Molecular Phylogenetics and Evolution 75 (2014) 91–102

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

journal homepage: www.elsevier.com/locate/ympev

Phylogenomics of the plant family Araceae

Claudia L. Henriquez [a,b,*], Tatiana Arias [d], J. Chris Pires [c], Thomas B. Croat [b], Barbara A. Schaal [a]

[a] Evolution, Ecology & Population Biology, Division of Biology, Washington University in St. Louis, One Brookings Drive, St. Louis, MO 63130, USA
[b] Missouri Botanical Garden, 4344 Shaw Blvd., St. Louis, MO 63110, USA
[c] Department of Biological Sciences, University of Missouri, Columbia, MO 65211, USA
[d] School of Biological Sciences, The University of Hong Kong, Pokfulam Road, Hong Kong

ARTICLE INFO

Article history:
Received 26 July 2013
Revised 15 February 2014
Accepted 18 February 2014
Available online 1 March 2014

Keywords:
Araceae
Massively parallel sequencing
Clade
Phylogeny

ABSTRACT

The biogeography, chromosome number evolution, pollination biology and evolutionary history of the plant family Araceae have recently become much clearer (Cabrera et al., 2008; Chartier et al., 2013; Cusimano et al., 2011, 2012; Nauheimer et al., 2012). However, phylogenetic ambiguity near the root of the tree precludes answering questions about the early evolution of the family. We use Illumina sequencing technology and reference based assembly to resolve the remaining questions in the deep phylogeny of Araceae. We sampled 32 genera and obtained 7 from GenBank (including an outgroup), representing 42 of 44 major clades described in Cusimano et al. (2011). A subsequent phylogenomic analysis based on mitochondrial data was performed to test congruence between plastid and mitochondrial data for phylogenetic inference. Plastid sequences produced strongly supported phylogenies. In contrast, mitochondrial phylogenies were weakly supported and incongruent with chloroplast data (Templeton test, $p \leqslant 0.0001$), although several smaller clades were recovered. New strongly-supported clades seen here are: (1) *Anubias* and *Montrichardia*, excluding *Calla*, form a clade that is sister to the *Zantedeschia* clade; (2) the South African genus *Zantedeschia* is sister to the Old World *Anchomanes* clade; and (3) within the *Zantedeschia* clade, *Philodendron* is sister to the rest. *Calla* and *Schismatoglottis* form a clade at the base of one of two major clades in Aroideae based on complete chloroplast sequences. Although statistical support is weak, morphological and cytological features support this topology.

1. Introduction

Araceae, or the Arum family, is a large and ancient monocot plant family most notable for its impressive morphological diversity, including the smallest known angiosperm and some of the largest vegetative and reproductive structures in the world (Simpson, 2006). The family consists of c. 3800 species in 118 genera, distributed mostly in the tropics but can range into temperate and, in the case of *Calla palustris*, circumboreal regions (Boyce and Croat, 2013; Ulrich et al., 2013). Members of Araceae occupy a wide array of ecological habitats from sea level to above 3000 m and range from submerged, emergent or free-floating aquatics, to epiphytic, climbing and terrestrial plants (Bown, 2000; Cabrera et al., 2008; Croat, 1988; Gonçalves, 2004; Gonçalves et al., 2007). Stems can be rhizomatous, cormose, tuberous or reduced to a thallus-like structure and leaves can be simple, highly divided or fenestrate (Mayo et al., 1997; Simpson, 2006). Araceae are distinguished from closely related families in having a great diversity of calcium oxalate crystals (raphides, druses, crystal sand, styloids and prismatics), possessing a spadix of small, bisexual or unisexual flowers, subtended by a spathe, and they lack ethereal oil cells (Grayum, 1990; Keating, 2003; Stevens, 2001 onwards).

Detailed classification of Araceae, established as a family in 1789 (Jussieu, 1789), began in the nineteenth century with the work of Heinrich Wilhelm Schott (1794–1865) and Adolf Gustav Engler (1844–1930). Schott's pre-Darwinian classification grouped genera based on inflorescences, flowers and fruits (Mayo et al., 1997; Nicolson, 1987). A modified version of this classification was used by Hooker (1883) who divided Araceae into 11 tribes, and later by Hutchinson (1973) who divided the family into 18 tribes (Grayum, 1990; Hooker, 1883; Hutchinson, 1973). Engler's new system of classification, which included hypotheses of evolutionary transitions of not only floral, but also of vegetative morphological and anatomical characters (Engler, 1920, Mayo et al., 1997, Nicolson, 1987), has been the framework for much subsequent work (Bogner, 1978; Bogner and Nicolson, 1991; Grayum, 1990; Hotta, 1970, Mayo et al., 1997, Nakai, 1943). Grayum's

* Corresponding author at: Evolution, Ecology & Population Biology, Division of Biology, Washington University in St. Louis, One Brookings Drive, Campus Box 1137, St. Louis, MO 63130, USA.
E-mail address: chenriquez@wustl.edu (C.L. Henriquez).

http://dx.doi.org/10.1016/j.ympev.2014.02.017
1055-7903/

1990 revision, based on a large survey of palynological characters, is notable in recognizing *Acorus* as separate from all other Araceae (Grayum, 1987, 1990).

Since the chloroplast restriction site data of French et al. (1995), molecular data have been used to infer evolutionary relationships at all levels in the family (Barabé et al., 2002; Cabrera et al., 2008; Chartier et al., 2013; Cusimano et al., 2011; Gauthier et al., 2008, Gonçalves et al., 2007, Nauheimer et al., 2012; Renner et al., 2004; Renner and Zhang, 2004; Rothwell et al., 2004; Tam et al., 2004; Wong et al., 2010). To date, the most comprehensive family-wide molecular data set consists of six chloroplast (*rbcL*, *matK*, partial *trnK* intron, partial *tRNA-Leu* gene, *trnL-trnF* spacer, and partial *tRNA-Phe* gene) and one nuclear (*PhyC*) markers (Cabrera et al., 2008; Chartier et al., 2013), and has been used to clarify the evolutionary history, biogeography, pollination biology and chromosomal evolution of Araceae (see also Cusimano et al., 2011; Nauheimer et al., 2012). Araceae has an inferred ancestral haploid chromosome number of $n = 16$ or $n = 18$, and began to diversify in the Early Cretaceous, approximately 122 Mya, as the breakup of Pangea was finalizing (Cusimano et al., 2012; Nauheimer et al., 2012). By the Cretaceous/Paleogene boundary all eight of the currently recognized subfamilies, including the duckweed subfamily Lemnoideae, were present and form a clade that is sister to a clade comprising all other members of the order Alismatales (Cabrera et al., 2008; Nauheimer et al., 2012; Tobe and Kadokawa, 2010). Evolutionary relationships among six of the eight subfamilies (Gymnostachydoideae, Orontioideae, Lemnoideae, Pothoideae, Monsteroideae and Lasioideae), all of which contain bisexually-flowered members, are well-supported (Cabrera et al., 2008; Cusimano et al., 2011; Nauheimer et al., 2012). The Unisexual Flowers clade, containing subfamilies Zamioculcadoideae and the highly diverse Aroideae (1573 species, 75 genera), diverged from the bisexual-flowered lineage during the Late Cretaceous approximately 90 Mya (Nauheimer et al., 2012). Low resolution of several deep nodes in the phylogeny of the Unisexual Flowers clade leaves open several important questions, including the position of the highly autapomorphic, bisexually-flowered genus *Calla* (Cabrera et al., 2008; Chartier et al., 2013; Cusimano et al., 2011; Ulrich et al., 2013). Although *Calla* is well-supported in a clade with two unisexually-flowered genera (*Montrichardia* and *Anubias*) in the nuclear tree from Chartier et al. (2013), the position of that clade at the base of Aroideae is not strongly supported and biogeographical and morphological features make this grouping dubious. *Calla* has spirally arranged perfect flowers that emerge acropetally, disulcate pollen, an inferred ancestral haploid chromosome number of $n = 18$ and a circumboreal, mainly European geographical distribution (Chartier et al., 2013; Stevens, 2001 onwards, Ulrich et al., 2013). *Anubias* is an African genus and *Montrichardia* is South American, but both share an inferred ancestral haploid chromosome number of $n = 12$. The only feature shared by all three is a helophytic habit, which occurs elsewhere in the family (Chartier et al., 2013; Cusimano et al., 2011; Grayum, 1990). Another generic placement that warrants further investigation is the weakly-supported sister relationship of the South African genus *Zantedeschia* ($n = 16$) with the strictly South American tribe Spathicarpeae ($n = 17$) (Cabrera et al., 2008; Chartier et al., 2013; Cusimano et al., 2011; Nauheimer et al., 2012). In addition, weakly-supported relationships among the smaller clades within the *Zantedeschia* clade are in need of further clarification.

With the advent of massively parallel sequencing, phylogenetic analyses can now be based on tens of thousands of nucleotides, which can greatly enhance our confidence in the resulting evolutionary hypotheses (Givnish et al., 2010; Steele et al., 2012; Xi et al., 2012). Sequencing of plastomes and mitogenomes includes de novo assemblies using complete genomic DNA and reference-based assemblies using DNA enriched for chloroplasts, and combinations thereof (Givnish et al., 2010; Steele et al., 2012). In addition to variation in the proportion of organellar and nuclear DNA used in creating libraries for sequencing, the suite of genomic tools now available to process and analyze the resulting deluge of genomic data permits the use of multiple software programs to corroborate results.

Phylogenomic studies in plants have generally focused on the chloroplast genome, whereas the mitochondrial genome, due to its complicated mutational dynamics, has been more commonly used in studies of structural variation, nucleotide substitution rates and horizontal gene transfer (Knoop et al., 2011; Mower et al., 2007; Palmer et al., 2000; Richards et al., 2009; Richardson et al., 2013; Xi et al., 2012). The low silent-site substitution rate of plant mitochondrial DNA, which has been shown to be one-third less than that of plant chloroplast DNA, plus the extensive RNA-editing and retroprocessing that occurs in this genome perhaps explain why, in plant phylogenetic studies, mitochondrial regions have typically been used in combination with plastid regions (Renner and Zhang, 2004; Seberg and Petersen, 2006; Seberg et al., 2012; Steele et al., 2012; Wolfe et al., 1987). In addition, previous studies have shown that phylogenies reconstructed from mitochondrial data are less resolved and incongruent with plastid data (Petersen et al., 2006, 2013). However, slow silent substitution rates are not consistent across the entire mitochondrial genome or among all plant lineages; mitochondrial genes from highly divergent plant genera have been shown to have substitution rates similar to that of the rapidly evolving mammalian mitochondrial genome (Mower et al., 2007; Palmer et al., 2000). The question remains whether large-scale datasets based on tens to hundreds of thousands of aligned nucleotides, representing both coding and non-coding regions, from the mitochondrial genome possess enough phylogenetic signal to resolve evolutionary relationships in plants at the family level.

Here we use Illumina sequencing technology with total genomic DNA and reference-based assembly of the chloroplast, using the programs Geneious 6.0.3 and Bowtie2, to resolve some of the major remaining questions in the current phylogeny of Araceae. A subsequent phylogenomic analysis of mitochondrial sequences obtained from reference-based assembly was performed to compare congruence of the mitochondrial phylogeny with the plastid phylogeny.

2. Materials and methods

2.1. Taxon sampling

For the chloroplast analysis, we sampled 32 genera of Araceae and obtained from GenBank the complete, annotated chloroplast genomes of 5 additional genera: *Colocasia esculenta* (Ahmed et al., 2012), *Lemna minor* (Mardanov et al., 2008), *Wolffiella lingulata*, *Wolffia australiana* and *Spirodela polyrhiza* (Wang and Messing, 2011). We included at least one representative from 42 of the 44 clades of Araceae named in Cusimano et al. (2011). For a list of genera included in this study and the higher taxa they represent refer to Table 1. The two taxa not sampled were Cryptocoryneae and Culcasieae, although larger clades within which they are nested were sampled; these are the Rheophytes clade and the *Homalomena* clade, respectively. Of the 11 phylogenetically isolated genera in Cusimano et al. (2011) (*Calla*, *Callopsis*, *Montrichardia*, *Anubias*, *Zantedeschia*, *Philonotion*, *Protarum*, *Pistia*, *Alocasia*, *Pinellia* and *Arisaema*), 4 genera (*Callopsis*, *Philonotion*, *Protarum*, *Pistia*) were not sampled. *Gymnostachys anceps*, the sole member of subfamily Gymnostachydoideae, was not sampled but its sister relationship with subfamily Orontioideae, here represented by *Orontium*, has been strongly supported in other studies (Cabrera et al., 2008;