTCCATCTGGAATCTTTGTTTC  | Kim (unpublished data)     |
|          | 3F         | CGTACAGTACCTCTTGTACTTACAGG    | Kim (unpublished data)     |
|          | 472F       | CCCCRTYCATCTATCTTGGTTC        | Yu et al. (2011)           |
|          | 1248R      | GCTTTRATAATGGAAAGATTTCTGC     | Yu et al. (2011)           |
| trnH-psbA| psbAF      | GTTATGCATAGCAGTATGCTC         | Sun et al. (1994)          |
|          | trnH2      | TAGAAATCCCGGTTGCGTCCGGTTTAC   | Sun et al. (1994)          |
| ycf      | ycfbrF     | TCTCGACGAAATACAGAGTGGTGGAAT  | Dong et al. (2015)         |
|          | ycfbrR     | ATACATGCAAAAGTGGATTGAAAAA     | Dong et al. (2015)         |
GenAlEx v. 6.503 (Peakall & Smouse, 2012). Many methods have been described to date to analyze microsatellite data to obtain the information of the genetic clusters. In this paper, four different clustering analyses were used to calculate the assignment of the individuals to genetic clusters: SNAPCLUST, STRUCTURE, DAPC and BAPS (Pinho et al., 2019).

SNAPCLUST was performed with the package “adegenet” v. 2.1.2 (Beaugin et al., 2018) in software R v. 3.6.3 (R Core Team, 2020). SNAPCLUST tries to achieve a maximum likelihood solution for the clustering problem, with a combination of geometric approaches and expectation-maximization (EM) likelihood optimization. The function “snapclust.choose.k” was used to estimate the optimal number of clusters. The maximum value of K was set to 22 and the Akaike information criterion (AIC; Akaike, 1998) was used as the criterion. The optimal number of clusters, which minimized the AIC value, was seven (Fig. S3). SNAPCLUST was then run for K = 7 using the Ward algorithm to determine initial group memberships and 50 different starting points for the EM algorithm. The convergence of the model was checked before subsequent analysis.

STRUCTURE v. 2.3.4 was used to analyze the genetic structure of the data (Pritchard et al., 2000). This program uses a Markov chain Monte Carlo approach to estimate an individual’s admixture from a pre-set number of clusters (K). The program was run for 1 000 000 steps with an additional burn-in of 100 000 steps using the admixture ancestry model with independent allele frequencies (IAF). The number of populations (K) ranged from one to 22, and the program was run 10 times for each value of K. The main pipeline of the software CLUMPKAK (Kopelman et al., 2015) was used with default options to assist with the representation and interpretation of the results of STRUCTURE. The best value of K was estimated with the best K mode of CLUMPKAK (Kopelman et al., 2015) with two methods: using Bayes rule with a uniform prior K and evaluation of the estimated log likelihood (ln(Pr(X|K))), as described by Pritchard et al. (2010), and with the distribution of the value ΔK described by Evanno et al. (2005). For the selected K, the cluster of runs with the highest mean likelihood was considered as optimal and the rest of the minor clusters were discarded (Jakobsson & Rosenberg, 2007).

A discriminant analysis of principal components (DAPC) was carried out with the package “adegenet” v. 2.1.2 (Jombart, 2008; Jombart et al., 2010) in software R v. 3.6.3 (R Core Team, 2020). DAPC is a multivariate analysis that performs a discriminant analysis on genetic data previously transformed with principal component analysis, and unlike the other methods, it does not assume a model on allele frequencies. Genetic clusters were identified with the function “find.clusters” with 1 000 000 iterations in each run, 100 different starting centroids and keeping all the principal components. The number of clusters was set to 10 since this was the number of clusters with the lowest value of the Bayesian information criterion (BIC; Schwarz, 1978) (Fig. S5). In order to decide on the number of principal components (PC) to retain for the DAPC, the function “optim.a.score” was used to calculate the α-score for every number of retained PCs. The α-score measures the trade-off between the loss of power of discrimination when including few PCs, and over-fitting the model when including too many PCs. The α-score was calculated with 1000 simulations per number of retained PC. The final DAPC was carried out with five PC retained, since this was the number of PC that maximized the α-score (Fig. S6), and with all five discriminant axes retained as well (Fig. S7).

BAPS v. 6.0 was used for population mixture and admixture analyses (Corander & Marttinen, 2006; Corander et al., 2008). BAPS relies on a two-step analysis. In the first step, the mixture analysis determines the number of clusters and individual assignments with a Bayesian stochastic partition approach. In the second step, the admixture analysis performs a simulation test for evaluating the statistical significance of the deviation from the no admixture. For the mixture analysis, clustering of individuals was performed 10 times for each maximum value of populations (5, 10, 15, 20, and 25). The admixture analysis was performed with the results of the mixture analysis with 100 iterations for the admixture coefficients for the individuals, 200 reference individuals from each population and 20 iterations for the admixture coefficients of the reference individuals.

3 Results

3.1 DNA barcoding results

The analyzed methods of identification showed very different results in barcoding efficacy (Fig. 1). NN, BCM, and NJ trees were the most efficient methods for species identification in *Tamarix*, with up to 78%, 70%, and 63% of correctly assigned individuals, respectively. For BCM, the change in the threshold had minor effects on the results, affecting only one or two individuals in some combinations of markers, in most cases the singletons *Tamarix hispida* Willd. and *Tamarix leptostachya* Bunge, which changed their result from incorrect to unidentified (Appendix A). Conversely, tID showed poor results independently of the threshold used, in the range of 0%–34% of correct identifications, and most of the individuals had ambiguous identifications. This low performance of tID could be explained by the small interspecific genetic distance of the studied regions. For all combinations of markers, the range of inter- and intra-specific distances overlapped, and no clear barcode gap could be found (Fig. S1).

Regarding NN and BCM, the combination of regions improved, in general, the percentage of correct identifications. From the five combinations with only one region, ycf1 and ITS were the ones that individually had a higher power of resolution, with 64% and 61% of correct identifications in NN, respectively (Fig. 1). Combinations of two markers increased the percentage of correct identifications up to the maximum value of 78% for NN and 70% for BCM. The combinations of two markers that revealed the highest percentages of correct identifications were matK+ITS (78% in NN and 69% in BCM) and ITS+ycf1 (75% in NN and 70% in BCM). The inclusion of additional markers did not improve the identification success since similar values were obtained for combinations of two markers and for combinations of three or more markers (Fig. 1).

The identification success varied widely among *Tamarix* species and the combinations of markers (Appendix A). The eight species *Tamarix africana* Poir., *Tamarix amplexicaulis*
Ehrenb., *T. aphylla*, *Tamarix dalmatica* B. R. Baum, *Tamarix hampeana* Boiss. & Heldr., *Tamarix minoa* J. L. Villar et al., *Tamarix nilotica* (Ehrenb.) Bunge, and *T. usneoides* could be identified with almost any combination of markers with NN and BCM, including matK + ITS and ITS + ycf1 (Table 3 and Appendices A2, A3). Conversely, only a few specific combinations of markers retrieved correct results for certain species. The two best-performing combinations, matK + ITS and ITS + ycf1, were able to correctly identify all the individuals of *Tamarix smyrnensis* Bunge (with both combinations), *Tamarix canariensis* Willd. (only with ITS + ycf1), *Tamarix arceuthoides* Bunge and *Tamarix parviﬂora* DC. (both only with matK + ITS) (Table 3 and Appendices A2, A3). In the case of *Tamarix boveana* Bunge, *T. gallica*, *Tamarix hohenackeri* Bunge, and *Tamarix tetragyna* Ehrenb., most of the identifications with the combinations matK + ITS and ITS + ycf1 were ambiguous or incorrect for both NN and BCM, but *T. boveana* identifications were correct with NN (Table 3 and Appendices A2, A3). In addition, no combination of markers was able to identify correctly all the individuals of any of these four species using BCM. Similarly, the four singletons *Tamarix chinenensis* Lour., *T. hispida*, *T. leptostachya*, and *Tamarix ramosissima* Ledeb. were incorrectly identiﬁed. Finally, the majority of the misidentifications, excluding the singletons, were between: (i) the species *T. arceuthoides*, *T. boveana*, *T. canariensis*, *T. gallica* and *T. tetragyna*, (ii) the species *T. hampeana* and *T. parviﬂora*, and (iii) the two species *T. africana* and *T. hohenackeri* (Table 3 and Appendix A).

The NJ trees showed similar results to NN and BCM algorithms. Individual regions showed low phylogenetic resolution, but their combination improved the resolution (Figs. 2, S2). The basal portion of the plastid and nuclear trees was quite in agreement, with a similar position of the species *T. africana*, *T. aphylla*, *T. dalmatica*, and *T. usneoides* (Fig. 2). These species appeared, at least in the plastid tree, in well-differentiated clades, and they were also easily identifiable with NN and BCM analyses. Many of the Mediterranean species, such as *T. boveana*, *T. gallica*, *T. parviﬂora*, and *T. tetragyna*, did not form independent groups in either of the two trees (Fig. 2).

Some remarkable discrepancies could be observed between ITS and plastid NJ trees. The three Mediterranean species *T. boveana*, *T. gallica*, and *T. tetragyna*, which were commonly misidentiﬁed with NN and BCM, were intermixed in the nuclear tree, whereas they also appeared grouped with other Asian and Mediterranean species (e.g., *T. arceuthoides*, *T. hampeana* *T. nilotica*, and *T. parviﬂora*, among others) in the plastid tree (Fig. 2). Besides, the studied samples of *T. parviﬂora* appeared in different positions in the ITS tree, with two individuals clustering with *T. hampeana* and one specimen of *T. parviﬂora* (field ID TparvBIar) as the external branch of the group formed by *T. dalmatica* and *T. minoa*. Another remarkable difference between these two
Table 3: Identification results for the best close match (BCM) analyses for every species for the combinations matK + ITS and ITS + ycf1

| Species       | Identification       | Species with confusion                          |
|---------------|----------------------|------------------------------------------------|
| T. africana   | Correct              | T. tetragyna**                                  |
| T. amplexicaulis | Correct             | T. boveana, T. canariensis, and T. tetragyna   |
| T. aphylla    | Correct              |                                                |
| T. arceuthoides | Correct (2*/1**)   | Incorrect (1)                                  |
| T. boveana    | Correct** (3) + Ambiguous (6*/3**) | T. gallica                                  |
| T. canariensis | Correct (4*/5**) + Incorrect* (1) | T. gallica*                                  |
| T. chinensis  | Incorrect            | T. ramosissima                                 |
| T. dalmatica  | Correct              |                                                |
| T. gallica    | Ambiguous (3) + Incorrect (3) | T. boveana, T. canariensis, and T. tetragyna |
| T. hampeana   | Correct              |                                                |
| T. hispida    | Incorrect            | T. africana* and T. canariensis**             |
| T. hohenackeri| Correct* (1) + Incorrect (1*/2**) | T. africana** and T. ramosissima            |
| T. leptostachya| Incorrect           | T. africana* and T. canariensis**             |
| T. minoa      | Correct              |                                                |
| T. nilotica   | Correct              |                                                |
| T. parviflora | Correct (3*/2**) + Incorrect** (1) | T. hampeana**                                |
| T. ramosissima| Incorrect            | T. hohenackeri                                |
| T. smyrnensis | Correct              |                                                |
| T. tetragna   | Ambiguous (2) + Incorrect (1) | T. arceuthoides and T. gallica               |
| T. usnoeides  | Correct              |                                                |

For these combinations of markers in BCM, both thresholds (1% and optimized) retrieved the exact same results. Numbers between brackets indicate the number of samples with each identification when more than one identification status was found within a species; *Applicable only for matK + ITS; **Applicable only for ITS + ycf1.

3.2 Microsatellite results

Private alleles were fairly common for the specimens and at the species level (Appendix C). Most of the analyzed samples, 56 out of 67, had no less than one private allele for each species. All species, but T. smyrnensis, showed at least one private allele, and up to a maximum of six private alleles in the western Mediterranean species T. africana and T. gallica. Every microsatellite, except for T190-32, revealed a minimum of one private allele for one species. However, in most cases, private alleles displayed low frequencies and hence, their effectiveness for the identification of species might be reduced. According to the DAPC results, private alleles had, in general, a low contribution to the variance of the discriminant axes. The alleles with a higher contribution in the DAPC were those non-private alleles that were well distributed in many species of the genus, but not in all of them (Appendix D). Nevertheless, certain private alleles had high frequencies in some species, such as those detected for the microsatellite T129-2 for T. amplexicaulis; T133-2 and T163-3 for T. aphylla; T129-2, T190-33, and T190-3 for T. hispida; T145-3 for T. nilotica; and T140-32 for T. usnoeides (Appendix C). Consequently, the detection of these private alleles could be used for the proper identification of certain species.

The number of genetic clusters identified by the four different methods varied considerably, ranging from 7 to 17. For SNAPCLUST, the AIC criterion retrieved 7 as the optimum number of clusters (Fig. S3). For STRUCTURE results, ΔK identified K = 2 as the optimum number of clusters, followed by K = 6 and K = 8 with similar values (Appendix E). For K = 2, no clear biological pattern could be observed between species (Fig. S4), whereas K = 6 and K = 8 showed small differences in the separation into two clusters of T. aphylla and T. usnoeides, and the separation into two groups of T. boveana and T. gallica/T. minoa/T. parviflora (Fig. S4). The k for which Pr(K = k) was highest was K = 8 (Appendix E); therefore, we selected it as the optimum number of clusters for the STRUCTURE analysis. For DAPC, the function “find.clusters” identified the optimum number of genetic clusters to minimize the value of BIC as 10 (Fig. S5). Finally, BAPS identified the optimum number of clusters as 17 with a probability of 0.990.

In the SNAPCLUST analysis, no admixture was detected in any individual, and almost all Tamarix species belonged to a single group (Fig. 3 and Appendix F). Only two clusters had a clear correspondence to a single species (T. amplexicaulis and T. boveana, respectively), whereas the majority of the species shared partially or entirely genetic clusters with other species. Most of the samples of T. africana clustered together, except for one sample, which formed a genetic cluster together with T. gallica and T. parviflora. Besides, one sample of T. canariensis appeared within the same cluster as
The rest of the clusters were composed by the species: (i) *T. aphylla*, *T. dalmatica*, *T. minoa*, *T. tetragyna*, and *T. usneoides*; (ii) *T. arceuthoides*, *T. canariensis*, and *T. nilotica*; and (iii) *T. chinensis*, *T. hampeana*, *T. hispida*, *T. hohenackeri*, *T. leptostachya*, *T. ramosissima*, and *T. smyrnensis*. The STRUCTURE results identified admixture for some of the samples. From the 10 different runs, three groups of runs could be identified by CLUMPAK. The major cluster (including 7 out of 10 runs), which had a higher likelihood, showed a clear genetic structure among the analyzed *Tamarix* species (Fig. 3 and Appendix G). The other two minor clusters were similar to the major one and only had minor changes (Fig. S4). STRUCTURE yielded similar results as SNAPCLUS, but with some remarkable differences. SNAPCLUS identified a large cluster that included *T. aphylla*, *T. dalmatica*, *T. minoa*, *T. tetragyna*, and *T. usneoides* that STRUCTURE fragmented in smaller clusters. The species *T. usneoides* and *T. aphylla* were separated into their own respective genetic clusters, whereas the other species appeared within other clusters. *Tamarix dalmatica* changed to the cluster of *T. africana*, *T. tetragyna* moved to the cluster of *T. boveana*, and *T. minoa* appeared within the cluster of *T. gallica*. Besides, some degree of admixture was revealed for certain Mediterranean individuals of *T. africana*, *T. canariensis*, *T. hohenackeri*, *T. minoa*, *T. parviflora*, *T. smyrnensis*, and *T. tetragyna*. In the case of *T. canariensis* and *T. minoa*, almost all their individuals (except for the individual of *T. canariensis* similar to *T. africana*) were identified.
Species DAPC also identified Smyrnensis and Nilotica and showed similar results as the STRUCTURE analysis, with this analysis broadly matched the 20 studied species.

The specimens of the Mediterranean species formed altogether a different genetic cluster from the species T. canariensis and another one from T. nilotica, and clustered with half of the individuals of T. africana. The plastid region ycf1 displayed the best identification success among all the five individual markers, which would be in line with its recent increase in DNA barcoding studies (Li et al., 2018). Our study has proved its effectiveness in taxonomically complex groups, such as the genus 

Finally, BAPS showed no significant admixture for any individual of any species. The 17 genetic clusters detected by this analysis broadly matched the 20 studied Tamarix species, although some exceptions were found (Fig. 3 and Appendix I). The genetic cluster of T. canariensis included one individual of T. africana and another one of T. gallica. The individuals of T. tetragyna appeared divided between two different genetic clusters from the species T. boveana and T. minoa, respectively. Finally, T. chinensis and T. ramosissima appeared in the same cluster, similar to the species T. hampeana, T. leptostachya and one individual from T. hohenackeri and other one from T. smyrnensis, which altogether formed the same cluster.

4 Discussion

4.1 DNA barcoding methods

DNA barcoding has proven to be a useful tool to differentiate many of the Tamarix species. A two-locus barcode has been suggested as the best combination to identify most species of this genus since the addition of more markers did not improve discrimination beyond the best-performing two-locus barcode, as Mishra et al. (2016) also stated. However, the best combination for Tamarix was not the standard barcode of rbcL + matK (CBOL Plant Working Group, 2009), but the combinations matK + ITS and ITS + ycf1. The combination of nuclear and plastid regions notably improved the identification success in the genus Tamarix, as was observed by Austerlitz et al. (2009). ITS has been proposed as a powerful DNA barcode region for many taxonomic groups since its rapid evolution allows to differentiate closely related, congeneric species (China Plant BOL Group, 2011). The plastid region ycf1 displayed the best identification success among all the five individual markers, which would be in line with its recent increase in DNA barcoding studies (Li et al., 2018). Our study has proved its effectiveness in taxonomically complex groups, such as the genus Tamarix, which would support it being described as “the most promising plastid DNA barcode of land plants” (Dong et al., 2015). However, previous phylogenetic studies have disclosed notable discrepancies between nuclear and plastid genes (Villar et al., 2019), probably due to ILS or hybridization processes among different evolutionary lineages.

Distance-based methods and tree-based methods retrieved similar results in DNA barcoding efficacy, as it has been observed for other plant groups [e.g., Poaceae and Fabaceae; Birch et al. (2017) and Wu et al. (2020), respectively]. Similar to these previous studies, the distance-based method NN yielded, in general, a better efficacy than BCM, which was also in consonance with Austerlitz et al. (2009). These authors stated that NN is a more reliable method in comparison to other methods, although the inclusion of a threshold, like in BCM, would help to avoid misidentification of unknown samples that do not have conspecifics in the reference library (Collins & Cruickshank, 2013). However, the other widely used methods based on thresholds, that is, tID, retrieved the worst results in the identification of Tamarix species, although the performance of this analysis has been documented similar to BCM (Birch et al., 2017). The use of distance thresholds in DNA barcoding has been thoroughly debated (Collins & Cruickshank, 2013, and references herein cited), and these authors recommended to use optimized thresholds generated directly from the data rather than using a
The use of SSR in species identification

The use of SSR proved to be able to identify those species, which were not easily differentiable using DNA barcoding. SSR are more variable than those nuclear and plastid regions normally used in DNA barcoding, and they have been applied with success for the genetic identification at the species level (Sharma et al., 2008; Wang et al., 2019). The use of different data sets (DNA barcoding or SSR) for the genetic identification of Tamarix species had a substantial impact on the observed results. Both types of data were capable to identify some of the species (e.g., T. amplexicaulis, T. aphylla, and T. usneoides, among others), but differed in the identification of certain Mediterranean species (e.g., T. boveana vs. T. gallica). The DNA barcoding approach was not able to distinguish between the species T. gallica and T. boveana, two western Mediterranean species that co-exist together in saline ecosystems throughout South-East Spain and northern Africa (Villar, 2017). Nevertheless, these two species are easily differentiated morphologically: T. gallica is characterized by the presence of pentamerous flowers and narrow inflorescences (4–5 mm) against T. boveana, which has tetramerous flowers and wider inflorescences (8–9 mm) (Baum, 1978). Besides, both species also show notable differential phenological features (Baum, 1978) and ecological characteristics (Moreno et al., 2019). Contrary to barcoding data, SSR data clearly separated them irrespective of the analysis used. Therefore, both molecular approaches were complementary to recognize these two common species throughout the western Mediterranean, although since all the samples of T. boveana came from South-East Spain, samples from northern Africa should be analyzed in future research to confirm these results. Similarly, DNA barcoding produced confusion between other species pairs, such as T. gallica and T. tetragyna, T. hohenackeri and T. africana, and no clear distinction even between T. gallica and T. canariensis, but SSR did. The use of SSR in the taxonomic identification of species is rare in literature, even though SSR can be superior to the commonly used sequence data for DNA barcoding (Wang et al., 2019). Consequently, the use of a combination of DNA barcoding and SSR seemed to differentiate the majority of the species of the genus Tamarix, especially those taxa from the Mediterranean region.

In addition, taking into account the type of analysis to be carried out is relevant since the final results might vary greatly. The different existing methods of SSR analyses retrieved different results, as Pinho et al. (2019) recently reported. Although the obtained data from the analyzed methods suggested relatively similar relationships between species, they markedly differed in the optimum number of genetic clusters finally identified. SNAPCLUST gave a lower number of genetic clusters than the rest of the methods, while BAPS retrieved a higher number. Pinho et al. (2019) found similar patterns between methods with simulated data sets, and these authors argued that BAPS overestimates the true number of clusters, and SNAPCLUST, DAPC, and STRUCTURE underestimate it. This pattern has also been detected in our data, and according to Pinho et al. (2019), the true number of clusters should probably be between 10, identified by DAPC, and 17 of BAPS. In our study, we have examined 20 species, and hence the true number of genetic clusters should be close to 20, but all four analyzed methods underestimated this value. Except for DAPC, the other three SSR analyses (SNAPCLUST, STRUCTURE and BAPS) rely on genetic models that assume either Hardy-Weinberg or linkage equilibrium (Pinho et al., 2019), which are not expected to be met by our data. Some species were only represented by one individual, and it is known that unequal sample sizes or substantial phylogenetic signals can lead to poor estimates of $K$ (Kalnowski, 2011; Neophytou, 2014; Pinho et al., 2019). Consequently, the limitations of the methods used for SSR analyses and the sampling procedure of the species may probably be the cause for the small number of clusters observed by the analyses.

The ideal method of analysis for the identification of Tamarix species based on SSR data should meet some additional particular requirements, such as the identification of potential hybridization processes. The presence of hybrid morphologies in Tamarix has been reported between native taxa and even between native and allochthonous taxa (Gaskin & Shafroth, 2005; Samadi et al., 2013; Mayonde et al., 2016). The four used methods of data analysis (BAPS, DAPC, STRUCTURE, and SNAPCLUST) are able to detect admixture between different genetic clusters (Pritchard et al., 2000; Corander & Marttinen, 2006; Jombart & Collins, 2017; Beugin et al., 2018), and hence, theoretically able to detect hybridization processes. Our data revealed that only STRUCTURE and DAPC were able to identify admixture between groups, but not BAPS and SNAPCLUST. In previous studies, BAPS and STRUCTURE were able to identify admixed individuals, but not SNAPCLUST and DAPC (Pinho et al., 2019) and Neophytou (2014) found that BAPS tended to underestimate the number of admixed individuals. Using the same data matrix, our results would suggest that the analyzed methods might have a different sensitivity to violations in genetic model assumptions or to the low numbers of individuals. However, scarce comparative studies on other groups of plants are available, and therefore, more comparative studies using different clustering methods would be needed to elucidate this aspect. Moreover, DAPC and STRUCTURE were the best ones for the genetic identification at the species level, with similar results. However, DAPC has some characteristics that make it preferable over STRUCTURE. DAPC is orders of magnitude faster than STRUCTURE and it is also as accurate in unraveling the underlying structuring (Jombart et al., 2010). Besides, DAPC is also a more versatile method since it does not rely on a particular population genetics model unlike other commonly used methods (Jombart et al., 2010). Nonetheless, the use of different methods when performing genetic clustering is highly encouraged since different methods could give widely different results (Pinho et al., 2019).
The separation of the Mediterranean species *T. gallica* and *T. canariensis* has not proved conclusive with our data of DNA barcoding and SSR. These two species are morphologically rather similar, with slight differences based on a glabrous rachis of the racemes, bracts oblong not exceeding the calyx, and petals elliptic 1.5–1.75 mm long in *T. gallica*, and a papillose rachis of the racemes, bracts linear-triangular equaling or exceeding the calyx, and petals obovate 1.25–1.5 mm long in *T. canariensis* (Baum, 1978). Moreover, both species share overlapping distribution areas and ecological characteristics in the western Mediterranean Basin (Baum, 1978; Cirujano, 1993). Although their taxonomic identity has been thoroughly discussed, recent phylogenetic studies suggested a clear separation between these two species (Villar et al., 2019, and references herein cited). Therefore, these authors maintained their taxonomic identification and they restricted the geographical area of distribution of *T. canariensis* only to the Canary Islands, while *T. gallica* showed a widespread distribution through European and North African territories along the western Mediterranean Basin. However, our data pointed out that these two species could not be clearly separated with SSR, suggesting a more complex genetic relationship between them than previously reported by Villar et al. (2019). The use of SSR would seem to have sufficient power to elucidate the genetic relationship between these two species given their high genetic variability. However, further specific analyses with an extensive population sampling throughout the whole range of the species would be needed to arrive at accurate conclusions. In addition, *T. gallica* also showed a close genetic relationship with *T. parviflora* since these two species could not be clearly separated by the SSR analyses (except for BAPS). However, unlike the case of *T. canariensis*, these two species have clear morphological differences: *T. gallica* has pentamerosum flowers and a caducous corolla, while *T. parviflora* has tetramerous flowers with a subpersistent corolla (Baum, 1978). Besides, these two species are grouped in separated phylogenetic clades (Villar et al., 2019) and they have different geographical distributions, with *T. gallica* occurring in the western Mediterranean and *T. parviflora* in the eastern Mediterranean (Baum, 1978; Villar, 2017). Given the differences between these two distinctive species, their identification with SSR data might probably improve by the addition of further specimens of *T. parviflora*.

Some species, such as *T. chinensis*, *T. hispida*, *T. hohenackeri*, *T. leptostachya*, and *T. ramosissima*, mainly from Asia, could not be differentiated by any of the methods here applied. The species *T. chinensis*, *T. hohenackeri*, and *T. ramosissima* are phylogenetically related, as they all appeared in the same well-supported clade (Villar et al., 2019). These three species exhibit a similar morphology with sessile leaves with a narrow base, persistent petals after anthesis, and stamens inserted between the nectariferous disc lobes (Baum, 1978). As a matter of fact, the morphological delimitations of these species could be diffuse (Baum, 1978; Yang & Gaskin, 2007), which would explain the low genetic differentiation observed among them. The other two species, *T. hispida* and *T. leptostachya*, could not be genetically differentiated either, though these two species also appeared in separated clades (Villar et al., 2019) and exhibit notably different morphologies. Although both species are characterized by a caducous corolla and stamens inserted above the nectariferous disc lobes, *T. hispida* differs from *T. leptostachya* in some vegetative and floral features related to the indumentum (hairy vs. glabrous), basal morphology of the leaves (cordate-auriculate vs. narrow), and length of the racemes (1.5–7 cm vs. 7–15 cm), among other characters (Baum, 1978). Nevertheless, the identification success for these five species is quite low, which could be caused by the low number of samples analyzed since SSR analyses require a minimum number of individuals to perform correctly (Hale et al., 2012). As a consequence, an increase in the number of individuals could improve the genetic separation and identification of these species.

### 4.3 Hybridization processes in *Tamarix*

Hybridization processes have previously been reported for some taxa of *Tamarix* (Gaskin & Schaal, 2003; Gaskin & Shafroth, 2005; Gaskin & Kazmer, 2009; Samadi et al., 2013; Mayonde et al., 2016). Our results are in agreement with these reports, and they suggest the existence of a hybrid origin for certain specific individuals assigned to the Mediterranean species *T. africana*, *T. canariensis*, *T. minoa*, and *T. Smyrnensis*, based on STRUCTURE and DAPC since these two methodologies were the only ones able to identify admixture between genetic clusters. Two samples of *T. africana* (field IDs Ebr07 and FrCg1.19) might have a hybrid origin between this species and *T. gallica* since these two species co-exist in the sampling locations of the two specimens. Moreover, the potential hybrid origin of these two individuals was also supported by the NJ tree data since they clustered with all the samples of *T. africana* in the plastid tree and with *T. gallica* in the ITS tree. Similarly, Terzoli et al. (2014) also reported the potential presence of hybrid individuals between these species *T. africana* and *T. gallica* from Italian populations. Another example of potential hybridization was identified for one sample of *T. canariensis* (field ID TGC4.21), whose genetic data show a clear mixture with *T. africana*. Once again, they both occur together in the sampling location, and this peculiar individual clustered with *T. canariensis* and *T. africana* in the plastid and ITS NJ trees, respectively. Finally, one sample of *T. Smyrnensis* (field ID 4690-06) seems also to have a hybrid origin, but it could not be properly identified with the current sampling.

Another interesting case of possible hybridization corresponded to the eastern Mediterranean species *T. minoa*. Based on STRUCTURE data, this species appeared as a clear mixture between two different groups of species: *T. boveana*/*T. tetragna* and *T. gallica*/*T. parviflora*. Two of these species (*T. boveana* and *T. gallica*) have a western Mediterranean distribution, but *T. tetragna* and *T. parviflora* have an eastern Mediterranean distribution, like *T. minoa*, suggesting that the two latter species might have played a relevant role in the origin of this taxon as putative parents. Villar et al. (2019) recently stated the possibility that the species *T. minoa* had a potential hybrid origin since it clustered with different taxa (*T. dalmatica* and *T. indica*) depending on the nuclear and plastid phylogenetic data, respectively. It is interesting to note that, based on the DAPC analysis, *T. minoa* appeared in the same genetic cluster as *T. dalmatica*. The discrepancy between these two analyses (DAPC and STRUCTURE) could be explained by the poor sampling of the species and populations involved in
the possible hybridization process of this peculiar species. Further studies would be needed to explain the hybrid origin and the putative parents of *T. minoa*. Nevertheless, the results here exposed, together with the phylogenetic relationships reported by Villar et al. (2019), would probably support the hypothesis of hybridization in speciation within *Tamarix* since hybridization has been reported as an important force in generating angiosperm species diversity (Soltis & Soltis, 2009). The use of codominant nuclear markers, such as SSR, is highly encouraged in those plant groups with reported hybridization processes, such as *Tamarix*. DNA barcoding analyses are usually based on plastid markers with maternal transmission (China Plant BOL Group, 2011), and hence it is expected to fail with hybrid samples. Nevertheless, overinterpreting STRUCTURE results would not be recommended since different scenarios of the genetic history could lead to undistinguishable STRUCTURE results (Lawson et al., 2018). Consequently, further research on this topic would be needed to elucidate the hybridization processes within the genus *Tamarix*.

5 Conclusion

DNA barcoding allows the genetic identification of certain species of the genus *Tamarix*. The two-locus barcodes *matK* + *ITS* and *ITS + ycf1* were the best-performing combinations, allowing up to 69% and 70%, respectively, correct identification with BCM. However, given the recent radiation of the genus, DNA barcoding fails in phylogenetically close groups, such as many Mediterranean species. In this context, the use of highly variable SSR can help in species identification. Our study shows that the combination of both types of data (DNA barcoding and SSR) improves success, and it is especially relevant in detecting hybridization, which is common in the genus. However, caution must be exercised when choosing the clustering methods for the SSR data, since different methods can yield very different results.

Acknowledgements

We would like to thank John F. Gaskin for providing plant material, Olivier Maurin for his helpful suggestions that greatly improved the manuscript, and the curators of the herbaria ABH, MO, and W. This research has been funded by the Project OAPN 354/2011 (Ministerio de Agricultura, Alimentación y Medio Ambiente), the FPU grant program (AP-2012-1954 and EST14/00383, Ministerio de Educación) and BEST/2019/155 grant (Generalitat Valenciana, Spain). This study is part of the Ph.D. thesis of A.T.

Data Accessibility

The raw data that support the findings of this study can be obtained from https://doi.org/10.5061/dryad.ttdzo8mow.

References

Akaie H. 1998. Information theory and an extension of the maximum likelihood principle. In: Parzen E, Tanabe K, Kitagawa G eds. Selected papers of Hirotugu Akaie. New York: Springer. 199–213.

Austerlitz F, David O, Schaeffer B, Bleakley K, Olteanu M, Leblois R, Veuille M, Laredo C. 2009. DNA barcode analysis: a comparison of phylogenetic and statistical classification methods. BMC Bioinformatics 10: S10.

Baum BR. 1978. The genus *Tamarix*. Jerusalem: Israel Academy of Sciences and Humanities.

Beugin MP, Gayet T, Pontier D, Devillard S, Jombart T. 2018. A fast likelihood solution to the genetic clustering problem. Methods in Ecology and Evolution 9: 1006–1016.

Birch JL, Walsh NG, Cantrill DJ, Holmes GD, Murphy D.J. 2017. Testing efficacy of distance and tree-based methods for DNA barcoding of grasses (Poaceae tribe Poaeae) in Australia. PLoS One 12: e0186259.

Brock JH. 1994. *Tamarix* spp. (salt cedar), an invasive exotic woody plant in arid and semi-arid riparian habitats of western USA. *Ecology and Management of Invasive Riverside Plants* 4: 28–44.

Brown SDJ, Collins RA, Boyer S, Lefort M-C, Malumbres-Olarte J, Vink CJ, Cruickshank RH. 2012. SPIDER: An R package for the analysis of species identity and evolution, with particular reference to DNA barcoding. *Molecular Ecology Resources* 12: 562–565.

CBOL Plant Working Group. 2009. A DNA barcode for land plants. *Proceedings of the National Academy of Sciences of the United States of America* 106: 12794–12797.

China Plant BOL Group. 2011. Comparative analysis of a large dataset indicates that internal transcribed spacer (ITS) should be incorporated into the core barcode for seed plants. *Proceedings of the National Academy of Sciences of the United States of America* 108: 19641–19646.

Cirujano S. 1993. *Tamarix* L. In: Castroviejo S, Aedo C, Cirujano S, Lainz M, Montserrat P, Morales R, Muñoz Garmendia F, Navarro C, Paiva J, Soriano C eds. *Flora iberica* 3. Madrid: Real Jardín Botánico, CSIC. 437–443.

Collins RA, Cruickshank RH. 2013. The seven deadly sins of DNA barcoding. *Molecular Ecology Resources* 13: 969–975.

Corander J, Marttinen P. 2006. Bayesian identification of admixture events using multilocus molecular markers. *Molecular Ecology* 15: 2833–2843.

Corander J, Marttinen P, Siren J, Tang J. 2008. Enhanced Bayesian modelling in BAPS software for learning genetic structures of populations. *MBC Bioinformatics* 9: 539.

Dong W, Xu C, Li C, Sun J, Zuo Y, Shi S, Cheng T, Guo J, Zhou S. 2015. *ycf1*, the most promising plastid DNA barcode of land plants. *Scientific Reports* 5: 8348.

Doyle JJ, Deyhe JL. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemical Bulletin* 19: 11–15.

Dunning LT, Savolainen V. 2010. Broad-scale amplification of *matK* for DNA barcoding plants, a technical note. *Botanical Journal of the Linnean Society* 164: 1–9.

Evanno G, Regnaut S, Goudet J. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. *Molecular Ecology* 14: 2611–2620.

Ford CS, Ayres KL, Toomey N, Haider N, van Alphen Stahl J, Kelly LJ, Wikström N, Hollingsworth PM, Duff RJ, Hoot SB, Cowan RS, Chase MW, Wilkinson MJ. 2009. Selection of candidate coding DNA barcoding regions for use on land plants. *Botanical Journal of the Linnean Society* 159: 1–11.

Gaskin JF. 2003. Molecular systematics and the control of invasive plants: a case study of *Tamarix* (Tamaricaceae). *Annals of the Missouri Botanical Garden* 90: 109–118.
Terrones A, Juan A. 2020. Seventeen new microsatellites for Tamarix gallica and cross-amplification in Tamarix species. Applications in Plant Sciences 8: e11317.

Terzoli S, Abbruzzese G, Beritognolo I, Sabatti M, Valentini R, Kuzminsky E. 2014. Genetic characterization of a Tamarix spp. germplasm collection in Italy. Botany 92: 360–369.

Twyford AD, Ennos RA. 2012. Next-generation hybridization and introgression. Heredity 108: 179–189.

Villar JL. 2017. Tamarix. Euro+Med Plantbase. The information resource for Euro-Mediterranean plant diversity [online]. Available from http://www2.bgbm.org/EuroPlusMed/ [accessed 19 February 2021].

Villar JL, Alonso MÁ, Juan A, Gaskin JF, Crespo MB. 2019. Out of the Middle East: New phylogenetic insights in the genus Tamarix (Tamaricaceae). Journal of Systematics and Evolution 57: 488–507.

Wang J, Luo J, Ma YZ, Mao XX, Liu JQ. 2019. Nuclear simple sequence repeat markers for DNA barcodes for identification of closely related Rhododendron species on the same mountain. Journal of Systematics and Evolution 57: 278–286.

Wu FF, Gao Q, Liu F, Wang Z, Wang Jl, Wang XG. 2020. DNA barcoding evaluation of Vicia (Fabaceae): Comparative efficacy of six universal barcode loci on abundant species. Journal of Systematics and Evolution 58: 77–88.

Yang Q, Gaskin JF. 2007. Tamaricaceae. In: Wu ZY, Raven PH, Hong DY eds. Flora of China. Beijing: Science Press; St. Louis: Missouri Botanical Garden Press. 13: 58–69.

Yu J, Xue JH, Zhou SL. 2011. New universal matK primers for DNA barcoding angiosperms. Journal of Systematics and Evolution 49: 176–181.

Supplementary Material

The following supplementary material is available online for this article at http://onlinelibrary.wiley.com doi/10.1111/jse.12830 suppinfo:

Fig. S1. Inter- and intra-specific pairwise distances among all the Tamarix individuals for the five individual regions (rbcL, matK, ITS, trnH-psbA, and ycf1), the two best performing two-locus combinations (matK + ITS and ITS + ycf1) and the combination of all five regions.

Fig. S2. Neighbor joining (NJ) trees of (A) rbcL, (B) matK, (C) trnH-psbA, and (D) ycf1. Numbers above each branch indicate bootstrap support (BS).

Fig. S3. Akaike information criterion (AIC) for each number of clusters (k) from 1 to 22 for the function “snapclust-choose.k” for SNAPCLUST.

Fig. S4. Estimated membership coefficients for each individual in each cluster of the STRUCTURE analysis for the major clusters of K = 2 and K = 6, and for the major and minor clusters for K = 8.

Fig. S5. Bayesian information criterion (BIC) for each number of clusters from 1 to 22 for the function “find.clusters” of the DAPC.  

Fig. S6. A-score for each number of retained principal components (PC) from 1 to 10 for the DAPC. The a-score was calculated with 1000 simulations per number of retained PC.

Fig. S7. Scatterplot of the first two discriminant axis (DA) of the discriminant analysis of principal components (DAPC). Prior grouping of individuals can be seen in Appendix H and it is indicated with colors and inertia ellipses. On the top right, cumulative variance is explained by the principal component analysis (PCA) eigenvalues. For the DAPC, only the first five principal components (PC) were retained, which are marked in darker gray and conserved 46.2% of variance. On the top left, eigenvalues of the discriminant analysis (DA).

Appendix A1. Summary of the barcoding analyses. On the following six sheets, raw data of each analysis are included.

Appendix A2. Raw results of the nearest neighbor (NN) analyses.

Appendix A3. Raw results of the best close match (BCM) analyses with 1% of threshold.

Appendix A4. Raw results of the best close match (BCM) analyses with an optimized threshold.

Appendix A5. Raw results of the threshold-based analyses (tID) with 1% of threshold.

Appendix A6. Raw results of the threshold-based analyses (tID) with an optimized threshold.

Appendix A7. Raw results of the neighbor Joining (NJ) trees.

Appendix A8. Raw microsatellite data.

Appendix A9. List of private alleles of the microsatellite data by species. Frequency values higher than 0.5 are highlighted.

Appendix D. Variance contribution of the different alleles to the five discriminant axes (LD) of the discriminant analysis of principal components (DAPC). Values higher or equal than 0.05 are highlighted. Private alleles are also indicated.

Appendix E. Estimation of the best number of populations K for the STRUCTURE results, based on the distribution of the value AK described by Evanno et al. (2005), and on the Bayes rule with a uniform prior K and evaluation of the estimated log likelihood (ln(Pr(X|K))), as described by Pritchard et al. (2010).

Appendix F. Group membership probabilities of every individual for the SNAPCLUST analysis.

Appendix G. Estimated membership coefficients for each individual in each cluster for the major cluster of the STRUCTURE analysis.

Appendix H. Posterior membership probabilities, and prior and posterior group assignments, of every individual for the DAPC.

Appendix I. Bayesian posterior mean estimate of the proportion of the genome represented by each cluster for every individual for the admixture BAPS analysis.