



Evolution of the *beta-amylase* gene in the temperate grasses: Non-purifying selection, recombination, semiparalogy, homeology and phylogenetic signal [☆]



Miguel Minaya ^a, Antonio Díaz-Pérez ^{a,b}, Roberta Mason-Gamer ^c, Manuel Pimentel ^d, Pilar Catalán ^{a,e,*}

^a Departamento de Ciencias Agrarias y del Medio Natural (Botánica), Escuela Politécnica Superior-Huesca, Universidad de Zaragoza, Ctra Cuarte km 1, 22071 Huesca, Spain

^b Instituto de Genética y Centro de Investigaciones en Biotecnología Agrícola (CIBA), Facultad de Agronomía, Universidad Central de Venezuela, 2101 Maracay, Aragua, Venezuela

^c Department of Biological Sciences, University of Illinois at Chicago, USA

^d Grupo de Investigación en Biología Evolutiva (GIBE), Facultad de Ciencias, Universidade da Coruña, Campus da Zapateira sn., 15071 A Coruña, Spain

^e Department of Botany, Institute of Biology, Tomsk State University, Lenin Av. 36, Tomsk 634050, Russia

ARTICLE INFO

Article history:

Received 18 February 2015

Revised 8 May 2015

Accepted 10 May 2015

Available online 29 May 2015

Keywords:

β-amylase phylogeny

Low copy nuclear genes

Phylogenetic incongruence

Pooideae

Pseudogenization

Purifying and diversifying selection

ABSTRACT

Low-copy nuclear genes (LCNGs) have complex genetic architectures and evolutionary dynamics. However, unlike multicopy nuclear genes, LCNGs are rarely subject to gene conversion or concerted evolution, and they have higher mutation rates than organellar or nuclear ribosomal DNA markers, so they have great potential for improving the robustness of phylogenetic reconstructions at all taxonomic levels. In this study, our first objective is to evaluate the evolutionary dynamics of the LCNG *β*-amylase by testing for potential pseudogenization, paralogy, homeology, recombination, and phylogenetic incongruence within a broad representation of the main Pooideae lineages. Our second objective is to determine whether *β*-amylase shows sufficient phylogenetic signal to reconstruct the evolutionary history of the Pooideae grasses. A multigenic (ITS, *matK*, *ndhF*, *trnTL*, and *trnLF*) tree of the study group provided a framework for assessing the *β*-amylase phylogeny. Eight accessions showed complete absence of selection, suggesting putative pseudogenic copies or other relaxed selection pressures; resolution of *Vulpia alopecuroides* 2x clones indicated its potential (semi) paralogy; and homeologous copies of allopolyploid species *Festuca simensis*, *F. fenas*, and *F. arundinacea* tracked their Mediterranean origin. Two recombination events were found within early-diverged Pooideae lineages, and five within the PACCMAD clade. The unexpected phylogenetic relationships of 37 grass species (26% of the sampled species) highlight the frequent occurrence of non-treelike evolutionary events, so this LCNG should be used with caution as a phylogenetic marker. However, once the pitfalls are identified and removed, the phylogenetic reconstruction of the grasses based on the *β*-amylase exon + intron positions is optimal at all taxonomic levels.

© 2015 Elsevier Inc. All rights reserved.

1. Introduction

Phylogenies based on molecular data help us to understand a variety of genomic evolutionary phenomena such as pseudogenization, gene duplication (paralogy), lineage sorting, recombination, transposition, and horizontal gene transfer (Mason-Gamer and Kellogg, 1997; Soreng and Davis, 2000; Catalán et al., 2004; Minaya et al., 2013). These evolutionary events can result in a

phylogenetic history that is more reticulate than bifurcating (cf. Kellogg, 2006), contradicting the premise that individual gene trees represent the true phylogenies of the organisms (cf. Linder and Rieseberg, 2004; Feliner and Rossello, 2007).

The inclusion of a mixture of orthologous and paralogous sequences in phylogenetic analyses can produce robust, yet erroneous, hypotheses of relationships (Sanderson and Doyle, 1992; Wendel and Doyle, 1998). Although a number of criteria have been suggested as evidence of orthology (e.g. Doyle, 1991; Doyle and Doyle, 1999; Cronn et al., 2002), the simplest approach to identifying orthologs is through sequence similarity and phylogenetic analyses. Orthologs are expected to be more similar and phylogenetically more closely related to each other than to their paralogs, and consequently, orthologs must form monophyletic clades and conform to generally accepted species trees (Buckler et al., 1997;

[☆] This paper was edited by the Associate Editor Stefanie M. Ickert-Bond.

* Corresponding author at: Escuela Politécnica Superior de Huesca, Universidad de Zaragoza, Ctra. Cuarte km 1, 22071 Huesca, Spain. Fax: +34 974239302.

E-mail addresses: miguel.minaya@gmail.com (M. Minaya), antonio.dz.pz@gmail.com (A. Díaz-Pérez), robie@uic.edu (R. Mason-Gamer), mpimentel@udc.es (M. Pimentel), pcatalan@unizar.es (P. Catalán).

Small et al., 2004). Like divergent paralogs, recombinants can also compromise phylogeny reconstruction (Sanderson and Doyle, 1992; Buckler et al., 1997). Recombination generates mosaic genes, where different regions have different phylogenetic histories (Doyle, 1996), especially when it involves non-orthologs or occurs across species boundaries. Thus, it violates the assumption, underlying many methods of phylogenetic analysis, that phylogenies can be represented as strictly bifurcating trees. For these reasons, the accurate detection and elimination of paralogous and recombinant sequences before phylogenetic reconstruction is highly relevant (cf. Buckler et al., 1997; Posada and Crandall, 2001).

The extent to which selection contributes to fixation, maintenance, or elimination of nucleotide mutations has been a long-standing problem in understanding molecular sequence evolution since the proposal of the neutral theory (Kimura, 1968). Here we investigate selection site-by-site, using the generalized branch-site methods of Kosakovsky-Pond et al. (2011) and fitting a distribution of substitution rates across sites (random effect model = REL; Nielsen and Yang, 1998), thus inferring the rate at which individual sites evolve given this distribution (Kosakovsky-Pond and Frost, 2005). Theory predicts that the number of substitutions leading to adaptive amino acid changes (non-synonymous mutations, d_N) should be significantly larger than the number of substitutions at synonymous sites (d_S) (Miyata and Yasunaga, 1980; Li, 1997). Conversely, if selection consistently weeds out amino acid changes, then d_N should be significantly smaller than d_S . Thus, an estimated d_N/d_S value near 1 suggests neutrality, and a d_N/d_S greater than or less than 1 indicates positive or negative selection, respectively (reviewed by Yang and Bielawski, 2000; Preston and Kellogg, 2006; Brunner et al., 2009).

The grass family (Poaceae) is one of the largest and most diverse families of the angiosperms, with approximately 10,000 species and 600–700 genera (Clayton and Renvoize, 1986; Watson and Dallwitz, 1992). Previous evolutionary studies of the Poaceae (e.g., GPWG, 2001, 2012; Davis and Soreng, 2007; Bouchenak-Khelladi et al., 2008; Schneider et al., 2009) have provided valuable insights into the evolutionary structure of grass genomes and their phylogeny. However, major phylogenetic relationships remain to be solved, especially at the tribal and generic levels. Low copy nuclear genes (LCNGs), unlike multicopy nuclear genes (e.g. ribosomal DNA genes: ITS, ETS; Alvarez and Wendel, 2003; Feliner and Rossello, 2007), are rarely subject to gene conversion or concerted evolution. Furthermore, they show higher sequence variation than organellar or rDNA genes, and have a great potential for improving the robustness of phylogenetic reconstructions at all taxonomic levels (cf. Sanderson and Doyle, 1992; Sang and Zhang, 1999; Sang, 2002; Small et al., 2004; Hughes et al., 2006). However, the real utility of LCNGs for phylogenetic reconstruction has been less encouraging than their potential advantages would predict. This is because of their evolutionary dynamics. LCNGs experience gene duplication and recombination (Wendel and Doyle, 1998; Small et al., 2004; Feliner and Rossello, 2007) and can be transferred among taxa through hybridization or other horizontal gene transfer events (e.g., Ghatnekar et al., 2006; Vallenback et al., 2008, 2010; Mahelka and Kopecký, 2010; Brassac et al., 2012). More practical issues include the lack of universal PCR primers and often-limited understanding of copy numbers and evolutionary rates.

The low-copy β -amylase gene (1-4- α -glucan maltohydrolyase) has rarely been used as a phylogenetic marker in angiosperms (Rajakpase et al., 2004), but a few studies have used it for phylogenetic reconstruction of some grass lineages (e.g. Wang et al., 1997; Ziegler, 1999; Mason-Gamer, 2005). This led to our interest in characterizing the molecular evolution and phylogeny of the β -amylase gene in the temperate grasses. The specific aims of this study were to: (i) explore the extent of purifying and diversifying

selection across the sampled β -amylase grass sequences; (ii) evaluate pseudogenization, paralogy, homeology, recombination, and phylogenetic discordance of β -amylase gene sequences within a broad representation of the main Pooideae lineages, with special emphasis on the Poeae – Aveneae and Loliinae lineages; (iii) examine phylogenetic bias among different sets of coding and non-coding positions [exon + intron positions, exons, first and second codon positions, third codon positions, and third codon positions + introns] by comparing each phylogenetic inference with a cpDNA-ITS tree of the Poaceae; and (iv) reconstruct the phylogenetic relationships of the study group based on the best-resolved and most congruent data set. Finally, we assess the potential for similar evolutionary events in other lineages of the grass family, and the suitability of this gene for the reconstruction of the Poaceae tree, based on our large sampling of outgroup (non-Poeae – Aveneae) lineages.

2. Material and methods

2.1. Taxon sampling and DNA analysis

The taxon sampling represents the main Pooideae lineages and close outgroups, with special emphasis on the Poeae – Aveneae (Supplementary Table S1). Sampling included 142 species classified in 88 genera and 7 subfamilies (GPWG, 2001, 2012; Soreng et al., 2007, and <http://www.tropicos.org/>): Centothecoideae: 1 species, Panicoideae: 6, Danthonioideae: 7, Chloridoideae: 10, Ehrhartoideae: 2, Bambusoideae: 1, and Pooideae: 115 (Lygeae/Nardeae: 2, Stipeae: 3, Diarrheneae: 1, Brachypodieae: 3, Bromaeae: 1, Triticeae: 25, Poeae – Aveneae: 80). Four of the six subfamilies of the outgroup PACMAD clade are sampled in this study (all but Arundinoideae and Micrairoideae). Amplified β -amylase sequences from 16 *Brachypodium* and Loliinae samples were cloned to check for potential paralogous and heterologous copies: *Brachypodium distachyon* (9 clones), *Festuca abyssinica* (2), *F. altaica* (7), *F. arundinacea* (7), *F. capillifolia* (9), *F. elegans* (7), *F. fenas* (7), *F. hystrix* (5), *F. paniciana* (2), *F. paniculata* (10), *F. rubra* (5), *F. scabra* (5), *F. scariosa* (8), *F. simensis* (7), *Micropyropsis tuberosa* (2), *Vulpia alopecuroides* (5). These taxa previously showed paralogous copies in other LCNG in Loliinae and close allies [GBSSI (Granule-Bound-Starch-Synthase I); Díaz-Pérez et al., 2014] and *Brachypodium* [CAL (Calmodulin-binding protein), DGAT (Diacyl Glycerol Acyl Transferase), GI (Gigantea); Catalán et al., 2012]. The origins of the plants studied, the locations of voucher specimens, and the GenBank sequence accession numbers are listed in Table S1. *Joinvillea ascendens* (Joinvilleaceae) was selected to root the tree because it is closely related to the grasses and clearly positioned in Poales (Campbell and Kellogg, 1987; Linder and Rudall, 1993).

Doyle and Doyle's (1987) CTAB method was used to isolate DNA from silica-gel-dried leaves and from fresh materials for most of the studied samples. For herbarium samples, DNA was extracted using the DNeasy® Plant Mini Kit (QIAGEN Ltd., West Crawley, UK) procedure. A 2370-nucleotide region (Fig. S1), extending from exons 2 to 5 of the β -amylase gene was amplified and sequenced using the primers 2a-for, 2b-for, 2c-for, 2d-for, 2g-for, 3a-for, 3a-bac, 4a-bac, 4b-for, 5a-bac, and 5b-bac, as in Mason-Gamer (2005). Procedures for DNA amplification, cloning and sequencing are indicated in Appendix 1.

The β -amylase sequences were aligned with Protein Multiple Sequence Alignment Software version 3.7 (MUSCLE) (Edgar, 2004). The alignments were manually adjusted using MacClade 4.08 OS X (Maddison and Maddison, 2008). Amino acid translations were used to guide the nucleotide alignments. While alignment was generally straightforward in the coding regions, there were

some intron regions where patterns of length variation made homology assessment difficult. We also found a Miniature Inverted repeat Transposable Element (MITE) in intron 4, which included highly variable and non-homologous transposable elements and their footprints (Minaya et al., 2013). Seven regions with ambiguous alignments and/or MITEs were excluded from subsequent analyses [Supplementary Fig. S1: intron 2 (positions 160–319, 329–370); intron 3 (560–740, 784–821, 847–880, 976–1269); intron 4 (1602–2083)].

2.2. Species tree reconstruction

A multigenic cpDNA-ITS tree of the Poaceae was used to assess the phylogenetic distribution of paralogous copies, recombinant sequences, and the existence of phylogenetic bias between the β -amylase reconstructions based on five different data subsets (see below). The Poaceae cpDNA-ITS tree (cf. Minaya et al., 2013) was constructed using the multicopy ribosomal *ITS1-5.8S-ITS2* region and the plastid *matK*, *ndhF*, *trnTL*, and *trnLF* regions from a collection of 142 grass taxa plus the close relative *Joinvillea* (Joinvillaceae) included as an outgroup (Table S1). Published primers and protocols were used to amplify and sequence the ITS, *trnTL*, and *trnLF* regions (Catalán et al., 2004; Quintanar et al., 2007), the *ndhF* gene (Catalán et al., 1997), and the *matK* gene (Döring et al., 2007). Character conflict was assessed for each pair of data sets using the Partition Homogeneity (PH) test (Incongruence Length Difference of Farris et al., 1994), conducted in PAUP* 4.0 beta 10 (Swofford, 2002) using heuristic searches of 100 random-order-entry replicates, with TBR and MulTrees ON.

Bayesian and Maximum parsimony (MP) based searches were performed for both the separate and the combined data sets using, respectively, MrBayes 3.2.2 (Altekar et al., 2004; Ronquist et al., 2011) and PAUP* v4.10b (Swofford, 2002) (procedures are provided in Appendix 2). All gaps were treated as missing data. Branch support was indicated on the combined tree with parsimony bootstrap (BS) and Bayesian posterior probability (PP) values.

2.3. Detection of potential β -amylase paralogues, phylogenetic conflict, recombinants, and non-purifying sequences or pseudogenes

Paralogous gene copies were inferred using the tree-based method of Altenhoff and Dessimoz (2012), from confirmed diploid accessions showing clones in two divergent positions of a phylogenetic tree. This method, which also allowed us to identify accessions with unexpected placement, is based on reconciling a species tree (e.g., multigenic cpDNA-ITS tree of the Poaceae) and the β -amylase gene tree, and by annotating all splits of the β -amylase tree as either duplications or speciation events, according to the placement and support of the cloned β -amylase sequences. Once the splits of the gene tree have been classified as speciation or duplication events, it is simple to infer whether the given accession shows orthologous or paralogous sequences, based on where the split between its gene copies appear. Since ambiguously aligned intron positions could lead to erroneous topologies, the β -amylase gene tree was based on a 588-bp matrix that included the exon positions of 142 grass species (totaling 223 grass sequences, including all the direct sequences and the clones; Table 1, β -amylase data set 1). We further used the species-overlap rule (Gabalión, 2008), a complementary method to corroborate potential β -amylase paralogy. According to Gabalión (2008), a node represents a duplication event if it is ancestral to a tree bifurcation (e.g. sister clades) that contains sets of sequences from the same species or accessions that overlap to some degree. Conversely, if the two partitions contain sets of species that are mutually exclusive, the node is considered to represent a speciation event.

The Recombination Detection Program (RDP4) version 4.16 (Martin et al., 2010) was used to analyze potential recombination events within the β -amylase sequences. The RDP, GENECONV, BootScan, MaxChi, Chimaera, SiScan, LARD, and 3SEQ methods implemented in RDP4 were used to identify recombinant and parental sequences and to calculate probability scores for potential recombination events. We used the default settings in all cases. Only potential recombinant events detected by at least two methods were considered significant. The Bonferroni multiple comparison correction test was performed to diminish the expected number of false positive results. Masking of similar sequences was allowed to increase the power of the recombination detection methods. For these analyses we used a matrix with only exon positions. This matrix included all Poaceae sequences sampled (142 grass species); however, to discard spurious variation from Polymerase Chain Reaction (PCR) artifacts, the cloned sequences from 16 species were collapsed into a single consensus sequence (or type) from closely related sequences following Díaz-Pérez et al. (2014). All intraspecific sequences with a p -distance lower than 0.01 (base differences per site) were collapsed into a single type using the program BioEdit v7.2.5 (Hall, 1999) (Table 1, β -amylase data set 2; 146 sequences, 588 positions, exons 2–5).

Two complementary criteria were used to detect pseudogenic copies, and one of them also detected potential types of selection pressure on the β -amylase sequences. First, we checked for the existence of Premature Terminal Codons (PTCs) among the coding positions of the 223 sampled β -amylase sequences (Table 1, β -amylase data set 1). According to Akhunov et al. (2013) some potential pseudogenic copies could accumulate mutations and eventually produce PTCs at the early stages of evolution.

Secondly, we assumed that most of the coding positions in a functional protein are constrained by purifying selection, rather than positive selection (Kimura, 1968), while pseudogenes should show neutral variation. Purifying and diversifying selection were investigated through the analysis of distribution of substitution rates across sites and branches under a Random Effect Likelihood (REL) model (Nielsen and Yang, 1998; Kosakovsky-Pond and Frost, 2005), where the distribution is represented by discrete classes defined by specific selection pressure ranges ($\omega = d_N/d_S$; d_N and d_S representing nonsynonymous/synonymous substitutions per nonsynonymous/synonymous sites, respectively). REL models assume that substitutions along a branch of a phylogenetic tree are characterized by a continuous-time stationary Markov process, defined by its instantaneous rate matrix (Q). Similar to nucleotide substitution models, the likelihood of a change along a branch is given by the transition matrix. Considering that selective pressures may vary over both sites and time, branch-site REL methods are able to estimate the proportion of sites with different ω values for all lineages of the tree (Kosakovsky-Pond et al., 2011). In this case, each branch-site combination has instantaneous rate and transition matrices. We restricted our analysis to terminal lineages to evaluate recent pressures associated with sequences from a subset of 65 BEP species, including all putative sequences under likely paralogy, homeology, recombination, and topological incongruence (Table 1, β -amylase data set 3; 69 sequences, 588 positions, exons 2–5). This was done to detect sequences with high proportions of sites under neutral or nearly neutral selection, indicating potential relaxed selection pressure (represented by $\omega = d_N/d_S \approx 1$) and evidence of pseudogenization. We performed the Branch-Specific Branch-Site REL (BS-REL) model, as suggested in Kosakovsky-Pond et al. (2008), to estimate β -amylase lineage-specific proportion of sites (Pr1, Pr2 and Pr3) showing rates circumscribed to three discrete classes ($\omega_1 \leq \omega_2 \leq 1 \leq \omega_3$), respectively. These classes are associated with the neutral model; classes ω_1 and ω_2 are less than or equal to 1, specifying stronger and weaker purifying selection, whereas ω_3 (unconstrained class)

Table 1

β -amylase data partitions used in this research. Each data set is enumerated and described according to their appearance in the main text.

β -amylase data partition	n° grass species/ sequences	Number of nucleotide positions	Characteristics	Analysis performed
1. Exon positions	142/223	588	Including all species and clones sampled	<ul style="list-style-type: none"> • Compared to the reference tree to find paralogous copies and topologically incongruent sequences • Check the existence of PTCs among the coding positions • Bayesian phylogeny (Fig. S2)
2. Exon positions	142/146	588	Including: <ul style="list-style-type: none"> • All species sampled. • Cloned sequences collapsed into a consensus type when their p-distance was lower than 0.01 	Recombination (RDP4) (Table 2)
3. Exon positions	65/69	588	Selection of representative sequences classified within the BEP clade, including: <ul style="list-style-type: none"> • Topologically incongruent accessions and accessions under potential paralogy, homeology, or recombination. • Cloned sequences collapsed into a consensus type when their p-distance was lower than 0.01. 	Branch-Site REL (DataMonkey) (Table 3)
4. Exon + intron positions	142/146	850	Including: <ul style="list-style-type: none"> • All species sampled. • Cloned sequences collapsed into a consensus type when their p-distance was lower than 0.01. 	<ul style="list-style-type: none"> • Bayesian phylogeny (Fig. 2) • NeighborNet partition network tree (Fig. 3)
5. Exon + intron positions	105/105	850	<ul style="list-style-type: none"> • Excluding 37 species (41 accessions including the collapsed cloned sequences) under unexpected placement, potential pseudogenization, semi-paralogy, homeology, or recombination (Fig. 2). 	<ul style="list-style-type: none"> • SH test (Table 4) • Bayesian phylogeny (Fig. S3)
5a. Exon positions	105/105	588		<ul style="list-style-type: none"> • Comparison of the Bayesian topology (Fig. 4)
5b. First and second codon positions	105/105	392	<ul style="list-style-type: none"> • Cloned sequences were collapsed into a consensus type when their p-distance was lower than 0.01. 	<ul style="list-style-type: none"> • PH and SH tests
5c. Third codon positions	105/105	196		<ul style="list-style-type: none"> • Comparison of the Bayesian topology • PH and SH tests
5d. Third codon positions + unambiguous introns	105/105	456		<ul style="list-style-type: none"> • Comparison of the Bayesian topology • PH and SH tests

is greater than 1, and is related to adaptive selection. According to the estimated ω_1 , ω_2 , ω_3 values for each branch, we assigned each of these values and its proportion to one out of five selection scenarios: (I) $\omega = 0-0.6$ (strong negative selection), (II) $\omega = 0.61-0.9$ (weak negative selection), (III) $\omega = 0.91-1.5$ (neutral selection), (IV) $\omega = 1.51-5$ (weak positive selection), and (V) $\omega > 5$ (strong positive selection). BS-REL was performed using the maximum likelihood-based tools implemented in the WEB interface DataMonkey (Pond and Frost, 2005).

2.4. Reconstruction of β -amylase phylogeny

A Bayesian tree and a NeighborNet partition network (Bryant and Moulton, 2004; Huson et al., 2010) were computed based on a β -amylase 850-bp data matrix that included the exon + unambiguous intron positions and 146 sequences (from 142 species) after collapsing the cloned sequences into a consensus type (Table 1, β -amylase data set 4). The Bayesian phylogeny and the NeighborNet partition network tree provide a graphical representation of phylogenetic relationships, including sequences with apparently anomalous placements on the tree (Huson and Bryant, 2006). These anomalies could have originated from incomplete lineage sorting, horizontal gene transfer, recombination, or hybridization. Within the NeighborNet partition network analysis, a pairwise p -distance matrix representing the proportion of base differences between sequences was computed first. Then, the NeighborNet algorithm was employed to generate circular partitions which were subsequently processed by the Equal Angle algorithm (Bryant and Moulton, 2004; Huson et al., 2010). Partition statistical support was generated through 1000 bootstrap pseudoreplicates. All analyses were performed using Splitstree version 4.12.3 (Huson and Bryant, 2006).

Once we had detected and excluded pseudogenes, paralogs, recombinants, and sequences that were unexpectedly placed relative to the cpDNA-ITS tree (37 total; see results for more details), we reconstructed the β -amylase phylogeny based on a 850-bp matrix with exons plus unambiguous intron positions from 105 species, including the different types of collapsed clones (Table 1, β -amylase data set 5). Because one of the aims of this study is to examine phylogenetic bias among β -amylase reconstructions based on different sets of coding and non-coding positions, the data matrix (β -amylase data set 5 in Table 1) was subdivided into four additional data matrices: (5a) only coding positions (588 bp); (5b) only first and second codon positions (392 bp); (5c) only third codon positions (196 bp); and (5d) third codon positions plus unambiguous intron positions (456 bp) (Table 1, β -amylase data sets 5a–5d).

The Partition Homogeneity (PH) test (Incongruence Length Difference (ILD) test of Farris et al., 1994), and the more conservative Shimodaira–Hasegawa (SH) nonparametric bootstrap test (Shimodaira and Hasegawa, 1999) were conducted on the β -amylase data set 5 (Table 1) and its four subdivisions (Table 1: β -amylase data sets 5a, 5b, 5c, 5d). Both tests were implemented in PAUP* v4.10b (Swofford, 2002). The PH test was used to assess character conflict among subsets 5a–5d, using heuristic searches of 100 random-order-entry replicates, with TBR and MulTrees ON. The SH test was used to determine whether the topology of the Bayesian tree derived from data set 5 (Table 1), which corresponds to the most likely tree (Section 3), was significantly different from the Bayesian trees derived from its four subdivisions (data sets 5a to 5d). The SH test was done with resampling estimated log-likelihood (RELL) optimization and 1000 bootstrap replicates. Since comparing more than two trees leads to a multiple comparison problem that cannot be solved by Bonferroni

corrections (cf. Shimodaira and Hasegawa, 1999), we ran the SH test comparing only two trees each time.

The independent Bayesian searches conducted for each of seven β -amylase data sets (Table 1, β -amylase data sets 1, 4, 5, 5a, 5b, 5c, 5d) were run with MrBayes 3.2.2 (Altekar et al., 2004; Ronquist et al., 2011), using *Joinvillea ascendens* to root the trees (see Appendix 2 for details). Tests of goodness of fit for alternative nucleotide substitution models were performed through the Hierarchical Likelihood Ratio Test (hLRTs), AIC and BIC tests in jModelTest 2 (Guindon and Gascuel, 2003; Durraba et al., 2012). The estimated GTR + I + gamma model was used for all Bayesian inferences performed in this study.

3. Results

3.1. The cpDNA-ITS species tree of the grasses

The cpDNA-ITS tree of the Poaceae (Fig. 1) provided a framework for evaluating paralogy and incongruence in the β -amylase data set, and for assessing potential phylogenetic bias among the four β -amylase subdivisions (Table 1: data sets 1, 4, 5, 5a, 5b, 5c, 5d). The multigenic data set consisted of nuclear ITS (134 sequences/609 nucleotides) and plastid *matK* (120/1259), *ndhF* (124/710), *trnTL* (116/789) and *trnLF* (127/975) sequences (Table S1).

The independent phylogenetic searches conducted with the five separate ITS, *matK*, *ndhF*, *trnTL*, and *trnLF* data sets yielded highly congruent topologies. The PH test detected incongruence among the plastid (cpDNA) data sets ($P < 0.01$), except for *matK* – *trnTL* ($P = 0.4$) and *ndhF* – *trnTL* ($P = 0.3$). Partition incongruence was also found between the plastid and nuclear data ($P = 0.01$). Despite this, the separate topologies were similar to the tree based on the combined nuclear and plastid datasets, with incongruence observed mostly among some terminal tips. Furthermore, those topological conflicts were not well supported. Since the incongruence did not affect the main results and conclusions of this research, we chose to conduct phylogenetic analyses on the combined dataset of the five sequenced genes. The BI tree (Fig. 1) and MP tree (Fig. 1; bootstrap support values) from the combined data set were highly congruent to each other and with those obtained for the grasses by previous authors (Bouchenak-Khelladi et al., 2008; Schneider et al., 2009; Blaner et al., 2015; Soreng et al., 2015). The MP statistics of the Poaceae cpDNA-ITS tree are shown in Table S2.

The 142 Poaceae species fell into the two traditionally described PACCMAD and BEP clades (e.g. GPWG, 2001, 2012; Sanchez-Ken et al., 2007; Bouchenak-Khelladi et al., 2008; Schneider et al., 2009; Blaner et al., 2015; Soreng et al., 2015). The PACCMAD clade was split into two strongly supported sister lineages, Centothecoideae–Panicoideae and Danthonioideae–Chloridoideae (e.g., Soreng and Davis, 1998; Bouchenak-Khelladi et al., 2008, and references therein). The BEP clade included the Ehrhartoideae, Bambusoideae, and Pooideae, which were grouped as described by Davis and Soreng (2007), Bouchenak-Khelladi et al. (2008), Schneider et al. (2009) and Triplett and Clark (2010). The Pooideae clade showed the successive divergences of the more ancestral Lygeae/Nardeae, Stipeae, Diarrheneae, and Brachypodieae lineages, and the separation of the more recently evolved core pooids (Bromeae – Triticeae and Poeae – Aveneae lineages). The relationships within the Poeae – Aveneae clade agreed with those proposed by Döring et al. (2007), Quintanar et al. (2007), and Schneider et al. (2009). The successively enlarged phylogenies of Loliinae obtained by Catalán et al. (2004, 2006) and Inda et al. (2008) were confirmed in our analyses; the Loliinae split into two main groups, the broad-leaved and fine-leaved Loliinae.

3.2. The β -amylase data set: recombination, pseudogenization and incongruence

Highly similar clones were grouped into consensus types according to the p -distance (< 0.01) criterion (Díaz-Pérez et al., 2014). Clones from twelve out of 16 species were grouped into a single type, suggesting that they represent the same orthologous copy (Fig. 2). However, clones from each of the remaining four species (*Brachypodium distachyon*, *Festuca fenas*, *F. arundinacea*, and *Vulpia alopecuros*) were classified into different types (p -distance > 0.01). Diploid *Brachypodium distachyon* clones 2 and 3 vs. clones 1 and 4–9 were phylogenetically close enough to be considered allelic copies. Diploid *Vulpia alopecuros* clones 1–3 vs. 4–5 were paraphyletic (Fig. 2), and their placement within the Aulaxyper clade was incongruent with the cpDNA-ITS reference tree (*V. alopecuros* clone 4–5, 0.89 PP/ $< 75\%$ BS; *V. alopecuros* clone 1–3, *F. iberica*, *F. rubra* clone 1–5, 1.00 PP/97% BS), rather than within the expected Loretia [*V. fasciculata*/*V. geniculata* clade (cf. Fig. 1)]. Though the Loretia group appeared in the β -amylase tree with low support, we hypothesize that the strongly grouped *V. alopecuros* clones within the Aulaxyper clade indicate a “semi-paralogous” copy (defined here as the detection of only one paralogous copy [Aulaxyper-type] but not the other [Loretia-type]). Related tetraploid *Festuca fenas* and hexaploid *F. arundinacea* clones grouped into two lineages, each consisting of one representative of each species, suggesting two homeologous copies (Fig. 2).

Seven recombination events were detected by two or more algorithms ($p < 0.05$) implemented in the program RDP4 in the exon-only β -amylase data set (Table 2 and Fig. 2). The recombinant sequences were from: (i) *Gynerium sagittatum* (Gynerieae); (ii) *Stipagrostis zeyheri* (Cortaderiinae/Danthoniinae); (iii) *Muhlenbergia tenuiflora* and *M. montana* (Muhlenbergiinae); (iv) *Setaria* sp. (Setariinae) and *Panicum miliaceum* and *P. virgatum* (Paniceae); (v) *Brachypodium distachyon* clones 2 and 3 (Brachypodieae); (vi) *Danthonia decumbens* and *Schismus barbatus* (Cortaderiinae/Danthoniinae); and (vii) *Diarrhena americana* (Diarrheninae). Five of the recombination events occurred in the PACCMAD clade, whereas two were associated with early-diverged Pooideae. No recombination events were detected in the core Pooideae clade. Most recombination events showed phylogenetically widely divergent minor and major parents involving different subtribes (Table 2 and Fig. 2), suggesting ancient hybridization events, although we cannot discard a wrong assignment of parents given the limited sampling sizes associated with the PACCMAD and the early divergent Pooideae lineages.

There were no PTCs (Akhunov et al., 2013) among the β -amylase coding sequences, so by this criterion, we did not find any evidence of pseudogenetic sequences. However, the Branch-Site REL (BS-REL) analysis detected different sequence positions under strongly negative, negative, neutral, positive, and strongly positive selection within exons (Table 3 and Fig. 2). Strong negative selection [$0 \leq \omega < 0.9$; range I (Table 3), dark blue bar (Fig. 2); 85.25% average value] was predominant across sequences and positions, followed by neutral selection [$0.91 \leq \omega \leq 1.5$; range III (Table 3), green bar (Fig. 2); 9.88%], and then by low percentages of strongly positive selection [$\omega > 5$; range V (Table 3), red bar (Fig. 2); 2.78%], negative selection [$0.9 \leq \omega \leq 0.61$; range II (Table 3), dark blue bar (Fig. 2); 1.7%] and positive selection [$0.5 \leq \omega \leq 1.51$; range IV (Table 3), pink bar (Fig. 2); 0.23%]. Fifty-eight out of 69 sequences tested with the BS-REL method showed ω values above the negative selection average (85.4%), most of them (57 sequences) with 0% neutral sites; whereas one sequence showed a considerable proportion of sites with weak negative selection pressure [*Colpodium drakensbergense* (100%); Table 3, Fig. 2]. Seven sequences, however, showed a high proportion of neutral sites,

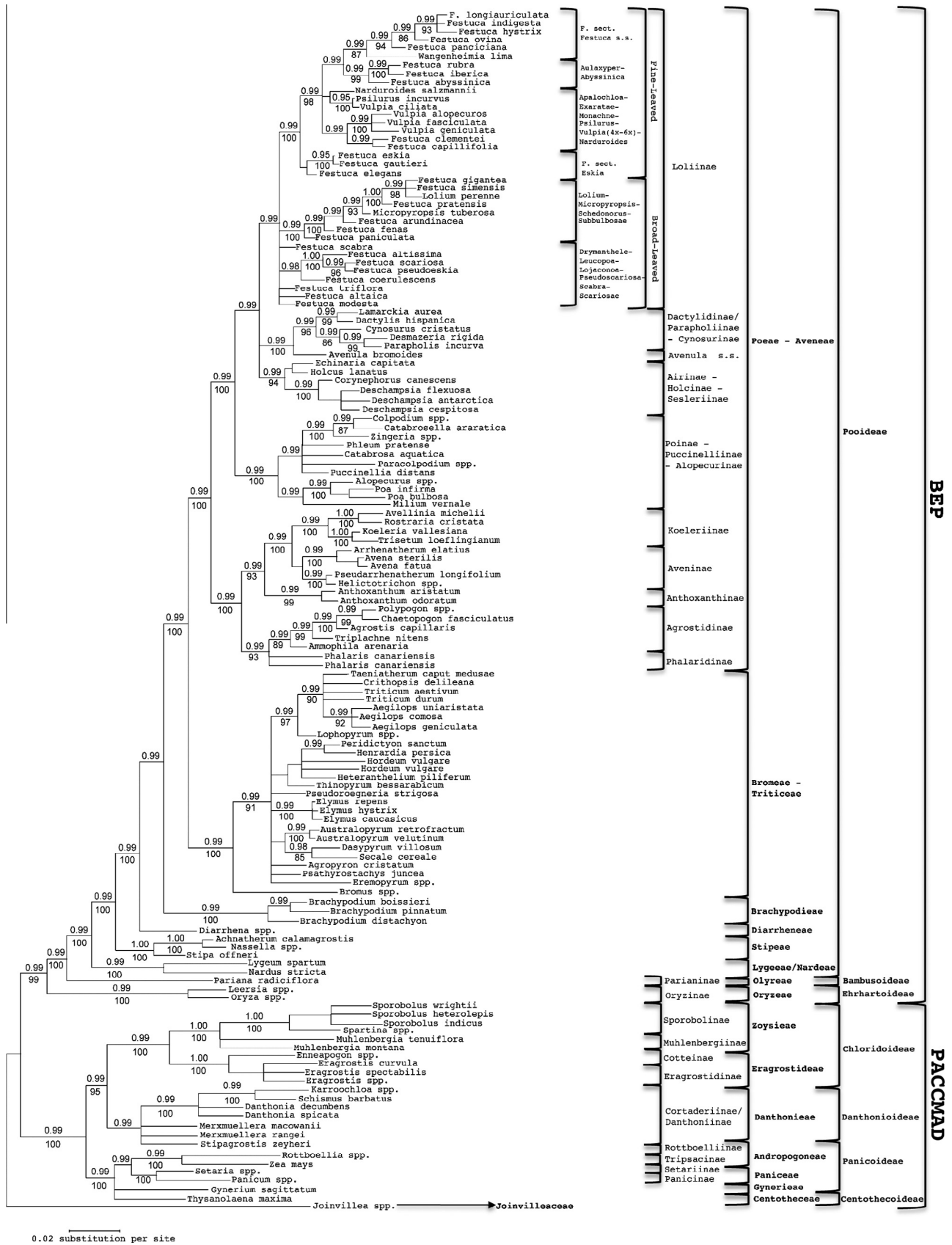


Fig. 1. Poaceae cpDNA-ITS tree. Bayesian Inference estimation based on nuclear (ITS) and plastid (*trnL*, *trnL*F, *ndhF*, *matK*) gene sequences (Table S1). Bayesian posterior probabilities of ≥ 0.90 and MP bootstrap support values of $\geq 75\%$ are shown above and below the nodes, respectively. Species assignment to sections, subgenera, tribes, subfamilies, and evolutionary clades are represented on the right side of the figure following Soreng et al. (2003, 2007), Torrecilla et al. (2003, 2004), Catalán et al. (2006), Quintanar et al. (2007), Bouchenak-Khelladi et al. (2008), Inda et al. (2008), Schneider et al. (2009); and Minaya et al. (2013). Identical criteria were followed in the remaining phylogenetic inferences presented in this paper.

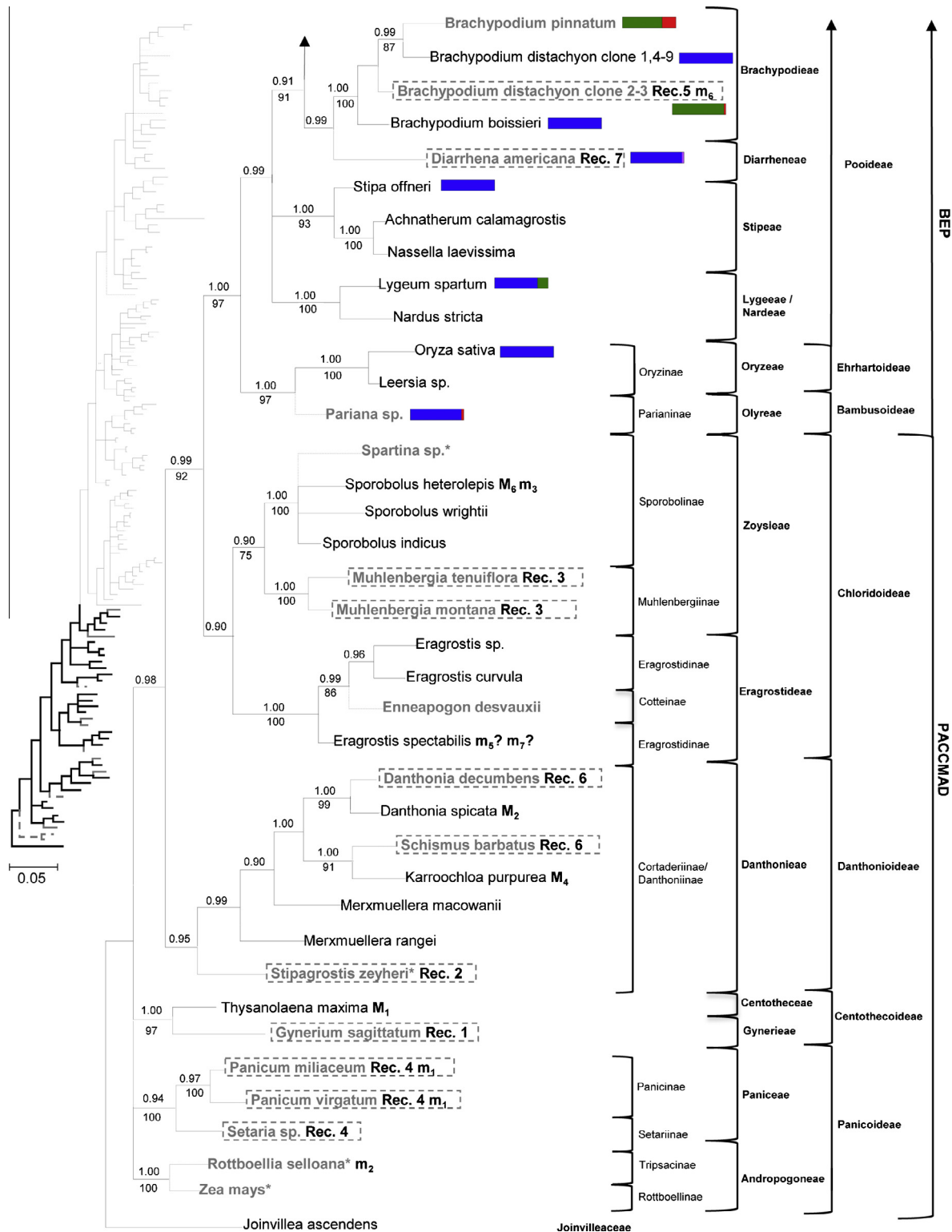


Fig. 2. Bayesian Majority Rule consensus tree of Poaceae based on the complete β -amylase exon + unambiguous intron data set. Bayesian posterior probabilities ≥ 0.90 and ML bootstrap support $\geq 75\%$ are shown above and below the branches, respectively. Species assignment to sections, subgenera, tribes, subfamilies, and evolutionary clades are represented on the right side of the figure. Gray branches, dashed boxes, and bold gray species names indicate recombination events detected in the exon-only data set (Table 1, β -amylase data set 2) by two or more methods implemented in the RDP4 program; Rec. n: indicates the number of recombination events (see Table 2). M_n : major parents in the recombination n; m_n : minor parents in the recombination n; $m_n?$: most likely parents in the recombination n. Pink and red colors on the horizontal bars, bold gray species names and gray branches indicate a high proportion of neutral, positive selected sites, and a complete absence of purifying selective positions, based on BS-REL method (see Table 3). Colors on the horizontal bars indicate strong negative selection (dark blue), weak negative selection (light blue), neutral selection (green), weak positive selection (pink), and strong positive selection (red). Gray boxes indicate homeologous clones of the allopolyploid species *Festuca fenas* and *F. arundinacea* and semi-paralogous clones of *Vulpia alopecuroides*. Bold gray species names and gray dotted branches indicate that these accessions are topologically incongruent when compared with their placement in the Poaceae reference tree (Fig. 1). Asterisks (*) indicate placements different in the phylogeny based on the exon-only data set (Table 1, β -amylase data set 1; Fig. S2) vs. exon + intron data set (Table 1, β -amylase data set 4; this figure). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

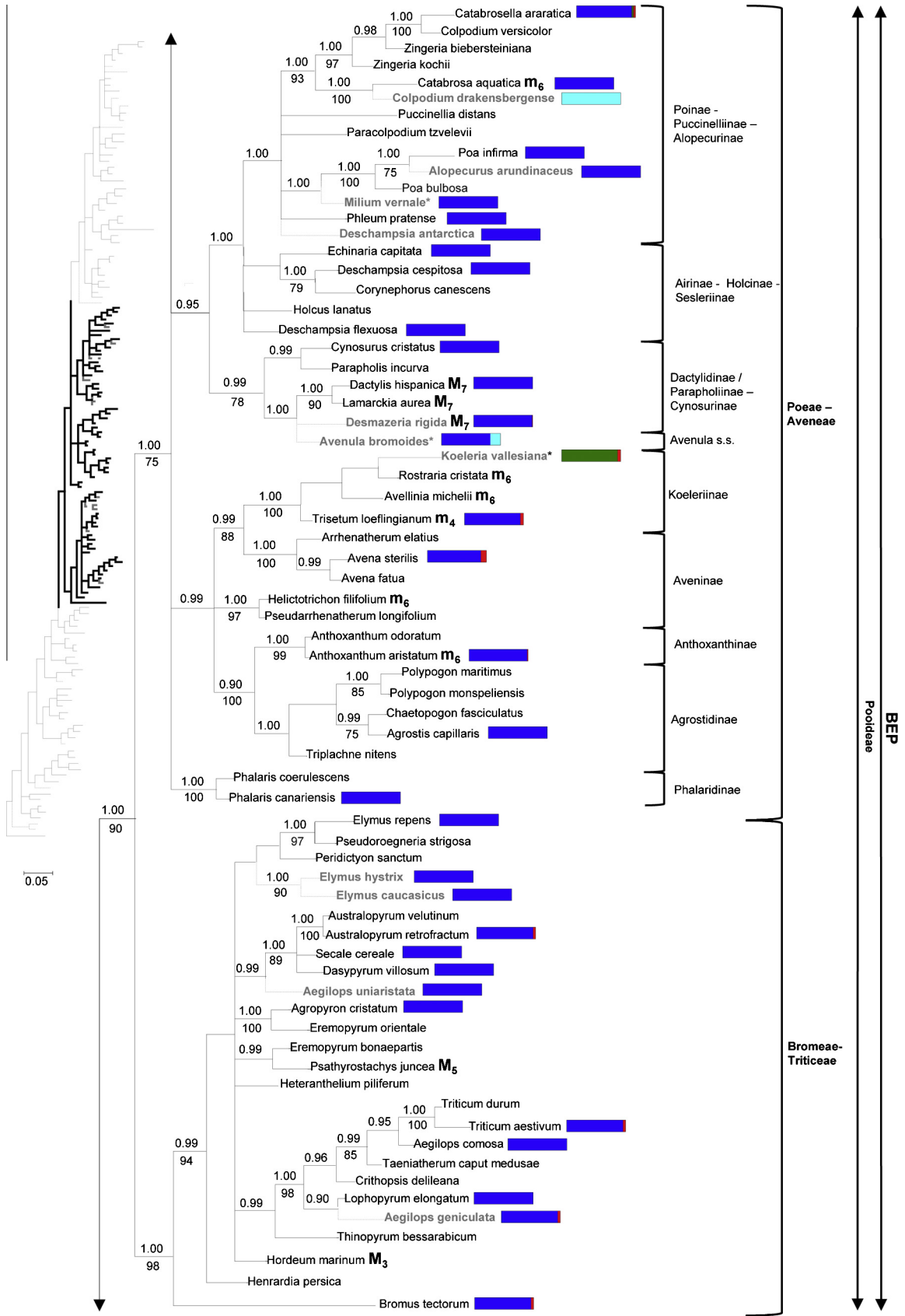


Fig. 2 (continued)

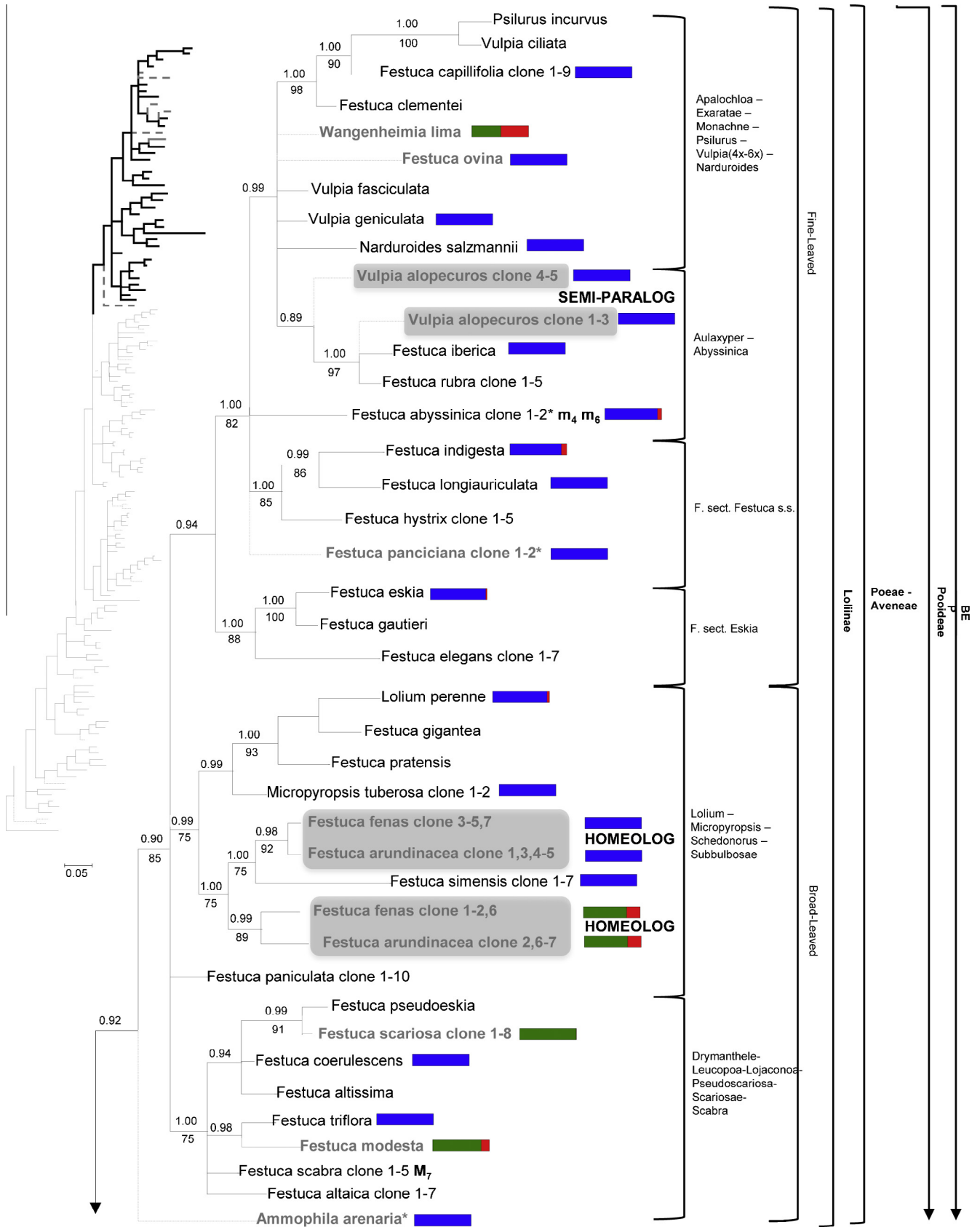


Fig. 2 (continued)

Table 2

Summary of recombination events identified in the β -amylase exon data set (588 positions: exons 2–5) by two or more methods implemented in the program RDP4. These events were detected through 142 species (146 sequences) of Poaceae that included the consensus sequences (p -distance < 0.01) of 16 cloned species (Table 1, β -amylase data set 2). Number of Rec. Event: Each recombination event is enumerated according to its statistical importance in the RDP4 analysis. Rec. sp.: Species under recombination.

Number of rec. event	Rec. sp.	Parental sequence(s)		Probability of significant tests for different detection methods in RDP4			
		Major	Minor	MaxChi	Chimaera	SiScan	3Seq
1.	<i>Gynerium sagittatum</i>	<i>Thysanolaena maxima</i>	<i>Panicum miliaceum</i> <i>Panicum virgatum</i>	1.36E–4	–	–	3.24E–2
2.	<i>Stipagrostis zeyheri</i>	<i>Dantohia spicata</i>	<i>Rottboellia seloana</i>	1.14E–3	4.29E–2	–	–
3.	<i>Muhlenbergia montana</i> <i>Muhlenbergia tenuifolia</i>	<i>Hordeum marinum</i>	<i>Sporobolus heterolepis</i>	2.34E–3	–	7.51E–5	–
4.	<i>Setaria</i> sp. <i>Panicum miliaceum</i> <i>Panicum virgatum</i>	<i>Karoochloa purpurea</i>	<i>Trisetum loeflingianum</i> <i>Festuca abyssinica</i> 1–2	–	2.98E–2	1.46E–4	–
5.	<i>Brachypodium distachyon</i> clone 2–3	<i>Psathyrostachys juncea</i>	Unknown (<i>Eragrostis spectabilis</i>)	4.39E–2	1.42E–2	–	–
6.	<i>Danthonia decumbens</i> <i>Schismus barbatus</i>	<i>Sporobolus heterolepis</i>	<i>Anthoxanthum aristatum</i> <i>Avellinia michelii</i> <i>Catabrosa aquatica</i> <i>Festuca abyssinica</i> <i>Helictotrichon filifolium</i> <i>Rostraria cristata</i> Unknown (<i>Eragrostis spectabilis</i>)	8.17E–4	–	3.96E–6	–
7.	<i>Diarrhena americana</i>	<i>Dactylis hispanica</i> <i>Desmazeria rigida</i> <i>Festuca scabra</i> <i>Lamarckia aurea</i>	Unknown (<i>Eragrostis spectabilis</i>)	–	2.58E–2	8.89E–3	–

low proportion of sites under positive selection, and a complete absence of positions under purifying selection [*Brachypodium pinatum* (72% neutral sites/28% positively selected sites), *F. arundinacea* clones 2, 6, and 7 (75%/25%), *F. modesta* (85%/15%), *F. fenas* clones 1, 2, and 6 (75%/25%), *F. scariosa* clones 1–8 (100%/–), *Koeleria vallesiana* (95%/5%), *Wangenheimia lima* (50%/50%); Table 3, Fig. 2]. These results indicate relaxed selection rates in their β -amylase gene copies, suggesting that they might be pseudogenes.

The Bayesian phylogenetic inference (BI) based on the β -amylase exon and exon + intron data sets (Table 1, β -amylase data sets 1 and 4, and Fig. S2 and Fig. 2, respectively) revealed 28 incongruent species when assessed in comparison to the reference tree (Fig. 1). Eight of these conflicting taxa (*Festuca paniciana*, *F. abyssinica*, *Koeleria vallesiana*, *Milium vernale*, *Rottboellia seloana*, *Stipagrostis zeyheri*, *Spartina* sp., and *Zea mays*) were only observed in the exon based-tree (Fig. S2); they were either congruent with the reference tree, or remained unresolved, in the exon + intron phylogeny (Fig. 2). Two of the incongruent sequences were associated with neutral selection (*Koeleria vallesiana* and *Wangenheimia lima*; Table 3) while another (*Stipagrostis zeyheri*) was associated with a recombination event (Table 2). The remaining incongruent sequences, which showed moderate to high support in the β -amylase phylogenies (Figs. 2 and S2), were *Enneapogon desvauxii* (PACCMAD), *Pariaria* sp. (Bambusoideae), *Aegilops geniculata*, *A. uniaristata*, *Elymus caucasicus*, and *E. hystrix* (Triticeae), and *Avenula bromoides*, *Desmazeria rigida*, *Deschampsia antarctica*, *Alopecurus arundinaceus*, *Colpodium drakensbergense*, *Ammophila arenaria*, *Vulpia alopecurus*, and *Festuca ovina* (Poeae – Aveneae).

The NeighborNet graph partitioned data into five statistically supported groups with moderate-to-high bootstrap support (BS > 70%) (Fig. 3). These groups were: (i) the Poeae Loliinae subtribe (BS = 78%); (ii) the remaining subtribes of Poeae – Aveneae (BS = 71% to 97%); (iii) Bromeae – Triticeae (BS = 97%); (iv) Olyreae + Nardeae + Lygeae + Stipeae + Diarrheneae + Brachypodiaceae (BS = 71 to 83%); and (v) PACCMAD + BEP early diverged lineages Oryzae + Olyreae (BS = 83%).

3.3. Beta-amylase phylogeny

The PH test did not detect incongruence (Table 4; $p > 0.05$) among the exon, first and second codon positions, third codon

positions, and third codon + intron positions data sets (Table 1, β -amylase data sets 5a–5d) after removal of the 37 conflicting species (41 sequences including the collapsed clones) with likely pseudogenization, semi-paralogy, homeology, recombination, or topological incongruence with the cpDNA-ITS tree. The SH pairwise topological tests (Table 4) used the exons + unambiguous introns tree as a reference (Table 1, β -amylase data set 5; Fig. S3); this tree was selected as the optimal β -amylase tree since it was better resolved and more robust than any of the alternative trees (Figs. 4A–D). It also showed the best fit to the reference tree (Fig. 1), and to previously published phylogenies of the grass family (Bouchenak-Khelladi et al., 2008; Schneider et al., 2009; Blaney et al., 2015; Soreng et al., 2015). The SH results indicated that all tested topologies (Fig. 4) were equally well supported by the data (Table 4; $p < 0.05^*$). A visual comparison of the topologies (Figs. 4A–D) showed that branching order was, in general, poorly resolved in all the alternative β -amylase trees. The Bayesian trees based on exons (Fig. 4A) and third codon plus intron positions (Fig. 4D) had higher resolution than did the trees based on either first and second codon positions (Fig. 4B) or third codon positions (Fig. 4C) positions. Surprisingly, the Bromeae – Triticeae lineages were better resolved in the tree based on third codon positions (Fig. 4C), and the fine-leaved Loliinae clade was better resolved in the tree based on third codon + intron positions (Fig. 4D). Branch lengths suggest that the third codon + intron positions (Fig. 4D) have evolved considerably faster in the Bromeae – Triticeae lineages than in the Loliinae subtribe. However, first and second codon positions (Fig. 4B) evolved faster in the fine-leaved Loliinae.

The most plausible β -amylase-based phylogeny of the grasses is the one based on exon plus unambiguous intron positions (Fig. S3), which contained 105 sequences (from 105 grass species), including the collapsed clones, after removal of the 41 conflicting sequences (Table 1, β -amylase data set 5). The samples were grouped into the two traditionally described BEP (1.00 PP/100% BS) and PACCMAD grass lineages. The paraphyletic PACCMAD clade showed the successive divergences of Centothecoideae (*Thysanolaena*), the monophyletic Danthonioideae (1.00/82), and Chloridoideae (1.00/88), in agreement with the reference tree (Fig. 1).

The BEP clade showed a well-supported sister relationship of Ehrhartoideae (*Oryza*, *Leersia*; 1.00/100) to Pooideae (1.00/87).

Table 3
Summary of the Branch-Site REL (BS-REL) results analyzed in the WEB interface DataMonkey. Maximum likelihood estimates of rate classes with $\omega_1 \leq 1$, $\omega_2 \leq 1$ and ω_3 (unconstrained) and proportion of sites (Pr1, Pr2 and Pr3, respectively) evolving at such rate classes along each branch. $\omega = d_N/d_S$, where d_N and d_S are the number of non-synonymous and synonymous mutations, respectively. Roman numerals (I to IV) correspond to the total percentage of sites (across ω_1 , ω_2 and ω_3) evolving according to distinct selection scenarios: (I) $\omega = 0-0.6$, Strongly Negative Selection (dark blue bar in Fig. 2); (II) $\omega = 0.61-0.9$, Weak Negative Selection (light blue bar in Fig. 2); (III) $\omega = 0.91-1.5$, Neutral Selection (green bar in Fig. 2); (IV) $\omega = 1.51-5$, Weak Positive Selection (pink bar in Fig. 2); and (V) $\omega > 5$, Strongly Positive Selection (red bar in Fig. 2). These analyses were done for 69 sequences (from 65 species) of Poaceae that included the consensus sequences (p -distance < 0.01) of 16 cloned species (Table 1, β -amylase data set 3).

Species	ω_1	Pr1	ω_2	Pr2	ω_3	Pr3	I	II	III	IV	V
<i>Aegilops comosa</i>	0	0	0	0	0	1	100%	0	0	0	0
<i>Aegilops geniculata</i>	0.06	0.96	0.06	0.03	10000	0.01	99%	0	0	0	1%
<i>Aegilops uniaristata</i>	0	0	0	0	0	1	100%	0	0	0	0
<i>Agropyron cristatum</i>	0	0	0	0	0	1	100%	0	0	0	0
<i>Agrostis capillaris</i>	0	0	0	0	0	1	100%	0	0	0	0
<i>Alopecurus arundinaceus</i>	0	0	0	0	0	1	100%	0	0	0	0
<i>Ammophila arenaria</i>	0	0	0	0	0	1	100%	0	0	0	0
<i>Anthoxanthum aristatum</i>	0	0.96	0	0.03	20.67	0.01	99%	0	0	0	1%
<i>Australopyrum retrofractum</i>	0	0.96	0	0.02	25.39	0.02	99%	0	0	0	2%
<i>Avena sterilis</i>	0	0.94	8.18E-07	0	8.42	0.06	94%	0	0	0	6%
<i>Avenula bromoides</i>	0.29	0.83	0.65	0.17	1.11	0	83%	17%	0	0	0
<i>Brachypodium distachyon clone 2-3</i>	1	0.97	1	0.02	834.46	0.01	0	0	99%	0	1%
<i>Brachypodium distachyon clone 1,4-9</i>	0	0	0	0	0	1	100%	0	0	0	0
<i>Brachypodium boissieri</i>	0.31	0.94	0.31	0.06	1.01	0	100%	0	0	0	0
<i>Brachypodium pinnatum</i>	1	0.72	1.00	4.57E-06	614.23	0.28	0	0	72%	0	28%
<i>Bromus tectorum</i>	0.37	0.97	0.81	0	33.03	0.03	97%	0	0	0	3%
<i>Catabrosa aquatica</i>	0	0	0	0	0	1	100%	0	0	0	0
<i>Catabrosella araratica</i>	0.06	0.98	0.91	0.00	38.01	0.02	98%	0	0.2%	0	1.8%
<i>Colpodium drakensbergense</i>	0.65	1	0.55	0	1.08	0	0	100%	0	0	0
<i>Cynosurus cristatus</i>	0	0	0	0	0	1	100%	0	0	0	0
<i>Dactylis hispanica</i>	0	0	0	0	0	1	100%	0	0	0	0
<i>Dasyphyrum villosum</i>	0	0	0	0	0	1	100%	0	0	0	0
<i>Deschampsia antarctica</i>	0	0	0	0	0	1	100%	0	0	0	0
<i>Deschampsia cespitosa</i>	0	0	0	0	0	1	100%	0	0	0	0
<i>Deschampsia flexuosa</i>	0	0	0	0	0	1	100%	0	0	0	0
<i>Desmazeria rigida</i>	0.28	0.99	0.28	0.00	137.74	0.01	99%	0	0	0	1%
<i>Diarrhena americana</i>	0	0.86	0	0.02	3.80	0.13	98%	0	0	2%	0
<i>Echinaria capitata</i>	0	0	0	0	0	1	100%	0	0	0	0
<i>Elymus caucasicus</i>	0	0	0	0	0	1	100%	0	0	0	0
<i>Elymus hystrix</i>	0	0	0	0	0	1	100%	0	0	0	0
<i>Elymus repens</i>	0.171564	1	0.248175	0	0.24	0	100%	0	0	0	0
<i>Festuca abyssinica clone 1-2</i>	0	0.94	0	0.00	6.54	0.06	94%	0	0	0	6%
<i>Festuca arundinacea clone 2,6-7</i>	1	0.75	1.00	0	1468.99	0.25	0	0	75%	0	25%
<i>Festuca arundinacea clone 1,3,4-5</i>	0	0	0	0	0	1	100%	0	0	0	0
<i>Festuca capillifolia clone 1-9</i>	0	0	0	0	0	1	100%	0	0	0	0
<i>Festuca coerulescens</i>	0	0	0	0	0	1	100%	0	0	0	0
<i>Festuca eskia</i>	0	0.99	0.88	0	49.19	0.01	99%	0	0	0	1%
<i>Festuca iberica</i>	0	0	0	0	0	1	100%	0	0	0	0
<i>Festuca indigesta</i>	0.00	0.90	0.89	0.00	8.76	0.10	90%	0	0	0	10%
<i>Festuca longiauriculata</i>	0	0	0	0	0	1	100%	0	0	0	0
<i>Festuca modesta</i>	1	0.85	0.99	1.85E-07	754.11	0.15	0	0	85%	0	15%
<i>Festuca ovina</i>	0	0	0	0	0	1	100%	0	0	0	0
<i>Festuca triflora</i>	0.34	0.75	0.39	0.25	0.98	0	100%	0	0	0	0
<i>Festuca fenas clone 1-2,6</i>	1	0.75	1.00	3.40E-05	1098.20	0.25	0	0	75%	0	25%
<i>Festuca fenas clone 3-5,7</i>	0	0	0	0	0	1	100%	0	0	0	0
<i>Festuca panciciana clone 1-2</i>	0	0	0	0	0	1	100%	0	0	0	0
<i>Festuca scariosa clone 1-8</i>	1	0.67	0.86	0	1.17	0.33	0	0	100%	0	0
<i>Festuca simensis clone 1-7</i>	0	0	0	0	0	1	100%	0	0	0	0
<i>Joinvillea ascendens</i>	0	0.86	0	0.00	4.85	0.14	86%	0	0	14%	0
<i>Koeleria vallesiana</i>	1	0.94	1	0.01	2541.96	0.05	0	0	95%	0	5%
<i>Lolium perenne</i>	0	0.96	0	0.01	15.34	0.03	97%	0	0	0	3%
<i>Lophopyrum elongatum</i>	0	0	0	0	0	1	100%	0	0	0	0
<i>Lygeum spartum</i>	0.06	0.79	0.96	0	1.46	0.21	79%	0	21%	0	0
<i>Milium vernale</i>	0	0	0	0	0	1	100%	0	0	0	0
<i>Micropyropsis tuberosa clone 1-2</i>	0	0	0	0	0	1	100%	0	0	0	0
<i>Narduroides salzmännii</i>	0.04	0.78	0.07	0.22	3.23	0	100%	0	0	0	0
<i>Oryza sativa</i>	0	0	0	0	0	1	100%	0	0	0	0
<i>Pariana sp.</i>	0.05	0.97	0.05	0.00	8.41	0.03	97%	0	0	0	3%
<i>Phalaris canariensis</i>	0.38	0.39	0.38	0.19	0.38	0.42	100%	0	0	0	0
<i>Phleum pratense</i>	0	0	0	0	0	1	100%	0	0	0	0
<i>Poa infirma</i>	0	0	0	0	0	1	100%	0	0	0	0
<i>Secale cereale</i>	0	0	0	0	0	1	100%	0	0	0	0
<i>Stipa offneri</i>	0	0	0	0	0	1	100%	0	0	0	0
<i>Trisetum loeflingianum</i>	0	0.97	0	0.00	7.04	0.03	97%	0	0	0	3%
<i>Triticum aestivum</i>	0	0.97	0	0.02	60.27	0.01	99%	0	0	0	1%
<i>Vulpia alopecuroides clone 4-</i>	0	0	0	0	0	1	100%	0	0	0	0
<i>Vulpia alopecuroides clone 1-3</i>	0	0	0	0	0	1	100%	0	0	0	0
<i>Vulpia geniculata</i>	0	0	0	0	0	1	100%	0	0	0	0
<i>Wangenheimia lima</i>	1	0.50	0.88	0	4299.40	0.50	0	0	50%	0	50%
Average percentage							85.25%	1.7%	9.88%	0.23%	2.78%

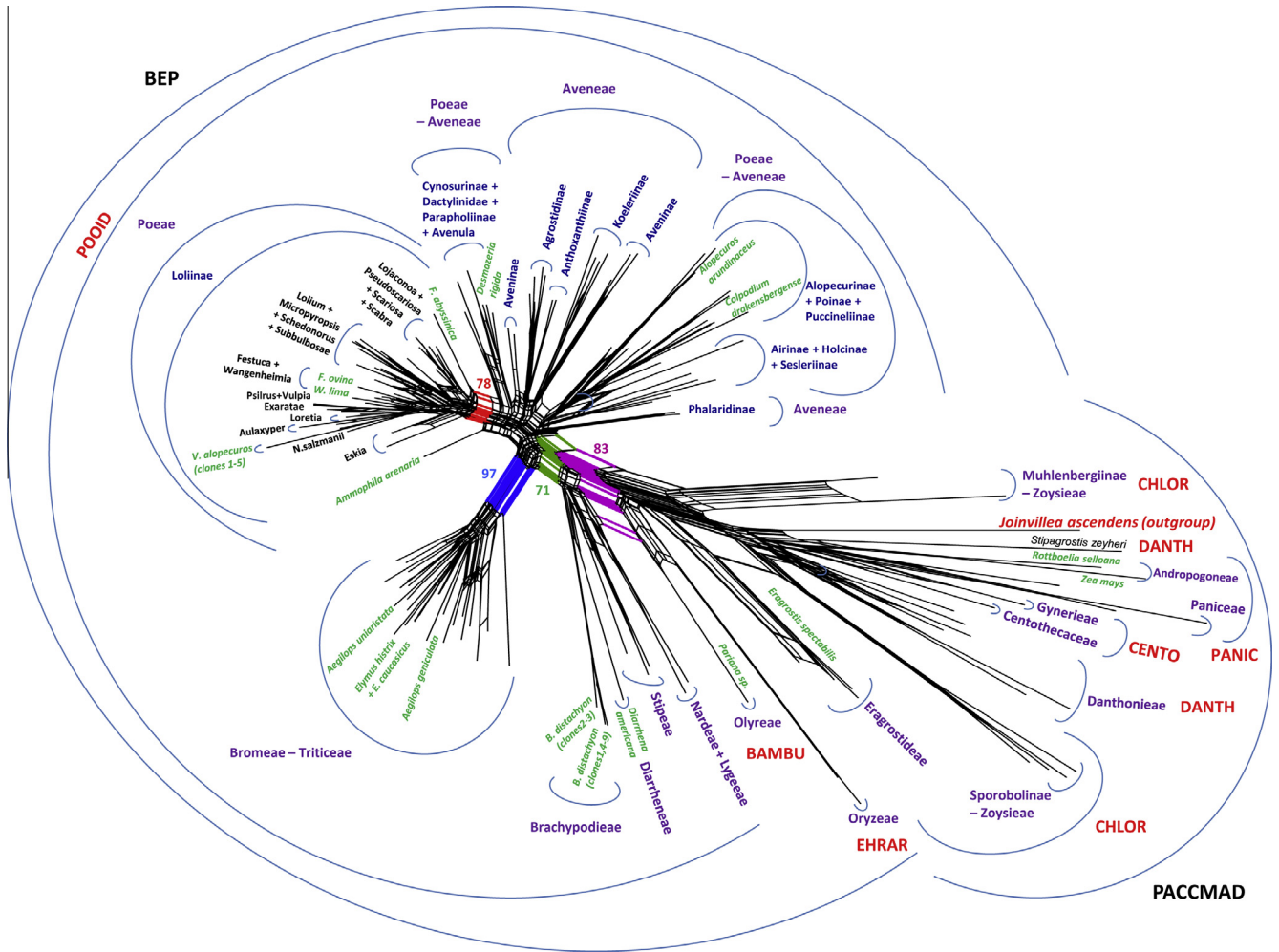


Fig. 3. NeighborNet partition network tree. Graphical representation of phylogenetic relationships considering possible incompatible and ambiguous signals in the grass β -amylase exon + intron sequences (Table 1, β -amylase data set 4). Major tribal and subtribal lineages correspond to those indicated in Figs. 1 and 2. Green accessions represent topologically incongruent lineages with respect to the reference cpDNA + ITS tree. Bootstrap support values are indicated for the four main splits: Pooideae (83%, purple); core pooids (71%, green); Bromoideae – Triticeae (97%, blue); Lolliinae (78%, red). Subfamily abbreviations: POOID (Pooideae); BAMBU (Bambusoideae); EHRAR (Ehrartoideae); CHLOR (Chloridoideae); DANTH (Danthioideae); CENTO (Centothecoideae); PANIC (Panicoideae). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 4

Results of the Partition Homogeneity (PH) test [H_0 : the compared data sets are incongruent] and the Shimodaira–Hasegawa test (SH) tests [H_0 : the topology of the most likely tree (Table 1, β -amylase data set 5: phylogeny with the highest likelihood score) is similar to the compared topology (Table 1, β -amylase data sets 5a–5d; Figs. 5a–5d)]. LnL: likelihood score; Diff. LnL: difference of –LnL between the compared trees; * the null hypotheses is accepted. These analyses were done using 105 sequences (from 105 species) of Poaceae that included the consensus sequences (p -distance < 0.01) of 16 cloned species (Table 1, β -amylase data set 5).

β -amylase subdivisions	(5a) Exon positions	(5b) First and second codon positions	(5c) Third codon positions	(5d) Third codon positions + introns
<i>Partition Homogeneity (PH) test</i>				
(5a) Exon positions	–	1.00	1.00	0.96
(5b) First and second codon positions		–	0.86	0.32
(5c) Third codon positions			–	1.00
(5d) Third codon positions + introns				–
<i>Shimodaira–Hasegawa test (SH) test</i>				
Bayesian topology compared	–LnL	Diff. –LnL	p -value	
(5) Exon + intron positions	13266.38892	(Best)	–	
(5d) Third codon positions + introns	13352.79645	86.40752	0.013*	
(5a) Exon positions	13423.06073	156.67181	<0.001*	
(5b) First and second codon positions	14233.37348	966.98456	<0.001*	
(5c) Third codon positions	14455.69237	1189.30345	<0.001*	

Divergence within Pooideae consisted of the early split of the sisters Lygeae/Nardeae (1.00/100), Stipeae (0.99/89), Brachypodieae (1.00/100) and the core Pooideae (1.00/93). The resolution of the Bromoideae – Triticeae clade was similar to that found in the

β -amylase exon-based tree (Fig. S2). The Poae s. l. lineages collapsed in a large polytomy (1.00/88). The Poae – Aveneae lineage was divided into six subtribes: (i) Phalaridinae (*Phalaris*; 1.00/100); (ii) Agrostidinae + Anthoxanthinae + Aveninae + Koeleriinae (0.93/-);

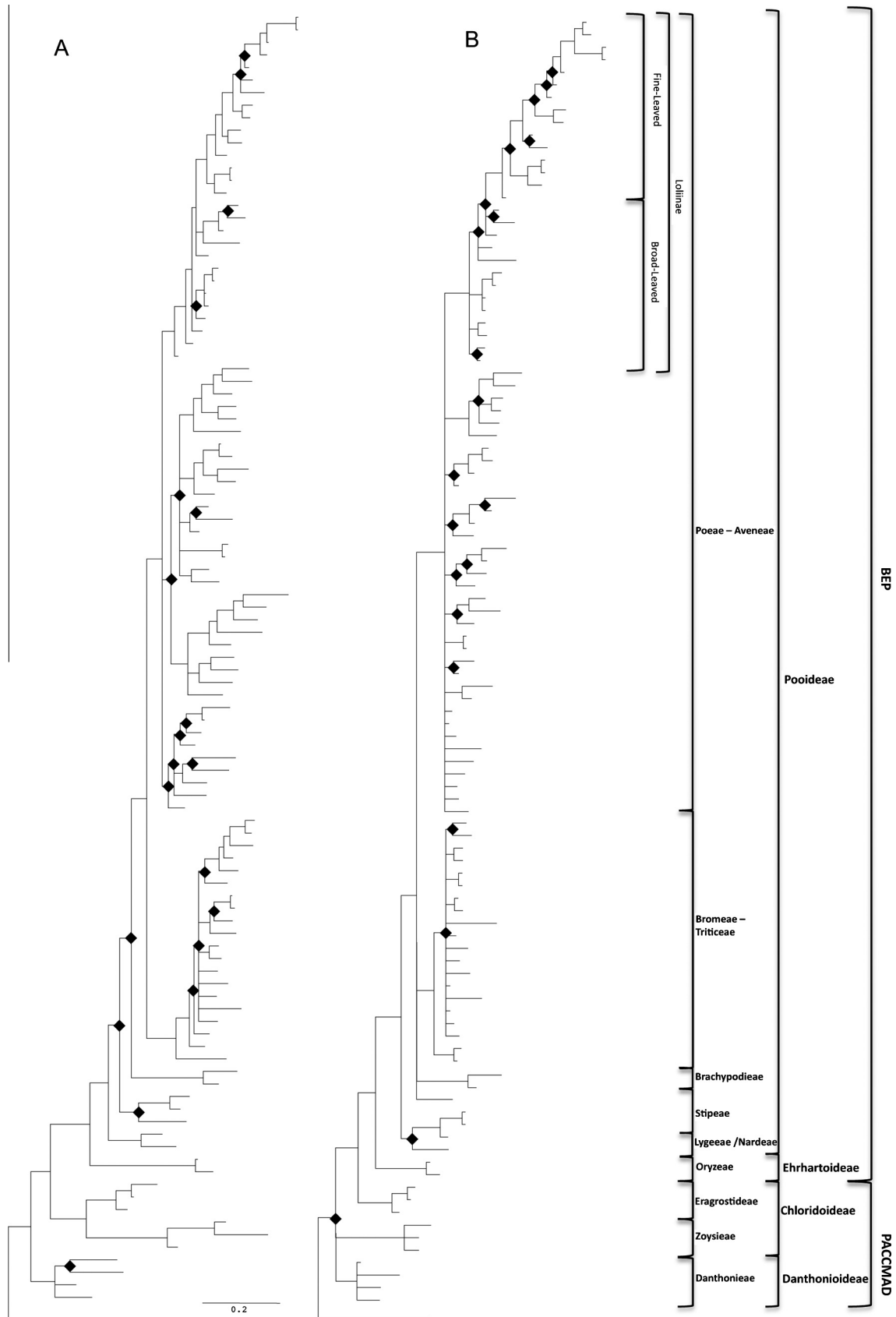


Fig. 4. Alternative Bayesian topologies of grasses based on different partitions of the β -amylase gene. (A) exon positions; (B) first and second codon positions; (C) third codon positions; and (D) third codon positions plus aligned unambiguous introns after removing the 37 grass species (41 accessions) with likely pseudogenization, recombination and/or topological incongruence (Table 1, β -amylase data sets 5a–5d, respectively). Diamonds indicate clades with low bootstrap (<75%) or posterior probability (<90%) support. Lineage assignments to tribes, subfamilies and clades are presented on the right side of the figures.

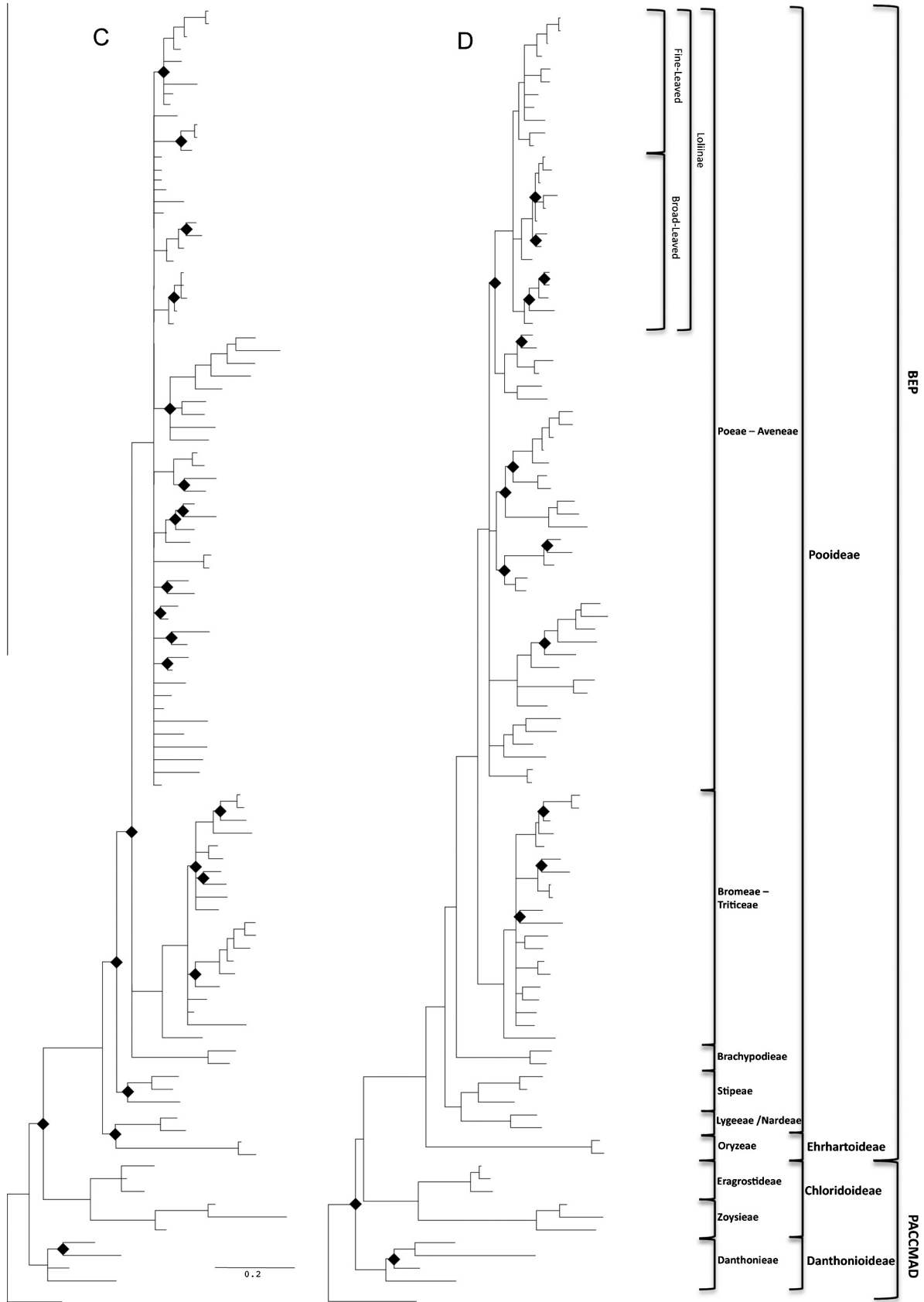


Fig. 4 (continued)

(iii) Dactyloideae/Parapholiinae – Cynosurinae (0.99/-); (iv) Airinae – Holcinae – Sesleriinae (0.97/-); (v) Poinae – Puccinelliinae – Alopecurinae (1.00/97); and (vi) Loliinae (1.00/98). The Loliinae clade collapsed into a basal polytomy of two broad-leaved lineages and the fine-leaved fescues. The broad-leaved group included representatives of (i) Drymanthele – Leucopoa – Lojaconoa – Psudoscariosa – Scariosa – Scabra clade (0.94/-); and (ii) *Lolium* – *Micropyropsis* – *Schedonorus* – *Subbulbosae* clade (0.99/98). The fine-leaved Loliinae split into three main clades: (i) the most basal *Eskia* (0.99/89); (ii) *Festuca* (1.00/95); and (iii) *Apalochloa* – *Exaratae* – *Monachne* – *Psilurus* – *Vulpia* (4x–6x) – *Narduroides* + *Aulaxyper* s.l. (1.00/99).

4. Discussion

4.1. Evolution of the β -amylase exon sequences in the grasses: relaxed selection pressure, pseudogenization, semi-paralogy, homeology, recombination and incongruence

Our study has shown the complex evolutionary dynamics of the LCNG β -amylase in the temperate grasses and close allies through the detection of 41 sequences (from 37 species) with likely pseudogenization, paralogy, homeology, recombination and/or phylogenetic incongruence (Fig. 2). Our failure to detect PTCs as evidence of pseudogenic copies was corroborated using the amino acid translations to guide the alignment of these coding sequences. However, since exons 1, 6, and 7, and parts of exons 2 and 5 were not sequenced, this criterion alone does not conclusively rule out the presence of pseudogenes. Using a complementary method to detect pseudogenization, we interpreted sequences with a large proportion of sites under neutral selection as an indicator of relaxation of selection pressure (cf. Kosakovsky-Pond et al., 2008, 2011). Thus, the eight sequences with a high proportion of neutral sites, and a complete absence of sites under purifying selection, are probably pseudogenes. The distribution of these copies among three separate temperate pooid lineages [Brachypodieae; Aveneae – Koeleriinae; and Poeae – Loliinae; Table 3, Fig. 2] indicates that pseudogenization and/or relaxation of selection pressure in β -amylase sequences could be largely extended across the Poaceae, and affect both orthologous and homeologous copies.

Differentiating orthologous and paralogous DNA sequences is essential for phylogenetic inference (Sanderson and Doyle, 1992; Wendel and Doyle, 1998). However, distinguishing paralogous copies which diverged after a duplication event from orthologs related by descent can be a challenge (Fu and Dooner, 2002; Small et al., 2004). Several publications have demonstrated paralogous copies in LCNGs such as the β -amylase (Kreis et al., 1987; Wang et al., 1997; Li et al., 2002; Mason-Gamer, 2005) and GBSSI (Díaz-Pérez et al., 2014) genes in the grasses. In our study, a single case of β -amylase semi-paralogy has been found in the fine-leaved Loliinae *Vulpia alopecuros*. The unexpected placement of the diploid *V. alopecuros* sequences within the *Aulaxyper* clade (Figs. 2 and S2) reflects a close relationship of this Loretia-type species with the *Festuca rubra* group (*Aulaxyper*). This supports a recent LCNG GBSSI-based study (Díaz-Pérez et al., 2014), which recovered strong phylogenetic relationships between members of the Loretia clade and different *Aulaxyper* taxa, supporting recent gene flow between the two clades; this favored the occurrence of intergeneric \times *Festulpia* hybrids and introgressed plants. The two *V. alopecuros* *Aulaxyper*-type β -amylase copies likely represent a fraction of a semi-paralogous sequence, even though the orthologous sequences of *V. alopecuros*, which should be nested within the Loretia clade, have not been found. This could be explained as a consequence of PCR bias favoring only one paralogous copy, or gene loss. Alternatively, gene flow, rather than paralogy, might

explain the unexpected placement of *V. alopecuros* clones, so introgressive hybridization could also be invoked to explain the two types of *V. alopecuros* clones (cf. Harper et al., 2011; Cui et al., 2015).

The broad-leaved Loliinae polyploids *F. fenas* and *F. arundinacea* have two types of β -amylase homeologous copies (Fig. 2). The resolution within the *Lolium* – *Micropyropsis* – *Schedonorus* lineage (Figs. 2, S2) confirms the origins of these allopolyploid *Schedonorus* species in the ancestral Western-Mediterranean region (Inda et al., 2014). Several studies (e. g. cross-hybridizations, Chandrasekharan et al., 1972; genome-mapping, Humphreys et al., 1995; phylogenetic analysis, Inda et al., 2008, 2014) support the allohexaploid origin of *Festuca arundinacea* from diploid *F. pratensis* and allotetraploid *F. fenas*. Inda et al. (2014) further suggested the putative contribution of *F. fenas* as the maternal progenitor of *F. arundinacea*. The β -amylase phylogeny revealed two divergent types of clones in *F. fenas*, which could have been derived from each of its two putative homeologous genomes, and, in turn, inherited by its descendant hybrid *F. arundinacea*. By contrast, the absence of *F. arundinacea* clones similar to the progenitor *F. pratensis* suggests that those paternal copies might have been lost in the allohexaploid, or could have gone undetected with the current sample.

The robust sister relationship of the afro-montane *F. simensis* to *F. fenas* and *F. arundinacea* reveals for the first time the potential origin of one of the progenitors of this putative allotetraploid. Inda et al. (2014) described a sister relationship of *F. simensis* with a potential diploid *Lolium*-like parent based on their ITS phylogeny, whereas their plastid *trnT* phylogeny suggested a Eurasian *Schedonorus* maternal genome donor. Based on these hypotheses and our current findings, the tetraploid *F. simensis* might have originated from a cross between a diploid western Mediterranean *F. fenas*-type maternal parent and a Eurasian diploid *Lolium*-type paternal parent, followed by genome duplication and a likely Pliocene colonization of the afro-montane belt (cf. Inda et al., 2014).

Recombination analyses of 146 β -amylase sequences detected seven recombination events (Table 2 and Fig. 2). Although the major and minor parental sequences of the recombinants could not be unambiguously identified, the seven recombinants suggest that cross-species recombination occurs frequently in the grass family. This might be explained in part by their allogamous pollination syndrome (cf. Lian et al., 2013), but could also be related to other unknown biological or evolutionary processes (cf. Minaya et al., 2013). Most of the potential parental sequences for the detected recombinants are phylogenetically close. For example, the major and minor parental sequences detected in *Gynerium sagittatum* were *Thysanolaena maxima* (Centotheceae) and *Panicum miliaceum* – *P. virgatum* (Paniceae). Since nuclear recombination is common among taxa of hybrid origin (Marques et al., 2010), the detection of recombinant sequences could reveal unexpected hybridizations. Alternatively, these results might reflect traces of more ancestral palaeohybridizations, as pointed out by Mason-Gamer (2008), Salse et al. (2008), and Mahelka and Kopecký (2010) in different grass groups. Two remarkable exceptions to the phylogenetically close parental sequences rule were observed in *Muhlenbergia* (Muhlenbergiinae), where the major parental sequence was detected in Triticeae (*Hordeum marinum*) and in *Panicum* (Paniceae), and the minor parental sequence in Koeleriinae (*Trisetum loeflingianum*) and Loliinae (*Festuca abyssinica*) (Table 2 and Fig. 2). Because the putative parental lineages are highly divergent, each characterized by high rates of hybridization, polyploidy (cf. Mason-Gamer, 2005, 2008; Catalán et al., 2006), Horizontal Gene Transfer (HGT) (Ghatnekar et al., 2006; Vallenback et al., 2008, 2010), and *Stowaway* MITES (Mason-Gamer, 2005; Minaya et al., 2013), these evolutionarily distant recombinations could have occurred by a number of

processes, including the combination of palaeo- or neo-homeologous copies.

Of the 28 species on the β -amylase tree that were incongruent with respect to their placement in the cpDNA-ITS reference tree (Fig. 1), two were associated with apparent pseudogenization and one with a recombination event. LCNs are prone to a diversity of evolutionary fates, so other possible explanations for the remaining conflicting sequences include incomplete lineage sorting, retention of alternative copies from ancestral polyploid genomes, horizontal gene transfer, differential gene copy losses (Fortune et al., 2007; Díaz-Pérez et al., 2014), or sampling artifacts in some lineages (i.e. PACCMAD and Bambusoideae).

4.2. Phylogenetic signals of β -amylase sequences within the Pooideae

The identification and removal of problematic sequences is key for using the β -amylase gene to disentangle the phylogenetic history of the Pooideae. A visual comparison of the alternative topologies obtained based on the exons, first and second codon positions, third codon positions, and third codon + intron positions data sets, after removal of the 41 conflicting sequences (Fig. 4), shows little incongruence among the four data sets, an observation corroborated by the PH test (Table 4). In addition, these four topologies are similar to the phylogeny based on the exon + intron data set depicted in Fig. S3 (taken as the reference β -amylase tree); this is confirmed by the results of the SH test (Table 4). Resolution also varies among the trees; for example, the phylogeny recovered for the Triticeae is better resolved by third codon positions than first and second codon positions (Fig. 4). The advantage of using third codon positions in phylogenetic reconstructions becomes particularly clear at low taxonomic levels or among lineages originated by rapid and recent divergence events (cf. Gaut et al., 1997; Small et al., 1998; Cronn et al., 2002; Christin et al., 2007). Nevertheless, our results demonstrate that other grass lineages do not follow this pattern. The recently evolved Loliinae clade was less resolved when it was based on third codon positions than first and second codon positions (Fig. 4). Similarly, the Loliinae lineages had shorter branch lengths in the Bayesian third codon positions tree than in the first and second codon positions reconstruction. In both cases, the uncoupling of first and second vs. third codon evolutionary rates observed suggests that locus-specific selection coefficients can vary widely among grass lineages. These observations highlight the importance of understanding how different partitions of the β -amylase gene data set have influenced the phylogenetic reconstruction of the grasses (cf. Christin et al., 2007).

The most plausible β -amylase phylogeny of Pooideae after removing the 41 problematic accessions is the one based on exon plus unambiguous intron positions (Fig. S3). The monophyly of the BEP clade is strongly supported (Figs. 2 and S3). Although all the molecular analyses published to date, including the β -amylase analysis, have supported the monophyly of Pooideae, the relationships found among its constituent groups have varied widely. In particular, the early-diverging tribal lineages are frequently unresolved or only weakly supported (e.g. Catalán et al., 1997; Soreng and Davis, 2000; GPWG, 2001). The basal divergence of the Lygeae/Nardeae, suggested on morphological grounds (e.g. Clayton and Renvoize, 1986; Soreng and Davis, 1998) has been corroborated in our analysis, and is in agreement with other molecular studies (e.g. Bouchenak-Khelladi et al., 2008). However, the placement of the Stipeae as sister to the Lygeae/Nardeae group conflicts with recent phylogenies that reconstruct Stipeae as a more recently evolved group (Bouchenak-Khelladi et al., 2008). A variety of molecular studies place Stipeae as a basal or sub-basal pooid lineage, but with a position that fluctuates according to the

taxa sampled or the genome analyzed (e.g. Hilu et al., 1999; Döring et al., 2007; Bouchenak-Khelladi et al., 2008).

The core Pooideae reconstruction shows a monophyletic Bromeae – Triticeae and a largely polytomic Poeae – Aveneae clade. For the Triticeae, our Bayesian inference supports a more restricted phylogenetic analysis of the β -amylase gene (Mason-Gamer, 2005). It is, however, not in full agreement with phylogenies based on other data sets; the numerous molecular analyses have failed to converge on a straightforward estimate of the relationships among the diploid genera of this tribe (cf. Kellogg et al., 1996; Mason-Gamer and Kellogg, 1996; Seberg and Frederiksen, 2001; Petersen and Seberg, 2002; and Mason-Gamer, 2005). The Poeae – Aveneae group includes the five main lineages of the core Aveneae group (Phalaridinae, Agrostidinae, Anthoxanthinae, Aveninae, and Koeleriinae), mostly agreeing with the phylogenies proposed by Soreng et al. (2003), Quintanar et al. (2007), and Saarela et al. (2010). The unresolved position of the *Pseudarrhenatherum/Helictotrichon* lineage is consistent with both a nested relationship within Aveneae (e.g. Quintanar et al., 2007), and with placement outside of *Avena* and *Arrhenatherum* (Couderc and Guédès, 1976; Röser et al., 2001).

The NeighborNet partition network analysis (Fig. 3) highlights the phylogenetic separation between the Loliinae and Aveneae subtribes. This separation was not observed in the Bayesian tree based on GBSSI (Díaz-Pérez et al., 2014), where there was little resolution within the BEP clade. The genetic distinctiveness of the Bromeae – Triticeae partition from Poeae – Aveneae in the NeighborNet tree seems to indicate a more basal phylogenetic position for Bromeae – Triticeae. The more basal BEP lineages are found in the fourth and fifth groups of the NeighborNet tree, suggesting that the Oryzeae + Olyreae lineage, placed in the fifth group of the NeighborNet tree, could be one of the first groups to diverge in the separation of the BEP and PACCMAD clades, as is also suggested by the GBSSI Bayesian tree (Díaz-Pérez et al., 2014).

The incongruent phylogenetic patterns described for 37 grass species (26% of the sampled species; Fig. 2) and the detection of various MITEs (Minaya et al., 2013) highlight the frequent occurrence of evolutionary events that violate the assumption that phylogenetic history is strictly tree-like, such as recombination, relaxed selection pressure, semiparalogy, incomplete lineage sorting, horizontal gene transfer, or differential gene copy losses in the β -amylase gene. Thus, β -amylase gene sequences should be applied to phylogenetic analyses with caution. However, once the pitfalls are identified and removed, the phylogenetic reconstruction of the temperate grasses and selected outgroups based on the β -amylase exon + intron positions is highly informative at all taxonomic levels.

Acknowledgments

We thank Majid Sharifi and Pedro Torrecilla for their help with β -amylase cloning and sequencing, Elizabeth Kellogg, Lynn Clark, Lynn Gillespie, and Jeff Saarela for providing samples for us, Emily Lemonds for writing assistance, and two anonymous reviewers for their valuable comments to an early version of the manuscript. The work was supported by two consecutive Spanish Ministry of Science and Innovation research Grant projects CGL2009-12955-C02-01 and CGL2012-39953-C02-01. MM was funded by a Spanish Ministry of Science and Innovation PhD FPI fellowship. Phylogenetic analyses were conducted at the cluster computer system of the BIFI Research Institute (University of Zaragoza, Spain).

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at <http://dx.doi.org/10.1016/j.ympev.2015.05.014>.

References

- Akhunov, E.D., Sehgal, S., Liang, H., Wang, S., Akhunova, A.R., Kaur, G., Li, W., Forrest, K.L., See, D., Simková, H., Ma, Y., Hayden, M.J., Luo, M., Faris, J.D., Dolezel, J., Gill, B.S., 2013. Comparative analysis of syntenic genes in grass genomes reveals accelerated rates of gene structure and coding sequence evolution in polyploid wheat. *Plant Physiol.* 161, 252–265.
- Altekar, G., Dwarkadas, S., Huelsenbeck, J.P., Ronquist, F., 2004. Parallel Metropolis coupled Markov chain Monte Carlo for Bayesian phylogenetic inference. *Bioinformatics* 20, 407–415.
- Altenhoff, A.M., Dessimoz, C., 2012. Inferring orthology and paralogy. In: Anisimova, M. (Ed.), *Evolutionary Genomics: Statistical and Computational Methods*. Springer Science+Business Media, Berlin, pp. 259–279.
- Alvarez, I., Wendel, J.F., 2003. Ribosomal ITS sequences and plant phylogenetic inference. *Mol. Phylogenet. Evol.* 29, 417–434.
- Blaner, A., Schneider, J., Röser, M.A., 2015. 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. *Mol. Phylogenet. Evol.* 87, 14–27.
- Bouchenak-Khelladi, Y., Salamin, N., Savolainen, V., Forest, V., Van der Bank, M., Chase, M.W., Hodkinson, T.R., 2008. Large multi-gene phylogenetic trees of the grasses (Poaceae): progress towards complete tribal and generic level sampling. *Mol. Phylogenet. Evol.* 47, 488–505.
- Brassac, J., Jakob, S.S., Blattner, F.R., 2012. Progenitor-derivative relationships of *Hordeum* polyploids (Poaceae, Triticeae) inferred from sequences of TOPO6, a nuclear low-copy gene region. *PLoS ONE* 7, e33808.
- Brunner, P.C., Keller, N., McDonald, A., 2009. Wheat domestication accelerated evolution and triggered positive selection in the β -Xylosidase enzyme of *Mycosphaerella graminicola*. *PLoS ONE* 11, e7884.
- Bryant, D., Moulton, V., 2004. Neighbor-Net: an agglomerative method for the construction of phylogenetic networks. *Mol. Biol. Evol.* 21, 255–265.
- Buckler, E.S., Ippolito, A., Holtsford, T.P., 1997. The evolution of ribosomal DNA: divergent paralogues and phylogenetic implications. *Genetics* 145, 821–832.
- Campbell, C.S., Kellogg, E.A., 1987. Sister group relationships of the Poaceae. In: Soderstrom, T.R., Hilu, K.W., Campbell, C.S., Barkworth, M.E. (Eds.), *Grass Systematics and Evolution*. Smithsonian Institution Press, Washington, pp. 217–224.
- Catalán, P., Kellogg, E.A., Olmstead, R.G., 1997. Phylogeny of Poaceae subfamily Pooideae based on chloroplast *ndhF* gene sequences. *Mol. Phylogenet. Evol.* 8, 150–166.
- Catalán, P., Torrecilla, P., López Rodríguez, J.A., Olmstead, R.G., 2004. Phylogeny of the festucoid grasses of subtribe Loliinae and allies (Poeae, Pooideae) inferred from *ITS* and *trnL-F* sequences. *Mol. Phylogenet. Evol.* 31, 517–541.
- Catalán, P., Torrecilla, P., López-Rodríguez, J.A., Müller, J., 2006. Molecular evolutionary rates shed new light on the relationships of *Festuca*, *Lolium*, *Vulpia* and related grasses (Loliinae, Pooideae, Poaceae). In: Bailey, J., Ellis, R.G. (Eds.), *Current Taxonomic Research on the British & European flora*. BSBI, London, pp. 45–70.
- Catalán, P., Müller, J., Hasterok, R., Jenkins, G., Mur, L.A., Langdon, T., Betekhtin, A., Siwinska, D., Pimentel, M., López-Alvarez, D., 2012. Evolution and taxonomic split of the model grass *Brachypodium distachyon*. *Ann. Bot-London* 109, 385–405.
- Chandrasekharan, P., Lewis, E.J., Borrill, M., 1972. Studies in *Festuca* 2. Fertility, relationships between species of sections *Vinovae* and *Scariosae*, and their affinities with *Lolium*. *Genetica* 43, 375–386.
- Christin, P., Salamin, N., Savolainen, V., Duvall, M.R., Besnard, G., 2007. C₄ Photosynthesis evolved in grasses via parallel adaptive genetic changes. *Curr. Biol.* 17, 1241–1247.
- Clayton, W.D., Renvoize, S.A., 1986. *Genera Gramineum: Grasses of the World*. Kew Bull. Series XIII, Royal Botanic Gardens, Kew.
- Couderc, J.M., Guédès, M., 1976. *Helictotrichon pallens* and *H. setifolia*. *Taxon* 25, 188.
- Cronn, R.C., Small, R.L., Haselkorn, T., Wendel, J.F., 2002. Rapid diversification of the cotton genus (*Gossypium: Malvaceae*) related by analysis of sixteen nuclear and chloroplast genes. *Am. J. Bot.* 89, 707–725.
- Cui, H., Sun, Y., Deng, J., Wang, M., Xia, G., 2015. Chromosome elimination and introgression following somatic hybridization between bread wheat and other grass species. *Plant Cell. Tiss. Org.* 120, 203–210.
- Darriba, D., Taboada, G.L., Doallo, R., Posada, D., 2012. JModelTest2: more models, new heuristic and parallel computing. *Nat. Methods* 9, 772.
- Davis, J.L., Soreng, R.J., 2007. A preliminary phylogenetic analysis of the grass subfamily Pooideae (Poaceae), with attention to structural features of the plastid and nuclear genomes, including an intron loss in GBSSI. In: Columbus, J.T., Friar, E.A., Porter, J.M., Prince, L.M., Simpson, M.G. (Eds.), *Monocots: Comparative Biology and Evolution*. Poales. Rancho Santa Ana Botanical Garden, Claremont, pp. 335–348.
- Díaz-Pérez, A., Sharifi-Tehrani, M., Inda, L.A., Catalán, P., 2014. Polyphyly, gene-duplication and extensive allopolyploidy framed the evolution of the ephemeral *Vulpia* grasses and other fine-leaved Loliinae (Poaceae). *Mol. Phylogenet. Evol.* 79, 92–105.
- Döring, E., Schneider, J., Hilu, K.W., Röser, M., 2007. Phylogenetic relationships in the Aveneae/Poeae complex (Pooideae, Poaceae). *Kew Bull.* 62, 407–424.
- Doyle, J.J., 1991. Evolution of higher-plant glutamine synthetase genes: tissue specificity as a criterion for predicting orthology. *Mol. Biol. Evol.* 8, 366–377.
- Doyle, J.J., 1996. Homoplasy connections and disconnections: genes and species, molecules and morphology. In: Sanderson, M.J., Hufford, L. (Eds.), *Homoplasy: the Recurrence of Similarity in Evolution*. Academic Press, San Diego, pp. 67–89.
- Doyle, J.J., Doyle, J.L., 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem. Bull.* 19, 11–15.
- Doyle, J.J., Doyle, J.L., 1999. Nuclear protein-coding genes in phylogeny reconstruction and homology assessment: some examples from Leguminosae. In: Gornall, R.J. (Ed.), *Molecular Systematics and Plant Evolution*. Taylor and Francis, London, pp. 229–254.
- Edgar, R.C., 2004. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *Bioinformatics* 5, 113.
- Farris, J.S., Källersjö, M., Kluge, A.G., Bult, C., 1994. Testing the significance of incongruence. *Cladistics* 10, 315–319.
- Feliner, G.N., Rossello, J.A., 2007. Better the devil you know? Guidelines for insightful utilization of nrDNA ITS in species-level evolutionary studies in plants. *Mol. Phylogenet. Evol.* 44, 911–919.
- Fortune, P.M., Schierenbeck, K.A., Ainouche, A.K., Jacquemin, J., Wendel, J.F., Ainouche, M.L., 2007. Evolutionary dynamics of Waxy and the origin of hexaploid *Spartina* species (Poaceae). *Mol. Phylogenet. Evol.* 43, 1040–1055.
- Fu, H., Dooner, H.K., 2002. Intraspecific violation of genetic colinearity and its implications in maize. *Proc. Natl. Acad. Sci. USA* 99, 9573–9578.
- Gabaldón, T., 2008. Large-scale assignment of orthology: back to phylogenetics? *Genome Biol.* 9, 235.
- Gaut, B.S., Clark, L.G., Wendel, J.F., Muse, S.V., 1997. Comparisons of the molecular evolutionary process at *rbcl* and *ndhF* in the grass family (Poaceae). *Mol. Biol. Evol.* 14, 769–777.
- Ghatnagar, L., Jaarola, M., Bengtsson, B.O., 2006. The introgression of a functional nuclear gene from *Poa* to *Festuca ovina*. *P. Roy. Soc. Lond. B. Biol.* 273, 395–399.
- G.P.W.G., The Grass Phylogeny Working Group, 2001. Phylogeny and subfamilial classification of the grasses (Poaceae). *Ann. Mo. Bot. Gard.* 88, 373–457.
- G.P.W.G., The Grass Phylogeny Working Group, 2012. New grass phylogeny resolves deep evolutionary relationships and discovers C4 origins. *New Phytol.* 193, 304–312.
- Guindon, S., Gascuel, O., 2003. A simple, fast, and accurate method to estimate large phylogenies by maximum-likelihood. *Syst. Biol.* 52, 696–704.
- Hall, T.A., 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucl. Acid. Symp. Ser.* 41, 95–98.
- Harper, J., Armstead, I., Thomas, A., James, C., Gasior, D., Bisaga, M., Roberts, L., King, I., King, J., 2011. Alien introgression in the grasses *Lolium perenne* (perennial ryegrass) and *F. pratensis* (meadow fescue): the development of seven monosomic substitution lines and their molecular and cytological characterization. *Ann. Bot-London* 107, 1313–1321.
- Hilu, K.W., Alice, L.A., Liang, H., 1999. Phylogeny of Poaceae inferred from *matK* sequences. *Ann. Mo. Bot. Gard.* 86, 835–851.
- Hughes, C.E., Eastwood, R.J., Bailey, C.D., 2006. From famine to feast? Selecting nuclear DNA sequence loci for plant species-level phylogeny reconstruction. *Philos. T. Roy. Soc. B* 361, 211–225.
- Humphreys, M.W., Thomas, H.M., Morgan, W.G., Meredith, M.R., Harper, J.A., Thomas, H., Zwierzykowski, Z., Ghesquière, M., 1995. Discriminating the ancestral progenitors of hexaploid *Festuca arundinacea* using genomic in situ hybridisation. *Heredity* 75, 171–174.
- Huson, D.H., Bryant, D., 2006. Application of phylogenetic networks in evolutionary studies. *Mol. Biol. Evol.* 23, 254–267.
- Huson, D.H., Rupp, R., Scornavacca, C., 2010. *Phylogenetic Networks, Concepts, Algorithms and Applications*. Cambridge University Press, Cambridge.
- Inda, L.A., Segarra-Moragues, J.G., Müller, J., Peterson, P.M., Catalán, P., 2008. Dated historical biogeography of the temperate Loliinae (Poaceae, Pooideae) grasses in the northern and southern hemispheres. *Mol. Phylogenet. Evol.* 46, 932–957.
- Inda, L.A., Sanmartín, I., Buerki, S., Catalán, P., 2014. Mediterranean origin and Miocene–Holocene Old World diversification of meadow fescues and ryegrasses (*Festuca* subgenus *Schedonorus* and *Lolium*). *J. Biogeogr.* 41, 600–614.
- Kellogg, E.A., 2006. Progress and challenges in studies of the evolution of development. *J. Exp. Bot.* 57, 3505–3516.
- Kellogg, E.A., Appels, R., Mason-Gamer, R.J., 1996. When gene trees tell different stories: the diploid genera of Triticeae (Gramineae). *Syst. Bot.* 21, 321–347.
- Kimura, M., 1968. Evolutionary rate at the molecular level. *Nature* 217, 624–626.
- Kosakovsky-Pond, S.L., Frost, S.D.W., 2005. Not so different after all: a comparison of methods for detecting amino acid sites under selection. *Mol. Biol. Evol.* 22, 1208–1222.
- Kosakovsky-Pond, S.L., Poon, A.F.Y., Leigh Brown, A.J., Frost, S.D.W., 2008. A maximum likelihood method for detecting directional evolution in protein sequences and its application to influenza A virus. *Mol. Biol. Evol.* 25, 1809–1824.
- Kosakovsky-Pond, S.L., Murrell, B., Fourment, M., Frost, S.D.W., Delpont, W., Scheffler, K., 2011. A random effects branch-site model for detecting episodic diversifying selection. *Mol. Biol. Evol.* 28, 3033–3043.
- Kreis, M., Williamson, M., Buxton, B., Pywell, J., Hejgaard, J., Svendsen, I., 1987. Primary structure and differential expression of β -amylase in normal and mutant barleys. *Eur. J. Biochem.* 179, 517–525.
- Li, W.H., 1997. *Molecular Evolution*. Sinauer, Sunderland.
- Li, C.D., Langridge, P., Zhang, X.Q., Eckstein, P.E., Rossnagel, B.G., Lance, R.C.M., Lefol, E.B., Lu, M.Y., Harvey, B.L., Scoles, G.J., 2002. Mapping of barley (*Hordeum vulgare* L.) *beta*-amylase alleles in which amino acid substitution determines *beta*-amylase isoenzyme type and the level of free *beta*-amylase. *J. Cereal Sci.* 35, 39–50.

- Lian, S., Lee, J., Cho, W.K., Yu, J., Kim, J., Choi, H., Kim, K., 2013. Phylogenetics and recombination analyses of tomato spotted wilt virus. *PLoS ONE* 8, e63380.
- Linder, C.R., Rieseberg, L.H., 2004. Reconstructing patterns of reticulate evolution in plants. *Am. J. Bot.* 91, 1700–1708.
- Linder, H.P., Rudall, P.J., 1993. The megagametophyte in *Anarthria* (Anarthriaceae, Poales) and its implications for the phylogeny of the Poales. *Am. J. Bot.* 80, 1455–1464.
- Maddison, D.R., Maddison, W.P., 2008. *MacClade 4.08*. Sinauer Associates Inc., Sunderland.
- Mahelka, V., Kopecký, D., 2010. Gene capture from across the grass family in the allohexaploid *Elymus repens* (L.) Gould (Poaceae, Triticeae) as evidenced by ITS, GBSSI, and molecular cytogenetics. *Mol. Biol. Evol.* 27, 1370–1390.
- Marques, I., Feliner, G.N., Draper, D., Martins-Louçao, M.A., Furtés-Aguilar, J., 2010. Unraveling cryptic reticulate relationships and the origin of orphan hybrid disjunct populations in *Narcissus*. *Evolution* 64, 2353–2368.
- Martin, D.P., Lemey, P., Lott, M., Moulton, V., Posada, D., Lefevre, P., 2010. RDP4: a flexible and fast computer program for analyzing recombination. *Bioinformatics* 26, 2462–2463.
- Mason-Gamer, R.J., 2005. The β -amylase genes of grasses and a phylogenetic analysis of the Triticeae (Poaceae). *Am. J. Bot.* 92, 1045–1058.
- Mason-Gamer, R.J., 2008. Allohexploidy, introgression, and the complex phylogenetic history of *Elymus repens* (Poaceae). *Mol. Phylogenet. Evol.* 47, 598–611.
- Mason-Gamer, R.J., Kellogg, E.A., 1996. Chloroplast DNA analysis of the monogenomic Triticeae: phylogenetic implications and genome-specific markers. In: Jauhar, P.P. (Ed.), *Methods of Genome Analysis in Plants*. CRC Press, Boca Raton, pp. 301–325.
- Mason-Gamer, R.J., Kellogg, E.A., 1997. Testing for phylogenetic conflict among molecular data sets in the tribe Triticeae (Gramineae). *Syst. Biol.* 45, 524–545.
- Minaya, M., Pimentel, M., Mason-Gamer, R., Catalán, P., 2013. Distribution and evolutionary dynamics of *Stowaway* Miniature Inverted repeat Transposable Elements (MITEs) in grasses. *Mol. Phylogenet. Evol.* 68, 106–118.
- Miyata, T., Yasunaga, T., 1980. Molecular evolution of mRNA: a method for estimating evolutionary rates of synonymous and amino acid substitutions from homologous nucleotide sequences and its application. *J. Mol. Evol.* 16, 23–36.
- Nielsen, R., Yang, Z., 1998. Likelihood models for detecting positive selected amino acid sites and applications to the HIV-1 envelope gene. *Genetics* 148, 929–936.
- Petersen, G., Seberg, O., 2002. Molecular evolution and phylogenetic application of *DMC1*. *Mol. Phylogenet. Evol.* 22, 43–50.
- Pond, S.L., Frost, S.D., 2005. Datamonkey: rapid detection of selective pressure on individual sites of codon alignments. *Bioinformatics* 21, 2531–2533.
- Posada, D., Crandall, K.A., 2001. Evaluation of methods for detecting recombination from DNA sequences: Computer simulations. *Proc. Natl. Acad. Sci. USA* 98, 13757–13762.
- Preston, J.C., Kellogg, E.A., 2006. Reconstructing the evolutionary history of paralogous APETALA1/FRUITFULL-Like genes in grasses (Poaceae). *Genetics* 174, 421–437.
- Quintanar, A., Castroviejo, S., Catalán, P., 2007. Phylogeny of the tribe Aveneae (Pooideae, Poaceae) inferred from plastid *trnT-F* and nuclear *ITS* sequences. *Am. J. Bot.* 94, 1554–1596.
- Rajapakse, S., Nilmalgoda, S.D., Molnar, M., Balnard, R., Ballard, R.E., Austin, D.F., Bohac, J.R., 2004. Phylogenetic relationships of the sweet potato in *Ipomoea* series *Batatas* (Convolvulaceae) based on nuclear β -amylase sequences. *Mol. Phylogenet. Evol.* 30, 623–632.
- Ronquist, F., Huelsenbeck, J.P., Teslenko, M., 2011. MrBayes, version 3.2 manual: Tutorials and Model Summaries <http://mrbayes.sourceforge.net/mb3.2_manual.pdf>.
- Röser, M., Winterfeld, G., Grebenstein, B., Hemleben, V., 2001. Molecular diversity and physical mapping of 5S rDNA in wild and cultivated oat grasses (Poaceae: Aveneae). *Mol. Phylogenet. Evol.* 21, 198–217.
- Saarela, J.M., Liu, Q., Peterson, P.M., Soreng, R.J., Paszko, B., 2010. Phylogenetics of the grass Aveneae-type plastid DNA clade (Poaceae: Pooideae, Poaeae) based on plastid and nuclear ribosomal DNA sequence data. In: Seberg, O., Petersen, G., Barfod, A.S. (Eds.), *Diversity, Phylogeny, and Evolution in the Monocotyledons: Proceedings of the Fourth International Conference on the Comparative Biology of the Monocotyledons and the Fifth International Symposium of Grass Systematics and Evolution*, Copenhagen, pp. 557–585.
- Salse, J., Bolot, S., Throude, M.I., Jouffe, V., Piegu, B., Quraishi, U.M., Calcagno, T., Cooke, R., Delseny, M., Feuillet, C., 2008. Identification and characterization of shared duplications between rice and wheat provide new insight into grass genome evolution. *Plant Cell* 20, 11–24.
- Sanchez-Ken, J.G., Clark, L.G., Kellogg, E.A., Kay, E.E., 2007. Reinstatement and emendation of subfamily Micrairoideae (Poaceae). *Syst. Biol.* 54, 183–196.
- Sanderson, M.J., Doyle, J.J., 1992. Reconstruction of organismal and gene phylogenies from data on multigene families: concerted evolution, homoplasy, and confidence. *Syst. Biol.* 41, 4–17.
- Sang, T., 2002. Utility of low-copy nuclear gene sequences in plant phylogenetics. *Biochem. Mol. Biol.* 37, 121–147.
- Sang, T., Zhang, D., 1999. Reconstructing hybrid speciation using sequences of low-copy nuclear genes: hybrid origins of five *Paeonia* species based on *Adh* gene phylogenies. *Syst. Bot.* 24, 148–163.
- Schneider, J., Döring, E., Hilu, K.W., Röser, M., 2009. Phylogenetic structure of the grass subfamily Pooideae based on comparison of plastid *matK* gene-3' *trnK* exon and nuclear *ITS* sequences. *Taxon* 58, 405–424.
- Seberg, O., Frederiksen, S., 2001. A phylogenetic analysis of the monogenomic Triticeae (Poaceae) based on morphology. *Bot. J. Linn. Soc.* 136, 75–97.
- Shimodaira, H., Hasegawa, M., 1999. Multiple comparisons of log-likelihoods with applications to phylogenetic inference. *Mol. Biol. Evol.* 16, 1114–1116.
- Small, R.L., Ryburn, J.A., Cronn, R.C., Seelanan, T., Wendel, J.F., 1998. The tortoise and the hare: choosing between noncoding plastome and nuclear *Adh* sequences for phylogenetic reconstruction in a recently diverged plant group. *Am. J. Bot.* 85, 1301–1315.
- Small, R.L., Cronn, R.C., Wendel, J.F., 2004. Use of nuclear genes for phylogeny reconstruction in plants. *Aust. Syst. Bot.* 17, 145–170.
- Soreng, R.J., Davis, J.L., 1998. Phylogenetics and character evolution in the grass family (Poaceae): simultaneous analysis of morphological and chloroplast DNA restriction character sets. *Bot. Rev.* 64, 1–85.
- Soreng, R.J., Davis, J.L., 2000. Phylogenetic structure in Poaceae subfamily Pooideae as inferred from molecular and morphological characters: misclassification versus reticulation. In: Jacobs, S.W.L., Everett, J. (Eds.), *Grasses: Systematics and Evolution*. CSIRO, Melbourne, pp. 61–74.
- Soreng, R.J., Peterson, P.M., David, G., Judziewicz, E.J., Zuloaga, F.O., Filgueiras, T.S., Morrone, O., 2003. *Catalogue of New World Grasses (Poaceae): IV. Subfamily Pooideae*. *Contr. U.S. Natl. Herb* 48, 1–730.
- Soreng, R.J., Davis, J.L., Voionmaa, M.A., 2007. A phylogenetic analysis of Poaceae tribe Poeae sensu lato based on morphological characters and sequence data from three plastid-encoded genes: evidence for reticulation, and a new classification of the tribe. *Kew Bull.* 62, 425–454.
- Soreng, R.J., Peterson, P.M., Romaschenko, K., Davidse, G., Zuloaga, F.O., Judziewicz, E.J., Filgueiras, T.S., Davis, J.L., Morrone, O., 2015. A worldwide phylogenetic classification of the Poaceae (Gramineae). *J. Syst. Evol.* 53, 117–137.
- Swofford, D.L., 2002. PAUP*: Phylogenetic analysis using parsimony and other methods, version 4b10. Sinauer Associates, Sunderland.
- Torrecilla, P., López-Rodríguez, J.A., Stancik, D., Catalán, P., 2003. Systematics of *Festuca* L. sects. *Eskia* Willk., *Pseudotropis* Kriv., *Amphigenes* (Janka) Tzvel., *Pseudoscariosa* Kriv. and *Scariosae* Hack. based on analysis of morphological characters and DNA sequences. *Plant Syst. Evol.* 239, 113–139.
- Torrecilla, P., López-Rodríguez, J.A., Catalán, P., 2004. Phylogenetic relationships of *Vulpia* and related genera (Poeae, Poaceae) based on analysis of its and *trnL-F* sequences. *Ann. Mo. Bot. Gard.* 91, 124–158.
- Triplett, J.K., Clark, L.G., 2010. Phylogeny of the temperate bamboos (Poaceae: Bambusoideae: Bambuseae) with an emphasis on *Arundinaria* and allies. *Syst. Bot.* 35, 102–120.
- Vallenback, P., Jaarola, M., Ghatnekar, L., Bengtsson, B.O., 2008. Origin and timing of the horizontal transfer of a PgiC gene from *Poa* to *Festuca ovina*. *Mol. Phylogenet. Evol.* 46, 890–896.
- Vallenback, P., Ghatnekar, L., Bengtsson, B.O., 2010. Structure of the natural transgene PgiC2 in the common grass *Festuca ovina*. *PLoS ONE*, e13529. doi: <10.1371/journal.pone.0013529>.
- Wang, S.M., Lue, W.L., Wu, S.Y., Huang, H.W., Chen, J.C., 1997. Characterization of a maize β -amylase cDNA clone and its expression during seed germination. *Plant Physiol.* 113, 403–409.
- Watson, L., Dallwitz, M.J., 1992 onwards. *The Grass Genera of the World: Descriptions, Illustrations, Identification, and Information Retrieval; Including Synonyms, Morphology, Anatomy, Physiology, Phytochemistry, Cytology, Classification, Pathogens, World and Local Distribution, and References*. Version: 12th August 2014. <<http://delta-intkey.com>>.
- Wendel, J.F., Doyle, J.J., 1998. Phylogenetic incongruence: window into genome history and molecular evolution. In: Soltis, D.E., Soltis, P.S., Doyle, J.J. (Eds.), *Molecular Systematics of Plants II*. Kluwer Academic Publishing, Norwell, DNA Sequencing, pp. 265–296.
- Yang, Z., Bielawski, J.P., 2000. Codon-substitution models for detecting molecular adaptation at individual sites along specific lineages. *Mol. Biol. Evol.* 19, 908–917.
- Ziegler, P., 1999. Cereal beta-amylase. *J. Cereal Sci.* 29, 195–204.