Molecular Phylogenetics and Evolution 87 (2015) 14–27

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

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

A multi-locus analysis of phylogenetic relationships within grass subfamily Pooideae (Poaceae) inferred from sequences of nuclear single copy gene regions compared with plastid DNA

Anne Hochbach, Julia Schneider, Martin Röser *

Martin Luther University Halle-Wittenberg, Institute of Biology, Geobotany and Botanical Garden, Neuwerk 21, 06108 Halle (Saale), Germany

ARTICLE INFO

Article history:
Received 24 September 2014
Revised 6 March 2015
Accepted 12 March 2015
Available online 21 March 2015

Keywords:
*mat*K
Nuclear single copy
Phylogeny
Poaceae
Pooideae
Topo6

ABSTRACT

To investigate phylogenetic relationships within the grass subfamily Pooideae we studied about 50 taxa covering all recognized tribes, using one plastid DNA (cpDNA) marker (*mat*K gene–3′*trn*K exon) and for the first time four nuclear single copy gene loci. DNA sequence information from two parts of the nuclear genes topoisomerase 6 (*Topo6*) spanning the exons 8–13 and 17–19, the exons 9–13 encoding plastid acetyl-CoA-carboxylase (*Acc1*) and the partial exon 1 of phytochrome B (*PhyB*) were generated. Individual and nuclear combined data were evaluated using maximum parsimony, maximum likelihood and Bayesian methods. All of the phylogenetic results show *Brachyelytrum* and the tribe Nardeae as earliest diverging lineages within the subfamily. The 'core' Pooideae (Hordeeae and the Aveneae/Poeae tribe complex) are also strongly supported, as well as the monophyly of the tribes Brachypodieae, Meliceae and Stipeae (except *PhyB*). The beak grass tribe Diarrheneae and the tribe Duthieeae are not monophyletic in some of the analyses. However, the combined nuclear DNA (nDNA) tree yields the highest resolution and the best delimitation of the tribes, and provides the following evolutionary hypothesis for the tribes: *Brachyelytrum*, Nardeae, Duthieeae, Meliceae, Stipeae, Diarrheneae, Brachypodieae and the 'core' Pooideae. Within the individual datasets, the phylogenetic trees obtained from *Topo6* exon 8–13 shows the most interesting results. The divergent positions of some clone sequences of *Ampelodesmos mauritanicus* and *Trikeraia pappiformis*, for instance, may indicate a hybrid origin of these stipoid taxa.

1. Introduction

Together with bamboos (subf. Bambusoideae) and rice-like grasses (subf. Ehrhartoideae) the subfamily Pooideae is a part of the major BEP clade within the Poaceae [cf. Grass Phylogeny Working Group (GPWG), 2001]. Pooideae encompass about one third of all grasses with approximately 3300 species in 150 genera (Clayton and Renvoize, 1986). The genera are distributed predominately in the cool temperate grasslands of both hemispheres and include some of the most important cereals such as wheat (*Triticum*), rye (*Secale*), barley (*Hordeum*) oat (*Avena*) and also forage plants (Clayton and Renvoize, 1986).

Traditional classifications of the subfamily Pooideae based on morphological and anatomical characters show no consistent results in terms of the number and delimitation of the tribes and the systematic position of several genera (e.g. Clayton and Renvoize, 1986; Macfarlane and Watson, 1980, 1982; Tzvelev, 1976, 1989; Watson and Dallwitz, 1992, 1992 onwards). The 'core' Pooideae [Bromeae, Hordeeae (syn. Triticeae), Aveneae and Poeae] described by Davis and Soreng (1993) are uniquely placed to this subfamily, whereas for example the tribes Ampelodesmeae, Anisopogoneae, Brachyelytreae, Diarrheneae, Lygeeae, Nardeae and Stipeae were treated under a separate subfamily Stipoideae (Watson and Dallwitz, 1992 onwards) or were partly classified under subfamily Arundinoideae (Watson and Dallwitz, 1992). Furthermore the tribes Brachyelytreae, Diarrheneae and Phaenospermateae were classified as members of the subfamily Bambusoideae (Clayton and Renvoize, 1986).

Phylogenetic studies based on structural characters and/or molecular data suggest a broader definition of the subfamily Pooideae including all these tribes (e.g. Bouchenak-Khelladi et al., 2008; Catalán et al., 1997; Davis and Soreng, 2007; Döring et al., 2007; Duvall et al., 2007; Hilu et al., 1999; Hsiao et al., 1999; Schneider et al., 2009, 2011; Soreng and Davis, 1998, 2000) and the recently described tribe Duthieeae (Schneider et al., 2011).

The majority of phylogenetic markers used in grasses and angiosperms in general are from the plastid genome and/or nuclear

* Corresponding author.
E-mail addresses: anne.hochbach@botanik.uni-halle.de (A. Hochbach), julia.schneider@botanik.uni-halle.de (J. Schneider), martin.roeser@botanik.uni-halle.de (M. Röser).

http://dx.doi.org/10.1016/j.ympev.2015.03.010
1055-7903/

ribosomal internal transcribed spacer (nrITS; Duarte et al., 2010 and references therein). Currently there are no investigations using nuclear single or low copy genes to study subfamily Pooideae. Only Mathews et al. (2000) and GPWG (2001) analyzed nuclear single copy genes (*PhyB* and GBSSI) for the whole grass family, but with a comparatively small set of pooid taxa. However, there are several investigations on grasses at the tribe or genus level (Brassac et al., 2012; Fan et al., 2007, 2009; Hand et al., 2010; Huang et al., 2002; Jakob and Blattner, 2010; Mason-Gamer, 2005, 2013; Sha et al., 2010; Triplett et al., 2012).

In this study, individual analyses of five DNA loci in a nearly congruent taxon set were used to address questions about the major phylogenetic structure of the subfamily Pooideae, the delimitation or definition of the tribes and even subtribes, as well as the utility of these markers for phylogenetic investigations. The chloroplast *mat*K gene–3′*trn*K exon region or a part of it has already been proven to be meaningful in several previous phylogenetic analyses of grasses (Blaner et al., 2014; Döring et al., 2007; Hand et al., 2010; Hilu et al., 1999; Schneider et al., 2009, 2011, 2012; Soreng et al., 2007) and was used here for comparative purposes. Two of the four nuclear single copy loci are part of the gene topoisomerase 6 (*Topo6*) subunit B, which is an important house-keeping gene and involved in the repair of double-strand breaks (Hartung et al., 2002). This gene contains a relatively high number of variable introns between 19 conserved exons and was used in previous investigations in barley, the Aveneae/Poeae tribe complex and the whole grass family (Blaner et al., 2014; Brassac et al., 2012; Jakob and Blattner, 2010; Wölk and Röser, 2014). In this study, the *Topo6* loci spanning the exons 8–13 and 17–19 were selected for phylogenetic analyses. The first contains a balanced distribution of exons and introns, whereas the second is characterized by a very high proportion of variable introns. Another nuclear gene in this study is the protein-coding gene *Acc1* (encoding plastid acetyl-CoA carboxylase), which proved to be suitable for phylogenetic analyses of the genus *Panicum* (Triplett et al., 2012), the tribe Hordeeae (Fan et al., 2007, 2009; Huang et al., 2002; Sha et al., 2010) and within the Aveneae/Poeae tribe complex (Hand et al., 2010). An approximately 800 bp region spanning exons 9–13 with a suitable share of exons and introns was selected for this study. Additionally we analyzed nuclear DNA sequences obtained from partial phytochrome B (*PhyB*). The *PhyB* gene is present in all seed plants and as photoreceptive signaling protein involved in light-sensitive processes such as germination control in the dark (Ludeña et al., 2011, and references therein). The used *PhyB* locus comprises a large part of conserved exon 1, which already found use across the grass family (Mathews et al., 2000). Several previous studies suggested that phylogenetic accuracy benefits from additionally included characters or genomic regions (Crawley and Hilu, 2012; Knoop and Müller, 2009). Therefore, all nDNA loci were additionally analyzed in a concatenated matrix.

2. Material and methods

2.1. Taxon sampling

In total, 331 sequences of 53 species and 5 different DNA loci were included in this study. *Fargesia nitida* and *Pleioblastus fortunei* from the related subfamily Bambusoideae were selected as outgroup taxa. Representatives of all the tribes of subfamily Pooideae acknowledged by GPWG (2001) and Schneider et al. (2009, 2011) were analyzed in this investigation. The molecular analyses were conducted on silica gel-dried leaf material collected in the field from living plants or on leaves from herbarium specimens. Information of origin, specimens voucher and ENA/GenBank sequence accession numbers of the analyzed taxa is given in Table 1.

2.2. Molecular methods

An amount of 20–45 mg leaf tissue was homogenized using Qbiogene FastPrep FP120 cell disrupter (Heidelberg, Germany). Genomic DNA was extracted with the NucleoSpin Plant Kit according to the manufacturer's protocol (Macherey–Nagel, Düren, Germany). DNA concentrations were checked with a NanoDrop spectrophotometer (Thermo Fisher Scientific, Waltham, USA).

In this study the plastid *mat*K gene–3′*trn*K exon and the four nuclear single copy gene loci *Topo6* exon 8–13, *Topo6* exon 17–19, *Acc1* exon 9–13 and *PhyB* were PCR-amplified. Primers used to amplify and subsequently sequence these gene regions are listed in Table 2 and the binding positions are illustrated in Fig. 1. PCR was carried out on an Eppendorf (Hamburg, Germany) mastercycler, following the protocol detailed by Schneider et al. (2009). Amplification of the *mat*K gene–3′*trn*K exon was performed for 3 min at 94 °C; followed by 30 s at 94 °C, 1 min at 50 °C and 2 min at 72 °C in a total of 34 cycles and a finishing stage at 72 °C for 10 min. The parameters for the two *Topo6* regions were: 3 min at 94 °C, followed by 35 cycles of 30 s at 94 °C, 1–3 min at 50 °C and 2–5 min at 68 °C, with a final step at 68 °C for 10 min. The PCR program for the amplification using *Acc1* primers was 3 min at 94 °C, followed by 35 cycles of 30 s at 94 °C, 2 min at 50 °C–54 °C, 1 min at 72 °C and the final extension at 72 °C for 10 min. *PhyB* settings were: 3 min at 94 °C; 35 cycles of 30 s at 94 °C, 30 s–3 min at 54–61 °C, 45 s–5 min at 72 °C and 10 min at 72 °C.

Amplicons were purified by using the NucleoSpin Extract II Kit (Macherey–Nagel) and served as template for the sequencing reaction conducted by StarSEQ (Mainz, Germany) or Eurofins MWG Operon (Ebersberg, Germany). Additionally, untreated PCR products were sent to LGC Genomics (Berlin, Germany) for purification and sequencing.

2.3. Cloning of PCR products

Amplification products with ambiguous peaks or DNA concentrations too low for direct sequencing had to be cloned. Cloning of PCR products was performed using the pGEM-T Easy Vector System of Promega (Mannheim, Germany). Ligation and transformation were carried out according to the technical manual. The plasmid DNA was isolated using the GeneJET Plasmid Miniprep Kit (Fermentas, St. Leon-Rot, Germany) according to the manufacturer's protocol and the sequencing reaction was carried out by Eurofins. Alternatively, the Eurofins Prepaid Plate Kit for plasmid clones was used for plasmid DNA isolation and sequencing.

2.4. Phylogenetic analyses

Chromatograms of forward and reverse strands were edited and aligned manually using Sequencher 4.6 (Gene Codes Corporation, Ann Arbor, Michigan, USA). Sequence positions for which alignment was considered ambiguous were coded according to the IUPAC Code. Consensus sequences for highly similar clone sequences were created to reduce the number of singletons in the alignment, because the differences are very likely PCR errors. Ambiguous alignment regions were excluded from the analyses (3. Results, Table 3). Potentially parsimony-informative indels were coded as binary matrix following the coding method of Simmons and Ochoterena (2000) and gaps were treated as missing data.

The phylogenetic analyses were run with MrBayes v.3.2.4 (Bayesian inference [BI]; Huelsenbeck and Ronquist, 2001; Ronquist et al., 2012), with raxmlGUI (Maximum Likelihood [ML]; Silvestro and Michalak, 2012) and with PAUP* 4.0b10 (Maximum Parsimony [MP]; Swofford, 2002) separately for each of the five loci dataset. The four regions of the nuclear genome