Published in Botanical Journal of the Linnean Society 163, issue 1, 14-32, 2010
which should be used for any reference to this work

New insights into the phylogenetics and biogeography of *Arum* (Araceae): unravelling its evolutionary history

ANAHÍ ESPÍNDOLA[1]*, SVEN BUERKI[2], MARIJA BEDALOV[3,4], PHILIPPE KÜPFER[4] and NADIR ALVAREZ[5]

[1]*Laboratory of Evolutionary Entomology, Institute of Biology, University of Neuchâtel, Rue Emile-Argand 11, CH-2009 Neuchâtel, Switzerland*
[2]*Department of Biodiversity and Conservation, Real Jardín Botánico, CSIC, Plaza de Murillo 2, 28014 Madrid, Spain*
[3]*Department of Botany, Faculty of Sciences, University of Zagreb, Marulićev Trg 20/II, HR-10000 Zagreb, Croatia*
[4]*Laboratory of Evolutionary Botany, Institute of Biology, University of Neuchâtel, Rue Emile-Argand 11, CH-2009 Neuchâtel, Switzerland*
[5]*Department of Ecology and Evolution, University of Lausanne, Biophore Building, 1015 Lausanne, Switzerland*

The heat- and odour-producing genus *Arum* (Araceae) has interested scientists for centuries. This long-term interest has allowed a deep knowledge of some complex processes, such as the physiology and dynamics of its characteristic lure-and-trap pollination system, to be built up. However, mainly because of its large distributional range and high degree of morphological variation, species' limits and relationships are still under discussion. Today, the genus comprises 28 species subdivided into two subgenera, two sections and six subsections. In this study, the phylogeny of the genus is inferred on the basis of four plastid regions, and the evolution of several morphological characters is investigated. Our phylogenetic hypothesis is not in agreement with the current infrageneric classification of the genus and challenges the monophyly of several species. This demonstrates the need for a new infrageneric classification based on characters reflecting the evolution of this enigmatic genus. To investigate the biogeography of *Arum* deeply, further spatiotemporal analyses were performed, addressing the importance of the Mediterranean basin in the diversification of *Arum*. Our results suggest that its centre of origin was the European–Aegean region, and that major diversification happened during the last 10 Myr.

ADDITIONAL KEYWORDS: character tracing – infrageneric systematics – Mediterranean biogeography – phylogenetic inferences.

INTRODUCTION

With 109 genera and over 3700 species described (Mayo, Bogner & Boyce, 1997) Araceae have a worldwide distribution and are found in a wide range of environments, from Arctic–Alpine (e.g. *Calla palustris* L.) to xerophytic (e.g. *Anthurium nizandense* Matuda), with most species occurring in the tropics. The family encompasses a large variety of life forms, from epiphytic to aquatic, attesting the numerous adaptive radiations that have occurred in this early Cretaceous family (Chase *et al*., 2006; Anderson & Janssen, 2009). A remarkable feature in Araceae is the evolution of heat production in several genera (Minorsky, 2003), especially those displaying pollination-related associations with arthropods, in which thermogenesis is associated with the emission of volatile compounds and the attraction of pollinators (Moodie, 1975).

*Corresponding author. E-mail: maria.espindola@unine.ch

One of the few Palaearctic representatives of Araceae is the herbaceous genus *Arum* L., which comprises 28 described species (Lobin *et al*., 2007; CATE project, 2010). Because of its characteristic flowers, showing adaptations for trapping pollinators (Gibernau, Macquart & Przetak, 2004), and its ability to produce heat and odours, this genus has fascinated not only scientists but also the wider public for centuries (Theophrastus, 370 BC; Hruby, 1910; Boyce, 1993). As shown by archaeological and historical records, several *Arum* spp. have been used by humans since ancient times for food (tubers), medicine (fruits, leaves, tubers), fashion (tuber starch) and even magic (leaf decoctions) (Prime, 1960).

The inflorescences of *Arum* consist of two parts: a spadix and a spathe (Boyce, 1993). The spadix displays the unisexual flowers and harbours adaptations involved in heat production, whereas the spathe is a modified bract surrounding the spadix. One of the distinctive synapomorphies of the genus is the partition of the spadix. The lower zone comprises both female (lower portion) and male (upper portion) flowers placed in a floral chamber, which is usually delimited by male-sterile flowers modified as hairs: the staminodes. Its apex is a smooth, subcylindrical, usually stipitate organ, known as the spadix appendix (Boyce, 1993). This structure is also recognized as an efficient thermogenetic organ with which the plant attracts pollinating arthropods with heat and production of volatile compounds. The combination of odour emission and hair presence at the top of the floral chamber (acting as a fence) is a key feature for the efficient trapping of arthropods during the female receptive period and until pollen release (Gibernau *et al*., 2004).

Historically, the genus was defined by Fuchs (1542) and later established by Linnaeus (1753). The different species were first circumscribed on the basis of morphology (Schott, 1832), and chromosome counts (Bedalov, 1981) led to the identification of different ploidies in the genus (di-, tetra- and hexaploids, $x = 14$; for a review, see Boyce, 1989). In the most recent revisions of the genus (Boyce, 1993, 1994, 2006; Bedalov & Küpfer, 2005), several morphological characters (tuber shape, flower disposition, growth period, spadix shape and structure of sterile flowers) have been used to build a classification comprising two subgenera, two sections and six subsections. The subgenus *Gymnomesium* (Schott) Engl. is monospecific, including only the Hercynian endemic *Arum pictum* L.f. Subgenus *Arum* Engl. includes sections *Arum* and *Dioscoridea* Bronner, the latter being divided into six subsections (Table 1).

This classification may be controversial, notably because: (1) several taxa have been defined on the basis of herbarium specimens (this approach may not be optimal in this group as important characters are observable only on fresh material; Boyce, 1989); (2) species having large distributions and studied locally were sometimes either simultaneously described under different names (e.g. *A. italicum* Mill.) or assigned to different taxa when they belonged to the same taxon (e.g. *A. cylindraceum* Gasp.) (Bedalov & Küpfer, 2005); (3) following this last point, as several species harbour a high level of intraspecific polymorphism, this may even trigger the splitting of widely distributed taxa (Boyce, 2006). Therefore, it is now an appropriate time to evaluate the systematics of *Arum* based on molecular evidence. Published molecular phylogenetic analyses including species of *Arum* have focused on the investigation of relationships at the family level and have lacked sampling and resolution at the infrageneric level (Cabrera *et al*., 2008; Mansion *et al*., 2008). In this article, we aim to produce a phylogenetic hypothesis for the genus *Arum* by sequencing four plastid regions suitable for addressing relationships at the infrageneric level based on 26 of the 28 described species. This will allow us to assess the validity of the current classification and to examine the evolution of several key characters. To decipher the evolutionary history of this early Miocene genus (Mansion *et al*., 2008), we perform spatiotemporal analyses to determine the events that played a central role in the radiation, dispersion and isolation of the different species (Sanmartín, Enghoff & Ronquist, 2001; Médail & Diadema, 2009). Finally, on the basis of our results, we suggest guidelines for a new infrageneric classification of the genus *Arum*.

MATERIAL AND METHODS

Sampling

Analyses were based on 64 specimens, representing 26 of the 28 described species and spanning all subgenera, sections and subsections of *Arum* (Table 1). On the basis of Mansion *et al*. (2008), *Dracunculus canariensis* Kunth, *D. vulgaris* Schott, *Biarum davisii* Turrill and *B. dispar* (Schott) Talavera were used as outgroup taxa. Samples were either provided by the DNA Bank of the Royal Botanic Gardens, Kew (UK) or extracted directly from dried plant material from herbaria or field collections (Appendix 1).

DNA extraction, amplification and sequencing

DNA of freshly collected material and herbarium samples was extracted using the DNeasy Plant Kit (Qiagen, Basle, Switzerland). The plastid regions *3′rps16-5′trnK*, *ndhA* intron, *psbD-trnT* and *rpl32-trnL* were amplified with the primers described in Shaw *et al*. (2007). Amplifications were performed in a

Table 1. Current subgeneric taxonomy of genus *Arum* L. Taxa with an '*' were not included in the present study.

Subgenus	Section	Subsection	Species
Gymnomesium			*A. pictum* L.f.
Arum	*Arum*		*A. byzantinum* Schott
			A. concinnatum Schott
			A. italicum Mill.
			A. maculatum L.
			A. megobrebi Lobin, M.Neumann, Bogner & P.C.Boyce
	Dioscoridea	*Alpina*	*A. cylindraceum* Gasp.
			A. lucanum Cavara & Grande
		Discroochiton	*A. apulum* (Carano) P.Boyce
			A. balansanum R.R.Mill
			A. besserianum Schott
			A. cyrenaicum Hruby
			A. elongatum Steven
			A. gratum Schott*
			A. hainesii Riedl*
			A. nigrum Vell.
			A. orientale M.Bieb.
			A. purpureospathum P.C.Boyce
			A. sintenisii P.C.Boyce
		Tenuifila	*A. jacquemontii* Blume
			A. korolkowii Regel
			A. rupicola Boiss.
		Hygrophila	*A. euxinum* R.R.Mill
			A. hygrophilum Boiss.
		Poeciloporphyrochiton	*A. dioscoridis* Sibth. & Sm.
			A. palaestinum Boiss.
		Cretica	*A. creticum* Boiss. & Heldr.
			A. idaeum Coust. & Gand.

master mix containing 0.5 × buffer, 150 mM deoxynucleoside triphosphate (dNTP), 0.7 mM $MgCl_2$, 0.3 μg μL^{-1} bovine serum albumin (BSA), 0.5 μM primers and 1 unit of *Taq* Polymerase (Promega, Dübendorf, Switzerland) made up to a final volume of 30 μL with purified MilliQ water. Reactions were run in a TGradient thermocycler (Biometra, Goettingen, Germany). Initial denaturation was programmed for 2 min 30 s at 95 °C, followed by 35 cycles at 95 °C for 35 s, 54–60 °C for 45 s, 72 °C for 1 min and a final extension of 8 min at 72 °C. The purification of PCR products and fluorescence sequencing were performed by Macrogen, Inc. (Seoul, South Korea) and Fasteris Life Sciences (Geneva, Switzerland) with the same primers as used for PCR amplification.

Sequence alignment and phylogenetic reconstructions

Automatically generated base-calls for all sequences were checked and edited using ChromasPro 1.41 (Technelysium Pty Ltd, Tewantin, Australia). For each plastid region, alignment was performed using the ClustalW algorithm implemented in Bioedit 7.0 (Hall, 1999), followed by minor manual corrections. After concatenation of the four regions, a matrix of 3723 bp was obtained. Gaps were further coded following the simple method of Simmons & Ochoterena (2000), as implemented in FastGap 1.2 (Borchsenius, 2009).

The numbers of constant (C), variable (V) and potentially parsimony-informative (PI) sites were calculated for each partition using PAUP* v4.0b10 (Swofford, 2002). Before computing total evidence trees, we tested for incongruence among the four regions by applying the partition homogeneity test as implemented in PAUP* v4.0b10 with 100 replicates (this test is equivalent to the incongruence length difference test of Farris *et al*., 1994; for convenience, it is referred to as the ILD test). Total evidence trees (*sensu* Kluge, 1989) were determined using both Bayesian inference and maximum parsimony (MP) approaches.

Bayesian Markov chain Monte Carlo (MCMC) methods were used to approximate the posterior prob-

ability distribution of the phylogenetic trees on the basis of the combined plastid dataset with four distinct partitions plus one partition for the gap information, by running MrBayes v.3.1.2 (Ronquist & Huelsenbeck, 2003). Model selection for the plastid DNA partitions was tested using MrAIC (Nylander, 2004) based on the Akaike information criterion (Akaike, 1973), and a restriction model was applied to the partition containing the coded gaps. Three independent runs with one cold and five heated chains were run for 5×10^7 generations each. Frequencies were sampled every 1000 generations and temperature was fixed to 0.5. The convergence of MCMC was tested by computing the potential scale reduction factor (PSRF; Gelman & Rubin, 1992) as implemented in MrBayes, and by determining the effective sample size for each parameter using Tracer v.1.4 (Rambaut & Drummond, 2004). Accordingly, the burn-in period was set to 10^7 generations until stationarity of the likelihood value was established among the runs, and 10 000 sample points were discarded (20% of the total number of trees). The remaining 40 001 trees from each run were pooled (120 003 trees in total) to estimate the posterior probability distribution of the phylogenetic inference. To yield a single phylogenetic hypothesis, the posterior distribution was summarized in the 50% majority-rule consensus tree (referred to as the half-compatible tree in MrBayes), with a Bayesian posterior probability (BPP) at each node indicating statistical support.

The combined dataset was further analysed under the MP criterion using the parsimony ratchet (Nixon, 1999) as implemented in PAUPrat (Sikes & Lewis, 2001). Based on recommendations by Nixon (1999), ten independent searches were performed with 200 iterations, and 15% of the parsimony-informative characters were perturbed using PAUP* version 4.0b10. The shortest equally most parsimonious trees were combined to produce a strict consensus tree. Node support was determined by computing decay indices (DIs) (Bremer, 1988) as implemented in TreeRot 3.0 (Sorenson & Franzosa, 2007). DI measures the number of extra steps in tree length required before a node collapses (Bremer, 1988; Baker & DeSalle, 1997).

Finally, the level of congruence between Bayesian and MP analyses was determined by computing the quartet distance (Estabrook, 1992) between the two topologies. Considering that the distances between the different topologies were small (see Phylogenetic inferences section in Results), the remaining analyses were only based on the Bayesian inference analysis.

Character evolution

Character tracing was performed on traits generally used in taxonomic studies of *Arum*. On the basis of the topology of the 50% majority-rule Bayesian analysis, the following categorical characters were mapped using Mesquite 2.6 (Maddison & Maddison, 2009) with accelerated transformation optimization (ACCTRAN) and unordered parsimony: tuber form (rhizomatous/discoid), flower type (flag/cryptic), spadix/spathe ratio (0–0.5; 0.5–1; > 1) and ploidy (di-, tetra-, hexaploid). Characters were obtained from the latest systematic studies performed on the species (Boyce, 1993, 2006; Bedalov & Küpfer, 2005; Lobin *et al*., 2007).

Dating and biogeographical analyses

Because the molecular clock hypothesis was rejected (data not shown), the 50% majority-rule Bayesian inference tree was rendered ultrametric using the penalized-likelihood method (Sanderson, 2002; hereafter PL), as implemented in the program r8s v.1.71 (Sanderson, 2004) by applying a smoothing value of 1000 and the truncated Newton algorithm. The most external outgroup, *B. davisii*, was pruned for the estimation of the divergence time as required by the program (see Sanderson, 2004). The following calibration points were applied (according to Mansion *et al*., 2008): (1) the root node (i.e. the most recent common ancestor of genera *Arum*, *Biarum* Schott and *Dracunculus* Mill.) was constrained to a maximum age of 30.2 Mya; (2) the most recent common ancestor of *Arum* and *Dracunculus* was constrained to a minimum age of 27.3 Mya; and (3) the stem group of *Arum* subgenus *Arum* was constrained to a minimum age of 16.1 Mya.

Areas were defined following different studies on the geological and biogeographical history of the Mediterranean basin and surrounding areas (Meulenkamp & Sissingh, 2003; Mansion *et al*., 2008; Ree & Sanmartín, 2009), and were set to a number of ten: East European, West European, Apennines, Aegean, Anatolian, Iranian, Arabian, North African, Macaronesian Islands and Caucasus (Fig. 3). The rules applied to define the area for each species were as follows: (1) if the origin of the sample was known, the sample was attributed to the area in which it was sampled; (2) if the origin was unknown, the sample was assigned to the area in which the plant is known to be distributed according to Boyce (1993, 2006) and the search engine of *Flora Europaea* (Flora Europaea, 2009) [in the case of *A. balansanum* R.R.Mill., *A. byzantinum* Schott, *A. sintenisii* (Engl.) P.C.Boyce, *D. vulgaris*, *D. canariensis* and *B. dispar*]; (3) if a sample belonged to a widely distributed and well-described species for which we did not possess samples from all the parts of the distributional area, it was assigned to its region of origin plus the remaining noncovered regions according to Boyce (1993) (only in the case of *A. italicum*).

Table 2. Sequenced regions, with corresponding total number and percentages of base pairs (bp), constant (C), variable (V) and parsimony informative (PI) sites.

Region	Total (bp)	Constant (C) sites	Variable (V) sites	Parsimony-informative (PI) sites
3′rps16-5′trnK	845 (100%)	791 (93.6%)	54 (6.4%)	17 (2.0%)
ndhA intron	1077 (100%)	1021 (94.8%)	56 (5.2%)	24 (2.2%)
psbD-trnT	1024 (100%)	993 (97.0%)	31 (3.0%)	13 (1.3%)
rpl32-trnL	777 (100%)	730 (94.0%)	47 (6.0%)	36 (4.6%)

Dispersal-vicariance analysis (DIVA) is a method for inferring the most parsimonious reconstruction of ancestral ranges on a given phylogenetic tree by minimizing the number of dispersal and extinction events that are needed to explain the current terminal distributions (Ronquist, 1997). The program DIVA (Ronquist, 2001) uses a three-dimensional cost matrix to estimate the cost of moving from the ancestor to each of the descendants (Ronquist, 1997). It allows two different scenarios for range inheritance at speciation nodes: (1) duplication or within-area speciation, when the ancestor is distributed in a single area and each of the two descendants inherits the entire ancestral range (e.g. A to A); (2) vicariance, when the ancestor occurs in two or more areas and each descendant inherits a nonoverlapping subset of the ancestral range (e.g. AB to A and B). Only one dispersal event per branch (between two ancestral nodes) is allowed in the model, except for terminal branches leading to widespread taxa, for which DIVA postulates multiple dispersal events. To account for polytomies in the 50% majority-rule Bayesian inference tree, five more exceptions were required in our analysis (see below). DIVAs were run with the maximum number of areas allowed at ancestral nodes constrained to two. Uncertainty in phylogenetic relationships was accounted for in DIVA by using an approach proposed by Nylander *et al*. (2008), which integrates DIVA parsimony-based reconstructions over a Bayesian MCMC sample of trees representing the posterior probability of the tree topology (hereafter referred to as Bayes-DIVA). Specifically, we sampled one tree for every 16 trees (7501 in total) from the MCMC 'post-burnin' sample and used R scripts available from the second author to summarize/average ancestral area reconstructions over all sampled trees for each node in the 50% majority-rule Bayesian inference, which was used as the reference. Only those trees containing the node of interest were summarized in estimating the probabilities for that node. This approach allows an estimation of the marginal probabilities of ancestral ranges for a given node whilst integrating over the uncertainty in the rest of the tree topology (Nylander *et al*., 2008). Ancestral areas and vicariance/dispersal events were recorded following Buerki (2009).

As several polytomies were found in the 50% majority-rule topology, the following rule was applied to solve incompatibilities between nodes and to estimate correct dispersal-vicariance events (that otherwise would violate DIVA assumptions): if the most probable area for a given node was incompatible (according to DIVA assumptions) with that of the next coming node or tip, it was combined with the following most probable area(s); this was performed until the ancestral areas of the node were congruent with the areas assigned to the following node or tip. In order to summarize the different dispersal events across the three geological epochs spanning the diversification of *Arum* (Miocene, Pliocene and Pleistocene), a pairwise matrix of dispersion was built for each epoch to address the links among the ten defined areas [this was performed using R (R Development Core Team, 2009), with scripts available on request from the second author]. When a branch spanned over more than one epoch, the proportion of the branch over each epoch was considered, and the fraction D of one single dispersal event in a given epoch was recorded ($0 < D < 1$). To summarize the results, arrows with variable widths (proportional to the number of dispersal events) were drawn on palaeogeographical maps corresponding to the three relevant epochs (Meulenkamp & Sissingh, 2003).

RESULTS

Phylogenetic inferences

The combined dataset consisted of 250 sequences. Aligned lengths were 845 bp for *3′rps16-5′trnK*, 1077 bp for the *ndhA* intron, 1024 bp for *psbD-trnT*, 777 bp for *rpl32-trnL* and 104 binary positions corresponding to coded gaps. The final matrix thus contained a total of 3827 characters (3723 nucleotides and 104 gap presence/absence). Values for C, V and PI for each partition are given in Table 2. Partition *rpl32-trnL* provides slightly more informative sites (36) than the other partitions. The partition homoge-

neity test was passed ($P = 0.07$), indicating that the information provided by the four plastid regions was congruent. Topologies obtained with Bayesian inference and MP algorithms (Fig. 1) were highly congruent (quartet distance of 0.114, meaning that approximately 89% of the components were compatible between the two trees) and defined five major supported clades (see Fig. 1). Topologies depict *A. pictum* as sister to the rest of the genus, confirming the definition of subgenera *Gymnomesium* and *Arum* (supported by BPP = 1 and DI = 10; Fig. 1). In subgenus *Arum*, the first branching clade I includes *A. palaestinum* Boiss. and the different subspecies of *A. dioscoridis* Sibth. & Sm. (BPP = 1, DI = 3). Clade II contains *A. concinnatum* Schott and *A. italicum* (for which monophyly was not contradicted in the Bayesian analysis) (BPP = 0.95, DI = 3). Clade III includes the two most eastern taxa, *A. jacquemontii* Blume and *A. korolkowii* Regel, *A. megobrebi* Lobin, M.Neumann, Bogner & P.C.Boyce, *A. rupicola* Boiss. and the two easternmost samples of *A. maculatum* L. included in this study, and a new species, hereafter referred to as *A.* sp. nov. (BPP = 0.99, DI = 1). In the MP topology, *A. creticum* Boiss. & Heldr. is also included in clade III as the first branching lineage (Fig. 1). Clade IV includes two subclades containing *A. sintenisii* P.C.Boyce and *A. hygrophilum* Boiss., on the one hand, and *A. byzantinum* Schott, *A. nigrum* Vell. and some specimens of *A. elongatum* Steven, on the other (BPP = 0.84, DI = 1). In the Bayesian inference topology, *A. idaeum* Coust. & Gand. is also included in clade IV as the first branching lineage (Fig. 1). Finally, clade V is poorly resolved and includes the remaining taxa: *A. maculatum* (western samples), *A. cylindraceum* Gasp., *A. orientale* M.Bieb., *A. besserianum* Schott, *A. balansanum* R.R.Mill, *A. purpureospathum* P.C.Boyce, *A. euxinum* R.R.Mill, *A. apulum* (Carano) P.Boyce, *A. cyrenaicum* Hruby and one representative of *A. elongatum* (BPP = 1, DI = 2). The relative position of clade V swapped depending on the phylogenetic algorithm, as it was sister to clade IV in the Bayesian inference tree, but sister to clade III in the MP tree (Fig. 1). Incongruence between the two topologies concerned (1) the positions of clade V relative to clades III and IV, and (2) the branching of *A. idaeum* and *A. creticum*.

Sections as defined by Boyce (1989) were not supported by the phylogenetic hypotheses. Except for the cases of subsections *Poeciloporphyrochiton* (clade I, *A. dioscoridis* and *A. palaestinum*) and *Tenuifila* (subclade in clade III, *A. jacquemontii*, *A. korolkowii* and *A. rupicola*), our topologies did not support the current infrageneric delimitation (Fig. 1, Table 1). Finally, the monophyly of several widespread species (e.g. *A. elongatum* and *A. maculatum*) was not supported (see Fig. 1).

Character evolution

The reconstruction of ancestral states for the four studied characters is shown in Figure 2. The trait that appears to be most constrained from the phylogenetic reconstruction is ploidy (Fig. 2A), with an ancestral character state corresponding to diploidy ($2n = 2x = 28$) and one single evolution towards hexaploidy. Tetraploidy evolved several times. The remaining characters (Fig. 2B, tuber shape; Fig. 2C, flower shape; Fig. 2D, spathe/spadix ratio) show a pattern of multiple independent events and are much less informative at the infrageneric level. An exception could be the evolution of the rhizomatous tuber shape which, although largely symplesiomorphic, seems to be correlated with the level of ploidy.

Biogeographical analysis

Reconstructed ancestral areas for the different nodes corresponding to the 50% majority-rule Bayesian tree are shown in Figure 3. The two most probable ancestral areas from the crown nodes of the genus were the Aegean and West European regions. Later nodes show that the Aegean and Anatolian regions were the only areas to harbour ancestral lineages of the genus for a long time. Overall, a substantial proportion of the dispersion of *Arum* lineages towards their current distribution areas seems to have happened after the late Miocene (*c.* 10 Mya).

The rates and direction of dispersal events at three different time-slices corresponding to the Miocene, Pliocene and Pleistocene are shown in Figure 4. During the Miocene (23–5.3 Mya, Fig. 4A), dispersion mainly happened from the Aegean area to Anatolia. Exchanges were also possible between the newly emerging Caucasus region and the Aegean and Anatolian areas. Colonization of the Iranian area seems to have happened only during this period. Colonization of Macaronesia also occurred at this time, but other dispersals towards this area were probably also possible during the early Pliocene (5.3–2.6 Mya, Fig. 4B). During this epoch, the genus extended its distribution for the first time onto the Arabian plate, and important dispersion events seem to have happened from the Aegean (and, to a lesser extent, from the Anatolian region) to Eastern Europe. The North African region was colonized during the most recent geological epoch (Pleistocene, 2.6 Mya–present, Fig. 4C) probably via two pathways: (1) from the Apennines through the southern tip of the Italian Peninsula; and (2) from the Arabian region through the Gulf of Suez. Numerous dispersals also occurred from the Aegean to the Anatolian area in the Pleistocene and from the Apennine region to Western and Eastern Europe and the Aegean. During this last epoch, exchanges seemed to have halted between the

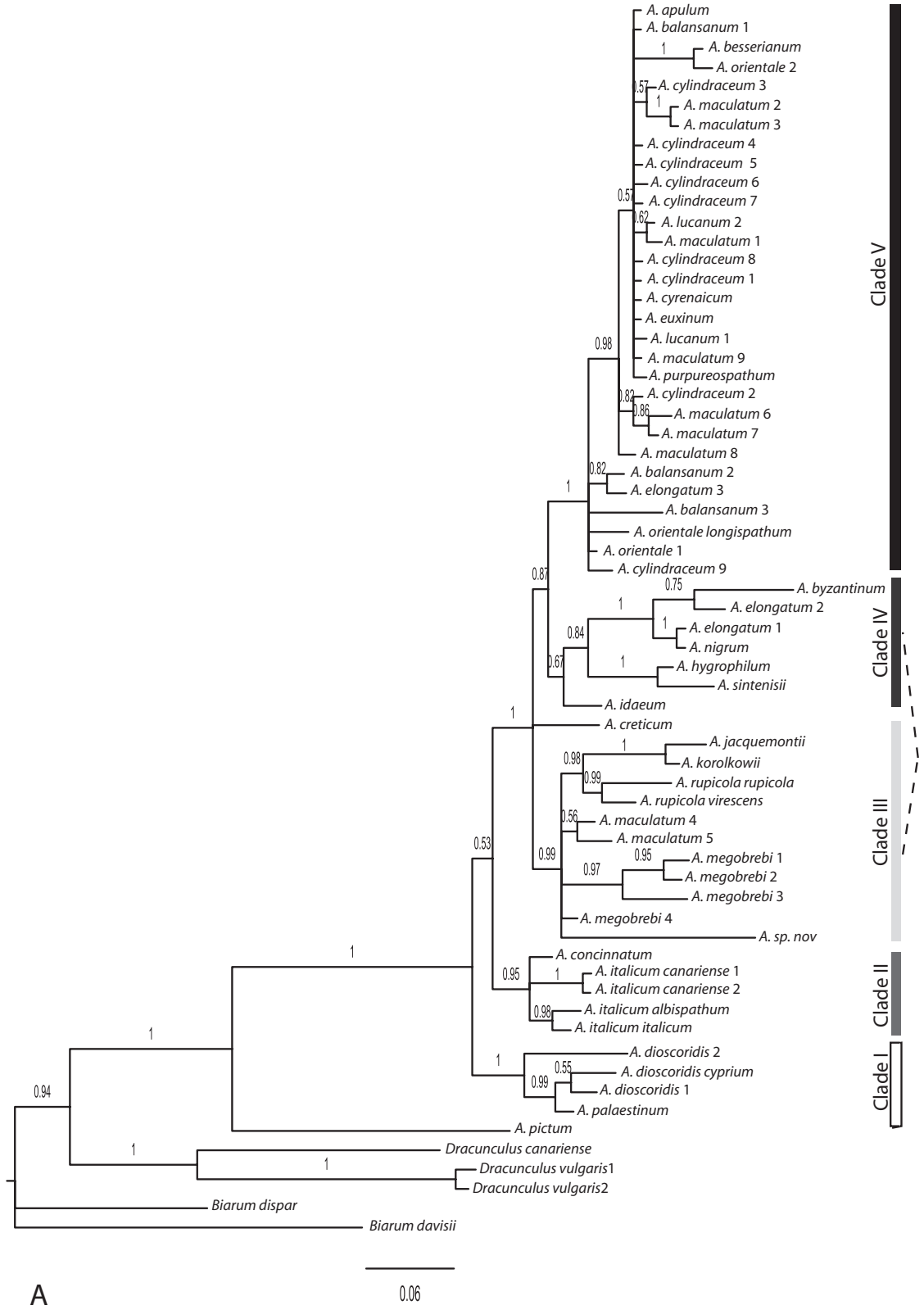


Figure 1. Inferred plastid phylogenies: A, Bayesian inference, half-compatible tree; B, maximum parsimony (MP), strict consensus tree. Values shown on the branches represent Bayesian posterior probability values (A) and decay indices (B). Vertical bars indicate major clades explained in the text.

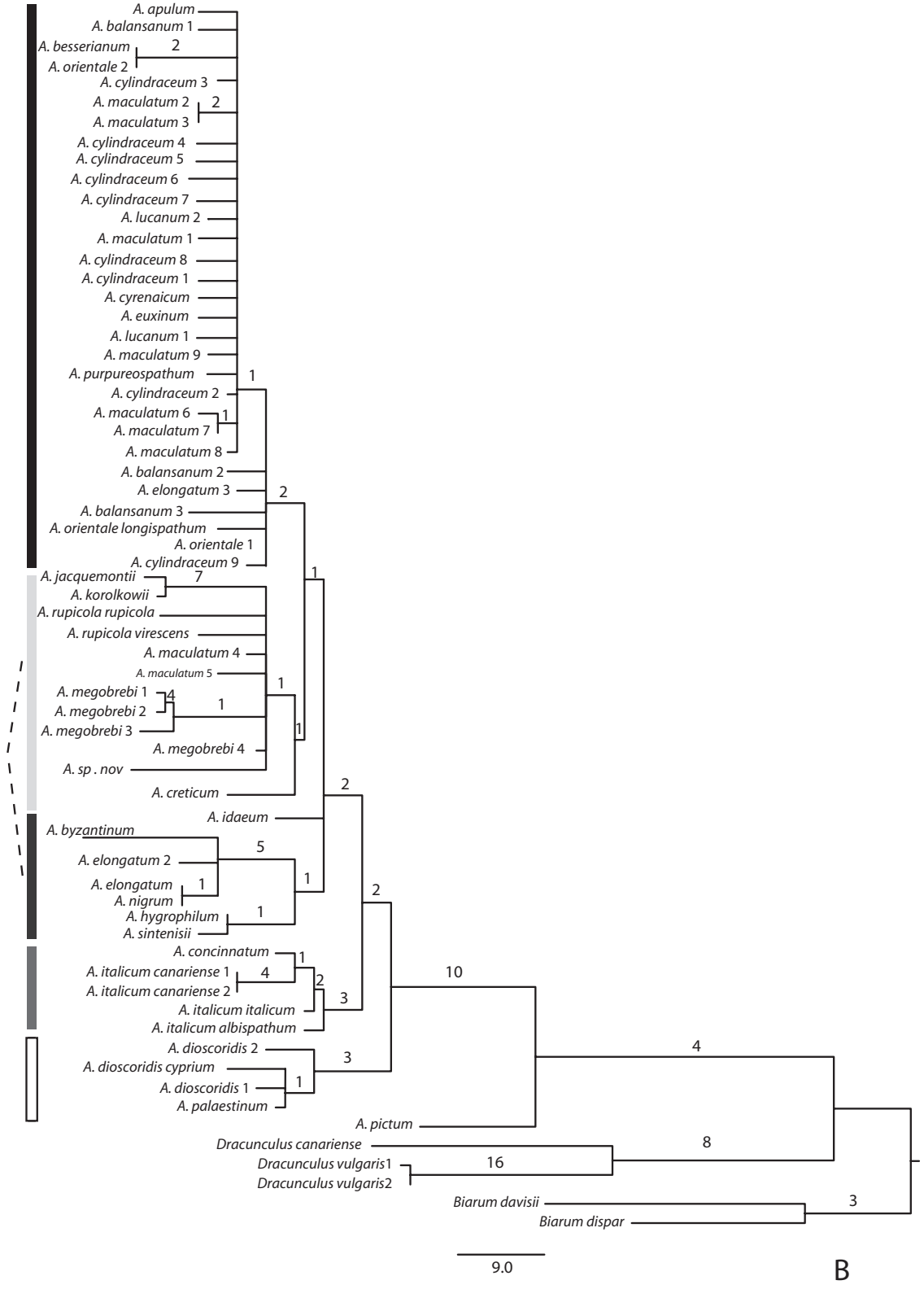


Figure 1. *Continued*

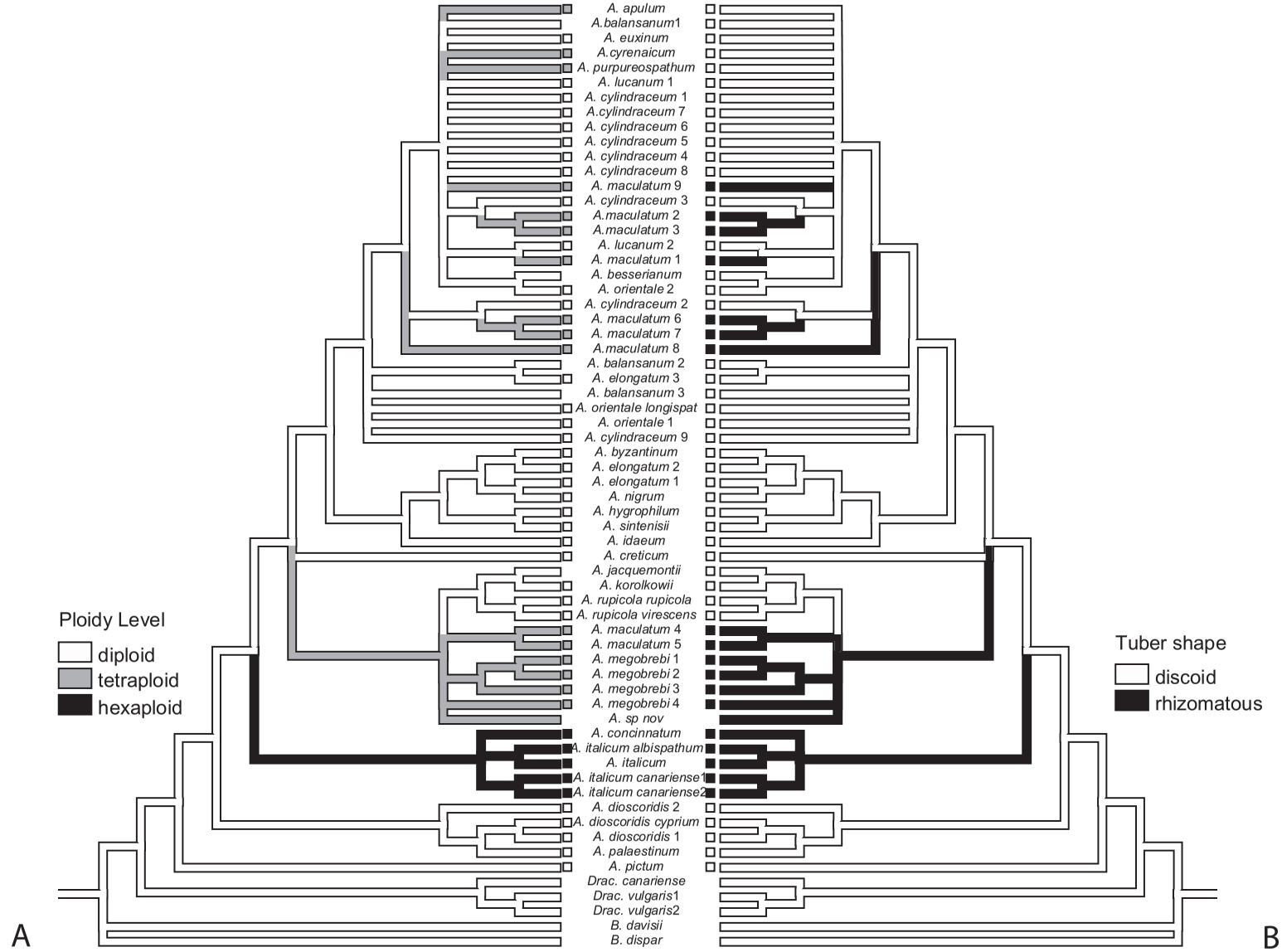


Figure 2. Evolution of categorical characters on the Bayesian topology: A, ploidy (2*x*, 4*x*, 6*x*); B, tuber shape (rhizomatous, discoid); C, flower type (flag, cryptic); D, spadix/spathe ratio (0–0.5; 0.5–1; > 1). Colours are explained in the legend of each figure. Missing squares at the level of terminal taxa indicate unknown characters.

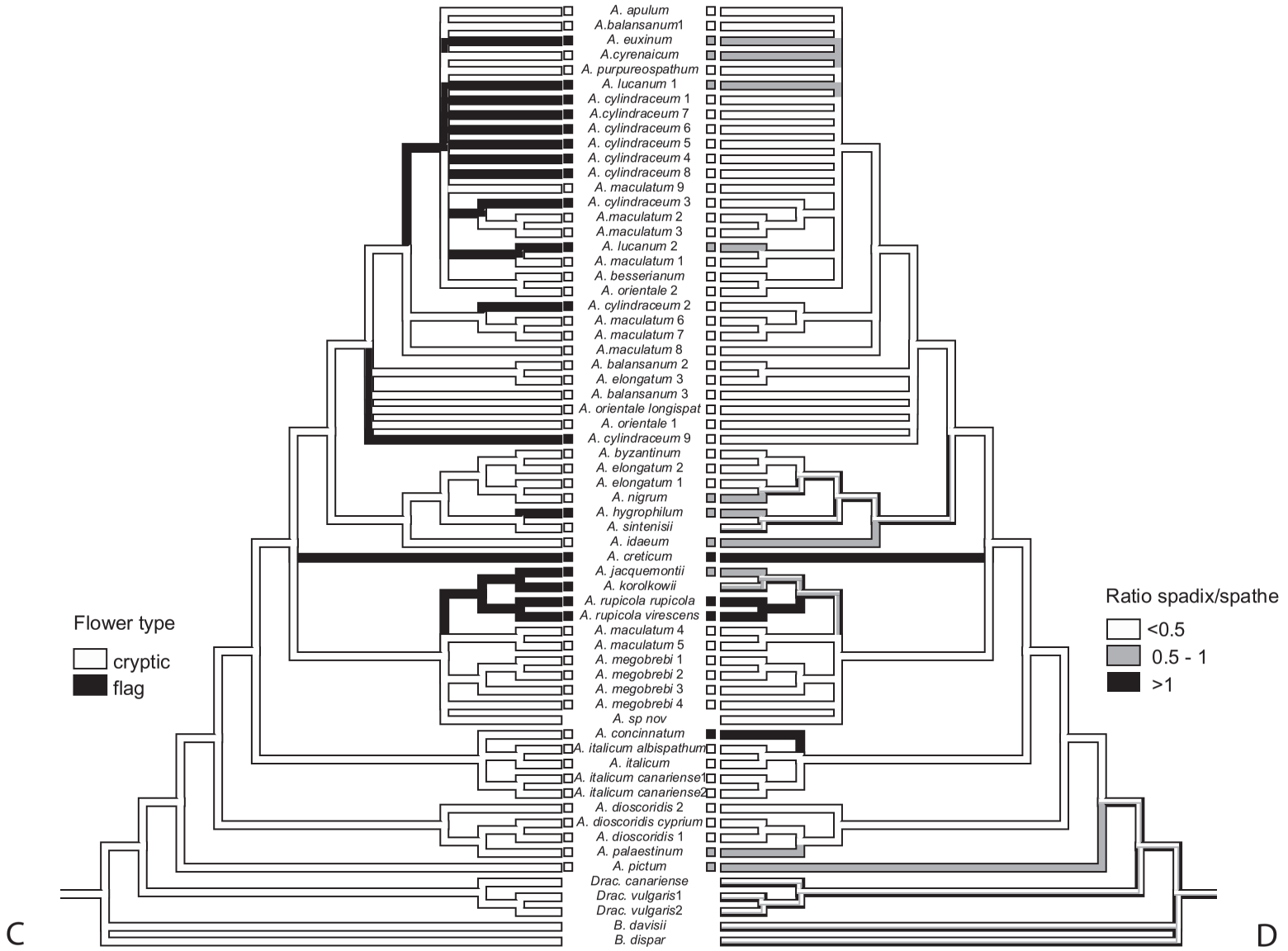


Figure 2. *Continued*

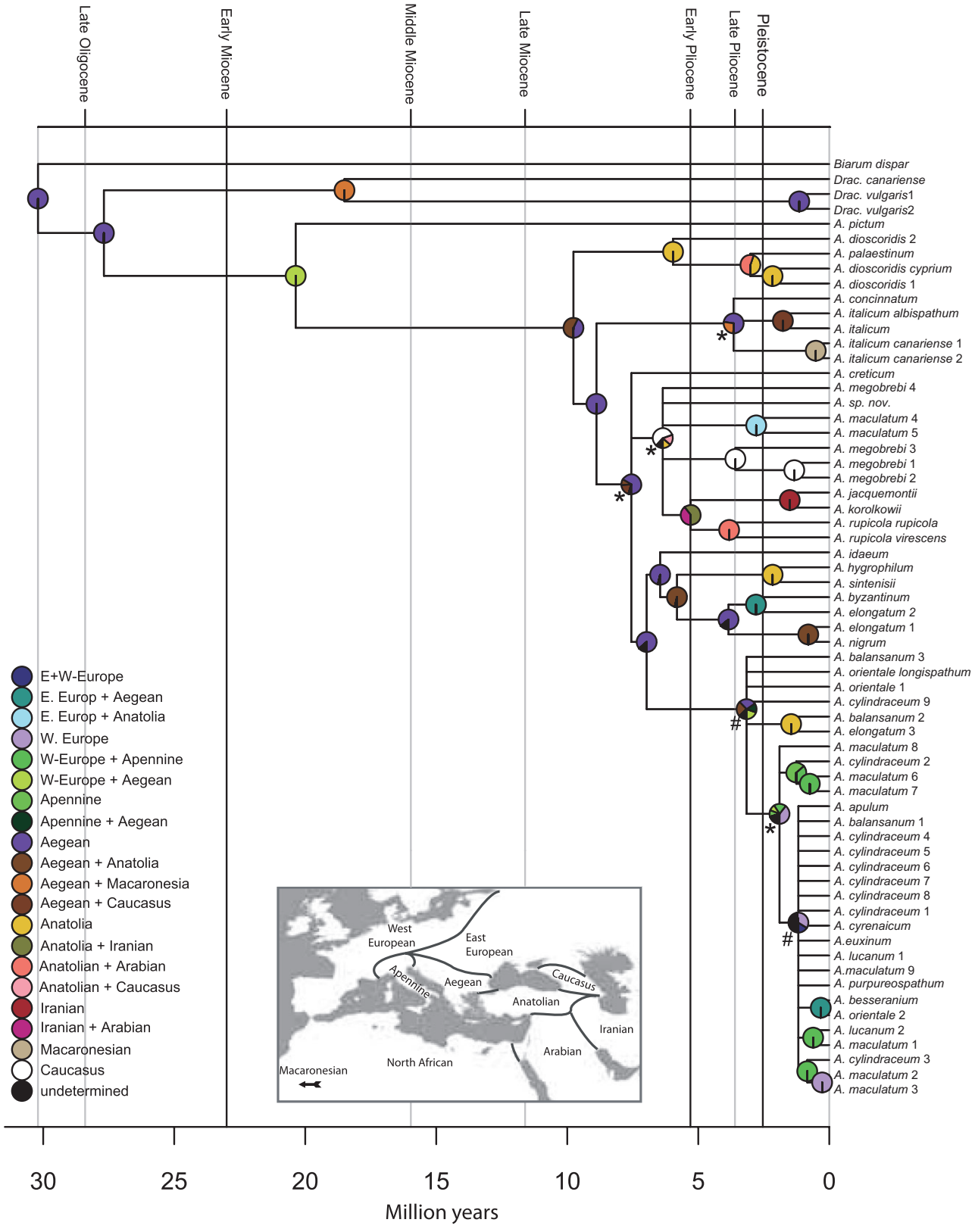

Late Oligocene
Early Miocene
Middle Miocene
Late Miocene
Early Pliocene
Late Pliocene
Pleistocene
Biarum dispar
Drac. canariense
Drac. vulgaris1
Drac. vulgaris2
A. pictum
A. dioscoridis 2
A. palaestinum
A. dioscoridis cyprium
A. dioscoridis 1
A. concinnatum
A. italicum albispathum
A. italicum
A. italicum canariense 1
A. italicum canariense 2
A. creticum
A. megobrebi 4
A. sp. nov.
A. maculatum 4
A. maculatum 5
A. megobrebi 3
A. megobrebi 1
A. megobrebi 2
A. jacquemontii
A. korolkowii
A. rupicola rupicola
A. rupicola virescens
A. idaeum
A. hygrophilum
A. sintenisii
A. byzantinum
A. elongatum 2
A. elongatum 1
A. nigrum
A. balansanum 3
A. orientale longispathum
A. orientale 1
A. cylindraceum 9
A. balansanum 2
A. elongatum 3
A. maculatum 8
A. cylindraceum 2
A. maculatum 6
A. maculatum 7
A. apulum
A. balansanum 1
A. cylindraceum 4
A. cylindraceum 5
A. cylindraceum 6
A. cylindraceum 7
A. cylindraceum 8
A. cylindraceum 1
A. cyrenaicum
A.euxinum
A. lucanum 1
A.maculatum 9
A. purpureospathum
A. besseranium
A. orientale 2
A. lucanum 2
A. maculatum 1
A. cylindraceum 3
A. maculatum 2
A. maculatum 3
E+W-Europe
E. Europ + Aegean
E. Europ + Anatolia
W. Europe
W-Europe + Apennine
W-Europe + Aegean
Apennine
Apennine + Aegean
Aegean
Aegean + Anatolia
Aegean + Macaronesia
Aegean + Caucasus
Anatolia
Anatolia + Iranian
Anatolian + Arabian
Anatolian + Caucasus
Iranian
Iranian + Arabian
Macaronesian
Caucasus
undetermined
West European
East European
Apennine
Aegean
Caucasus
Anatolian
Iranian
Arabian
North African
Macaronesian
30
25
20
15
10
5
0
Million years

Figure 3. Ancestral areas assigned by dispersal-vicariance analysis (DIVA) to each node of the Bayesian topology. Colours represent ancestral areas (see legend). All areas with a probability < 0.1 were pooled and treated as a single undetermined area (in black). Nodes having been treated with special rules are indicated by '*' or '#': '*' indicates that the second most probable area has been combined with the first in order to sketch compatible scenarios; '#' indicates that more than two ancestral areas have been combined to obtain the congruence of the nodes and tips. Scale corresponds to million years from present. Map shows areas defined for the biogeographical inference.

Aegean and Eastern Europe, despite the neighbouring position of these two regions.

DISCUSSION

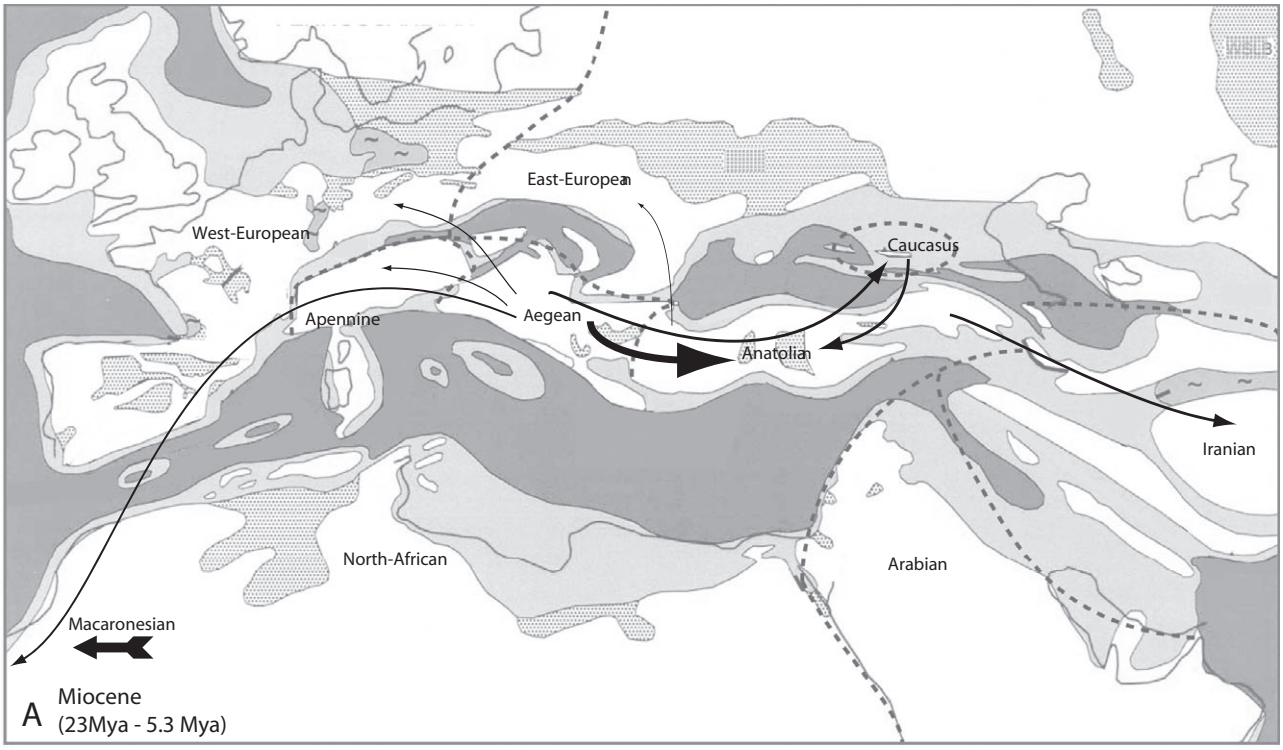
Infrageneric relationships and species' identity

Classical taxon definition and circumscription in the genus *Arum* (Boyce, 1993, 2006) only partially match our phylogenetic hypothesis. As shown in Figure 1, the identity of sections, subsections (*sensu* Boyce, 1989, 1993) and species is strongly challenged, and it seems obvious that there is a 'gap' between the current classification and the genetic identities of the taxa. Our analyses, however, confirm the validity of the two subgenera, as *A. pictum* (subgenus *Gymnomesium*) is the first branching lineage of *Arum* (as shown previously by Mansion *et al*., 2008). This result is supported by floral (staminodes present, but no pistillodes) and phenological (flowering in autumn and not in spring as in the rest of the species) characters restricted to *A. pictum*. For more than a century, the peculiarities of this Hercynian endemic have been recognized, and several authors have attempted to place it in a different genus (*Gymnomesium* Schott, 1855). The long branch separating this lineage from the other species (with a divergence estimated to be sometime between the early and middle Miocene; Fig. 3) favours the hypothesis of a palaeorelictual identity of *A. pictum* (as proposed by Mansion *et al*., 2008). Within subgenus *Arum*, only subsection *Poeciloporphyrochiton* (Fig. 1, clade I) is retrieved by our phylogenetic hypothesis. It appears as the first branching lineage in the subgenus, confirming that *A. dioscoridis* and *A. palaestinum* are closely related and placed in an external position, as proposed formerly by several authors (Hruby, 1910; Boyce, 1989, 1993). Another exception could be subsection *Tenuifila*, which is nested in clade III in the Bayesian topology, and might still be considered as a valid entity (see below). No other subsection is compatible with our results.

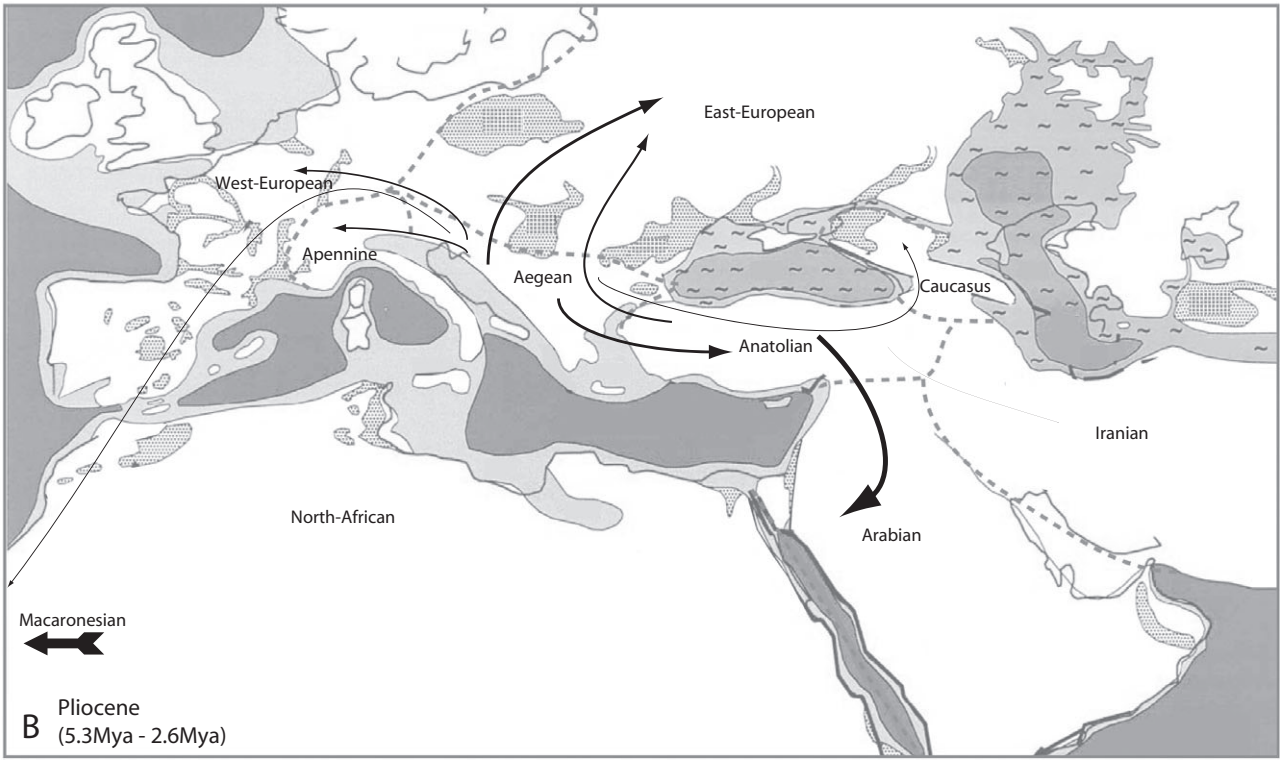
Both Bayesian and MP topologies argue in favour of the monophyly of hexaploid taxa (Fig. 1, clade II; Fig. 2A), with *A. italicum* specimens clustering together with *A. concinnatum*. The insular *A. italicum* ssp. *canariense* (Webb & Berthel.) P.Boyce is genetically differentiated from the 'continental' subspecies from which it diverged during the Pliocene (Fig. 3). As the monophyly of *A. italicum* is not retrieved in the MP analysis, a more thorough analysis (e.g. using genomic screening markers) should be performed to confirm the status of *A. concinnatum*.

Although the phylogenetic relationships among the three remaining clades (III, IV and V) are not yet resolved (i.e. the topology varies according to the phylogenetic algorithm), their respective monophyly is relatively well supported with DI $\geq$ 1 and BPP > 0.95 (with the exception of clade IV, which shows a lower support of 0.84; see Fig. 1). Current molecular data do not allow the discussion of the phylogenetic relationships of *A. creticum* and *A. idaeum*, two species with overlapping distributions in Crete, which are either placed in a polytomy at the base of these three clades or as the first branching lineages of clade III (MP topology) and clade IV (Bayesian topology), respectively. These species are morphologically similar (open floral chambers, sweet or weak odour production vs. closed floral chamber and clear lure-oriented odour production in the other species) and were included in subsection *Cretica* (Boyce, 1989). Relationships among clades III, IV and V should also be examined carefully as our results do not allow conclusions to be drawn regarding the position of clade V, as it appears as either sister to clade III (MP topology) or to clade IV (Bayesian inference tree).

The strongly supported clade III (excluding *A. creticum*; BPP = 0.99; DI = 1) comprises all members of subsection *Tenuifila* (i.e. *A. rupicola*, *A. jacquemontii* and *A. korolkowii*, which form a well-supported monophyletic group in the Bayesian topology; BPP = 0.98) and all representatives of *A. megobrebi*, two specimens of *A. maculatum* from the easternmost edge of the distribution and one sample from the Caucasus area. The latter should be considered as a new species (referred to as *A.* sp. nov.). The placement of *A. maculatum* samples in clade III is unexpected as the two specimens found here are highly divergent phylogenetically with respect to the European representatives (clade V) of this widely distributed species (Fig. 1). As a consequence, they might merit treatment as a different species if further morphological studies confirm this status by identifying synapomorphies. Clade III has a biogeographical coherence as the taxa included are found in the eastern part of the



A Miocene
(23Mya - 5.3 Mya)



B Pliocene
(5.3Mya - 2.6Mya)

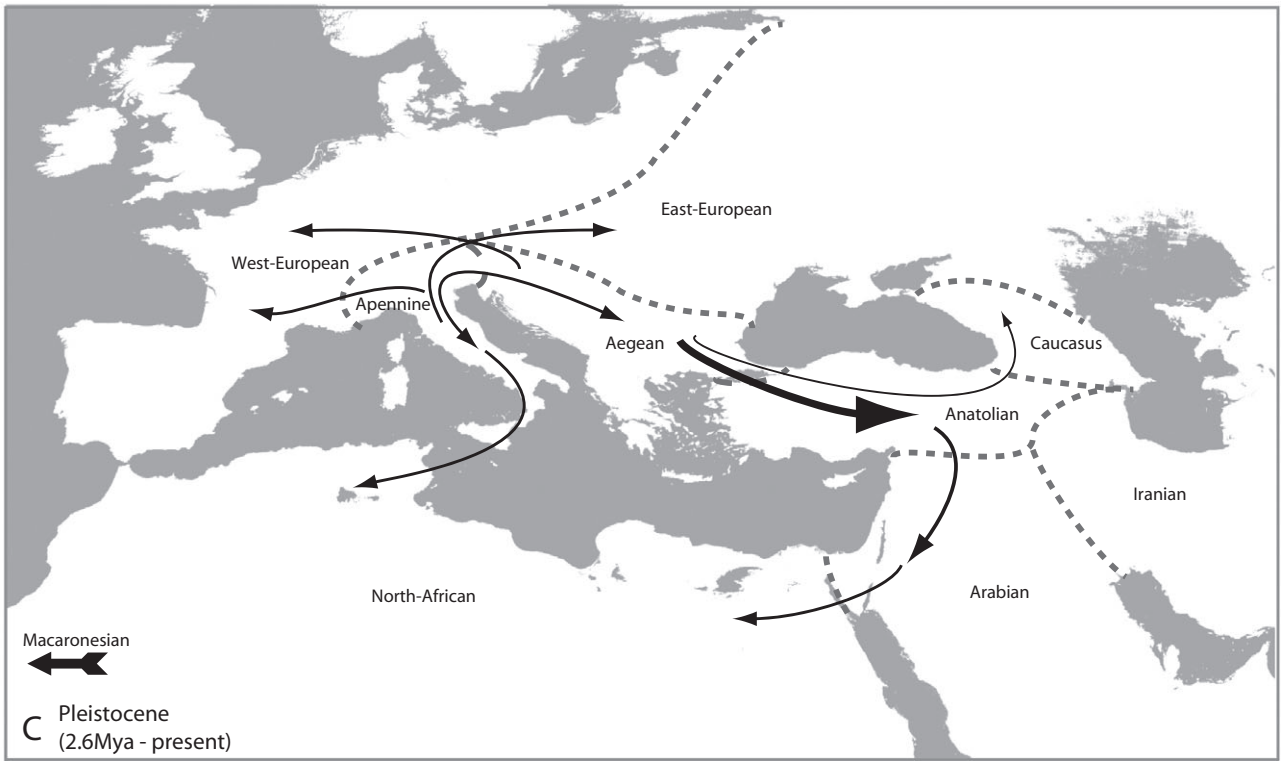


Figure 4. Dispersion events at three time slices: A, Miocene; B, Pliocene; C, Pleistocene (maps A and B, modified from Meulenkamp & Sissingh, 2003; with permission of the editors). The widths of arrows are proportional to dispersal rates. Broken lines indicate boundaries between biogeographical zones.

distribution area of the genus (from northern Turkey to the Iranian region), confirming the role played by this area in the diversification of *Arum* (as proposed previously by Bedalov & Küpfer, 2005).

Clade IV (excluding *A. idaeum*; BPP = 0.84; DI = 1) comprises two specimens of *A. elongatum*, which appears to be paraphyletic with respect to *A. nigrum*, and possibly *A. byzantinum* (in the Bayesian topology), although the position of the latter is not well supported (Fig. 1). More generally, the morphology typical for *A. elongatum* seems to be quite labile as one specimen is also found in clade V (see below). The other representatives of this clade are the closely related *A. sintenisii* (endemic to Cyprus) and the oriental *A. hygrophilum*, the former probably having diverged from the latter in the late Pliocene after a dispersal followed by an insular differentiation. This result was already predicted by Boyce (2006).

Finally, clade V is by far the least resolved, encompassing closely related taxa that diverged during the second half of the Pliocene and the Pleistocene (Fig. 3), most having colonized the Apennines and temperate habitats in Western and Eastern Europe (Fig. 4). In this clade, differentiation among specimens is weak and all species sampled more than once are paraphyletic (Fig. 1). Uncovering the relationships among taxa within this clade would require further analyses based on, for example, genomic screening. This might help to address the status of widely distributed taxa, such as *A. cylindraceum* and the 'European' *A. maculatum*, with that of narrow endemics such as *A. apulum* and *A. purpureospathum*. Another case of interest is the well-supported group composed of *A. besserianum* and one specimen of *A. orientale* (BPP = 1; DI = 2; Fig. 1). *Arum orientale* is still poorly defined as attested by successive revisions during the last 15 years (Boyce, 1994, 2006; Bedalov & Küpfer, 2005). This taxon was first described as a species with several subspecies, present in Crimea and extending to the eastern part of the Balkans (Boyce, 1993). However, several morphological characters point to a close relationship with *A. besserianum* distributed in Ukraine and Poland (P. Küpfer, pers. observ.). Consequently, this taxon certainly encompasses different paraphyletic lineages, and both its status and that of other taxa (e.g. *A. balansanum*) should be investigated using novel genomic techniques coupled with taxonomy.

Therefore, we have observed that two different patterns arise when testing the monophyly of species for which more than one specimen was collected. On the one hand, some are well supported by our analyses:

this is the case for *A. dioscoridis* (clade I), *A. rupicola* (clade III) and, to a lesser extent, *A. megobrebi* (clade III) (Fig. 1). On the other hand, some species are clearly polyphyletic (comprising specimens from lineages that diverged as early as the Pliocene), such as, for instance, *A. maculatum* (clades III and V) and *A. elongatum* (clades IV and V) (Fig. 1). Finally, the case of *A. italicum* (clade II) is somewhat intermediate, as the monophyly is not contradicted by the Bayesian topology, whereas the species is paraphyletic (with the inclusion of *A. concinnatum*) in the MP topology.

Character evolution: characterizing the identity of the most recent common ancestor

Among all the investigated characters, only ploidy seems to be related to the evolution of the genus (Fig. 2A). The remaining traits (spathe/spadix ratio, flower type, shape of the tuber) show patterns of multiple independent evolution and a looser correlation with the evolutionary history of *Arum*. Our results therefore support the hypothesis that the level of ploidy might constitute an informative character for the systematics of the genus, as first proposed by Bedalov & Küpfer (2005), and address the diploid status of the most recent common ancestor (a trait shared by both *A. pictum* and taxa within clade I). However, the abrupt transition from diploidy to hexaploidy (in clade II) seems to be unlikely and might require the existence of a transitional and yet extinct or undiscovered tetraploid form. The advantages of polyploids in terms of survival have been addressed recently in *Arum*, as artificial crossings between distinct species yielded polyploid hybrids that were 'robust and maintain themselves in cultivation without apparent difficulties' (Bedalov & Küpfer, 2005). Interestingly, the fact that clade V comprises both a substantial proportion of tetraploid lineages and an important number of recognized taxa could support the idea of an increased fitness in polyploids, facilitative for the radiation of this group (Fig. 3) (for a review of the ability of polyploids to colonize a wider range of habitats, see Prentis *et al*., 2008). However, before arriving at any conclusion, it is important that the phylogenetic relationships among the specimens of this clade are clarified.

One local phylogenetic constraint on the flower type (flag vs. cryptic) was addressed in the subclade corresponding to the *Tenuifila* subsection, with all taxa sharing a flag flower, whereas the ancestral state for this trait within *Arum* was a cryptic flower. As there is a strong association between cryptic flowers and attract-and-lure pollination strategies (Boyce, 1989; Gibernau *et al*., 2004), the latter should be considered as the ancestral pollination mode in the genus. It is important, however, to mention that the polyphyletic status of this character is not surprising, as it is related to reproductive structures, which, in *Arum*, appear to be highly correlated with fast-evolving pollination syndromes (Chouteau, Gibernau & Barabé, 2008). The evolution of this character would thus reflect more strongly the ecological processes that species have independently undergone rather than the evolutionary history of the genus.

Finally, our results confirm that the ancestral state of the *Arum* tuber shape was discoid, as proposed by Bedalov & Küpfer (2005), and that the appearance of the first rhizomatous species happened at the same time as the transition from diploidy to hexaploidy (clade II, Fig. 1). Although the transition from a discoid to a rhizomatous tuber occurred several times, there seems to be a trend towards a correlation between ploidy and tuber shape: all rhizomatous species are polyploid. In contrast, several polyploid species (*A. apulum*, *A. cyrenaicum* and *A. purpureospathum*) have discoid tubers. As sections within *Arum* were classically defined on the basis of this homoplasious character, there is a strong need to consider morphological characters from other plant parts to build a new classification of *Arum* compatible with our molecular evidence.

Arum through space and time

Bayes-DIVA provides strong support for an Aegean/Western European origin of the genus (Fig. 4) sometime in the early Miocene (*c.* 20 Mya). However, assuming that the earliest diverging lineage (now composed of only *A. pictum*) originated and survived in the Hercynian islands long before all other *Arum* spp. arose (according to the palaeorelictual hypothesis proposed by Mansion *et al*., 2008), the ancestral area corresponding to the rest of the genus is the Aegean region (Fig. 3). This zone has acted as a natural laboratory allowing the diversification of lineages, sometime in the late Miocene (Figs. 3, 4). The Aegean also appears to be a main source of dispersal events throughout the evolutionary history of *Arum*. Its central position with respect to the other areas in which the genus is present today could have facilitated this. Most dispersal events recorded during the middle to late Miocene occurred from the Aegean to the Anatolian region (Fig. 4A). Later, the emergence of the Iranian plate allowed its colonization once a landbridge was established with the Anatolian plate in the late Miocene (Meulenkamp *et al*., 2000). During this period, no dispersals were observed towards the Arabian plate that was still isolated from the northern lands by a marine transgression (Meulenkamp & Sissingh, 2003). Once the Caucasian archipelago emerged (and possibly after its uplift and contact with the Northern Anatolian region), further dispersals to

and from this region occurred in the late Miocene. At this same period, long-distance dispersals are recorded from the Aegean to the Macaronesian regions.

The first dispersal to the Arabian zone occurred more recently, during the Pliocene, when a sea regression (Meulenkamp & Sissingh, 2003) allowed this land to come into contact with the Anatolian region (Fig. 4B). The regression of the western Para-Tethys (following the uplift of the European plates) could also have permitted the dispersal from the Aegean to Eastern Europe. Exchanges continued between the Aegean and Anatolia and through these two zones to the Caucasus and Eastern Europe.

During more recent times (Pleistocene), new dispersals from the Aegean to Anatolia were recorded (Fig. 4C), probably facilitated by the Mediterranean regressions characterizing Quaternary climatic oscillations (Peulvast *et al*., 2000). At this time, North Africa was colonized twice through the Arabian plate and through the Apennines (Fig. 4C).

Although exchanges mainly occurred longitudinally (east–west) across land paths at the periphery of the seas during the early history of the genus, the pace of dispersion and diversification in *Arum* increased after the peri-Mediterranean region was unified (i.e. after the emergence of the Arabian and Iranian plates, the uplift of the Caucasus and the regressions of Tethys and Para-Tethys) (Meulenkamp *et al*., 2000; Meulenkamp & Sissingh, 2003).

The colonization of Macaronesia requires special treatment. The taxon inhabiting this region (*A. italicum* ssp. *canariense*) appears to have arrived there during the late Miocene or early Pliocene, in agreement with the timing of colonization already observed in several other endemic taxa of these islands (Carine *et al*., 2004). This ancient dispersal contrasts with the more recent colonization of North Africa. This discordance could be a result of either a first colonization of North Africa, having allowed the dispersal towards Macaronesia through mid-distance dispersal and further extinction of this lineage, or a long-distance dispersal directly from the Northern Peri-Tethys. Considering the morphology of the seeds (Mayo *et al*., 1997), this latter hypothesis could be possible only in association with animals; birds have already been proposed as the main dispersers of *A. maculatum* (Snow & Snow, 1988), which could also be true for *A. italicum* (Méndez, 1997).

Towards a new classification of *Arum*?

Because our phylogenetic reconstruction strongly contradicts the current systematics of the genus, the need for a new classification is evident (i.e. a large number of homoplasies are suggested by the tracing of the characters currently used in the delimitation of sections and species in our topology; Fig. 2). However, we recommend caution in formally proposing a new infrageneric classification until nonmolecular synapomorphies supporting the main clades are identified.

The two current subgenera, *Arum* and *Gymnomesium*, are supported by our analyses, although the status of the latter might be reconsidered, given the high level of phylogenetic differentiation of this monospecific subgenus. The characteristic morphology, development, distribution and, as shown in this study, phylogenetic position displayed by *A. pictum* could indicate that it would be more correct to place it in the monospecific genus *Gymnomesium* (as formerly proposed by Schott, 1855).

The two formerly defined sections within subgenus *Arum* are not supported by the phylogenetic analyses and, based on molecular evidence, we recommend a division of the subgenus into five sections (corresponding to clades I–V). Subsection *Poeciloporphyochiton* (corresponding to clade I) should be elevated to the rank of section, whereas new synapomorphies should be recovered for the other clades. The sectional classification of *A. creticum* and *A. idaeum* should also be investigated more thoroughly as our phylogenetic hypotheses only weakly associate them with clades III and IV, respectively. Although this study demonstrates the importance of ploidy as a putative synapomorphy in the case of clade II, a broad survey of morphological characters is strongly recommended.

Finally, the paraphyletic status of widespread species requires additional analyses to be performed with more variable markers in order to validate these findings. Nonetheless, our results already argue for a revision of species such as *A. maculatum* and *A. elongatum* in which major splits have been identified. Future taxonomic revisions should carefully consider characters not related to pollination, as lineages seem to be able to adapt quickly to changes in pollinator availability, leading to floral character convergence in distinct clades.

ACKNOWLEDGEMENTS

The authors would like to thank the different persons and institutions that kindly provided the material used in this study: the DNA Bank of the Royal Botanic Gardens, Kew (UK), as well as Dr Cusimano (University of Münich, Germany), Dr Lobim and Mr Neumann (Botanical Garden, University of Bonn, Germany). Discussions with Dr Bogner (Botanical Garden Munich, Germany) also largely contributed to the design of the manuscript. This study was funded by the Swiss National Science Foundation (project no. 3100A0-116778) and the Swiss Academy of Sciences (SCNAT+).

REFERENCES

Akaike H. 1973. Information theory and an extension of the maximum likelihood principle. In: Petrox BN, Caski F, eds. *2nd International Symposium on Information Theory*. Budapest: Akademiai Kiadó, 267–281.

Anderson C, Janssen T. 2009. Monocots. In: Hedges SB, Kumar S, eds. *The timetree of life*. Oxford: Oxford University Press, 203–212.

Baker RH, DeSalle R. 1997. Multiple sources of character information and the phylogeny of Hawaiian drosophilids. *Systematic Biology* **46:** 654–673.

Bedalov M. 1981. Cytotaxonomy of the genus *Arum* (Araceae) in the Balkans and the Aegean area. *Botanische Jahrbücher für Systematik, Pflanzengeschichte und Pflanzengeographie* **102:** 183–200.

Bedalov M, Küpfer P. 2005. Studies on the genus *Arum* (Araceae). *Bulletin de la Société Neuchâteloise des Sciences Naturelles* **128:** 43–70.

Borchsenius F. 2009. *Fastgap 1.2*. Distributed by the author. Aarhus, Denmark: Department of Biological Sciences, University of Aarhus. Available at http://192.38.46.42/aubot/fb/FastGap_home.htm (accessed 4 February 2010).

Boyce PC. 1989. A new classification of *Arum* with keys to the infrageneric taxa. *Kew Bulletin* **44:** 383–395.

Boyce PC. 1993. The genus *Arum.* A Kew magazine monograph. Royal Botanic Gardens, Kew, London. Norwich: HMSO Books.

Boyce PC. 1994. The *Arum* species of Greece and Cyprus. *Annales Musei Goulandris* **9:** 27–38.

Boyce PC. 2006. *Arum* – a decade of change. *Aroideana* **29:** 132–139.

Bremer K. 1988. The limits of amino acid sequence data in angiosperm phylogenetic reconstruction. *Evolution* **42:** 795–803.

Buerki S. 2009. Worldwide biogeography and systematics of Sapindaceae: a molecular and taxonomic survey combining large data sets and novel methodological approaches. Unpublished DPhil Thesis, University of Neuchâtel, Switzerland.

Cabrera LI, Salazar GA, Chase MW, Mayo SJ, Bogner J, Dávila P. 2008. Phylogenetic relationships of aroids and duckweeds (Araceae) inferred from coding and noncoding plastid DNA. *American Journal of Botany* **95:** 1153–1165.

Carine MA, Russell SJ, Santos-Guerra A, Francisco-Ortega J. 2004. Relationships of the Macaronesian and Mediterranean floras: molecular evidence for multiple colonizations into Macaronesia and back-colonization of the continent in *Convolvulus* (Convolvulaceae). *American Journal of Botany* **91:** 1070–1085.

CATE project. 2010. Creating a taxonomic science – January 2010; continuously updated. Available at http://www.cate-project.org (accessed 4 February 2010).

Chase MW, Fay MF, Devey DS, Maurin O, Rønsted N, Davies TJ, Pillon Y, Petersen G, Seberg O, Tamura MN, Asmussen CB, Hilu K, Borsch T, Davis JI, Stevenson DW, Pires JC, Givnish TJ, Sytsma KJ, McPherson MA, Graham SW, Rai HS. 2006. Multigene analyses of monocot relationships: a summary. *Aliso* **22:** 63–75.

Chouteau M, Gibernau M, Barabé D. 2008. Relationships between floral characters, pollination mechanisms, life forms and habitats in Araceae. *Botanical Journal of the Linnean Society* **156:** 29–42.

Estabrook GF. 1992. Evaluating undirected positional congruence of individual taxa between two estimates of the phylogenetic tree for a group of taxa. *Systematic Biology* **41:** 172–177.

Farris JS, Källersjö M, Kluge AG, Bult C. 1994. Testing significance of congruence. *Cladistics* **10:** 315–319.

Flora Europaea. 2009. *Flora Europaea database*. Edinburgh, UK: Royal Botanical Gardens Edinburgh. Available at http://rbg-web2.rbge.org.uk/FE/fe.html

Fuchs L. 1542. *Historia Stirpium*.

Gelman A, Rubin DB. 1992. Inference from iterative simulation using multiple sequences. *Statistical Science* **7:** 457–472.

Gibernau M, Macquart D, Przetak G. 2004. Pollination in the genus *Arum* – a review. *Aroideana* **27:** 148–166.

Hall TA. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series* **41:** 95–98.

Hruby J. 1910. Le genre *Arum*. Aperçu systématique avec considérations spéciales sur les relations phylogénétiques des formes. *Bulletin de la Société Botanique de Genève* **4:** 330–371.

Kluge AG. 1989. A concern for evidence and a phylogenetic hypothesis of relationships among *Epicrates* (Boidae, Serpentes). *Systematic Zoology* **38:** 7–25.

Linnaeus C. 1753. *Species plantarum 2*. Stockholm: 964–968.

Lobin W, Neumann M, Bogner J, Boyce PC. 2007. A new *Arum* species (Areae, Araceae) from NE Turkey and Georgia. *Willdenowia* **37:** 445–449.

Maddison WP, Maddison DR. 2009. *Mesquite: a modular system for evolutionary analysis. Version 2.6*. Distributed by the authors. Vancouver, BC, Corvallis, OR: University of British Columbia and Oregon State University. Available at http://mesquiteproject.org (accessed 4 February 2010).

Mansion G, Rosenbaum G, Schoenenberger N, Bacchetta G, Rosselló JA, Conti E. 2008. Phylogenetic analysis informed by geological history supports multiple, sequential invasions of the Mediterranean basin by the angiosperm family Araceae. *Systematic Biology* **57:** 269–285.

Mayo SJ, Bogner J, Boyce PC. 1997. *The genera of Araceae*. Kew: Royal Botanical Garden.

Médail F, Diadema K. 2009. Glacial refugia influence plant diversity patterns in the Mediterranean Basin. *Journal of Biogeography* **36:** 1–19.

Méndez M. 1997. Sources of variation in seed mass in *Arum italicum* (Araceae). *International Journal of Plant Sciences* **158:** 298–305.

Meulenkamp JE, Sissingh W. 2003. Tertiary palaeogeography and tectonostratigraphic evolution of the Northern and Southern Peri-Tethys platforms and the intermediate domains of the African–Eurasian convergent plate boundary zone. *Palaeogeography, Palaeoclimatology, Palaeoecology* **196:** 209–228.

Meulenkamp JE, Sissingh W, Calvo JP, Daams R, Londeix L, Cahuzak B, Kovac M, Marunteanu M, Ilynia LB, Khondkarian SO, Scherba IG, Roger J, Platel J-P, Hirsch F, Sadek A, Abdel-Gawad GI, Yeddi RS, Yaich C, Bouaziz S. 2000. Late Tortonian (8.4–7.2 Ma). In: Crasquin, S (coord.), *Atlas Peri-Tethys. Palaeogeographical maps. Explanatory notes*. Paris: CCGM-CGMW, 195–201.

Minorsky PV. 2003. The hot and the classic. *Plant Physiology* **132:** 1779–1780.

Moodie GEE. 1975. Heat production and pollination in Araceae. *Canadian Journal of Botany* **54:** 545–546.

Nixon KC. 1999. The parsimony ratchet, a new method for rapid parsimony analysis. *Cladistics* **15:** 407–414.

Nylander JAA. 2004. *Mraic.pl*. Distributed by the author. Uppsala, Sweden: Evolutionary Biology Centre, Uppsala University. Available at http://www.abc.se/~nylander (accessed 4 February 2010).

Nylander JAA, Olsson U, Alström P, Sanmartín I. 2008. Accounting for phylogenetic uncertainty in biogeography: a Bayesian approach to dispersal-vicariance analysis of the thrushes (Aves: *Turdus*). *Systematic Biology* **57:** 257–268.

Peulvast J-P, Vergne V, Vanney J-R, Barrier E. 2000. Last Glacial Maximum (late Pleistocene, 22000–18000 years BP). In: Crasquin, S (coord.), *Atlas Peri-Tethys. Palaeogeographical maps. Explanatory notes*. Paris: CCGM-CGMW, 209–219.

Prentis PJ, Wilson JRU, Dormontt EE, Richardson DM, Lowe AJ. 2008. Adaptive evolution in invasive species. *Trends in Plant Science* **13:** 288–294.

Prime CT. 1960. *Lords and ladies*. London: Collins.

R Development Core Team. 2009. *R: a language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

Rambaut A, Drummond AJ. 2004. *Tracer v1.3*. Distributed by the authors. Edinburgh, UK: Institute of Evolutionary Biology, University of Edinburgh. Available at http://tree.bio.ed.ac.uk/software/tracer (accessed 4 February 2010).

Ree RH, Sanmartín I. 2009. Prospects and challenges for parametric models in historical biogeographical inference. *Journal of Biogeography* **36:** 1211–1220.

Ronquist F. 1997. Dispersal-vicariance analysis: a new approach to the quantification of historical biogeography. *Systematic Biology* **46:** 195–203.

Ronquist F. 2001. *Diva 1.2*. Distributed by the author. Tallahassee, FL: School of Computational Science, Florida State University. Available at http://www.ebc.uu.se/systzoo/research/diva/diva.html (accessed 4 February 2010).

Ronquist F, Huelsenbeck JP. 2003. MRBAYES 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19:** 1572–1574.

Sanderson MJ. 2002. Estimating absolute rates of molecular evolution and divergence times: a penalized likelihood approach. *Molecular Biology and Evolution* **19:** 101–109.

Sanderson MJ. 2004. *R8s, program and documentation*. Distributed by the author. Tucson, AZ: Department of Ecology and Evolutionary Biology, University of Arizona. Available at http://loco.biosci.arizona.edu/r8s (accessed 4 February 2010).

Sanmartín I, Enghoff H, Ronquist F. 2001. Patterns of animal dispersal, vicariance and diversification in the Holarctic. *Biological Journal of the Linnean Society* **73:** 345–390.

Schott HW. 1832. Arum. In: Schott HW, Endlicher SL, eds. *Meletemata botanica. Arch. Bot. (Paris)* **1:** 246–252.

Schott HW. 1855. Gymnomesium. *Oesterreichische Botanische Wochenblatt* **5:** 17.

Shaw J, Lickey EB, Schilling EE, Small RL. 2007. Comparison of whole chloroplastic genome sequences to choose noncoding regions for phylogenetic studies in angiosperms: the tortoise and the hare III. *American Journal of Botany* **94:** 275–288.

Sikes DS, Lewis PO. 2001. *PAUPRat, Beta software, version 1: PAUP implementation of the parsimony ratchet*. Distributed by the authors. Storrs, CT: Department of Ecology and Evolutionary Biology, University of Connecticut. Available at http://users.iab.uaf.edu/~derek_sikes/software2.htm (accessed 4 February 2010).

Simmons MP, Ochoterena H. 2000. Gaps as characters in sequence-based phylogenetic analyses. *Systematic Biology* **49:** 369–381.

Snow BK, Snow DW. 1988. *Birds and berries*. Calton, UK: T. and A.D. Poyser.

Sorenson MD, Franzosa EA. 2007. *Treerot, version 3*. Distributed by the authors. Boston, MA: Boston University, Available at http://people.bu.edu/msoren/TreeRot.html (accessed 4 February 2010).

Swofford DL. 2002. PAUP*. Phylogenetic analysis using parsimony (*and other methods*)*. Version 4b10. Sunderland, MA: Sinauer Associates.

Theophrastus. 370 B.C. *The enquiry into plants*. English translation by Sir A. Hort, 1916. London: W. Heinemann.

NOTE ADDED IN PROOF

The authors of this study attest that at the time of the acceptance of our manuscript in *Botanical Journal of the Linnean Society* no phylogenetic studies of the genus *Arum* (Araceae) had been published. In the meantime, Linz *et al.* have published a phylogeny of *Arum* in the latest issue of *Taxon* (59: 405–415), i.e. two months before the release of our study. These studies were conducted independently and, although sharing the same group of interest, the sampling, approach and conclusions are consequently slightly different. For the sake of the understanding of the evolution of *Arum*, we encourage readers to compare the present study with that of Linz *et al.*

Linz J, Stökl J, Urru I, Krügel T, Stensmyr MC, Hansson BS. 2010. Molecular phylogeny of the genus *Arum* (Araceae) inferred from multi-locus sequence data and AFLPs. *Taxon* **59:** 405–415.

APPENDIX 1

Sample sources and GenBank accession numbers for all analysed samples. BG-Basle, Botanical Gardens, Basle, Switzerland; RBG-Kew, Royal Botanic Gardens, Kew, UK; pr. coll., private collection; NA, not available. Source information is provided in the following order: collector name, collection, voucher reference.

Sample	Source	Country of origin	GenBank accessions			
			3rps16-trnK	*ndhA*	*psbD-trnT*	*rpl32-trnL*
Arum apulum	Chase, RBG-Kew, 11022K	Italy	GU370965	GU371038	GU371101	GU371165
Arum balansanum1	Chase, RBG-Kew, 11009K	Turkey	GU370966	GU371049	GU371112	GU371174
Arum balansanum2	Küpfer, pr. coll., 09.06.01 4	Turkey	GU370967	GU371065	GU371128	GU371189
Arum balansanum3	Koenen, Haller, pr. coll., NA	Turkey	GU370968	GU371074	GU371138	GU371198
Arum besserianum	Bedalov, pr. coll., 1507	Ukraine	GU370969	GU371061	GU371124	GU371186
Arum byzantinum	Küpfer, pr. coll., 09.06.02 1	Turkey	GU370970	GU371063	GU371126	GU371188
Arum concinnatum	Chase, RBG-Kew, 11014K	Crete	GU370973	GU371060	GU371123	GU371185
Arum creticum	Chase, RBG-Kew, 11037K	Crete	GU370974	GU371068	GU371132	GU371193
Arum cylindraceum1	Chase, RBG-Kew, 11013K	Romania	GU370982	GU371028	GU371091	GU371155
Arum cylindraceum2	Espíndola & Revel, pr. coll., cylGIO1	Italy	GU370975	GU371080	GU371144	GU371204
Arum cylindraceum3	Espíndola & Revel, pr. coll., cylGYL3	Denmark	GU370976	GU371081	GU371145	GU371205
Arum cylindraceum4	Espíndola & Revel, pr. coll., cylMUS4	Poland	GU370977	GU371082	GU371146	GU371206
Arum cylindraceum5	Espíndola & Revel, pr. coll., cylPAD5	Romania	GU370978	GU371083	GU371147	GU371207
Arum cylindraceum6	Espíndola & Revel, pr. coll., cylSIS5	France	GU370979	GU371084	GU371148	GU371208
Arum cylindraceum7	Espíndola & Revel, pr. coll., cylTRA4	Austria	GU370980	GU371085	GU371149	GU371209
Arum cylindraceum8	Espíndola & Revel, pr. coll., cylTRP3	Macedonia	GU370981	GU371086	GU371150	GU371210
Arum cylindraceum9	Neumann, pr. coll., I21/05	Sicily	GU371026	GU371073	GU371137	–
Arum cyrenaicum	Chase, RBG-Kew, 11030K	Libya	GU370983	GU371072	GU371136	GU371197
Arum dioscoridis ssp. *cyprium*	Chase, RBG-Kew, 11021K	Turkey	GU370985	GU371077	GU371141	GU371201
Arum dioscoridis1	Chase, RBG-Kew, 11015K	Turkey	GU370986	GU371078	GU371142	GU371202
Arum dioscoridis2	Küpfer, pr. coll., 09.05.30 4	Turkey	GU370984	GU371064	GU371127	–
Arum elongatum1	Boyce, RBG-Kew, 1990-2019	Turkey	GU370989	GU371079	GU371143	GU371203
Arum elongatum2	Bedalov, pr. coll., 16.5.03	Ukraine	GU370990	–	GU371131	GU371192
Arum elongatum3	Küpfer, pr. coll., 09.05.09 1	Turkey	GU370991	GU371069	GU371133	GU371194
Arum euxinum	Chase, RBG-Kew, 11019K	Turkey	GU370992	GU371029	GU371092	GU371156
Arum hygrophilum	Chase, RBG-Kew, 11027K	Israel	GU370993	GU371030	GU371093	GU371157
Arum idaeum	Boyce, RBG-Kew, 1993-1895	Crete	GU370994	GU371031	GU371094	GU371158
Arum italicum ssp. *albispathum*	Chase, RBG-Kew, 11020K	Georgia	GU370995	GU371035	GU371098	GU371162
Arum italicum ssp. *canariense1*	Chase, RBG-Kew, 11032K	Madeira	GU370996	GU371032	GU371095	GU371159
Arum italicum ssp. *canariense2*	Chase, RBG-Kew, 11031K	Madeira	GU370997	GU371034	GU371097	GU371161
Arum italicum ssp. *italicum*	Boyce, RBG-Kew, 1978-4984	Greece	GU370998	GU371033	GU371096	GU371160
Arum jacquemontii	Boyce, RBG-Kew, 1969-5385	Afghanistan	GU370999	GU371036	GU371099	GU371163
Arum korolkowii	Boyce, RBG-Kew, 1994-3354	Tadzhikistan	GU371000	GU371037	GU371100	GU371164
Arum lucanum1	Boyce, RBG-Kew, 1987-1133	Italy	GU371001	GU371039	GU371102	GU371166
Arum lucanum2	Boyce, RBG-Kew, 1991-887	Italy	GU371002	GU371051	GU371114	GU371176
Arum maculatum1	Chase, RBG-Kew, 11161K	England	GU371003	GU371040	GU371103	GU371167
Arum maculatum2	NA, RBG-Kew, MJC 0002	England	GU371004	GU371041	GU371104	GU371168
Arum maculatum3	Boyce, RBG-Kew, 1990-475	Wales	GU371005	GU371042	GU371105	GU371169
Arum maculatum4	Boyce, RBG-Kew, 1990-2018	Turkey	GU371006	GU371043	GU371106	GU371170
Arum maculatum5	Bedalov & Küpfer, pr. coll., 1914b	Crimea	GU371007	GU371066	GU371129	GU371190
Arum maculatum6	Espíndola & Zryd, pr. coll., macKIL3	England	GU371008	GU371087	GU371151	GU371211
Arum maculatum7	Espíndola & Zryd, pr. coll., macLAC2	Italy	GU371009	GU371088	GU371152	GU371212
Arum maculatum9	Espíndola & Zryd, pr. coll., mac SFG7	France	GU371011	GU371090	GU371154	GU371214
Arum megobrebi1	Bedalov, pr. coll., 1432	Georgia	GU371012	GU371058	GU371121	GU371183
Arum megobrebi2	Bedalov, pr. coll., 1436	Georgia	GU371013	GU371059	GU371122	GU371184
Arum megobrebi3	Bedalov, pr. coll., 1442	Georgia	GU371014	GU371071	GU371135	GU371196
Arum megobrebi4	Neumann, pr. coll., 24219	Georgia	GU371015	GU371075	GU371139	GU371199
Arum nigrum	Boyce, RBG-Kew, 1992-2083	Montenegro	GU371016	GU371044	GU371107	GU371171
Arum orientale ssp. *longispathum*	Bedalov & Küpfer, pr. coll., 13.05.02 98/15	NA	GU371018	GU371070	GU371134	GU371195
Arum orientale1	Boyce, RBG-Kew, NA	NA	GU371019	GU371045	GU371108	GU371172
Arum orientale2	Bedalov & Küpfer, pr. coll., 284 99/61	Macedonia	GU371017	GU371067	GU371130	GU371191
Arum palaestinum	Chase, RBG-Kew, 11016K	NA	GU371020	GU371046	GU371109	GU371173
Arum pictum	Chase, RBG-Kew, 11024K	Balearic	GU371021	GU371047	GU371110	–
Arum purpureospathum	Chase, RBG-Kew, 11023K	Crete	GU371022	GU371048	GU371111	–
Arum rupicola ssp. *rupicola*	Chase, RBG-Kew, 11036K	Jordan	GU371023	GU371052	GU371115	GU371177
Arum rupicola ssp. *virensens*	Chase, RBG-Kew, 11034K	Turkey	GU371024	GU371050	GU371113	GU371175
Arum sintenisii	NA, BG-Basle, NA	NA	GU371025	GU371062	GU371125	GU371187
Arum sp. nov.	Neumann, pr. coll., 26940	Russia	GU371027	GU371076	GU371140	GU371200
Biarum davisii	NA, RBG-Kew, 2002-2839	NA	GU370971	GU371057	GU371120	GU371182
Biarum dispar	NA, RBG-Kew, 1991-123	Italy	GU370972	GU371056	GU371119	GU371181
Dracunculus canariensis	NA, RBG-Kew, 1982-2008	NA	GU370987	GU371053	GU371116	GU371178
Dracunculus vulgaris1	NA, RBG-Kew, 86 – 3894	NA	–	GU371054	GU371117	GU371179
Dracunculus vulgaris2	NA, RBG-Kew, NA	England	GU370988	GU371055	GU371118	GU371180