Molecular Phylogenetics and Evolution 53 (2009) 726–733

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

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

Flanking regions of monomorphic microsatellite loci provide a new source of data for plant species-level phylogenetics

Lars W. Chatrou [a,*], M. Pilar Escribano [b], Maria A. Viruel [b], Jan W. Maas [c], James E. Richardson [d], José I. Hormaza [b]

[a] Nationaal Herbarium Nederland, Wageningen Branch, and Wageningen UR, Biosystematics Group, Generaal Foulkesweg 37, 6703 BL, The Netherlands
[b] Estación Experimental La Mayora-CSIC, Algarrobo-Costa, Málaga 29750, Spain
[c] Nationaal Herbarium Nederland, Utrecht Branch, Heidelberglaan 2, 3584 CS Utrecht, The Netherlands
[d] Royal Botanic Garden Edinburgh, 20A Inverleith Row, Edinburgh EH3 5LR, United Kingdom

ARTICLE INFO

Article history:
Received 12 January 2009
Revised 16 July 2009
Accepted 23 July 2009
Available online 29 July 2009

Keywords:
Microsatellite flanking regions
Species-level phylogenetics
Neutral evolution

ABSTRACT

Well-resolved phylogenetic trees are essential for us to understand evolutionary processes at the level of species. The degree of species-level resolution in the plant phylogenetic literature is poor, however, largely due to the dearth of sufficiently variable molecular markers.

Unlike the common genic approach to marker development, we generated DNA sequences of monomorphic nuclear microsatellite flanking regions in a phylogenetic study of *Annona* species (Annonaceae). The resulting data showed no evidence of paralogy or allelic diversity that would confound attempts to reconstruct the species tree. Microsatellite flanking regions are short, making them practical to use, yet have astounding proportions of variable characters. They have 3.5- to 10-fold higher substitution rates compared to two commonly used chloroplast markers, have no rate heterogeneity among nucleotide positions, evolve in a clock-like fashion, and show no evidence of saturation. These advantages are offset by the short length of the flanking regions, resulting in similar numbers of parsimony informative characters to the chloroplast markers.

The neutral evolution and high variability of flanking regions, together with the wide availability of monomorphic microsatellite loci in angiosperms, are useful qualities for species-level phylogenetics. The general methodology we present here facilitates to find phylogenetic markers in groups where microsatellites have been developed.

1. Introduction

Increased focus on species-level phylogenetics in angiosperms has encouraged the pursuit of molecular markers that are capable of resolving phylogenetic relationships at lower taxonomic levels, i.e. have a mutation rate that is fast enough to produce sufficient variation (Crawford and Mort, 2004). The need for such markers resonates within the literature (Bailey et al., 2004; Choi et al., 2006; Crawford and Mort, 2004; Whittall et al., 2006), as only a small percentage of the published species phylogenetic trees in plants are fully resolved (Hughes et al., 2006).

Chloroplast markers have been an important source of data for plant phylogenetics. Apparent advantages of the application of chloroplast markers are its relative abundance in plant total DNA and the relatively conservative mutation rates, facilitating extraction and amplification using conserved primer binding sites. Furthermore, chloroplast markers are essentially single-copy. This avoids the reconstruction of erroneous organismal phylogenies due to the application of paralogous gene copies, which may be a problem when applying nuclear markers (Bailey et al., 2003; Baker et al., 2000; Sanderson and Shaffer, 2002).

The features that simplify the application of cpDNA markers at the species level, are however traded off against less desirable qualities for organismal species-level phylogenetics (Sang, 2002). Chloroplast markers generally evolve at rates that are too slow to provide sufficient phylogenetically informative characters over recent time spans (Richardson et al., 2001), even after considerable data collection (Perret et al., 2003; Pirie et al., 2006). It is not to say that not fully resolved phylogenies are meaningless. As long as critical nodes are well-supported they can serve to pinpoint biogeographical phenomena (e.g. Erkens et al., 2007b), or to falsify current classification based on morphological characters (e.g. Shaw and Small, 2004). Only few papers with fairly large chloroplast data sets have generated reasonably well-resolved and robustly

* Corresponding author. Fax: +31 317 484917.
E-mail addresses: Lars.Chatrou@wur.nl (L.W. Chatrou), pilescri@eelm.csic.es (M.P. Escribano), marian.viruel@eelm.csic.es (M.A. Viruel), J.W.Maas@uu.nl (J.W. Maas), J.Richardson@rbge.ac.uk (J.E. Richardson), ihormaza@eelm.csic.es (J.I. Hormaza).

1055-7903/$ - see front matter © 2009 Elsevier Inc. All rights reserved.
doi:10.1016/j.ympev.2009.07.024

supported species-level phylogenies (e.g. Clarkson et al., 2004). Furthermore, chloroplast markers are uniparentally inherited, usually maternally in angiosperms, and therefore provide only part of the evidence for the evolutionary development of a lineage if hybridization and introgression have taken place (Chase et al., 2003; Sunnucks, 2000; Vriesendorp and Bakker, 2005).

The search for useful markers for plant species-level systematics has predominantly yielded markers from genic regions, or, in the case of noncoding DNA, at short distances from genic regions. There is a growing body of literature on single- or low-copy nuclear genes that provide sufficient informative characters and do not complicate organismal phylogeny reconstruction with paralogous copies (Edwards et al., 2008; Emshwiller and Doyle, 1999; Sang et al., 1997; Small et al., 2004; Whittall et al., 2006). However, it has also become clear that rates of nucleotide substitution of these markers may differ significantly among lineages, and even among closely related species (Hughes et al., 2006). Therefore, attempts to extrapolate the utility of these markers for resolving species-level relationships outside the taxonomic group for which they have been developed may not succeed. These difficulties have led some researchers to suggest that a universal approach should be abandoned in favor of a lineage-specific one (Small et al., 2004).

However, an alternative to a gene-based approach to the development of variable nuclear markers involves search strategies that focus on randomly amplified regions throughout the genome (Bailey et al., 2004; Hughes et al., 2006). The high variability, abundance, uniform and genome-wide distribution, and neutral evolution of one of these, viz. microsatellites (Ellegren, 2004), make them potentially useful at the species level. However, their polymorphic nature brings about analytical problems, related to the translation of allele sizes to distance-based characters that are susceptible to incorrect homology assessment (Matsuoka et al., 2002; Primmer and Ellegren, 1998). We avoid this drawback by only focusing on the nucleotide sequences of the flanking regions alongside the microsatellite repeat region, not on the repeat region. A further factor possibly complicating phylogeny reconstruction is the presence of multiple alleles, i.e. of variation that does not necessarily have a one-to-one relationship to the organismal phylogeny, for example because of incomplete lineage sorting. The distinction between paralogous and orthologous microsatellite copies is less of a complicating factor. Microsatellites, including the flanking regions, usually represent unique and therefore orthologous loci (Sunnucks, 2000), although duplication events affecting microsatellite loci have been reported (Antunes et al., 2006; Zhang and Rosenberg, 2007).

Thus, an optimal microsatellite flanking region marker for plant species-level phylogenetics has a rate of substitution that allows resolving shallower relationships, is represented by orthologous copies, and is monomorphic within species, populations and individuals. We present examples of such markers from the plant family Annonaceae, and outline the potential for the broader applicability of this approach in other clades of angiosperms. We have taken orthology-by-default of as a starting point of our study, further hypothesizing that a neutral, highly variable marker system such as microsatellites, including the flanking regions, evolves at a fast enough rate to elucidate relationships among species of *Annona* (Annonaceae). In this plant family, the application of chloroplast sequence data has produced phylogenies that are poorly resolved at the species level, despite the gathering of large amounts of data (Erkens et al., 2007a; Mols et al., 2004). *Annona* is paraphyletic with respect to *Rollinia*, and the two genera together comprise a clade of approximately 175 species. Species of *Rollinia* were synonymized into *Annona* recently (Rainer, 2007). Here we provide the first published phylogenetic support for this taxonomic decision, as the former species of *Rollinia* (*Annona cuspidata*, *A. herzogii*, *A. mucosa*, *A. neochrysocarpa*) appear as a well-supported clade within *Annona* from the analyses we present here. Although our taxon sampling reflects one-eighth of the species diversity in *Annona*, covering the entire morphological diversity as well as the geographical distribution of the genus, additional taxon sampling would be needed to confidently corroborate the inclusion of *Rollinia* into *Annona*.

To assess the utility of microsatellite flanking regions, we need to address the following issues: (1) Can we produce flanking region sequences that are monomorphic within individuals? (2) Can we confirm the assumed orthology of the flanking region sequences? (3) What is the transferability of microsatellite regions across species of *Annona* and other Annonaceae? (4) What is the strength of the phylogenetic signal at the species level?

2. Materials and methods

2.1. Taxon sampling

For this study we sampled 24 species: 22 species of *Annona* and two species of *Asimina* as outgroup species (Table 1). Richardson et al. (2004) have shown that *Asimina* is sister to *Annona*. The samples of *Annona* represent the complete geographical range of the genus, as well as the considerable morphological (particularly floral) variation.

2.2. Character sampling

Two chloroplast markers, *rbcL* and *trnLF*, were sequenced. These markers have commonly been applied in phylogenetic analyses of Annonaceae (e.g. Couvreur et al., 2008; Erkens et al., 2007a; Pirie et al., 2006). These markers are generally considered to be useful at taxonomic levels above that of species. In a family-wide analysis the relationships among nine species of *Annona* were fairly well resolved but generally poorly supported, based on these two chloroplast markers only (Richardson et al., 2004). The microsatellite loci were selected based on a screening with the first 15 microsatellite loci that were developed in cherimoya (*Annona cherimola*) (Escribano et al., 2004). Seven of them (LMCH4, 5, 6, 9, 10, 11, and 14) produced amplification bands in the eight species studied initially (*Annona* sp. nov., *A. glabra*, *A. montana*, *A. muricata*, *A. oligocarpa*, *A. reticulata*, *A. senegalensis*, and *Rollina cuspidata* [now *Annona cuspidata*, Rainer, 2007]). No amplification was obtained for two loci (LMCH1 and LMCH13). For two additional loci (LMCH7 and LMCH8) amplification was obtained only with *A. montana* and *A. glabra*. LMCH9 and LMCH10 were selected for this study because they showed clear and monomorphic single-allele amplification bands in all the species tested.

2.3. DNA extraction, PCR amplification and sequencing

Total genomic DNA was extracted following a protocol adapted from the CTAB method (Doyle and Doyle, 1987), as described in Pirie et al. (2006). PCR conditions and primers for the chloroplast markers were standard, and are identical to Pirie et al. (2006). PCR products were purified using QIAquick PCR purification kits (Qiagen), and sequenced with the PCR primers. PCR conditions and primers for the microsatellites LMCH9 and LMCH10 are according to Escribano et al. (2004).

PCR products were resolved in 3% high-resolution agarose (Metaphor, FMC Bioproducts, Rockland, ME) gel electrophoresis in SB buffer at 5 V/cm.

Sequencing reactions had a total volume of 10 μl contained 0.5 μl DYEnamic ET Terminator (Amersham Pharmacia Biotech), 3.5 μl ET Terminator dilution buffer (Amersham Pharmacia Biotech), and 2–4 μl of DNA template, depending on the