



Development of a Core Collection for *Tetraclinis articulata* Using ISSR Markers and Maximization Strategy

Makkaoui Meryem^{1,2} · Abbas Younes⁴ · Medraoui Leila² · Alami Mohammed² · El Antry-Tazi Salwa¹ · Filali-Maltouf Abdelkarim³

Received: 6 March 2022 / Accepted: 18 February 2023 / Published online: 14 March 2023
© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2023

Abstract

Tetraclinis articulata (Vahl) Masters (Cupressaceae) is a woody tree species indigenous to North Africa. It is also found occasionally in Malta and Spain showing significant adaptability to different bioclimatic conditions. Due to its economic and ecologic values, this tree is affected by human and environmental disturbance leading to habitat fragmentation and genetic diversity loss. In this study, we aimed to promote the conservation of *T. articulata*, by measuring the allelic richness of a representative sample of 129 germplasms. Results showed differences in allele distribution. Inter-simple sequence repeat (ISSR) markers and the advanced maximization strategy implemented in PowerCore software were deployed to establish the first core collection of *T. articulata*. A core collection containing 10.07% out of the 129 *T. articulata* germplasms was established. The representativeness of the core collection was evaluated and validated by different methods. The genetic diversity parameters or molecular traits (the observed number of alleles (Na), the effective number of alleles (Ne), Shannon's information index (*I*), and Nei's expected heterozygosity (He) showed similarities between the core collection and the entire collection averaging at Na = 1.712, Ne = 1.398, *I* = 0.395, and He = 0.253 and Na = 1.970, Ne = 1.363, *I* = 0.382, and He = 0.238, respectively. The analysis of molecular variance indicates that most of the genetic variation in *Articulata* core collections resided within populations (96%), with lesser amounts among populations (4%). Two clusters were identified with STRUCTURE and validated with unweighted pair group method with arithmetic mean dendrogram. The identified collection will be useful for conservation and breeding programs.

Keywords *Tetraclinis articulata* · Conservation · Allelic richness · Core collection · Advanced maximization strategy

Key messages

- ISSR markers are a suitable tool for developing core collection for *Tetraclinis articulata*.
- The core collection resulted in 13 germplasms with the maximum of genetic diversity.
- The allelic richness shows differences throughout regions.
- PowerCore by applying the advanced maximization (M) strategy has proved a high efficiency in capturing pivotal alleles.

✉ Filali-Maltouf Abdelkarim
a.filalimaltouf@gmail.com

Makkaoui Meryem
meryemmakkaoui1989@gmail.com

Abbas Younes
abbayouns@gmail.com

Medraoui Leila
l.medraoui@gmail.com

Alami Mohammed
alamimohammed07@gmail.com

El Antry-Tazi Salwa
elantry.salwa@gmail.com

Introduction

The existence for forest genetic resources is increasingly important to fulfill current and future needs, such as increasing productivity and anticipating pest and disease

¹ Forest Research Center, Silviculture Department, Molecular Biology Laboratory, PO Box 763, Rabat, Morocco

² Faculty of Sciences, Microbiology and Molecular Biology Laboratory, Mohammed V University, PO Box 1014, Rabat, Morocco

³ Hassan II Academy of Science and Technology, Rabat, Morocco

⁴ Polydisciplinary Faculty, Polyvalent Team R&D-Mghila, Sultan Moulay Slimane University, PO Box, 592, Beni Mellal, Morocco

epidemic (Nirsatmanto 2020). Indeed, they are important in maintaining biodiversity, protecting soil, and adapting forests to climate change. As the effects caused by human activities and climate change, there is a high risk of predisposing forest genetic resources to decline and loss. Efforts are ongoing to resolve trade-offs inherent in the demand for timber and the demand for other goods and services (e.g., recreation) (Aggestam et al. 2020).

According to FAO report, Africa and Asia are two regions with the highest forest species diversity and almost half of the forest species reported by countries are threatened (Beardmore et al. 2014). The conservation of forest species is the most rational way to keep fauna and flora ecosystem in balance. In conserving forest genetic resources, we are concerned primarily with conserving genetic diversity within and among populations of a species (Rajora and Mosseler 2001). However, plant species differ enormously in biological traits and environmental requirements, making it unrealistic to apply a single system to all species (Peñas et al. 2016). Moreover, establishment and management of a collection with many individuals, especially for perennial woody plants, need spacious areas, skillful labor, and expertise in plant identification which may be a high-cost operation (Brown 1989; Hintum et al. 2000). Commonly, it is not feasible to maintain the entire gene pool; therefore, so-called core collections are established (De Beukelaer et al. 2018). A core collection is a subset of accessions which were selected by eliminating closely related samples while still capturing the genetic diversity of the original set of accessions (Frankel 1984).

For tree species with a wide natural distribution range, the purpose of constructing a core collection is to protect the diversity of germplasm resources for high-quality characteristics that are easily identified in the wild (e.g., growth traits) while for tree species with a limited natural range, the purpose is to protect the diversity of germplasm resources (Wang et al. 2021). Many various core collection selection methods have been proposed, and each of them differs in defining the selection criteria for a respective core collection establishment (Kelblerová et al. 2022). Originally, phenotypic data containing both morphological and agronomical traits were used to create core collections, whereas nowadays, molecular markers as neutral tools for measuring genetic variation have become the tool of choice (Soleimani et al. 2020). They have been used for creation of core collection in various crops such as *Triticum aestivum* (Balfourier et al. 2007) (Pascual et al. 2020), *Glycine max* (L.) Merr. (Jeong et al. 2019), *Juglans regia* L. (Mahmoodi et al. 2019), *Olea europaea* L. (El Bakkali et al. 2013), *Capsicum annuum* L. (Zewdie et al.

2004), *Setaria italica* (L.) Beauv. (Upadhyaya et al. 2008), *Saccharum officinarum* L. (Balakrishnan et al. 2000), and *Oryza sativa* L. (Zhang et al. 2011). For most forest tree species, there are only few core germplasm collections based on genotypic or phenotypic traits or both, such as *Picea abies* (Kelblerová et al. 2022), *Pinus yunnanensis* (Wang et al. 2021), *Eucalyptus cloeziana* (Lv et al. 2020), *Ceratonia siliqua* L. (Di Guardo et al. 2019), *Dalbergia odorifera* (Liu et al. 2019), *Quercus suber* L. (Laakili et al. 2018), *Cunninghamia lanceolata* (Duan et al. 2017), *Argania spinosa* (Mouhaddab et al. 2016), and *Cryptomeria japonica* (Miyamoto et al. 2015).

Using molecular markers, for creating a core collection reflecting directly the variation at the genetic level, is preferred because each entry in the core is chosen on the basis of its allelic potential, irrespectively from its phenotypic expression (Mahmoodi et al. 2019). Indeed, the cost, time, and effort required for phenotypic characterization, especially in a woody perennial species collection, are much greater than required for an assessment using molecular tools (El Bakkali et al. 2013). In light of this situation, few studies were devoted to the genetic diversity and the structure of *T. articulata* using inter-simple sequence repeat (ISSR) markers (Makkaoui et al. 2020; Sánchez-Gómez et al. 2013) and amplified fragment length polymorphism (AFLP) and SSR markers (García-Castaño et al. 2021). Among different marker systems, ISSR are mostly dominant genetic markers that also include the AFLP, random amplified polymorphic DNA (RAPD) markers, and their derivatives (Ng and Tan 2015). The method provides highly reproducible results and generates abundant polymorphisms (Machkour-M'Rabet et al. 2009) and can rapidly differentiate the closely related individuals (Panahi and Ghorbanzadeh Neghab 2013). These characteristics allow ISSR markers to be suitable for developing core collection for *T. articulata*.

Two main methods have been used to assist the establishment of core collections: the M strategy (for maximization) and PowerCore. The M strategy or maximizing method is based on using genetic markers to sample the core collection while maximizing allele richness at each marker locus (Schoen and Brown 1993); it is worth noting that the M strategy not only determines the number of accessions that should come from different groups but also identifies the accessions that should be included (Hintum et al. 2000). PowerCore on the other hand selects the accessions with higher diversity representing the total coverage of marker alleles and trait states present in the entire collection (Kim et al. 2007) (Belaj et al. 2012). To the best of our knowledge, there are no studies focused on constructing a core

collection for *Tetraclinis articulata* based on molecular marker data and maximizing method.

Tetraclinis articulata (Vahl) Masters consists of a monotypic Mediterranean genus of the Cupressaceae family, located in northern Africa and southern Spain (Rourke 1991). In Morocco, the extent of *T. articulata* reaches 80% of the total tetraclinaie area (HCEFLCD 2004). Like other woody species, *T. articulata* is a multifunctional conifer offering ecological and economical functions: high yield, high pest resistance, and a very desirable wood (Fidah et al. 2015), appreciated for its natural beauty and its quality for marquetry (El Bouhtoury-Charrier et al. 2009). The chemical composition of essential oil extracted from *T. articulata* showed antimicrobial activity against *Staphylococcus aureus* and *Micrococcus luteus* (Bourkhiss et al. 2007). At present, Arar tree is in an elevated risk of losing habitat area and genetic material from its gene pool caused by fires, invasive species, storms, and droughts combined with greater competition for land with other sectors, including agriculture and urbanization. Accordingly, *T. articulata* requires particular attention for conservation throughout in situ and ex situ conservation, germplasm banks, and core collections. In light of this situation, only two studies were devoted to study the genetic diversity of *T. articulata* using ISSR markers (Makkaoui et al. 2020; Sánchez-Gómez et al. 2013).

In order to construct a core subset of *T. articulata* by understanding the genetic diversity pattern and structure, 15 ISSR markers are used to assess the genetic diversity of 129 germplasm of *T. articulata* originated from ten wild populations across the entire distribution area in Morocco (Makkaoui et al. 2020). Analyzing differences in genetic diversity of the core collection and through the entire population is a useful approach for testing the effectiveness of the core collection (Wang et al. 2011). Hence, choosing the suitable evaluation measurement depends on the purpose of the core collection. Core collections can be created with two general purposes in mind: (1) maximizing the total genetic diversity in a core (2) maximizing the representativeness of the genetic diversity of the whole collection in a core collection (Marita et al. 2000). A variety of measures have been used to evaluate the core collection (Wang et al. 2021). Genetic diversity parameters, such as total alleles, Na, Ne, H, I, and PCoA, are frequently used to evaluate the genetic diversity of the core collection (Li et al. 2006; Wang et al. 2011). Other criteria are also used to evaluate and validate the representativeness of the core collection such as the analysis of molecular variance (AMOVA) and the unweighted pair group method with arithmetic (UPGMA) based on genetic distances. In this context, the main objective of this study was to provide a methodical strategy for the construction of a core collection for *T. articulata* germplasm resources and evaluate their genetic diversity and structuring based on variations as revealed by ISSR markers.

Material and Methods

Sampling and Assessing Suitability DNA for ISSR Amplification

The initial collection is composed of 129 samples from ten natural forests across the entire distribution area in Morocco (Makkaoui et al. 2020). Using a commercial kit (DNA mini kit, Bioline, USA), genomic DNA was extracted from 129 lyophilized and ground leaves (50-mg powder), according to the manufacturer's instructions. Purified DNA quality was verified by spectrophotometry (ND-2000, NanoDrop, USA) at 260 and 280 nm and by electrophoresis on 1% agarose gel. The DNA concentration was adjusted to a 50 ng/μL and stored at – 20 °C for ISSR amplification.

A total of 51 primers were initially screened. Fifteen of them, which generated discernable and bright bands, were employed for the analysis of 129 samples (Makkaoui et al. 2020). The ISSR amplifications were carried out in a total volume of 25 μL containing 2.5 mM MgCl₂, 2 mM dNTP, 4 μM of primer, 1 unit Taq DNA polymerase, and 100 ng of total DNA. The reaction mixture was subjected to amplification in GenAmp® thermal cycler (Applied Biosystems, CA, USA) following the program: initial denaturation for 7 min at 94 °C, followed by 39 cycles of 30 s at 94 °C, 45-s annealing temperature 2 min at 72 °C, and 7 min final extension at 72 °C. A control was run by replacing genomic DNA with ddH₂O. Amplified products were electrophoresed in 2.8% agarose gel in 1X TAE buffer at 120 V for 2 h and then stained with ethidium bromide for 20 min. Gels were visualized under ultraviolet and photographed with Enduro™ GDS (Labnet, USA). The size of amplified fragments was estimated using 1-kb ladder molecular size standards (Bioline, USA) (Makkaoui et al. 2020).

Allelic Richness Distribution and Structure of the Full Collection

ISSR amplified fragments were scored for band presence (1) or absence (0), and a binary qualitative data matrix was constructed as previously described (Makkaoui et al. 2020). The allelic richness—also known as the neutral variability—and the polymorphism of the 10 populations of *Articulata* were mapped using the DIVA-GIS software (Hijmans et al. 2001); it was also used to map frequently alleles, locally common alleles, rare alleles, and private alleles (Fig. 1).

To explore spatial genetic structure of *T. articulata* in Morocco, the spatial principal component analysis (sPCA) (Fig. 2) was performed using all loci.

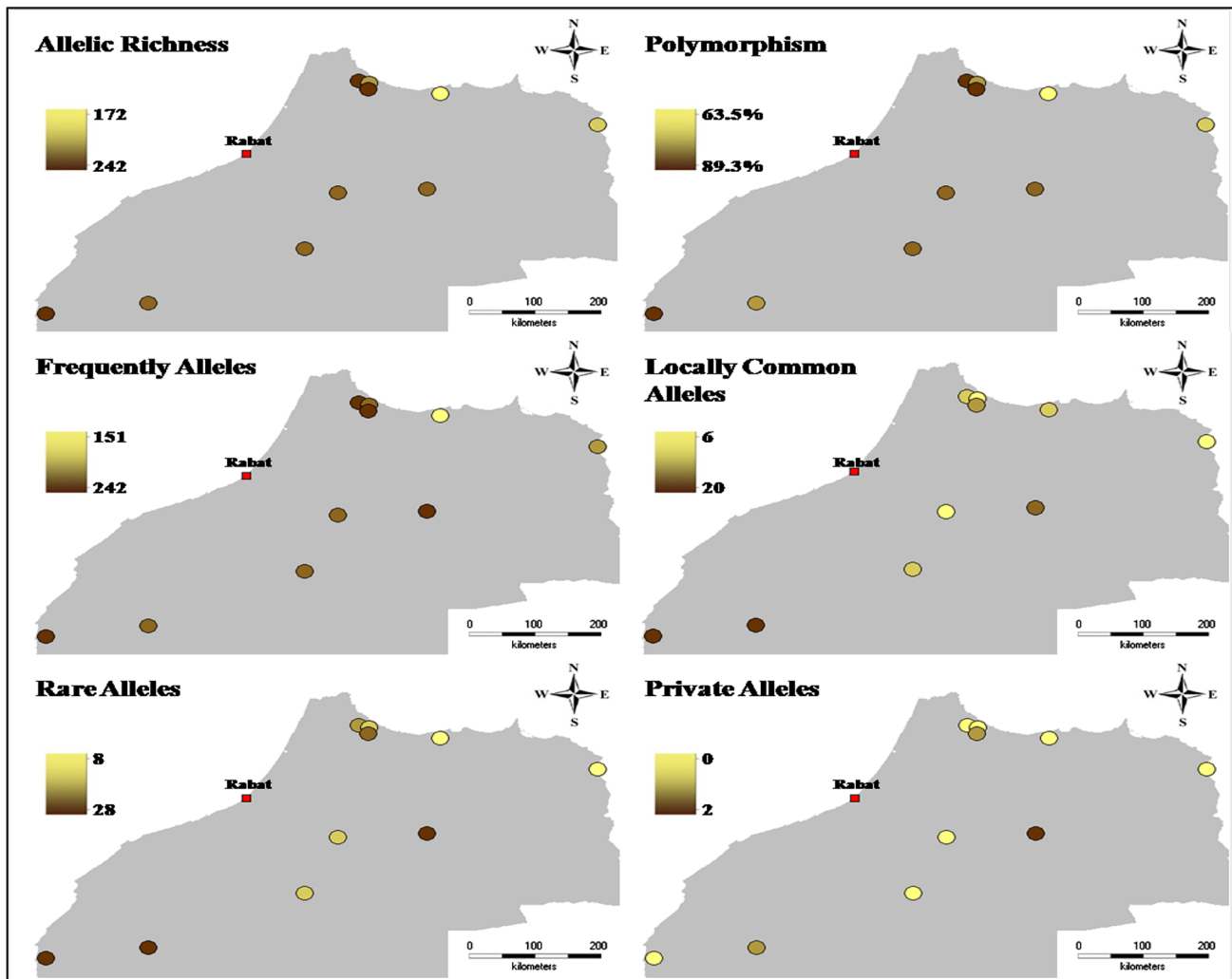


Fig. 1 Spatial variation of different genetic parameters, represented at a resolution of 30-s grid cells and circular areas of 10 km

The ten original germplasm resources were grouped into two major regions, according to geographical proximity, for further analysis:

1. North and East regions: Oued Laou, Dar ben Karrich, Chefchaouen, Al Hoceima, and Oujda
2. Central and South regions: Oulmes, Fes-Boulemane, Beni-Mellal, Essaouira, and Marrakech

Core Collection for *T. articulata*

Core Collection Construction

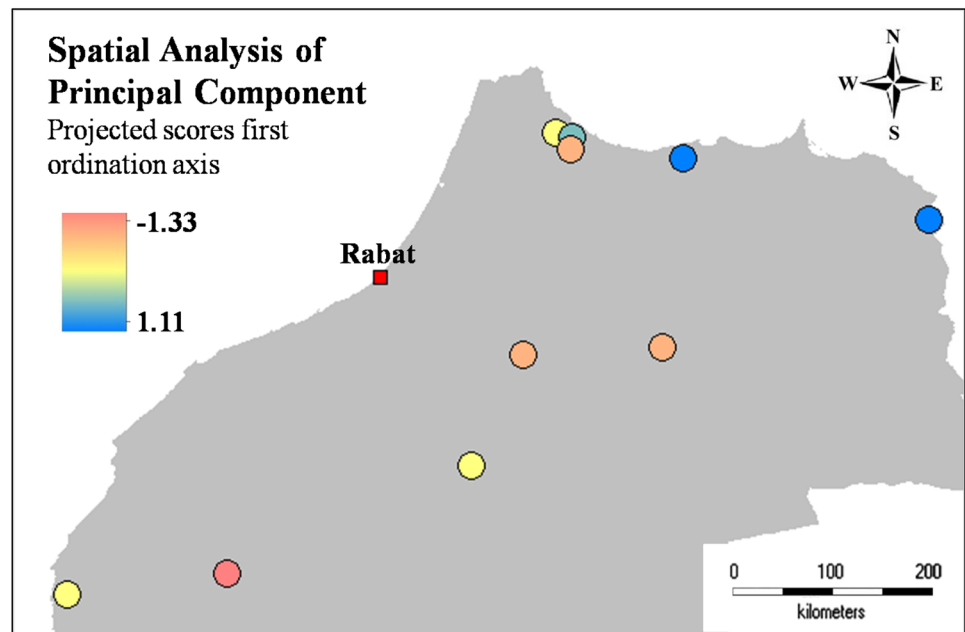
In the present study, the maximizing strategy implemented in PowerCore software (Kim et al. 2007) is used to construct the core subset. The M strategy (maximization or marker allele richness) maximizes the number of alleles at each marker by eliminating redundant accessions (Ndjondjop

et al. 2017; Schoen and Brown 1993). PowerCore selects from the whole collection the accessions with the highest diversity through a modified heuristic algorithm (Mahmoodi et al. 2019). Thus, the size of the final core collection is not known a priori and it depends on the levels of variability and redundancy present in the collection (Belaj et al. 2012).

Genetic Diversity Analysis

Using GenAlEx v6.5 software and assuming the Hardy–Weinberg equilibrium, the genetic diversity parameters (molecular traits) (the observed number of alleles (N_a), the effective number of alleles (N_e), the Shannon information index (I), the expected heterozygosity (H_e), the observed heterozygosity (uH_e), the percentage of polymorphic loci ($%P$), and allelic frequencies were calculated for the primary collection and the core collection GenAlEx v6.5 was also used to describe the AMOVA, the genetic

Fig. 2 Scores of sampled trees as projected on the first ordination axis of the biplot of a spatial analysis of principal components



differentiation among and within populations using 999 permutations, and principal coordinate analysis (PCoA). The mean *t*-test and variance *F*-test, respectively, were employed to detect whether the differences of the molecular traits and the AMOVA between the core subset and the primary collection were significant.

Genetic Structure of the Core Collection

A dendrogram of the genetic relationships among the core collection germplasms was constructed based on the unweighted pair group method with arithmetic mean (UPGMA). The genetic structure of the core collection was performed using STRUCTURE software version 2.1 (Pritchard et al. 2000) and STRUCTURE Harvester (Girardot et al. 2016), based on model testing for *k*-values ranging from 1 to 14 with 20 independent runs per *k* value. For each run, the burn-in was set to 50,000 followed by 100,000 MCMC (Markov Chain Monte Carlo) iterations. Then, the number of clusters ΔK was estimated (Evanno et al. 2005). The core collection will permit to have a representative of the whole genetic diversity of *T. articulata* in Morocco.

Results

Allelic Richness and Polymorphism Distributions of the Whole Collection

The software DIVA GIS was used to map the polymorphism and the allelic richness of the entire collection. In general, an elevated level of polymorphism exceeding 50% was

detected. The allelic richness is a straightforward measure of genetic diversity that is commonly used in studies based on molecular markers and aim at selecting populations for conservation (Petit et al. 1998). The high number of alleles present in Chefchaouen (242) and Al Hoceima shows the lowest allelic richness (172) (Fig. 1). In the whole collection, four populations retain both locally common and rare alleles, respectively: Essaouira, Fes-Boulemane, Chefchaouen, and Dar ben Karrich. Thus, rare alleles are essential in conservation because they represent unique evolutionary products and could provide the species with advantageous properties to cope with eventual environmental shifts (Peñas et al. 2016); they do not have a satisfactory representation in our study (0 to 2 alleles). The sPCA is a multivariate method devoted to the identification of spatial genetic patterns (Jombart et al. 2008). In our case, the sPCA clustered the collection in five groups: Oujda, Al Hoceima; Essaouira, Beni-Mellal, DB Karrich; Chefchaouen, Oulmes, Fes-Boulemane; and Marrakech, Oued Laou (Fig. 2).

Genetic Diversity of Core Collection

Selected Core Collection Individuals

Based on the results of the PowerCore analysis using a heuristic algorithm, 271 bands were obtained using 15 ISSR markers (Fig. 3), for 129 *T. articulata* germplasm and can be represented in a collection of 13 germplasms. The heuristic method is used as part of a global procedure that guarantees to find the optimum solution for a problem (Martí and Reinelt 2011). The 13 germplasms belong to eight provenances and are grouped on four

Table 1 List of core collection germplasms

Origin	Germplasm code	Latitude (north)	Longitude (west)	Number of loci
Essaouira	TS10	31.513430	−09.66729	113
Oujda	TO3	34.848250	−02.417910	56
Oujda	TO2	34.883440	−02.399370	57
Chefchaouen	TC4	35.181979	−05.103987	115
Chefchaouen	TC15	35.181971	−05.103979	109
Al Hoceima	TA17	35.095605	−04.074488	56
Al Hoceima	TA1	35.095597	−04.074480	64
Marrakech	TM6	31.354682	−08.235024	103
Marrakech	TM2	31.354678	−08.235020	83
Fes-Boulemane	TF2	33.443310	−04.421290	101
Fes-Boulemane	TF7	33.348550	−04.259050	13
DBK	TDBK3	30.479602	−10.603568	109
Oulmes	TOM6	33.332420	−05.532530	96

regions, with the maximum genetic diversity, constituting the most representative core collection of *T. articulata* in Morocco, noticing that germplasms from Beni-Mellal and Oued Laou are not represented in the core collection. The allelic richness varied from 13 (TF7) to 115 (TC4) (Table 1) and the core collection sampling proportion is 10.07% (Table 2).

The Genetic Diversity of the Core Collection and the Whole Collection

The estimates of genetic diversity (molecular traits) for the core collection and the whole collection are shown in Table 3. There were no significant differences in the mean of the effective number of alleles (N_e), the Shannon information index (I), and the expected heterozygosity (H_e). Moreover, the values of N_e , I , and H_e presented by the core collection are higher than those presented by the entire collection (1.398, 0.395, and 0.253 and 1.363, 0.382, and 0.238, respectively). The t -test ($P < 0.005$) of the observed number of alleles exhibited significant differences between the core collection and the entire collection. N_a is lower in the core collection than this presented by the primary collection (1.398 and 1.712, respectively). The

variance F -test ($P < 0.005$) showed no significant differences in the mean genetic diversity parameters (N_a , N_e , I , H_e), between the core subset and the primary collection.

Genetic Structure of the Core Collection

Complementary measures are utilized to describe the genetic structure of 13 selected germplasms of *T. articulata*. Based on Nei's genetic distances. The PCoA revealed three main clusters: TA1-TA17-TF7-TO2-TO3, TC15-TF2-TOM6-TM2-TM6, and TC4-TDBK3-TS10 (Fig. 4). STRUCTURE software version 2.1 (Pritchard et al. 2000) including the Bayesian method was also used to determine the structure and genetic relations among 13 germplasms of *T. articulata*, identified via STRUCTURE Harvester (Girardot et al. 2016). The K -value was used to estimate the number of clusters of the selected collection. The results indicated that ΔK reached a sharp peak when $K = 2$ (Fig. 5), suggesting the presence of two main groups in the core collection (Fig. 4). In addition, there was a small peak observed at $k = 9$ (Fig. 5), which indicates another informative population structure (Fig. 5). Hence, the core collection individuals were classified into two populations: TA1, TA17, TF7, TO2, and TO3 and TC5, TC15, TDBK3, TF2, TOM6, TM2, TM6, and TS10. To further determine the relationships among the core collection, UPGMA clustering was conducted

Fig. 3 ISSR amplification pattern of selected *T. articulata* germplasms using the primer BTH1 (A). ISSR amplification pattern using primer BTH11 for germplasms from the entire collection (B)

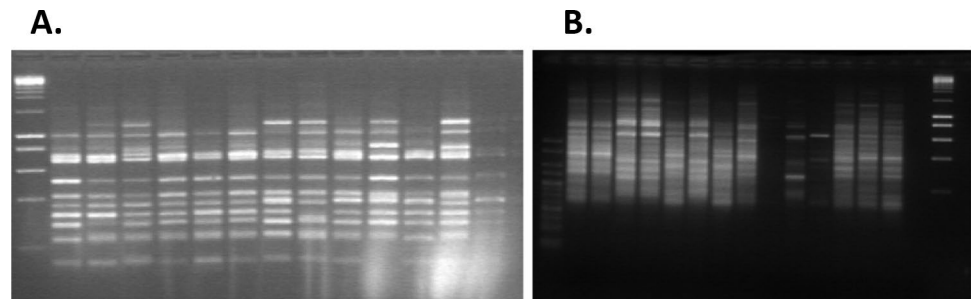


Table 2 Genetic diversity per locus in core collection and entire collection

Locus	Core collection				Entire collection				Ratio (core/entire)			
	Band Freq	<i>I</i>	<i>He</i>	<i>uHe</i>	Band Freq	<i>I</i>	<i>He</i>	<i>uHe</i>	Band Freq	<i>I</i>	<i>He</i>	<i>uHe</i>
BTH1	0.23	0.27	0.17	0.21	0.23	0.32	0.20	0.20	1.01	0.84	0.89	1.04
BTH2	0.29	0.27	0.18	0.21	0.25	0.33	0.20	0.21	1.15	0.82	0.86	1.00
BTH3	0.30	0.32	0.21	0.25	0.29	0.38	0.24	0.24	1.01	0.84	0.89	1.04
BTH4	0.29	0.31	0.20	0.24	0.17	0.26	0.15	0.16	1.73	1.17	1.33	1.53
BTH5	0.31	0.32	0.21	0.25	0.29	0.36	0.23	0.24	1.10	0.89	0.91	1.06
BTH6	0.24	0.26	0.17	0.21	0.31	0.39	0.25	0.25	0.80	0.67	0.70	0.83
BTH9	0.28	0.32	0.21	0.25	0.29	0.40	0.25	0.26	0.94	0.79	0.82	0.96
BTH10	0.36	0.32	0.22	0.25	0.31	0.40	0.26	0.26	1.16	0.80	0.84	0.96
BTH11	0.30	0.29	0.19	0.23	0.29	0.37	0.24	0.24	1.03	0.79	0.82	0.94
BTH12	0.35	0.28	0.18	0.21	0.41	0.48	0.32	0.33	0.85	0.57	0.57	0.66
BTH13	0.31	0.32	0.21	0.25	0.25	0.34	0.21	0.22	1.22	0.94	0.99	1.16
TH1	0.26	0.26	0.17	0.20	0.25	0.34	0.21	0.22	1.07	0.77	0.81	0.93
TH2	0.31	0.33	0.22	0.26	0.30	0.38	0.24	0.25	1.02	0.86	0.89	1.04
B1	0.32	0.29	0.19	0.23	0.30	0.40	0.25	0.26	1.10	0.72	0.77	0.88
B2	0.26	0.29	0.19	0.23	0.27	0.36	0.23	0.23	0.97	0.81	0.86	1.00
Average	0.29	0.30	0.20	0.23	0.28	0.37	0.23	0.24	1.08	0.82	0.86	1.00

I Shannon’s information index, *He* expected heterozygosity or Nei’s genetic diversity, *uHe* unbiased expected heterozygosity

using Nei’s unbiased genetic distance matrix. The dendrogram clearly shows two groups (A and B). Within group A, three individuals (TS10, TDBK3, TF1) clustered in subgroup (A1) and five individuals (TM1, TM2, TC1, TC2, TOM6) clustered in subgroup (A2). Within group B, four individuals (TF2, TO1, TO2, TA2) formed a subgroup while TA1 formed another subgroup (Fig. 6). Based on genetic distances, there was a similarity between PCoA and UPGMA in discriminating *T. articulata* germplasms.

Correlation Between the Core Collection and the Whole Collection

The number of amplification fragments produced by the 15 ISSR primers in the core collection and the full collection is shown in Table 2. The amount of polymorphism rate in the core collection and the whole collection shows

differences, 98.52% and 85.61%, respectively. However, no significant difference through mean *F*-test ($P < 0.005$) but significant difference is observed by the mean *T*-test ($P < 0.005$) (Table 3). The band frequencies per locus in the core collection are lower from those in the entire collection suggesting that all band types are presented in the core collection. The Shannon index (*I*), the observed heterozygosity (*He*), and the expected heterozygosity (*uHe*) are calculated respectively for both core and entire collections. Of 15 ISSR primers, the observed heterozygosity (*He*) and the expected heterozygosity (*uHe*) for BTH 6 primer were lower in the core collection, accounting 6% of all the ISSR primers used. The AMOVA carried out for the core subset and the entire collection is summarized in Table 5; the variation among population is 4%, 9% respectively; and the variation within population is 96% and 91% respectively, showing a difference in the variation among population (Table 4). By the mean of *T*-test ($P < 0.005$) and *F*-test ($P < 0.005$), there

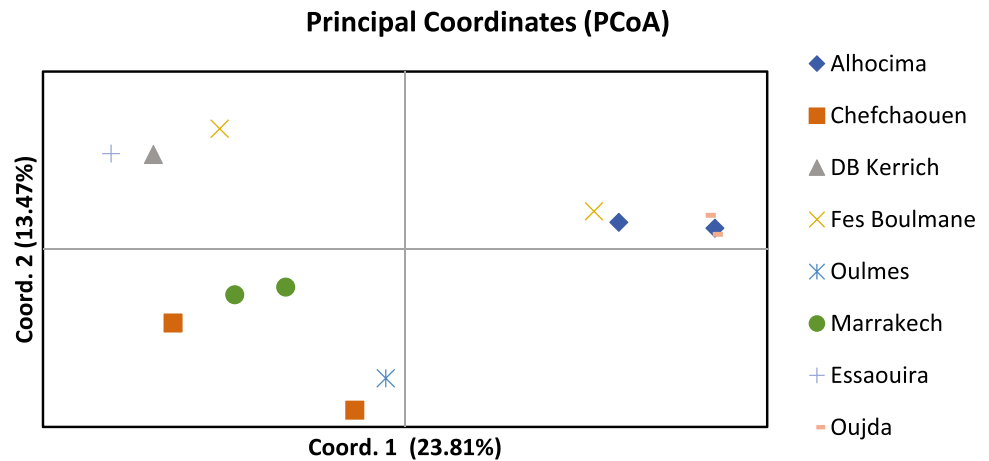
Table 3 Results of mean *t*-test and variance *F*-test between the primary collection and core subset of *T. articulata* based on molecular traits

Pop	<i>N</i>	<i>Na</i>	<i>Ne</i>	<i>I</i>	<i>He</i>	<i>uHe</i>	% <i>P</i>
Core collection	13	1.712	1.398	0.395	0.253	0.274	85.61%
Entire collection	129	1.970	1.363	0.382	0.238	0.240	98.52%
<i>T</i> -test	NS	*	NS	NS	NS	NS	NS
Variance	368.5789	3.9345	5.3264	6.1251	5.6830	6.1690	4.4931
<i>F</i> -test	*	NS	NS	NS	NS	NS	NS

N number of germplasm, *Na* observed number of alleles, *Ne* effective number of alleles, *I* Shannon’s information index, *uHe* unbiased expected heterozygosity, *He* expected heterozygosity, %*P* percentage of polymorphic loci

*significantly different at 0.005 probability level; NS not significant at 0.005 probability level

Fig. 4 Principal coordinate analysis of *Tetraclinis articulata* populations of ISSR based on pair-wise Nei's (1973) genetic distances



are no significant differences between the core subset and the primary collection (Table 5).

Discussion

Areas of Moderate and High Level of Allelic Richness

Measures of genetic diversity based on allelic richness are considered very important in the field of genetic

conservation, and marker-assisted methods for maximizing the number of alleles conserved have been shown to be effective (Schoen and Brown 1993). Besides, the allelic richness consists on estimating the number of alleles expected in samples of specified size, using the rarefaction method applied in ecology (Foulley and Ollivier 2006). In previous investigation, populations of *T. articulata* show low level of inbreeding ($G_{st}=0.13$) which is confirmed by the high level of heterozygosity (0.221); this one plays an important role in the response to environmental changes (Slatkin and

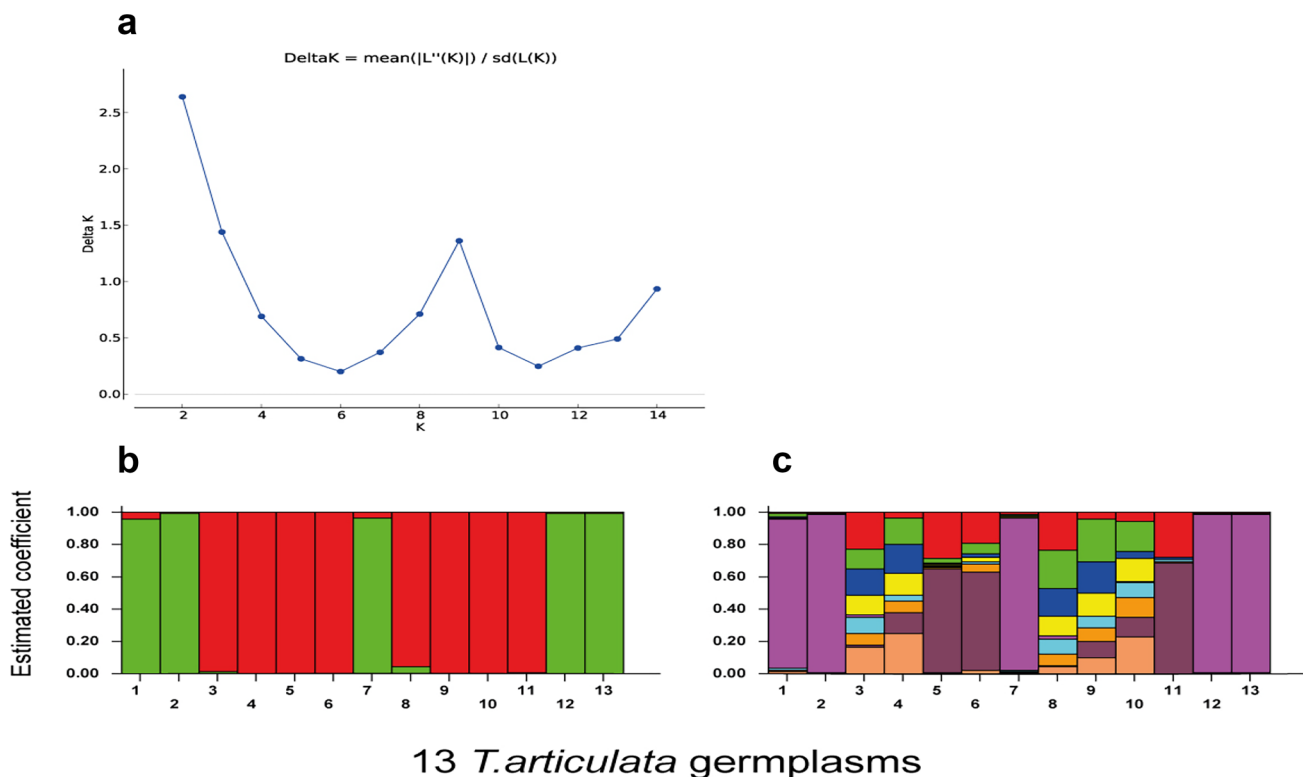
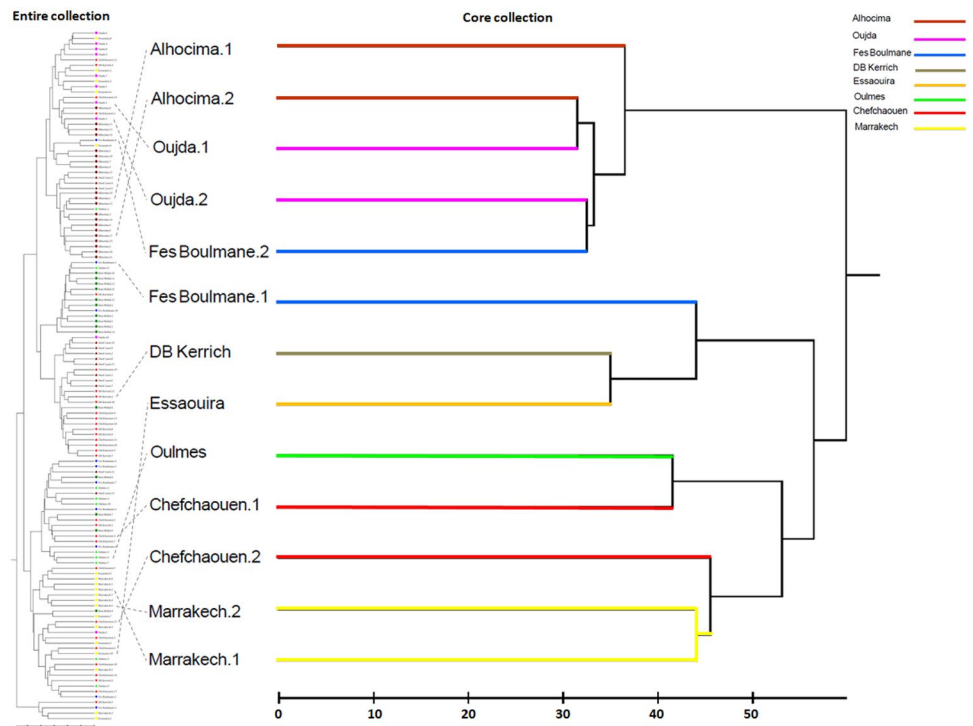


Fig. 5 Results of STRUCTURE analysis for 13 *T. articulata* germplasms based on ISSR markers. **a** Estimation of population using $\ln P(D)$ -derived delta K with cluster number (K) ranged from one to

14. **b** Genetic STRUCTURE of 13 germplasms when $k=2$. **c** Genetic STRUCTURE of 13 germplasms when $k=9$

Fig. 6 UPGMA dendrogram of core collection (13 germplasms) and entire collection (129 germplasms) of *T. articulata* constructed based on genetic similarity coefficients of ISSR data in the core collection and entire collection, respectively



Barton 1989) and indicate that the species did not endure drastic genetic bottlenecks (Makkaoui et al. 2020). Inbreeding may occur because of reduction and fragmentation of natural stands and cultivated areas, increasing the risk of allele loss, which eventually leads to genetic erosion (Lowe et al. 2005). On the other side, the Moroccan collection of *T. articulata* shows differences in allele distribution; populations of Essaouira, Fes-Boulemane, and Marrakech hold the highest number of both locally common and rare alleles. The priority for conservation should be given to populations that retain locally common alleles; these alleles occur in high frequency in a limited area and can indicate the presence of genotypes adapted to specific environments (Frankel et al. 1995).

Evaluation of the Core Collection of *T. articulata* Germplasm

One of the major goals of establishing a core collection is to collect a representative germplasm of the complete collection for in situ and ex situ conservations. Our study is aimed at selecting individuals representing the whole collection of *T. articulata*; this type of core collection called CC-I is especially suitable, for

situations requiring an overview of the genetic diversity of the accessions of the whole collection (Odong et al. 2013). Thus, the present core collection is created based on 15 ISSR markers and assisted by maximizing strategy (M) implemented in PowerCore software; it allows all alleles to be captured in a minimum number of accessions (Kim et al. 2007). This is the first such attempt in the Moroccan forestry. The analysis results in a core collection with 13 germplasms forming eight gene pools of *T. articulata* (Table 1). Oued Laou and Beni-Mellal regions were not included in the core collection because they do not contain a useful significant fraction useful for conservation strategy. The selected individuals have most of the genetic diversity and richness; also, they captured all the alleles (100%) existing in the entire collection (Table 2). The M strategy is the most appropriate approach for selecting entries with the most diverse alleles and eliminating redundancy (Franco et al. 2006). The percentage of germplasms selected by the algorithm is 10.07%, which is of reasonable size as previous studies proposed 5–20% core sizes (Thachuk et al. 2009; Hintum et al. 2000). However, there is no perfect proportion or fixed size for all core collections, as different species have different research needs because of the breadth and complexity of germplasm

Table 4 AMOVA of *T. articulata* core collection populations ($P < 0.001$)

Source	Degree of freedom	Sum of squares	Mean squares	Estimate variance	% total variation
Among Pops	3	177.141	59.047	2.211	4%
Within Pops	9	469.167	52.130	52.130	96%
Total	12	646.308		54.341	100%

Table 5 Comparison of AMOVA of *T. articulata* between the core collection and the entire collection populations ($P < 0.001$)

Source	Core collection					Entire collection				
	df	SS	MS	Est. Var	%	df	SS	MS	Est. Var	%
Among Pops	3	177,141	59,047	2,211	4%	3	499,720	166,573	4,445	9%
Within Pops	9	469,167	52,130	52,130	96%	125	5417,893	43,343	43,343	91%
Total	12	646,308		54,341	100%	128	5917,612		47,788	100%

resources (Wang and Li 2022). The small size of our core subset may be explained by the high redundancy and the presence of clonal origin (El Bakkali et al. 2013) or by the mode of reproduction (Noirot et al. 2003; Hintum et al. 2002). Various criteria were used to evaluate and validate the relevance of the established core collection. The evaluation of a core collection with the same data set that was used to create it ignores this simple but very important point (Odong et al. 2013). This is quite important, especially in the case of molecular marker data where the key assumption is that by maximizing diversity in a given set of marker loci, the diversity of genes of interest will also be maximized (Odong et al. 2013). By means of *T*-test and *F*-test showing no significant difference between the core subset and the entire collection, the selected collection could represent the diversity of the original germplasm. Hence, the number of effective alleles, Shannon index, and the observed heterozygosity of the selected collection were higher than those in the entire collection, an expected result, since genetic diversity increases with the elimination of genetically similar (or redundant) accessions (Agrama et al. 2009). The high variance within population revealed by the AMOVA in the core subset and the entire collection may explain the post-fire resprouting capacity of *T. articulata* (Hadjaj-Aoul et al. 2009; Terras et al. 2008), allowing the persistence of the ancestral gene pool and increasing the generation time of this species (Premoli and Steinke 2008). In other hand, based on the DNA amplification of ISSR marker system, the genetic distance between the selected germplasm of distinct geographical regions was revealed. The clustering methods UPGMA and PCoA are used to convert the large and complicated data set into the simplest graphical representations (Pandian et al. 2018). PCoA and UPGMA clustered the 13 germplasm according to their genetic distance and genetic similarity. The clustering of the selected germplasm did not only reflect their geographic locations or their climatic zones but also reflect the interaction of different processes (Sánchez-Gómez et al. 2013). Each cluster contains germplasm from locations with maritime and continental effects, i.e., Essaouira, Dar Ben Karrich and Fes-Boulemane; Oujda, Al Hoceima, and Fes-Boulemane. Moreover, the same group contains geographically distant individuals, i.e., Chefchaouen and Marrakech, proving the extent of genetic relationship between the genotype of the same species (Pandian et al. 2018). The present distribution of the selected individuals of *T. articulata* is probably occurred by human activities (trade and deliberate introductions for its wood) or naturally wind-mediated pollen

or seed dispersal (García-Castaño et al. 2021). The STRUCTURE program revealed another pattern of the selected collection, suggesting groups of geographically close individuals, i.e., Al Hoceima, Oujda, and Fes-Boulemane; Chefchaouen, Oulmès, and Marrakech. It was noted that the germplasm belonging to the north (Al Hoceima, Oujda) may have an Iberian origin (Sánchez-Gómez et al. 2013). Therefore, the three grouping methods used in this study (UPGMA, PCoA, and STRUCTURE) did not manage to significantly group the selected core subset in identical clusters. From another point of view, the weak spatial genetic pattern and the low number of fossil records in the southern shore of the Mediterranean are major impediments for phylogeographical reconstructions in *T. articulata* (Anderson et al. 2009). Different hypotheses could explain the complex structure of *T. articulata*; i.e., fossil evidence indicates that *Tetraclinis* Masters spread in the past from North America to Europe (Kvaček et al. 2000) or a genetic affinity between European and African populations (Kvaček et al. 2000).

Conclusion

To the best of our knowledge, the present study is the first to establish a core collection of *T. articulata* based on ISSR molecular markers and the M strategy implemented in PowerCore software. The synergy of these approaches revealed an important potential to locate germplasm with high genetic diversity that could serve accelerating breeding programs and exploring genetic diversity for conservation of *T. articulata* germplasm.

However, one type of data may not be sufficient to accurately assess the genetic variability of a germplasm and to develop effective core collections that capture the maximum genetic diversity of a germplasm (Yibo et al. 2010). Therefore, both molecular marker and morphological data should be considered when constructing a core collection (Wang et al. 2011). Moreover, core collection based on medicinal traits may also give insights about the representation of Arar tree diversity.

Acknowledgements First of all, this paper is dedicated to the memory of Mr. Hajib Said ex-director of C.R.F (Centre de la Recherche Forestière) Rabat-Morocco, who sadly passed away 02 November 2020. Secondly, the field prospection of this precious plant was made possible with the assistance of personal from CRF to whom we are grateful. Last, we gratefully knowledge Dr. Chaimaa Yatrib for her cooperation.

Author Contribution All authors contributed to the study conception and design. Material preparation, data collection, and analysis were performed by Makkaoui Meryem, Abbas Younes, Alami Mohammed, and Filali-Maltouf Abdelkarim. The first draft of the manuscript was written by Makkaoui Meryem and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Data Availability Data of this research are available upon reasonable request and are stored in hard disc.

Declarations

Consent to Participate Not applicable; this research does not contain any human participant.

Consent for Publication Not applicable. This research does not contain any human research data.

Conflict of Interest The authors declare no competing interests.

References

- Aggestam F et al (2020) Can nature conservation and wood production be reconciled in managed forests? A review of driving factors for integrated forest management in Europe. *J Environ Manage* 268:110670. <https://doi.org/10.1016/j.jenvman.2020.110670>
- Agrama H, Yan W, Lee F, Fjellstrom R, Chen MH, Jia M, McClung A (2009) Genetic assessment of a mini-core subset developed from the USDA rice genebank. *Crop Sci* 49:1336–1346
- Anderson CL, Channing A, Zamuner AB (2009) Life, death and fossilization on Gran Canaria—implications for Macaronesian biogeography and molecular dating vol 36. Wiley Online Library
- Balakrishnan R, Nair NV, Sreenivasan TV (2000) A method for establishing a core collection of *Saccharum officinarum* L. germplasm based on quantitative-morphological data. *Genet Resour Crop Evol* 47:1–9. <https://doi.org/10.1023/A:1008780526154>
- Balfourier F et al (2007) A worldwide bread wheat core collection arrayed in a 384-well plate. *Theor Appl Genet* 114:1265–1275. <https://doi.org/10.1007/s00122-007-0517-1>
- Beardmore T, Loo J, Mathews S, McGovern K, Nikiema A, Souvannavong O, ... Cavers S (2014) The state of the world's forest genetic resources. Rome: Food and Agriculture Organization. [Google Scholar]
- Belaj A et al (2012) Developing a core collection of olive (*Olea europaea* L.) based on molecular markers (DARs, SSRs, SNPs) and agronomic traits. *Tree Genet Genomes* 8:365–378. <https://doi.org/10.1007/s11295-011-0447-6>
- Bourkhiss M, Hnach M, Bourkhiss B, Ouhssine M, Chaouch A (2007) Composition chimique et propriétés antimicrobiennes de l'huile essentielle extraite des feuilles de *Tetraclinis articulata* (Vahl). *Masters du Maroc Revue Internationale des Sciences et Technologie* 3:232–242. <https://doi.org/10.4314/afsci.v3i2.61267>
- Brown A (1989) Core collections: A practical approach to genetic resource management. *Genome* 31:818–824. <https://doi.org/10.1139/g89-144>
- De Beukelaer H, Davenport GF, Fack V (2018) Core Hunter 3: flexible core subset selection. *BMC Bioinform* 19:203. <https://doi.org/10.1186/s12859-018-2209-z>
- Di Guardo M et al (2019) Genetic structure analysis and selection of a core collection for carob tree germplasm conservation and management. *Tree Genet Genomes* 15:41. <https://doi.org/10.1007/s11295-019-1345-6>
- Duan H et al (2017) Genetic Characterization of Chinese Fir from Six Provinces in Southern China and Construction of a Core Collection. *Scientific Reports* 7:13814. <https://doi.org/10.1038/s41598-017-13219-0>
- El Bouhtoury-Charrier F, Hakam A, Famiri A, Ziani M, Charrier B (2009) Wood characterization of *Tetraclinis articulata* and evaluation of its resistance against ligninolytic fungi. *IRG/WP*, 09-10697
- El Bakkali A, Haouane H, Moukhli A, Costes E, Van Damme P, Khadari B (2013) Construction of core collections suitable for association mapping to optimize use of Mediterranean olive (*Olea europaea* L.) genetic resources. *PLOS ONE* 8:e61265. <https://doi.org/10.1371/journal.pone.0061265>
- Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol Ecol* 14:2611–2620. <https://doi.org/10.1111/j.1365-294X.2005.02553.x>
- Fidah A, Rahouti M, Kabouchi B, Ziani M, El Bouhtoury-Charrier F, Famiri A (2015) Natural durability of *Tetraclinis articulata* (Vahl) Matsres woods against wood decay fungi: Laboratory test. *Wood Res* 60:865–872
- Foulley JL, Ollivier L (2006) Estimating allelic richness and its diversity. *Livest Sci* 101:150–158. <https://doi.org/10.1016/j.livprods.2005.10.021>
- Franco J, Crossa J, Warburton ML, Taba S (2006) Sampling strategies for conserving maize diversity when forming core subsets using genetic markers. *Crop Sci* 46:854–864
- Frankel OH (1984) Genetic perspectives of germplasm conservation. In: Arber W, Illemensee K, Peacock WJ, Starlinger P (eds) *Genetic manipulation: impact on man and society*. Cambridge University Press, Cambridge, pp 161–170
- Frankel OH, Brown AH, Burdon JJ (1995) The conservation of plant biodiversity. Cambridge University Press
- García-Castaño JL, Balao F, Lorenzo MT, Véla E, Hadjadj-Aoul S, Mifsud S, Terrab A (2021) A complex genetic structure of *Tetraclinis articulata* (Cupressaceae) in the western Mediterranean. *Bot J Linn* 197:420–438. <https://doi.org/10.1093/botlinnean/boab030>
- Girardot C, Scholtalbers J, Sauer S, Su SY, Furlong EEM (2016) Je, a versatile suite to handle multiplexed NGS libraries with unique molecular identifiers. *BMC Bioinform* 17:419. <https://doi.org/10.1186/s12859-016-1284-2>
- Hadjaj-Aoul S, Chouieb M, Loisel R (2009) Effect of environmental factors on natural regeneration of *Tetraclinis articulata* (Vahl) Masters in Oran region (Algeria) *Revue internationale d'écologie méditerranéenne* 35:19–31
- HCEFLCD (2004) Forêts en chiffres : Répartition géographique des formations forestières. Rabat
- Hijmans RJ, Guarino L, Cruz M, Rojas E (2001) Computer tools for spatial analysis of plant genetic resources data: 1. DIVA-GIS Plant Genetic Resources Newsletter 127:15–16
- Hintum, T. V, Brown, A., & Spillane, C. (2000). Core collections of plant genetic resources: Bioersivity International.
- Hintum, T. V., Hamilton, N. S., Engels, J., & Treuren, R. v. (2002). Accession management strategies: splitting and lumping. Paper presented at the Managing plant genetic diversity. Proceedings of an international conference, Kuala Lumpur, Malaysia, 12-16 June 2000.
- Jeong N et al (2019) Korean soybean core collection: Genotypic and phenotypic diversity population structure and genome-wide association study. *PLOS ONE* 14:e0224074. <https://doi.org/10.1371/journal.pone.0224074>
- Jombart T, Devillard S, Dufour AB, Pontier D (2008) Revealing cryptic spatial patterns in genetic variability by a new multivariate method. *Heredity* 101:92–103

- Kelblerová R, Dvořák J, Korecký J (2022) Genetic Diversity Maximization as a Strategy for Resilient Forest Ecosystems: a Case Study on Norway Spruce. *Forests* 13:489
- Kim KW et al (2007) PowerCore: a program applying the advanced M strategy with a heuristic search for establishing core sets. *Bioinform* 23:2155–2162. <https://doi.org/10.1093/bioinformatics/btm313>
- Kvaček Z, Manchester SR, Schorn HE (2000) Cones, seeds, and foliage of *Tetraclinis salicornioides* (Cupressaceae) from the Oligocene and Miocene of western North America: a geographic extension of the European Tertiary species. *Int J Plant Sci* 161:331–344
- Laakili A et al (2018) Diversity and spatial genetic structure of natural Moroccan *Quercus susber* L. assessed by ISSR markers for conservation. *Physiol Mol Biol Plants* 24:643–654. <https://doi.org/10.1007/s12298-018-0538-z>
- Li C, Pan D, Mao X, Tu C, Zhou H, Fan Z, Li X (2006) The genetic diversity of Gaozhou wild rice analyzed by SSR. *Chin Sci Bull* 51:562–572. <https://doi.org/10.1007/s11434-006-0562-1>
- Liu FM, Zhang NN, Liu XJ, Yang ZJ, Jia HY, Xu DP (2019) Genetic diversity and population structure analysis of *Dalbergia odorifera* germplasm and development of a core collection using microsatellite markers. *Genes* 10. <https://doi.org/10.3390/genes10040281>
- Lowe AJ, Boshier D, Ward M, Bacles CFE, Navarro C (2005) Genetic resource impacts of habitat loss and degradation; reconciling empirical evidence and predicted theory for neotropical trees. *Heredity* 95:255–273. <https://doi.org/10.1038/sj.hdy.6800725>
- Lv J et al (2020) Genetic diversity analysis of a breeding population of *Eucalyptus cloeziana* F. Muell. (Myrtaceae) and extraction of a core germplasm collection using microsatellite markers. *Ind Crops Prod* 145:112157. <https://doi.org/10.1016/j.indcrop.2020.112157>
- Machkour-M'Rabet S, Henaut Y, Dor A, Pérez-Lachaud G, Pélissier C, Gers C, Legal L (2009) ISSR (Inter Simple Sequence Repeats) as molecular markers to study genetic diversity in Tarantulas (Mygalomorphae). *J Arachnol* 37:10–14. <https://doi.org/10.1636/A08-27.1>
- Mahmoodi R, Dadpour MR, Hassani D, Zeinalabedini M, Vendramin E, Micali S, Nahandi FZ (2019) Development of a core collection in Iranian walnut (*Juglans regia* L.) germplasm using the phenotypic diversity. *Sci Hortic* 249:439–448. <https://doi.org/10.1016/j.scienta.2019.02.017>
- Makkaoui M, Abbas Y, El Antry-Tazi S, Medraoui L, Alami M, Rabani S, Filali-Maltouf A (2020) Genetic diversity of ten Moroccan populations of *Tetraclinis articulata* as revealed by Inter Simple Sequence Repeat (ISSR) markers. *Bois et Forêts des Tropiques* 345:13–23. <https://doi.org/10.19182/bft2020.345.a31927>
- Marita JM, Rodriguez JM, Nienhuis J (2000) Development of an algorithm identifying maximally diverse core collections. *Genet Resour Crop Evol* 47:515–526. <https://doi.org/10.1023/A:1008784610962>
- Martí R, Reinelt G (2011) Heuristic methods. In: *The linear ordering problem: exact and heuristic methods in combinatorial optimization*. Springer Berlin Heidelberg, Berlin, Heidelberg, pp 17–40. https://doi.org/10.1007/978-3-642-16729-4_2
- Miyamoto N, Ono M, Watanabe A (2015) Construction of a core collection and evaluation of genetic resources for *Cryptomeria japonica* (Japanese cedar). *J For Res* 20. <https://doi.org/10.1007/s10310-014-0460-3>
- Mouhaddab J et al (2016) Assessing genetic diversity and constructing a core collection of an endangered Moroccan endemic tree [*Argania spinosa* (L.) Skeels]. *Moroccan J Biol* 13:1–12
- Ndjiondjop MN et al (2017) Genetic variation and population structure of *Oryza glaberrima* and development of a mini-core collection using DArTseq. *Front Plant Sci* 8. <https://doi.org/10.3389/fpls.2017.01748>
- Nei M (1973) Analysis of gene diversity in subdivided populations. *Proceedings of the National Academy of Sciences* 70(12):3321–3323. <https://doi.org/10.1073/pnas.70.12.3321>
- Ng WL, Tan SG (2015) Inter-simple sequence repeat (ISSR) markers: are we doing it right? *ASM Sci J* 9:30–39
- Nirsatmanto A (2020) Impact of tree breeding on managing forest genetic resources: a review with a particular reference to Acacias and Eucalyptus grown in Indonesia. *IOP Conf Ser: Earth and Environ Sci* 522:012002. <https://doi.org/10.1088/1755-1315/522/1/012002>
- Noirot M, Anthony F, Dussert S (2003) A method for building core collections. In: Hamon P, Perrier X, Glaszmann JC (eds.) *Genetic diversity of cultivated tropical plants*. CIRAD, Montpellier (FRA), pp 67–75
- Odong TL, Jansen J, van Eeuwijk FA, van Hintum TJJ (2013) Quality of core collections for effective utilisation of genetic resources review, discussion and interpretation. *Theor Appl Genet* 126:289–305. <https://doi.org/10.1007/s00122-012-1971-y>
- Panahi B, Ghorbanzadeh Neghab M (2013) Genetic characterization of Iranian safflower (*Carthamus tinctorius*) using inter simple sequence repeats (ISSR) markers. *Physiol Mol Biol Plants* : an Int J Func Plant Biol 19:239–243. <https://doi.org/10.1007/s12298-012-0155-1>
- Pandian S, Marichelvam K, Satish L, Ceasar SA, Pandian SK, Ramesh M (2018) SPAR markers-assisted assessment of genetic diversity and population structure in finger millet (*Eleusine coracana* (L.) Gaertn) mini-core collection. *J Crop Sci Biotechnol* 21:469–481. <https://doi.org/10.1007/s12892-018-0034-0>
- Pascual L, Fernández M, Aparicio N, López-Fernández M, Fité R, Giraldo P, Ruiz M (2020) Development of a Multipurpose Core Collection of Bread Wheat Based on High-Throughput Genotyping Data *Agronomy* 10:534
- Peñas J, Barrios S, Bobo-Pinilla J, Lorite J, Martínez-Ortega M (2016) Designing conservation strategies to preserve the genetic diversity of *Astragalus edulis* Bunge, an endangered species from western Mediterranean region. *PeerJ* 4. <https://doi.org/10.7717/peerj.1474>
- Petit RJ, El Mousadik A, Pons O (1998) Identifying populations for conservation on the basis of genetic markers. *Conserv Biol* 12:844–855. <https://doi.org/10.1111/j.1523-1739.1998.96489.x>
- Premoli AC, Steinke L (2008) Genetics of sprouting: effects of long-term persistence in fire-prone ecosystems. *Mol Ecol* 17:3827–3835
- Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics* 155:945–959
- Rajora OP, Mosseler A (2001) Challenges and opportunities for conservation of forest genetic resources. *Euphytica* 118:197–212
- Rourke JP (1991) *Tetraclinis articulata*, a hitherto unrecorded naturalised alien conifer in South Africa *Bothalia* 21:62–64
- Sánchez-Gómez P, Jiménez J, Vera J, Sánchez-Saorín F, Martínez J, Buhagiar J (2013) Genetic structure of *Tetraclinis articulata*, an endangered conifer of the western Mediterranean basin. *Silva Fenn* 47:1–14. <https://doi.org/10.14214/sf.1073>
- Schoen DJ, Brown AH (1993) Conservation of allelic richness in wild crop relatives is aided by assessment of genetic markers. *Proc Natl Acad Sci USA* 90:10623–10627. <https://doi.org/10.1073/pnas.90.22.10623>
- Slatkin M, Barton NH (1989) A comparison of three indirect methods for estimating the average levels of gene flow. *Evolution* 43:1349–1368. <https://doi.org/10.1111/j.1558-5646.1989.tb02587.x>
- Soleimani B et al (2020) Comparison between core set selection methods using different Illumina marker platforms: a case study of assessment of diversity in wheat. *Front Plant Sci* 11. <https://doi.org/10.3389/fpls.2020.01040>
- Terras M, Labani A, Benabdeli K, Adda-Hanifi N (2008) Dynamique phytoécologique du Thuya de Berbérie face à l'incendie forêt méditerranéenne XXIX:33–40
- Thachuk C, Crossa J, Franco J, Dreisigacker S, Warburton M, Davenport GF (2009) Core hunter: an algorithm for sampling genetic resources based on multiple genetic measures. *BMC Bioinform* 10:1–13. <https://doi.org/10.1186/1471-2105-10-243>
- Upadhyaya HD, Pundir RPS, Gowda CLL, Gopal Reddy V, Singh S (2008) Establishing a core collection of foxtail millet to enhance the utilization of germplasm of an underutilized crop

- Plant Genetic. Resources 7:177–184. <https://doi.org/10.1017/S1479262108178042>
- Wang X, Cao Z, Gao C, Li K (2021) Strategy for the construction of a core collection for *Pinus yunnanensis* Franch. to optimize timber based on combined phenotype and molecular marker data. *Genet Resour Crop Evol* 68:3219–3240. <https://doi.org/10.1007/s10722-021-01182-9>
- Wang Y, Li Y (2022) Population Genetics and Development of a Core Collection from Elite Germplasm of *Xanthoceras Sorbifolium* Based on Genome-Wide SNPs. *Forests* 13:338. <https://doi.org/10.3390/f13020338>
- Wang Y, Zhang J, Sun H, Ning N, Yang L (2011) Construction and evaluation of a primary core collection of apricot germplasm in China. *Sci Hortic* 128:311–319
- Yibo D, Xinwu P, Qianhua Y, Hongjin W, Xujing W, Shirong J, Yufa P (2010) Ecological, morphological and genetic diversity in *Oryza rufipogon* Griff. (Poaceae) from Hainan Island. *China Genet Resour Crop Evol* 57:915–926
- Zewdie Y, Tong N, Bosland P (2004) Establishing a core collection of *Capsicum* using a cluster analysis with enlightened selection of accessions. *Genet Resour Crop Evol* 51:147–151. <https://doi.org/10.1023/B:GRES.0000020858.96226.38>
- Zhang H et al (2011) A core collection and mini core collection of *Oryza sativa* L. *China Theor Appl Genet* 122:49–61. <https://doi.org/10.1007/s00122-010-1421-7>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.