Phylogeography and population genetics of *Acanthophyllum squarrosum* complex (Caryophyllaceae) in the Irano-Turanian region

Masoumeh Mahmoudi Shamsabad, Mostafa Assadi & Laura Parducci

To cite this article: Masoumeh Mahmoudi Shamsabad, Mostafa Assadi & Laura Parducci (2019) Phylogeography and population genetics of *Acanthophyllum squarrosum* complex (Caryophyllaceae) in the Irano-Turanian region, Systematics and Biodiversity, 17:4, 412-421, DOI: 10.1080/14772000.2019.1590476

To link to this article: https://doi.org/10.1080/14772000.2019.1590476

Published online: 20 Jun 2019.

Article views: 21
Research Article

Phylogeography and population genetics of *Acanthophyllum squarrosum* complex (Caryophyllaceae) in the Irano-Turanian region

MASOUMEH MAHMOUDI SHAMSABAD¹, MOSTAFA ASSADI² & LAURA PARDUCCI³

¹Department of Plant Biology, Tarbiat Modares University, Tehran, 14115-154, Iran
²Research Institute of Forests and Rangelands, Agricultural Research Education and Extension Organization (AREEO), Tehran, P. O. Box 13185-116, Iran
³Department of Ecology and Genetics, Evolutionary Biology Centre, Uppsala University, Sweden

(Received 30 August 2018; accepted 13 February 2019)

*Acanthophyllum squarrosum* and two closely related species, *A. heratense* and *A. laxiusculum* (Caryophyllaceae), form a complex that covers parts of subalpine steppes of the Irano-Turanian (IT) region. In this study, we explored the genetic structure and phylogeography of this complex based on partial sequences of two chloroplasts (*psbA–trnH* and *rpl32–trnL (UAG)) and two nuclear (EST24 and nrITS) DNA regions. We analysed 80 individuals from eight populations and detected 12 chloroplast haplotypes, 16 and eight nuclear alleles in EST24 and nrITS sequences, respectively. Phylogenetic trees and haplotype networks did not show distinct genetic groups in the complex and this could be explained by incomplete lineage sorting or introgression between species. Divergence time analysis revealed a Quaternary origin for *A. squarrosum* complex at approximately 1.8 million years ago (Mya) and the neutrality test results indicated that this complex experienced a recent population expansion. AMOVA analysis of the chloroplast regions showed a significant genetic differentiation among populations and low genetic differentiation within populations, but opposite results were found with nuclear markers, implying introgression between *A. squarrosum* complex populations.

**Key words:** Acanthophyllum, Irano-Turanian region, phylogeny, phylogeography, population genetics, species complex

Introduction

The genus *Acanthophyllum* C. A. Mey (Caryophyllaceae) is an important element of subalpine steppe flora in central to south-west Asia and is represented by over 60 species, of which 23 are endemic to Iran (Schiman-Czeika, 1988). The genus includes small, woody base herbaceous perennials with spiny leaves and cushion forms that grow on sandy, stony, or rocky slopes and hills of open habitats (Ghaffari, 2004; Mahmoudi Shamsabad, Vaezi, Memariani, & Joharchi, 2012; Pirani et al., 2014; Schiman-Czeika, 1988). The Irano-Turanian (IT) region is considered the centre of diversification of *Acanthophyllum* species.

The IT region represents one of the evolutionary and biological diversity hotspots in the Old World, and is considered as the centre of origin of xerophytic taxa for neighbouring regions (Manafzadeh, Salvo, & Conti, 2014; Noroozi et al., 2018). The region has a complex geographic history and Quaternary climatic oscillations played an important role in shaping current floristic compositions (Djamali et al., 2012; Manafzadeh, Staedler, & Conti, 2017). In this geographic context, phylogeography and population genetic analyses can provide an understanding of how abiotic factors, such as climatic oscillations and geological activities, have influenced species distributions and population demography (Djamali et al., 2012; Manafzadeh, Staedler, & Conti, 2017). However, considering rich and high endemcity of this phytogeographic region, very few population-level studies using novel molecular methods have been conducted.

The *A. squarrosum* complex includes three closely related species: *A. squarrosum* Boiss., *A. heratense* Schiman-Czeika, and *A. laxiusculum* Schiman-Czeika,
which have overlapping distribution ranges over a wide geographic range of south-west Asia including Iran, Afghanistan, Pakistan, Azerbaijan, and Armenia (Fig. 1) (Schiman-Czeika, 1988). These species belong to section Oligosperma Schischk. (Schiman-Czeika, 1988) and relationships among them are not well resolved using chloroplast and nuclear markers but all three species nested in ‘Clade I’ of Pirani et al. (2014). The A. laxiusculum differs from A. squarrosum and A. heratense in having pedunculate flowers, and trichome density is used as the distinguishing character between A. squarrosum and A. heratense (Schiman-Czeika, 1988). However, trichome density and peduncle length are highly variable traits in Acanthophyllum populations (Basiri-Esfahani, Bidi, Assadi, & Rahimi-Nejad, 2011) making the taxa very difficult to distinguish based on morphological characters. The taxonomic status of these three species has been revised several times based on morphological characters. For example, a dwarf form of A. squarrosum (f. nana Aitch) has been upgraded to the species level by Schiman-Czeika (1988) and named A. heratense, while Basiri-Esfahani et al. (2011), based on morphological characters considered A. laxiusculum to be indistinct from A. heratense. Finally, in a recent multivariate analysis study based on morphological characters, A. squarrosum and A. heratense were considered part of the same species complex (Mahmoudi Shamsabad, Ejtehadi, Vaezi, & Joharchi, 2013).

The nuclear genomes are diploid or higher ploidy and have features such as recombination and heterozygosity, biparental inheritance by seed and pollen, and frequent intra-specific genetic variation in most single-copy or low-copy genes. In contrast, the chloroplast genome is haploid and non-recombining that is typically transmitted maternally only through seeds in angiosperms and within most species, genetic variations in chloroplast DNA (cpDNA) are typically low at the intra-specific level (Schaal, Hayworth, Olsen, Rauscher, & Smith, 1998). Thus, the analyses of gene loci from both the chloroplast and the nuclear genome would provide a broad view of the genetic structure of plant populations and better reveal their

**Fig. 1.** Haplotype distributions and networks of the eight Acanthophyllum populations analyzed in this study using cpDNA (1) nuclear ITS (2) and nuclear EST24 (3) markers. Pie charts show haplotypes frequencies. Yazd (A), Bandar-Abbas (B), Kerman (C), Semnan (D), Khorasan (E), Azerbaijan (I), Esfahan (J), Tehran (K).
Materials and methods
Sampling and PCR amplifications
In total, 80 specimens from eight locations were collected from the IT region, covering the majority of the *A. squarrosum* complex distribution range. At least one specimen of every population was deposited in the Herbarium of the Research Institute of Forests and Rangelands, Tehran (TARI) as a voucher. Sampling information is summarized in Table S1 (see online supplemental material, which is available from the article’s Taylor & Francis Online page at http://dx.doi.org/10.1080/14772000.2019.1590476).

Total genomic DNA was extracted from dried leaf tissues using a QIAGEN Plant Mini Kit according to the manufacturer’s protocol (QIAGEN GmbH, Hilden, Nordrhein-Westfalen, Germany). After screening several molecular regions, we selected two cpDNA regions (*psbA*-trnH and *rpl32*-trnL (UAG)); Shaw, Lickey, Schilling, & Small, 2007), one single-copy nuclear region (EST24; Toprak et al. 2016), and the nrITS region amplified by the primer pairs P17/26S-82R (Popp & Oxelman, 2001). Due to the poor amplification of the EST24 locus in some population we designed a new forward primer for this region (5’-TGGAGCCATTGTTTGTG-3’).

All amplification reactions were carried out in a volume of 20 µL containing 10–20 ng template DNA, 4.0 µL of HF Buffer 5×, 1.5 µL of MgCl2 (25 mM), 1 µL of dNTPs (2.5 mM each), 1 µL of DMSO, 0.4 µL primer (10 mM), 0.2 µL of Phusion High-Fidelity DNA Polymerase, and double-distilled water. Amplifications were run with an initial cycle of 30 s at 98°C, followed by 34 cycles of 10 s at 98°C, 30 s at 65°C, 30 s at 72°C, and a final 5 min step at 72°C as extension. PCR products were purified using a Qiaquick PCR purification kit (Qiagen, Germany), and sent for sequencing on an ABI PRISM 3730 XL automated sequencer using Macrogen service (Amsterdam, the Netherlands).

Sequence analysis
The nucleotide sequences were aligned using Muscle (Edgar, 2004) followed by manual adjustment and visual inspection in MEGA version 7.0 (Kumar, Stecher, & Tamura, 2016). We combined the two cpDNA regions to gain enough variable sites in the subsequent analysis. The best substitution models were selected using jModeltest version 2.1.7 (Darriba, Taboada, Doallo, & Posada, 2012). For the cpDNA matrix GTR + G model, for nrITS matrix GTR + I + G model and for EST24 matrix HKY + G model were selected. Phylogenetic relationships among species were inferred using Bayesian inference (BI) method employing *A. diezianum* Hand.-Mzt. and *A. korshinskyi* Schischk. as outgroups. BI was performed in MrBayes version 3.1.2 (Ronquist & Huelsenbeck, 2003). Each Markov chain was started from a random seed number and ran for 10 million MCMC generations in four parallel runs with every 1000th cycle sampled from the chain.

The BEAST version 2.4.5 (Bouckaert et al., 2014) was used to estimate the divergence time among species. The data matrix was composed of two unlinked partitions: nrITS sequences and *rpl32-trnL* (UAG) sequences. A Bayesian Markov chain Monte Carlo approach with an uncorrelated log-normal relaxed molecular clock was run for 10⁷ generations using the GTR + G as a substitution model for both partitions. The prior probability of the clock rate was set to a truncated normal distribution with a mean of 2.15 × 10⁻⁹, ranging from 0.38 × 10⁻⁹ to 7.83 × 10⁻⁹ mutations per site per year for nrITS (Kay, Whittall, & Hodges, 2006) and 1.01 × 10⁻⁹ mutations per site per year for cpDNA matrix (Graur & Li, 2000). Tracer v1.6 (Rambaut, Suchard, Xie, & Drummond, 2014) was used to check for sufficient sampling and convergence to the stationary distribution.

We used DnaSP version 5.0 (Librado & Rozas, 2009) to identify haplotypes from aligned DNA sequences, to measure genetic diversity within and among populations, to estimate nucleotide diversity (*Pi*), haplotype diversity (*Hz*), and the number of nucleotide differences among haplotypes. We used a U statistic test by Permut v1.2.1 (http://www.pierroton.inra.fr/genetics/labo/Software/Permut). A higher value of
Phylogeography and population genetics of Acanthophyllum squarrosum Complex

Table 1. Genetic diversity and differentiation parameters for cpDNA and nDNA markers.

| Markers | Pi   | H_T  | H_d  | N_ST | G_ST | H_S | FST |
|---------|------|------|------|------|------|-----|-----|
| Cp DNA  | 0.0032 | 0.167 | 0.887 | 0.23 | 0.18 | 0.738 | 0.716 |
| EST24   | 0.0165 | 0.142 | 0.966 | 0.28 | 0.167 | 0.829 | 0.28 |
| ITS     | 0.0024 | 0.177 | 0.687 | 0.219 | 0.164 | 0.663 | 0.258 |

NST than GST indicates the presence of a genetic differentiation among populations (Pons & Petit, 1996). The DnaSP was also used to study the demography of the populations. The Tajima’s D (Tajima, 1989) and Fu and Li’s F (Fu, 1997) values were calculated to detect departures from population equilibrium, and the pairwise mismatch distribution was used to test for population expansion. The shape of the frequency of pairwise differences is expected to be multimodal in samples drawn from populations at demographic equilibrium, while it is normally unimodal in populations having experienced a recent demographic or range expansion (Excoffier, 2004; Ray, Currat, & Excoffier, 2003; Slatkin & Hudson, 1991). We also used Arlequin version 3.5.2 (Excoffier & Lischer, 2010) to compute the sum of squared deviations (SSD) between the observed and expected mismatch distributions; a significant SSD indicates a departure from a model of population expansion. Analysis of molecular variance (AMOVA) and pairwise comparisons of Fst were run in Arlequin software. Finally, phylogeographic relationships among haplotypes were inferred from a maximum parsimony median-joining network calculated in Network version 4.2.0.1 software (Forster, Forster, & Watson, 2007).

Results

Genetic variation and haplotype distribution
The length of psbA–trnH region ranged from 257 to 265 base pairs (bp) due to the presence of a 5 bp indel, while the rpl32–trnL (UAG) length ranged from 678 to 685 bp due to one 7 bp indel. The combined plastid DNA matrix showed a consensus length of 951 bp across the 80 individuals. Twelve chloroplast haplotypes were identified (Table S1, see supplemental material online); of those seven (hap 2, hap 4, hap 7, hap 8, hap 10, hap 11, and hap 12) were population-specific (Fig. 1.1). Haplotype 3 was the most common one shared by seven populations (A, B, C, D, E, K, and J), whereas haplotype 5 was found in five populations (A, B, E, J, and I). Total haplotype (Hd) and nucleotide diversities (Pi) were 0.887 and 0.0032, respectively (Table 1). The highest values of haplotype diversity and nucleotide diversities in chloroplast sequences (Hd = 0.86 and Pi = 0.0028) were found in population E in the north-east of Iran, a region that is considered the centre of diversification of Acanthophyllum.

The nrITS matrix aligned across 80 individuals and had a total length of 448 bp. Eight haplotypes were derived from nrITS sequences; of those two haplotypes (hap 3 and hap 8) were population-specific (Fig. 1.2). The most common haplotype was haplotype 1, which was shared by three populations (E, D, and K). Total haplotype and nucleotide diversities were 0.687 and 0.0024, respectively (Table 1). The highest values of haplotype diversity and nucleotide diversities were found in population E (Hd = 0.74854 and Pi = 0.00335).

The EST24 marker was developed based on EST libraries of Silene uralensis (Rupr.) Bocquet (Toprak et al., 2016), and was very variable (Hd = 0.966 and Pi = 0.0165). The aligned matrix surveyed across 45 individuals (35 individuals gave poor amplifications and were therefore excluded) had a total length of 333 bp. Sixteen haplotypes were identified for the EST24 region and of these 10 (hap 1, hap 4, hap 5, hap 6, hap 7, hap 9, hap 10, hap 12, hap 13, and hap 14) were population-specific. The most common haplotype was haplotype 3, which was shared by seven populations (A, B, C, E, K, J, and I) and located at the centre of the network (Fig. 1.3).

Phylogeny and divergence time
The cpDNA and nDNA phylogenetic trees bear large polytomies and do not show any geographic or morphological structuring (Fig. 2). CpDNA and nDNA phylogenetic analyses detected several subclades that however did not correspond to the morphological classification but partially corresponded to geographic distance. However, no clear clade structure was revealed by cpDNA and nDNA data. Divergence time analysis revealed a Quaternary origin for A. squarrosum complex at ~1.8 Mya (95% HPD: 1.019–2.352 Mya) (Fig. 3).

Population structure
Analysis of Molecular Variance (AMOVA) using cpDNA markers showed that 72.68% of the genetic variation was among populations and 27.3% within populations. In contrast, nuclear EST24 data indicated 20.61% of the genetic variation was among populations and 79.39% within populations. Similarly, nrITS data showed that variation among populations was lower than variation within populations (25.88% and 74.12%, respectively). Moreover, data from genetic diversity and differentiation analyses revealed a genetic differentiation across all populations using all three markers (NST > GST; P < 0.05) (Table 1).
Neutrality test

Estimates of the neutral test (Tajima’s D and Fu’s Fs) and the mismatch distribution parameter (SSD) yielded significant negative values rejecting a scenario of population equilibrium in the Acanthophyllum populations and suggesting instead a recent population expansion (Table 2). A plot of the frequency of
all pairwise differences (mismatch distribution) for both cpDNA and nDNA data is presented in Fig. 4. The results showed that the mismatch distributions fitted a unimodal curve in both datasets, also indicating a recent expansion in the *Acanthophyllum* populations.
Discussion

The flora of IT Region is characterized by a high number of endemic genera and species (Takhtajan, 1986). Within this region, the richest flora is in the Iranian Plateau, with remarkable examples of the endemicity such as *Acantholimon* Boiss. with 120 species.

| Markers | Fu’s and Li’s F | Fu’s Fs | Tajima’s D | SSD/P value |
|---------|----------------|---------|------------|-------------|
| Cp DNA  | -3.5           | -13.3   | -1.69      | 0.0129/0.03 |
| EST24   | -0.147         | -31.8   | -2.03      | 0.0051/0.01 |
| ITS     | -0.869         | -13.7   | -1.09      | 0.0091/0.03 |

Table 2. Parameters of neutrality tests and demographic analysis.

![Fig. 4. Distribution of the number of pairwise nucleotide differences in *A. squarrosum* complex for combined cpDNA (1), nuclear EST24 (2) and nuclear ITS (3) haplotypes.](image)
(Plumbaginaceae), *Acanthophyllum* 60, and *Dionysia* Fenzl ~30 (Primulaceae), most having cushion forms (Takhtajan, 1986). Historical geography, topography with high mountains (for example Alborz and Zagros mountains), and climate promoted biodiversity in this region.

Species delimitation in closely related plant taxa can be challenging due to rapid and recent phylogenetic divergence, high phenotypic plasticity, introgression and partial barriers to gene flow between taxa (Shaffer & Thomson, 2007). In phylogenetic trees based on haplotypes, shared haplotypes might indicate hybridization or incomplete lineage sorting (Chiang et al., 2009). In this study, phylogenetic trees and haplotype networks provided no evidence for distinct genetic groups in the *A. squarrosum* complex, possibly because of incomplete lineage sorting and/or introgression following secondary contacts after speciation. Introgression and hybridization in Caryophyllaceae have been described previously in a number of studies (Fior, Karis, Casazza, Minuto, & Sala, 2006; Petri, Pfeil, & Oxelman, 2013; Pfeil, Toprak, & Oxelman, 2017; Pirani et al., 2014), and had important roles in diversification of *Acanthophyllum* species (Ghaffari, 2004).

In the *A. squarrosum* complex, we found higher genetic differentiation among populations relative to the intrapopulation variation in cpDNA, while nDNA data showed higher intrapopulation differentiation relative to interpopulation. This discordance might be explained by uniparental inheritance and transmission of cpDNA by only seed compared with biparental inheritance and transmission by seed and pollen in nDNA, which offer opportunities for nDNA to obtain more genetic components among populations by gene flow. In addition, recombination in the nuclear genome may have played an important role in decreasing interpopulation differentiation through nDNA data in this complex.

Divergence times indicated by BEAST analyses for the *A. squarrosum* complex were found to be quite recent in the Quaternary (1.8 Mya), suggest an important role of climate for the diversification event of this complex and that incomplete lineage sorting is a potential cause for shared haplotypes among populations. Moreover, the neutrality test and result of mismatch distribution analysis all suggest that *A. squarrosum* complex experienced a recent population expansion which is also reported in *Taxus wallichiana* (Liu et al., 2013), *Cycas revoluta*, and *C. taitungensise* (Chiang et al., 2009). It is possible therefore that the presence of a complex topography in the IT region in combination with climatic oscillations and geological activities of the Quaternary period created a wide range of environmental niches facilitating formation and persistence of isolated *Acanthophyllum* populations.

**Conclusion**

The analyses of chloroplast and nuclear loci did not detect distinct genetic groups in *A. squarrosum* complex and suggest that incomplete lineage sorting and/or introgression following secondary contacts after Quaternary speciation can explain most of the shared genetic variation observed between these three *Acanthophyllum* species in the IT region.

**Acknowledgements**

We thank Mohammad Vatanparast and Iraj Mehregan for helpful comments on the earlier versions of the manuscript. Financial support to MM was provided by the Ministry of Science Research and Technology of Iran, the International Association for Plant Taxonomy (IAPT), the Swedish Phytogeographical Society, and the Sernander Foundation.

**Disclosure statement**

No potential conflict of interest was reported by the authors.

**Supplemental data**

Supplemental data can be accessed here: https://dx.doi.org/10.1080/14772000.2019.1590476.

**References**

Basiri-Esfahani, S., Bidi, B., Assadi, M., & Rahimi-Nejad, M. R. (2011). A taxonomic study of *Acanthophyllum* C. A. Mey. (Caryophyllaceae) in Iran. *Iranian Journal of Botany*, 17, 24–39.

Bouckaert, R., Heled, J., Kühnert, D., Vaughan, T., Wu, C. H., Xie, D., … Drummond, A. J. (2014). BEAST 2: A software platform for Bayesian evolutionary analysis. *Public Library of Science Computational Biology*, 10, 4. doi:10.1371/journal.pcbi.1003537

Chiang, Y.-C., Hung, K.-H., Moore, S.-J., Ge, X.-J., Huang, S., Hsu, T.-W., … Chiang, T. Y. (2009). Paraphyly of organelle DNAs in Cycas Sect. Asiorientales due to ancient ancestral polymorphisms. *BioMed Central Evolutionary Biology*, 9, 161. doi:10.1186/1471-2148-9-161

Darriba, D., Taboada, G. L., Doallo, R., & Posada, D. (2012). JModelTest 2: More models, new heuristics and parallel computing. *Nature Methods*, 9, 772. doi:10.1038/nmeth.2109
Shaw, J., Lickey, E. B., Schilling, E. E., & Small, R. L. (2007). Comparison of whole chloroplast genome sequences to choose noncoding regions for phylogenetic studies in angiosperms: The Tortoise and the hare III. *American Journal of Botany, 94,* 275–288. doi:10.3732/ajb.94.3.275

Slatkin, M., & Hudson, R. R. (1991). Pairwise comparisons of mitochondrial DNA sequences in stable and exponentially growing populations. *Genetics, 129,* 555–562. doi:10.1093/hmg/7.3.399

Tajima, F. (1989). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics, 123,* 585–595. doi:PMC1203831

Takhtajan, A. (1986). *Floristic regions of the world.* Berkeley: University of California Press.

Toprak, Z., Pfeil, B. E., Jones, G., Marcussen, T., Ertekin, A. S., & Oxelman, B. (2016). Species delimitation without prior knowledge: DISSECT reveals extensive cryptic speciation in the Silene aegyptiaca complex (Caryophyllaceae). *Molecular Phylogenetics and Evolution, 102,* 1–8. doi:10.1016/j.ympev.2016.05.024

**Associate Editor: Nadia Bystriakova**