



OPEN ACCESS

EDITED BY

Mi-Jeong Yoo,
Clarkson University, United States

REVIEWED BY

Hannes Becher,
University of Edinburgh, United Kingdom
Zhiqiang Wu,
Kunpeng Institute of Modern Agriculture at
Foshan, China

*CORRESPONDENCE

Katarzyna Krawczyk
✉ katarzyna.krawczyk@uwm.edu.pl

RECEIVED 09 August 2023

ACCEPTED 23 November 2023

PUBLISHED 08 December 2023

CITATION

Krawczyk K, Paukszto Ł, Maździarz M
and Sawicki J (2023) The low level of
plastome differentiation observed in
some lineages of Poales hinders
molecular species identification.
Front. Plant Sci. 14:1275377.
doi: 10.3389/fpls.2023.1275377

COPYRIGHT

© 2023 Krawczyk, Paukszto, Maździarz and
Sawicki. This is an open-access article
distributed under the terms of the [Creative
Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The
use, distribution or reproduction in other
forums is permitted, provided the original
author(s) and the copyright owner(s) are
credited and that the original publication in
this journal is cited, in accordance with
accepted academic practice. No use,
distribution or reproduction is permitted
which does not comply with these terms.

The low level of plastome differentiation observed in some lineages of Poales hinders molecular species identification

Katarzyna Krawczyk*, Łukasz Paukszto, Mateusz Maździarz
and Jakub Sawicki

Department of Botany and Nature Protection, University of Warmia and Mazury in Olsztyn,
Olsztyn, Poland

Chloroplast genomes are a source of information successfully used in various fields of plant genetics, including molecular species identification. However, recent studies indicate an extremely low level of interspecific variability in the plastomes of some taxonomic groups of plants, including the genus *Stipa* L., which is a representative of the grass family. In this study we aimed to analyze the level of chloroplast genome diversity within particular genera as well as the effectiveness of identifying plant species in the Poaceae family and the other representatives of Poales order. Analysis of complete plastid genome alignments created for 96 genera comprising 793 species and 1707 specimens obtained from the GenBank database allowed defining and categorizing molecular diagnostic characters distinguishing the analyzed species from the other representatives of the genus. The results also demonstrate which species do not have any species-specific mutations, thereby they cannot be identified on the basis of differences between the complete chloroplast genomes. Our research showed a huge diversity of the analyzed species in terms of the number of molecular diagnostic characters and indicated which genera pose a particular challenge in terms of molecular species identification. The results show that a very low level of genetic diversity between plastomes is not uncommon in Poales. This is the first extensive research on super-barcoding that tests this method on a large data set and illustrates its effectiveness against the background of phylogenetic relationships.

KEYWORDS

chloroplast genome, super-barcoding, Poaceae, Poales, species identification, plastome variability

1 Introduction

Plastid genomes are widely employed in plant evolutionary research, encompassing areas such as phylogenetics, taxonomic implications, population genetics, and phylogeography (Huang et al., 2019) (Abdullah et al., 2021). The reason for which plastomes are willingly used in this applications are the same that predestinate it to be applied in molecular identification of species, i.e. mostly uniparental inheritance, relatively high level of polymorphism, and presence in multiple copies in a cell, which is of particular importance in the analysis of material with limited availability and herbaceous material (Palmer, 1985; Henriquez et al., 2014; Ramsey and Mandel, 2019). The method of species identification based on variability of nucleotide sequences, called DNA barcoding, has been developing dynamically since 2003 (Hebert et al., 2003) and in the case of its application in plants, from the very beginning (Kress et al., 2005) the main subject of research as a source of potential barcode sequences was the chloroplast genome. Looking for the most effective, universal and at the same time easy to amplify barcode sequence, the variability of many coding and non-coding candidate barcodes within the plastome was analyzed over the years. Initially, loci widely used in phylogenetic implications were tested for their usefulness in distinguishing representatives of individual plant species, among others: *trnH-psbA* (Kress et al., 2005), *trnL* (Taberlet et al., 2007), *atpF-atpH*, *matK*, *rbcL*, *rpoB*, *rpoC1*, *psbK-psbI* (CBOL Plant Working Group1 et al., 2009), *atpF-atpH* (Lahaye et al., 2008), *trnL-trnF* (Ferri et al., 2015), *accD*, *ndhJ* (Ford et al., 2009). In the following years of research, the utility of numerous regions within the plastome was investigated, also including its fragment-by-fragment variability, regardless of function (Xu et al., 2017; Krawczyk et al., 2018; Jiao et al., 2019; Šlipiko et al., 2020; Sawicki et al., 2021).

Numerous studies have shown that unfortunately none of candidate sequences meet the requirements of the universal barcode for plants, therefore a research trend focused around searching for the best combination of two or more single loci in the so-called multi-locus barcodes (Hollingsworth et al., 2011; Li et al., 2015). Currently, the gold standard in barcoding, the core-barcode with the highest discriminatory power, recommended by the CBOL Plant Working Group (CBOL Plant Working Group1 et al., 2009) is the *matK-rbcL* combination, which length in *Arabidopsis thaliana* oscillates around 1440 bp (Hollingsworth et al., 2011). Again, it is not ideal, mainly due to the low PCR efficiency of *matK* amplification in some taxonomic groups, especially gymnosperms and cryptogams (CBOL Plant Working Group1 et al., 2009). In the absence of a universal, highly effective barcode for all plants, some researchers have turned to developing specific barcodes, designed for lower taxonomic groups - such as families or even genera, forgoing the universality of the method in favor of its effectiveness (Saddhe and Kumar, 2017; Zhang et al., 2021). However, the most prospective research trend that is currently developing seems to be the super-barcoding (Parks et al., 2009), the method which analyzes the variability of complete plastid genomes. Thanks to the huge development of high-throughput sequencing methods in recent years, and also the

reduction of their cost, the analysis of whole-plastome variability is increasingly used in various studies, also including molecular identification of closely related land plant species (Kress, 2017; Fu et al., 2019; Chen et al., 2020; Li et al., 2021; Zhang et al., 2021; Guo et al., 2022) or even varieties within species (Ji et al., 2020; Jiang et al., 2023). The application of super-barcodes helps to increase efficiency of species identification and at the same time circumvents possible issues with low PCR efficiency (Li et al., 2015; Jiang et al., 2023).

The analysis of the complete plastome instead of standard barcodes, which individual length does not exceed 1 kb, may seem a little bit like using a sledgehammer to cut a nut, nevertheless, this approach is fully justified, as despite the high discriminative power of super-barcodes, in the case of some recently divergent species even they are not 100% effective. An example is the research on *Fritillaria* (Liliaceae) where super-barcode provided discriminatory power at the level of 87.5% (Wu et al., 2021). Much lower effectiveness of super-barcoding was reported for the Lauraceae family, where the use of the whole plastome sequence increased the efficiency of species identification only to 60% while the discrimination success obtained with the standard plant DNA barcodes was 40-50% (Liu et al., 2021). Another taxonomic group, which is a challenge in terms of both classical and molecular identification of species, due to high variability of features within genera, is the family of grasses (Poaceae) (Hand et al., 2013; Qiu et al., 2019). Studying the genetic variability of feather grasses (*Stipa* L. Poaceae), a strikingly low level of variability in plastome sequences between species pairs amounting from 0% to 0.3% (mean 0.01%) was observed (Krawczyk et al., 2018; Krawczyk et al., 2022). Confronting this result with significantly higher mean nucleotide variability values observed in *Taxillus* (Loranthaceae) (1.3%), *Dalbergia* (Fabaceae) (0.86%) or *Populus* (Salicaceae) (0.36%) plastomes (Su et al., 2021), the question arises whether the low level of differentiation observed in *Stipa* is a common phenomenon within grasses or whether it is exceptional in this taxonomic group. Despite the great sequence similarity of plastome genomes in feather grasses, the identification success in the genus reached 94.74% of species analyzed in the study (Krawczyk et al., 2018). This observation raises another question, whether low genetic variability translates into low success of molecular species identification. These questions seem to be relevant in the context of evolutionary genomics itself, but also from the point of view of the molecular species identification method. As in recent years apart from the methods based on the phylogenetic tree or genetic distance in barcoding studies, the development of methods based on molecular diagnostic characters (MDCs) (Merckelbach and Borges, 2020) can be observed, the need arises to study the effectiveness of this tool in different taxonomic groups so that in the future it can be determined what number of MDCs may be regarded a success of species identification.

To analyze the effectiveness of the super-barcoding method on a wide set of data, as well as taking into account that the essence of the barcoding method is the use of sequence resources available in public databases, data available in the GenBank database (Sayers et al., 2022) were used in the study.

This study aims to examine and compare the level of chloroplast genome diversity within particular genera as well as the effectiveness of identifying plant species in the Poaceae family against the background of the Poales order in relation to their phylogenetic relationships. The significant advantage of the availability of data on the chloroplast sequences of grasses over other families from Poales is the result of the emphasis placed in research on plants from the Poaceae family, due to the fact that grasses are important in the food, fodder, bioenergy, construction and horticultural industries, are the main component of meadows and bamboo forests, therefore are of great economic and ecosystem importance (Hodkinson et al., 2007; Soreng et al., 2017).

Recently published studies on the evolution of Pooideae indicated that one of the most important factors influencing the evolution of this group was the transition from closed to open habitats, that happened at the turn of Eocene and Oligocene. This event was associated with significant upshift of diversification rate in three Pooideae sublineages: core Pooideae, supertribe Stipodeae and supertribe Nardodae (Zhang et al., 2022). Moreover, the results of research showed that the occupation of open habitat might have subsequently facilitated the radiation and expansion of Pooideae. Taking into account the impact of the change of the habitat on the diversification of this taxonomic group, we also decided to analyze whether the currently occupied habitat type by species from the studied genera from Poales is related to the currently observed level of genetic diversity within the genus.

2 Materials and methods

2.1 Sequence acquisition and alignment

The data used in this research comes entirely from the GenBank database (Sayers et al., 2022), therefore the data collection is conditioned by the availability of complete chloroplast genome sequences in this public, open access database. The sequences were downloaded from GenBank in February 2022, and their accession numbers are given in Supplementary Table 1. In total, 1707 chloroplast genomes from 793 species representing 96 genera from Poales order were analyzed in the study. The vast majority of them, i.e. 90 genera, represent the Poaceae family, while Bromeliaceae and Cyperaceae are represented by two genera and Eriocaulaceae and Typhaceae only by one genus, respectively. Only those genera for which data on complete plastomes were available for at least three species were included in the analysis, hence the smallest number of species in the genus in the studied data set is three. The richest genus in terms of the number of analyzed species was *Andropogon*, for which plastome sequences were available for as many as 41 species. The analyzed data set is also very diverse in terms of the number of examined individuals per species ranging from one to 115 (*Brachypodium distachyon*). In the case of genera represented by very numerous sequences (*Oryza rufipogon* and *O. sativa*), the dataset used for further analysis was limited to 50 sequences due to the computational capabilities of the software

used. Before comparative analyses, the IR regions for all the plastome sequences were identified and removed by IRScope (Amiryousefi et al., 2018) and seqinr R package (Charif and Lobry, 2007). The MAFFT (Katoh and Standley, 2013) software with default parameters was used to create multiple sequence alignment (MSA) for each genus. For each of the alignments created for the genus, mean pairwise sequence identity expressed as a percentage and representing the proportion of identical residues in the aligned sequence pairs was calculated. From each MSA for a genus consensus sequence was generated. The consensus sequences of each dataset were collected and realigned to obtain the final MSA for Poales. The both types of MSAs were used for downstream analyses, the genera MSA for MDCs identification and final MSA for phylogeny reconstruction. The phylogenetic Poales tree was built using the Maximum Likelihood (ML) method implemented in the IQ-TREE 2 (Minh et al., 2020). The *Sparganium* consensus sequence was selected as a root of the tree.

2.2 The analysis of molecular diagnostic characters

The MDCs analysis was performed using the FastaChar v. 0.2.4 software (Merckelbach and Borges, 2020). The analysis was performed by comparing each species with the other members of the same genus included in the study. As a result, a list of diagnostic features of individual species was obtained, generated in the form of spreadsheets. The FastaChar enables analysis of molecular diagnostic characters (MDCs) defined as so-called pure diagnostic sites (Sarkar et al., 2008) which are polymorphic sites (commonly SNPs) in the nucleotide alignment for which all the members of the query taxon have a given variant that is not present in any member in the reference taxa. Sites which are polymorphic within the query taxon were not included in the analysis. The number of MDCs for each of the analyzed species was compiled and the average number of MDCs for the genus was calculated (Supplementary Table 2). This value was also calculated in relation to the length of the alignment.

MDCs obtained from FastaChar software were evaluated by custom R script. In the first part of the script, the algorithm focused on converting xls results to data frame format and combining data from each genus to one matrix. After the preparation step, the script assumes identification and classification of MDCs to different types of variations. The main criteria of MDCs classification were length of consecutive nucleotides within the MDCs event and type of substitution or indel. According to that assumption the data was divided into deletion, insertion, single mixed variation, multi substitution, multi deletion, multi insertion and multi mixed variation. MDCs with more than one continuous nucleotide were characterized as multi variations. The mixed type was described in case of occurrence substitution and indel within a single MDCs event in the genera.

The minimum, maximum and mean of MDCs, family membership, habitats and percentage of identity within genera

were visualized on circular plots created by R bioconductor packages, e.g. ggtreeExtra (Xu et al., 2021), phyloseq (Revell, 2012), circlize (Gu et al., 2014), ggtree (Yu et al., 2017).

2.3 Statistical analyses

Considering the differentiated number of species analyzed within individual genera, we checked whether there was a relationship between the number of species studied in the genus and the sum of MDCs in a genus. It was also tested whether the obtained results were significantly biased by the number of individuals analyzed within each species. The distributions of the studied dependent variables were tested for compliance with the normal distribution using the Shapiro-Wilk test. In all cases (sum of MDCs in a genus; MDCs per alignment length $\times 10^{-4}$; average number of MDCs in a genus), the distributions of the result variables were significantly deviated from the normal distribution (Shapiro-Wilk test, $p < 0.00001$) and positively skewed (skewness > 2.14). Therefore, in the analysis of the relationship between the studied variables and the predictors, generalized linear models (GLM) were used, allowing for testing the relationships between the predictors and the outcome variable. For the result variables (sum of MDCs in a genus; MDCs per alignment length, average number of MDCs in a genus) belonging to the count type, the Poisson distribution was adopted. Data describing the number of species in the genus, the number of individuals in the genus and mean pairwise sequence identity within a genus were logarithmic scaled. Statistics were calculated according to the algorithms of the Statistica 13.3 package.

The representativeness of the analyzed data was also examined by verifying what percentage of species distinguished within a given genus covers the data set we analyzed. When determining the number of species belonging to each genus, the list of species on the World Flora Online database (World Flora Online, 2023) was used.

2.4 MDCs in relation to the habitat type

Each genus analyzed in the study has been characterized in terms of the dominant type of habitats occupied by the representing species. The genera were assigned to one of three habitat categories: open, woodland or mixed basing on the available databases and publications (Florabase, 1998; AusGrass2, 2014; Grass Base - The Online World Grass Flora, 2014; Useful Tropical Plants Database, 2014; Nowak et al., 2020; Coastal Plain Plants, 2021; Lady Bird Johnson Wildflower Center, 2023; Native Plant Trust, 2023; The PLANTS Database et al., 2023; World Flora Online, 2023). The open habitats category includes grasslands, sand dunes, fields, wetlands, grassy hillsides and mountain slopes. Species occurring in the forest undergrowth as well as plants forming bamboo forests were classified to the woodland category. The mixed category includes species occurring in forest margins, savannas, thickets, ruderal habitats and species growing in both woodland and open areas.

The comparison of the mean pairwise sequence identity [%] calculated for each genus with the distinguished type of environment was tested with the Kruskal-Wallis test, and the differences between pairs of studied variables were tested with Dunn's multiple comparisons test.

3 Results

3.1 MDC against the background of the phylogenetic tree

The length of the 96 analyzed sequence alignments created for each particular genus ranged from 113,256 bp in *Eremopyrum* to 144,526 bp in *Carex* (Supplementary Table 2). Phylogenetic reconstruction based on consensus sequences generated from these alignments resulted in a well resolved phylogenetic tree (Supplementary Figure 1), with family Poaceae clearly divided into two sister clades, one comprising species from Bambusoideae, Oryzoideae and Pooideae subfamily (BOP clade) and the other one comprising representatives of Panicoideae, Aristidoideae, Chloridoideae, Micrairoideae, Arundinoideae and Danthonioideae (PACMAD clade), while the inner clades comprising individual subfamilies were highly supported.

The number of MDCs that distinguish a species from the other representatives of the same genus was extremely variable ranging from zero in numerous species to 8,672 in *Carex siderosticta*. The *Carex* genus was also characterized by the highest sum of MDCs calculated for the whole genus reaching 32,195. An average number of MDCs in a genus ranged from zero in *Secale* to 3,071 in *Eriocaulon*. (Supplementary Table 2).

A significant positive association was found between the sum of MDCs and the number of species in the genus [parameter = 0.028 ± 0.0002 (SEM), Wald stat. = 18036.572, $p < 0.001$; model: log-likelihood function = -168023.665, $\chi^2 = 15872.17$, $p < 0.00001$] (Figure 1). This indicates that, in general, an increase in the number of species analyzed within a genus results in the ability to define more MDCs. In turn, a significant negative relationship was found between the cumulative number of MDCs normalized to the alignment length and the number of individuals in the species [parameter = -0.244 ± 0.007 (SEM), Wald stat. = 1074.946, $p < 0.001$; model: log-likelihood function = -29290.708, $\chi^2 = 2166.708$, $p < 0.00001$] (Figure 2). This implies that a higher number of individuals studied within a genus leads to a reduced capacity to define MDCs.

The number of species studied within individual genera in some cases was highly representative, reaching up to 100% of species distinguished within the genus (*Ampelocalamus*, *Avena*, *Campeioestachys*, *Eremopyrum*, *Littledalea*, *Orinus*, *Oryza*, *Tripidium*), although, in the 73 analyzed genera, the availability of data allowed us to analyze less than half of species described (Figure 3; Supplementary Table 2). However, the representativeness of the analyzed data set in the individual genera does not coincide with their variability, expressed by the number of MDCs (Figure 3).

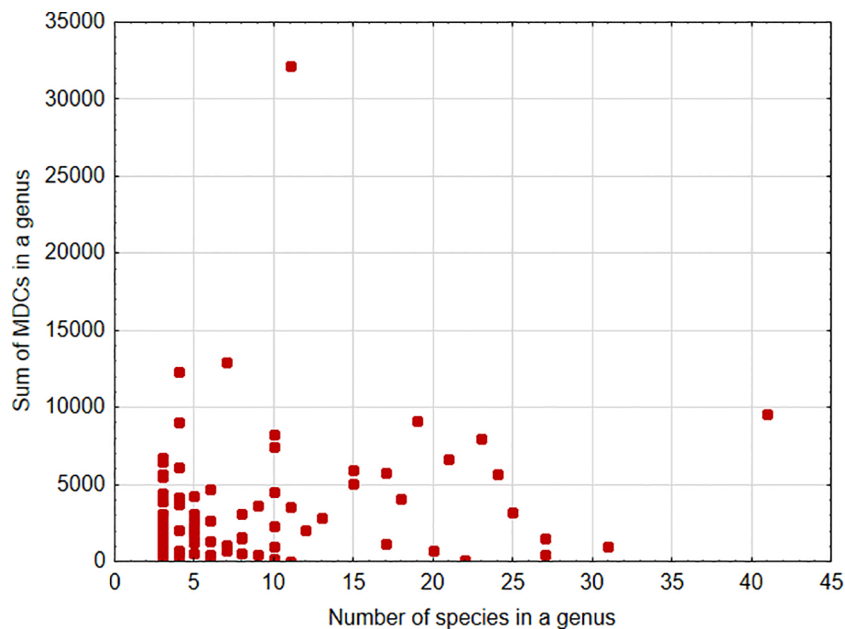


FIGURE 1
Sum of MDCs calculated for the genus versus number of species analyzed within the genus.

The obtained results showed that in nearly half (46%) of the 794 species over 100 species-specific MDCs were observed (Figure 4). On the other hand 93 species (12%) had only from one to ten MDCs in the whole chloroplast genome. Moreover, as many as 132 species (17%) did not have any MDC, therefore they couldn't be identified on the basis of differences in the complete cpDNA sequence (Figure 4). The analysis of the observed number of MDCs in

individual genera showed that the studied genera turned out to be very diverse in terms of the share of species in which not a single MDC was defined. (Figure 5). In 61 out of 96 analyzed genera, MDCs were observed in each of the studied species. The share of the remaining 35 species without private SNPs ranged from 5% (*Eragrostis*) to 91% (*Triticum*) or even 100% (*Secale*) (Supplementary Table 2). It is worth noting that in the case of 15

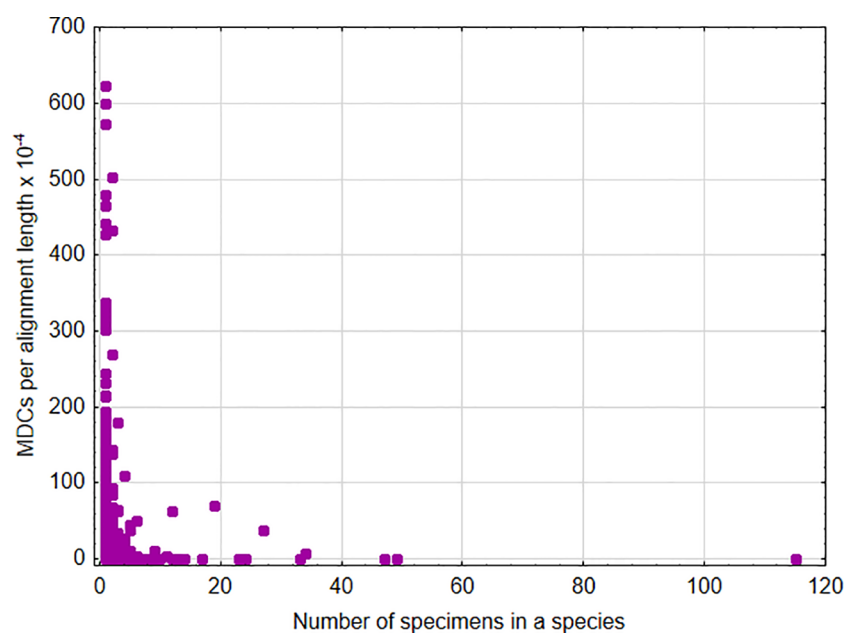


FIGURE 2
Number of MDCs observed in a species normalized by the length of alignment versus number of analyzed specimens in a species.

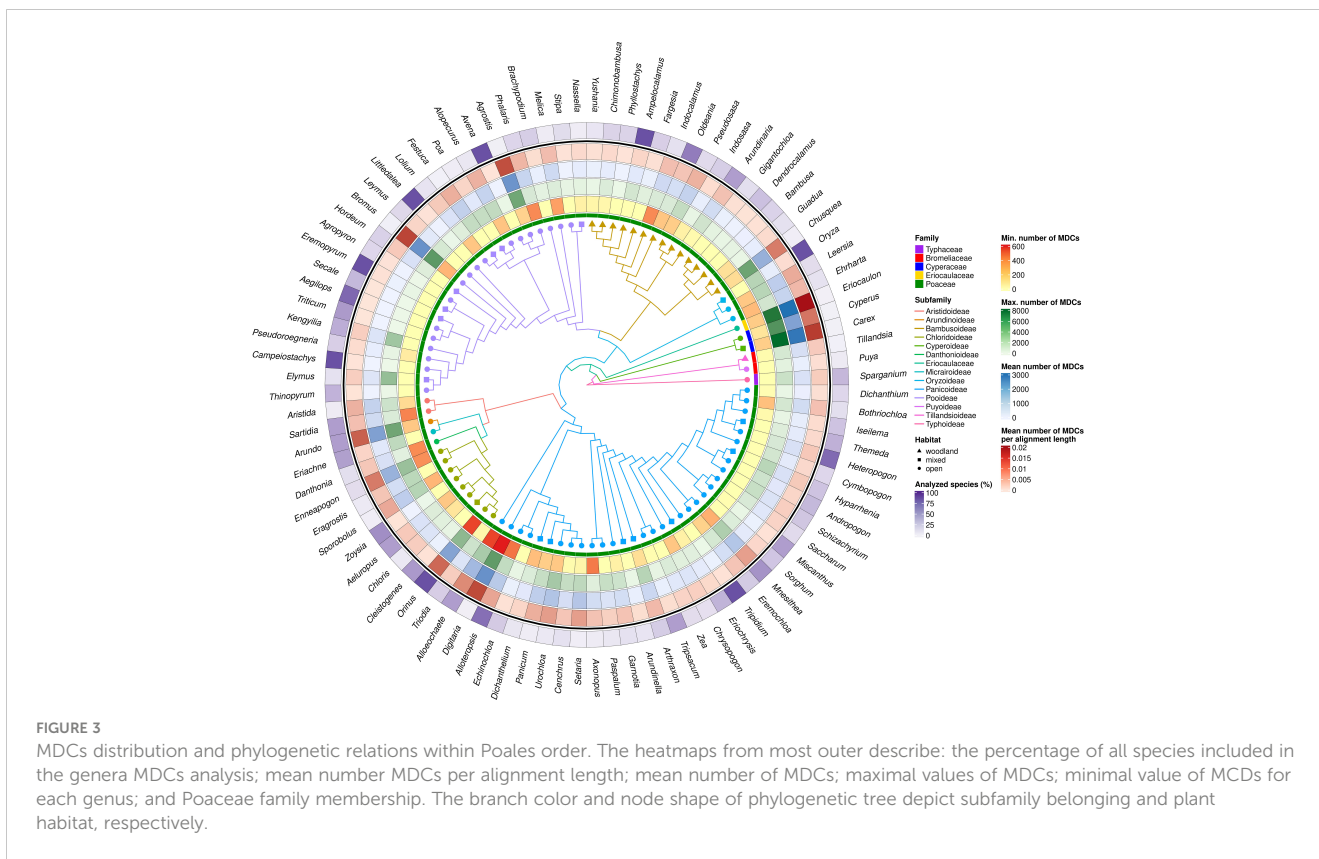


FIGURE 3
MDCs distribution and phylogenetic relations within Poales order. The heatmaps from most outer describe: the percentage of all species included in the genera MDCs analysis; mean number MDCs per alignment length; mean number of MDCs; maximal values of MDCs; minimal value of MDCs for each genus; and Poaceae family membership. The branch color and node shape of phylogenetic tree depict subfamily belonging and plant habitat, respectively.

genera, the share of species impossible to identify exceeded 25% (Figure 5). In terms of percentage of species with no MDCs in a genus, *Stipa* ranked among the 13 genera most difficult to identify (Supplementary Table 2).

Analyzing the distribution of genera with no MDCs on the phylogenetic tree, it can be seen that some of them with an exceptionally high share of species without MDCs: *Secale*, *Triticum*, *Aegilops*, *Campepistachys*, *Thinopyrum*, *Eremopyrum*,

Elymus, *Kengyilia* belong to the phylogenetically youngest part of the clade formed by the subfamily Pooideae (Figure 3). Also the genera *Miscanthus*, *Saccharum* and *Sorghum* are closely related and form a single clade within the subfamily Panicoideae. In addition, in its sister clade, there are also present genera in which some species do not have MDCs, namely: *Schizachyrium*, *Andropogon*, *Hyparrhenia*, *Heteropogon* and *Themeda*. A similar picture can be seen in the Bambusoideae for *Gigantochloa*, *Dendrocalamus* and

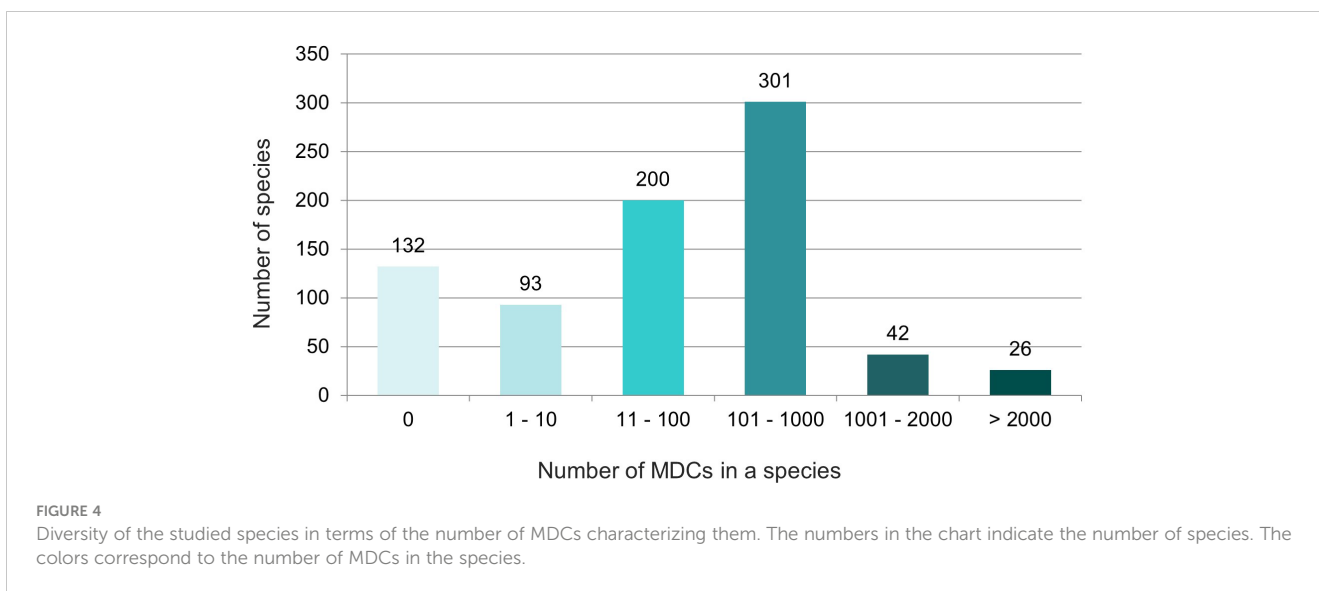


FIGURE 4
Diversity of the studied species in terms of the number of MDCs characterizing them. The numbers in the chart indicate the number of species. The colors correspond to the number of MDCs in the species.

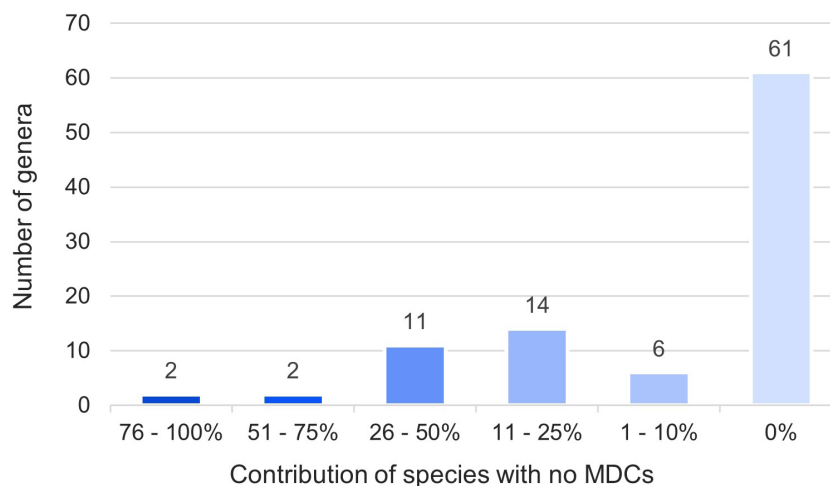


FIGURE 5

Share of species in a genus impossible to identify. The numbers in the chart indicate the number of types in a given range. The colors correspond to the percentage of species with no MDCs in each genera.

Bambusa. The other genera with no MDCs detected are found in different parts of the tree, without a clear relationship with phylogeny (Figure 3).

In turn, the studied genera characterized by high interspecific diversity, and therefore with high mean number of MDCs, but above all a high minimum MDCs number, belong to the Cyperaceae and Eriocaulaceae families, but are also observed in the basal part of the Panicoideae clade (*Alloeochoete*, *Digitaria*, *Alloteropsis*). The remaining cases of genera characterized by high MDCs level are located in different parts of the phylogenetic tree.

3.2 Mean pairwise genetic identity versus MDCs number

A significant negative relationship was found between the average number of MDCs in the genus and % pairwise identity [parameter = $-0.055 \pm 0.0004(\text{SEM})$, Wald stat. = 16454.907, $p < 0.00001$; Model: log-likelihood function = -29127.302 , $\chi^2 = 10773.301$, $p < 0.00001$] (Figure 6). This result clearly shows that genetic distance calculated as a percentage of mean pairwise sequence variability within a genus largely goes together with the

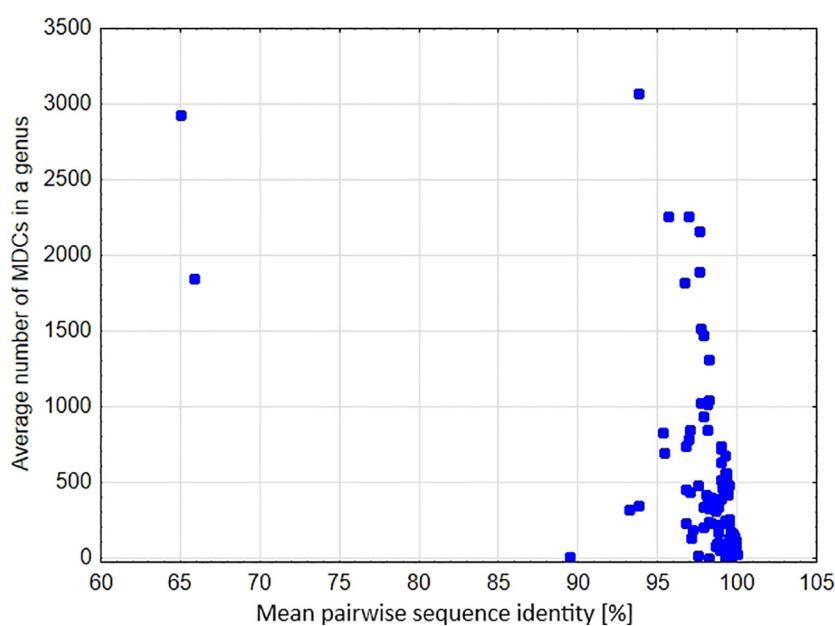
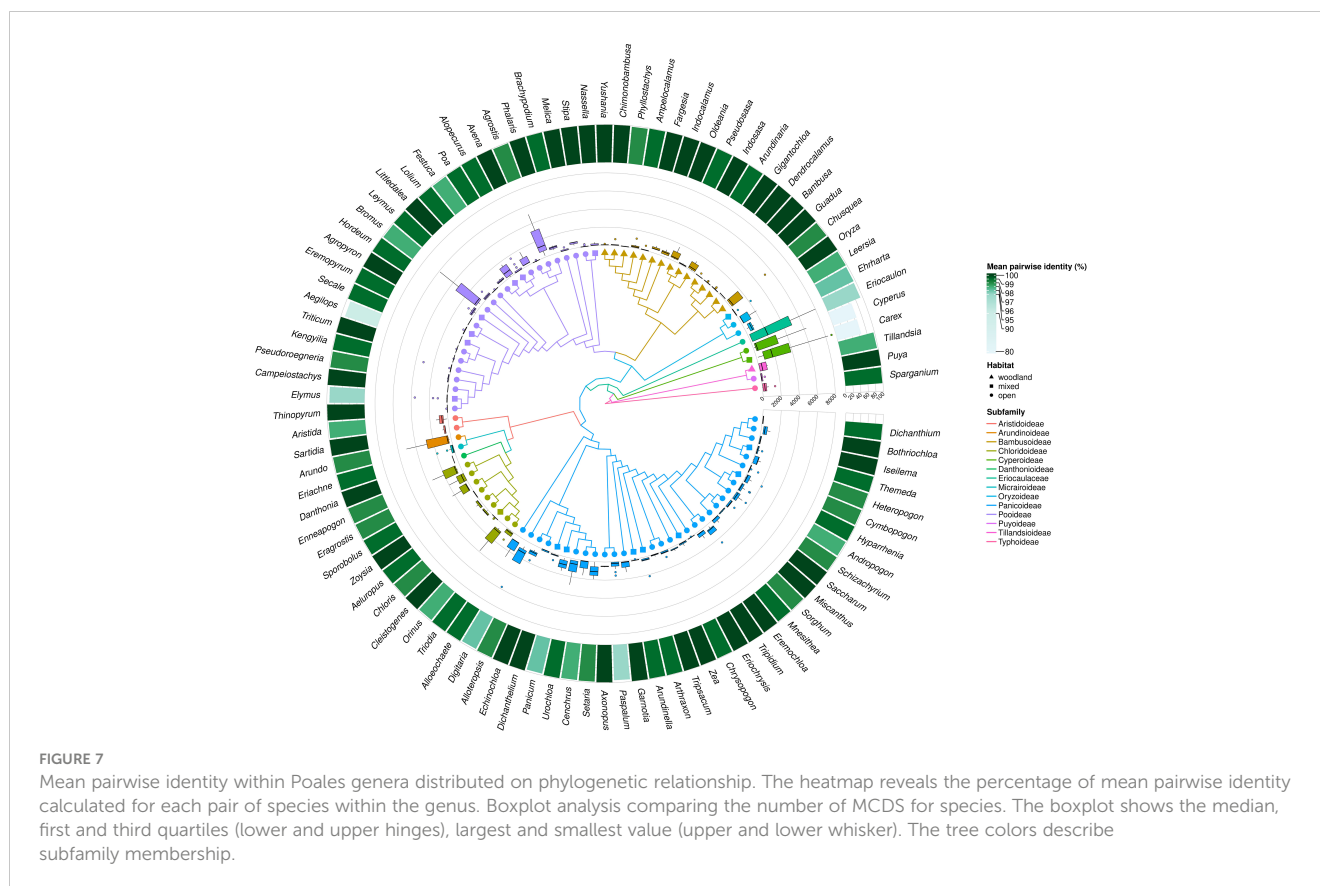


FIGURE 6

Mean pairwise sequence identity in a genus versus the average number of MDCs calculated for the genus.



effectiveness of molecular identification of species in a genus. This is also perfectly visible in the example of the genera *Carex*, *Cyperus* and *Eriocaulon*, where high variability of the aligned sequences coincide with high mean and maximum values of MDCs in a genus (Figure 7). In turn, there are cases of the opposite situation, as in *Aegilops*, *Elymus* and *Paspalum*, where the high genetic differentiation of sequences does not correspond with identification efficiency. On the other hand, some genera characterized by high sequence similarity: *Phalaris* (99.2%), *Axonopus* (99.3%), *Danthonia* (99.4%) had high minimal mean MDCs numbers (Figures 3, 5, 7), thanks to which molecular identification of their species was not a challenge. A surprising situation was also observed in three genera from the subfamily Bambusoideae: *Guadua*, *Indosasa* and *Yushania*, where despite 99.9% sequence identity, the success of species identification in each of these genera was 100%, moreover, the number of MDCs for the species was in the range of 107-132, 46-216 and 32-149, respectively.

3.3 Types of mutations generating MDCs

Among species-specific mutations generating MDCs in barcode analysis there were substitutions, single and multi-nucleotide insertions or deletions. Moreover, when comparing query species to other representatives of the genus, a situation was also observed when in query species in a given alignment position there was for example nucleotide A, while in other species there was C, G or T or

a deletion in this place, so this type of MDCs we defined as mixed variation. The analysis of the type of MDCs observed in the studied data set showed that in most species (65%) the dominant source of MDCs were indel and mixed variations, in 17% of species these were substitutions, while in 18% of cases their share was equal (Supplementary Table 3). The predominance of MDCs of substitution type generally applies to a few species in the genus, however, in the case of *Sparganium*, *Ehrharta* and *Leersia*, their share was predominant in 67% of species of the genus, and in *Puya* and *Bambusa* in 50% of species. In the Poaceae family, the genus *Stipa* stands out in this respect, in which substitutions prevailed in 29% of the analyzed species.

The analysis of the number of MDCs generated by different types of mutations showed that 27.1% of them are in substitution type while as much as 68.1% of observed MDCs in the data set came from multi-nucleotide indels or mixed variations. Detailed data on the contribution of each of the types of mutations defined as MDCs for analyzed species are provided in Supplementary Table 3.

3.4 Type of habitat

The analysis of the level of genus variability in relation to the type of habitat: open, closed or mixed, occupied by species representing a given genus showed that there was a significant difference in mean pairwise identity between the types of environments (Kruskal-Wallis test: $H(2, N = 96) = 6.936$; $p = 0.031$), with the difference relating to the mean pairwise identity between the open and woodland

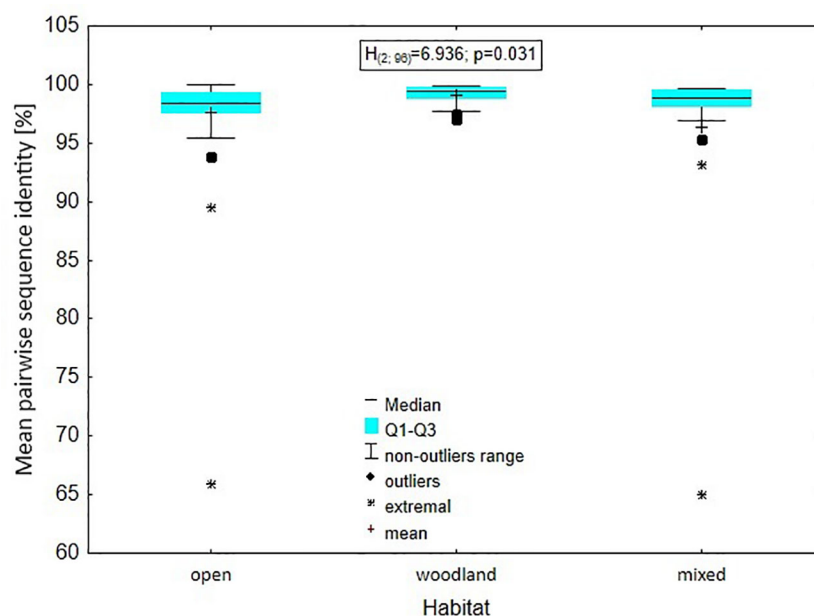


FIGURE 8

Level of genus variability in relation to the type of habitat: open, closed or mixed, occupied by species representing a given genus.

environments. A higher percentage was found in the woodland environment (Dunn test: $z = 2.613$, $p = 0.027$). Between mixed and woodland this difference was within the inference error (Dunn test: $z = 1.911$, $p = 0.168$); the situation was similar when testing the differentiation between open vs. mixed environments (Dunn test: $z = 0.242$, $p = 1.000$) (Figure 8).

4 Discussion

4.1 Efficacy of species identification within Poales

The results show that a very low level of genetic diversity between plastomes is not uncommon in Poales, since in 18 out of 96 examined genera mean pairwise sequence identity exceeded 99.5%. Furthermore, many of the species studied are impossible to distinguish even when analyzing the entire chloroplast genome due to the lack of species-specific mutations. A particularly large number of cases where super-barcoding failed to discriminate species have been observed in the Poaceae family. Representatives of the remaining families included in this study were generally characterized by high variability of chloroplast genomes, on the basis of which the molecular identification of species was successfully carried out. Although in the genus *Puya* from the family Bromeliaceae and in *Sparganium* from the Typhaceae, the identification success was also not complete.

Unfortunately, application of super-barcodes to identify species from Poales is not yet very common, which hinders a broader discussion on interspecific variability of this group of plants. However, entire chloroplast genome sequences are being used to reconstruct evolutionary history at the genus level. So far, this type

of research has been carried out for the *Bambusa* genus (Wysocki et al., 2015), *Festuca* (Hand et al., 2013; Qiu et al., 2019), *Lolium* (Hand et al., 2013), and *Stipa* (Krawczyk et al., 2018). Data on the variability of chloroplast genomes within these genera correspond well with the results of our study reporting similar values of genetic diversity.

4.2 Number of MDCs against the background of phylogenetic relationships

The result of phylogenetic reconstruction inferred in our study is only fragmentary, due to the availability of data for analysis, nevertheless, the presented phylogenetic tree is well-resolved and congruent in topology with recent phylogenetic studies on Poales (Saarela et al., 2015; Soreng et al., 2017; Gallaher et al., 2022; Wu et al., 2022). Our study shows that the level of genetic diversity and number of species-specific mutations are extremely variable within different lineages of Poales. There may be many reasons for this phenomenon as the mechanism of plastome evolution is very complex, including: gene losses, gene duplications, gene rearrangements, gene transfers between organelles, horizontal gene transfer, repeat sequences and point mutations (Bennetzen, 2007; Jansen and Ruhlman, 2012; Xu et al., 2015; Wu et al., 2022).

Within Poales, the most dynamically evolving group was the evolutionary branch leading to the grass family. The result of the accelerated rate of evolution was that the common ancestor of the Poaceae family lost from plastome introns in two genes: *clpP* and *rpoC1* and three coding regions: *accD*, *ycf1* and *ycf2* (Guisinger et al., 2010; Xu et al., 2015; Liu et al., 2020). The lack of these genome fragments, especially the *ycf1* gene, which is particularly useful in molecular identification of species of land plants, being a

source of numerous species-specific mutations (Dong et al., 2015), may be one of the reasons for the generally low interspecific diversity of plastomes observed in the grass family.

The mutation rate may also be influenced by selective pressure, which is varied in individual evolutionary lines of Poales (Guisinger et al., 2010; Piot et al., 2018; Krawczyk et al., 2022). Literature data also suggest that accelerated rates of both genome rearrangements and nucleotide substitutions in the chloroplast genome may possibly be caused by aberrant DNA repair mechanisms which result from mutations in nuclear-encoded DNA repair and/or replication genes (Jansen et al., 2007; Guisinger et al., 2010; Guisinger et al., 2011).

4.3 Super-barcoding in Poales versus other groups of plants

So far, there are also not many publications reporting the number of species-specific mutations in complete plastome sequences. Although the super-barcoding method is gaining in popularity as well as usage of molecular identification of species by defining MDCs, these two methodological approaches rarely go hand in hand in plant research. Regarding plastomes, we have scant data on selected liverworts, which show that between two species representing the genus *Marchantia* (*M. polymorpha* and *M. palacea*) 4,076 MDCs have been defined, while in the genus *Conocephalum* (*C. conicum* and *C. salebrosum*) 5,878 were observed (Sawicki et al., 2020). In turn, in the genus *Pellia* and *Apopellia* the number of detected plastid MDCs varied from 0 to 4,717 (Pauksztó et al., 2023), including 159–1,369 MDCs identified for cryptic species within the genus *Apopellia*. Comparison of these results with those obtained for Poales, where the highest average value of MDCs for the genus was 3,071 (in *Eriocaulon*), while the highest observed value for the species was over 32 k bp (in *Carex*), shows that the number of MDCs in Poales is extremely volatile, albeit at a fairly low level on average.

In the case of vascular plants, we can refer to data on the comparison of three species representing the genus *Pulsatilla*, where species-specific mutations ranged from 56 to 432 (Szczeńska and Sawicki, 2015). In the study on *Pulsatilla*, for each species, the number of specific substitutions was at least twice the number of species-specific indels. In Poales we observed almost three-fold advantage of indel and mixed MDCs over substitutions, as well as the overwhelming advantage of long indels over single ones, which is a good reflection of the rapid evolutionary changes that shaped the variability of Poales representatives. Long indels, which generate most of observed MDCs, have a particularly large impact on the efficiency of species identification within this group of plants.

4.4 Reasons for the low effectiveness of super-barcoding

Observations of low success of plant species identification with super-barcoding in some cases can be explained by the high level of genetic sequence similarity within a genus, however the variability

of plastomes is not always consistent with their diagnostic capability in molecular fingerprinting. This means that despite a large share of variable sites found within a plastome sequence, it may be characterized by a low share of private mutation sites, and thus - poor efficiency of species identification. Therefore, variability does not always prove the effectiveness of distinguishing taxa, because it does not have to be species-specific.

When looking for the reasons for the low efficiency of molecular identification of species by DNA barcoding in selected groups among Poaceae, it should be noted that many species in this group evolved by polyploidization, since most grass species examined against ploidy are polyploids (Rice et al., 2015). The phenomenon of polyploidization can lead to problems related to the identification of species by DNA barcoding based on the plastid genome. This is due to the fact that the newly formed autopolyploids are characterized by an identical chloroplast genome as the parent species. Differences within the plastid genome may accumulate over time, however, at least in the early generations following the polyploidization process, the aforementioned difficulties in species identification will occur. Recently formed allopolyploids will, in turn, have the chloroplast genome of only one of the species, which makes it impossible to distinguish them from the taxon from which the given plastid genome originates using the DNA barcoding method (Cowan and Fay, 2012). Unfortunately, for the vast majority of species studied in this study, the literature does not provide the number of chromosomes on the basis of which the contribution of polyploid species to the genus could be assessed. Thus, we do not have a complete picture of the share of polyploid species in individual genera. However, based on the data available in The Chromosome Counts Database (CCDB) (Rice et al., 2015), it can be concluded that the genera in which the success of species identification is particularly low include species with varying degrees of polyploidy. For example, the genus *Triticum* includes di-, tri-, tetra- and hexaploid species, the genera *Aegilops* and *Avena*: di-, tetra-, hexaploids, and *Oryza* di-, tetra-, hexa- and octoploids.

Another important mechanism of species evolution that reduces the effectiveness of species identification based on the comparison of chloroplast sequences is the phenomenon of hybrid origin of species and introgression across species barriers (Soltis and Soltis, 1991; Li et al., 2015). There are numerous known cases of both ancient and recent hybridizations and subsequent genomic introgressions between grass species, to name just a few genera analyzed in the study: *Lolium* (Jenkins, 1986), *Phyllostachys* (Lin et al., 2010), *Stipa* (Nobis et al., 2019; Baiakhmetov et al., 2020), *Urochloa* (Tomaszewska et al., 2023), *Aegilops*, *Secale*, *Triticum* and *Sorghum* (Kole, 2011).

An important factor hindering species identification based on the plastome may be a phenomenon of chloroplast capture through interspecies hybridization, which yields a plant with a new genetic combination of nuclear and chloroplast genomes (Rieseberg and Soltis, 1991). Hybridization phenomena and probable plastome capture have probably occurred many times in Poales, such as in the Triticeae between *Thinopyrum* and *Pseudoroegneria* or *Aegilops* and *Triticum* (Petersen et al., 2006; Bernhardt et al., 2017), as well as between species within the genus *Stipa* (Nobis et al., 2019) or *Zea*

(Morales-Briones et al., 2018; Wu et al., 2022). Another reason for the whole plastome lateral transfer between species may be asexual modes like natural grafting, which may occur when two plant roots contact each other (Stegemann and Bock, 2009; Stegemann et al., 2012), or in the case of transfer using a parasitic vector (Stegemann et al., 2012; Gao et al., 2014).

4.5 The influence of the occupied habitat on interspecies variability

The assessment of plastome variation and species-specific diversity concerning the habitats in which the studied species currently reside revealed the significant impact of this ecological factor on the development of interspecies diversity within Poales. In the genera which representatives occupy only open habitats, the diversity is significantly greater than in the case of genera occurring in woodlands and mixed habitats, which is probably related to the influence of the habitat type on the biology of pollination and seed dispersal. The influence of the occupied habitat on shaping the variability of Pooideae at earlier stages of the development of this phylogenetic line was previously demonstrated by Zhang and colleagues (Zhang et al., 2022). According to their study the transition from closed to open habitat which occurred at the turn of the Eocene and Oligocene coincided with an upshift of diversification rate in core Pooideae. The results showed that members of three Pooideae lineages transitioned parallelly to open habitats: the most recent common ancestor (MRCA) of supertribe Stipodae, the MRCA of the core Pooideae, and the MRCA of supertribe Nardodae. This event also coincided with a cooling of the climate, and a massive increase in gene and genomic duplications (Zhang et al., 2022). This indicated physiological and morphological innovations which subsequently contributed to plant adaptation to new or stressful environments. The combination of internal genetic events and external environmental stimuli resulted in the radiation of Pooideae into multiple successful lineages with wide expansion to new territories in temperate regions (Soltis and Soltis, 2016).

4.6 Non-biological factors affecting molecular identification of species

In addition to biological factors, the effectiveness and reliability of the DNA barcoding method are influenced by two issues that should be considered. The first is the level of knowledge of a given genus and the compliance of its division into species adequate to genetic variability. The taxonomic division within the genus is often contentious, and dependent on the adopted definition of the biological species. The status of taxa at the level of species and subspecies changes dynamically, as well as the affiliation of species to particular genera. This is perfectly illustrated by a study of the genus *Cephalotaxus*, where the effectiveness of species identification using the super-barcoding method was compared depending on the taxonomic division used (Wang et al., 2022). The second issue is the presence in GenBank of reference sequences that have been

deposited for mislabeled taxa. The scale of this phenomenon is unknown, although to some extent it affects the results of comparative analyses, probably also in the case of these studies. The problem of correct species identification on the example of *Oryza* genus and mislabeled plant material from seed banks was pointed out by Zhang and colleagues (Zhang et al., 2021). Situations when we analyze sequences incorrectly assigned to a given species result in the fact that we try to find species-specific mutations for genetically distant sequences, which leads to an artificial reduction in the number of defined MDCs. However, in order to capture such situations, each sequence that differs significantly from other representatives of the species should be verified at the source, and thus reach the material herbarium. In turn, the elimination of outlier sequences from the dataset does not seem legitimate either, unless we are sure that it is a mislabeled species, because there is a risk that in turn we could artificially reduce genetic diversity in some cases and increase the number of species-specific characters. This leads to the conclusion that regardless of whether we use the method of species identification based on the phylogenetic tree, on the analysis of genetic distance or on the analysis of MDCs, the basis for the reliability of the DNA barcoding method is the adequate taxonomic order in the studied group and excellent knowledge of the species affiliation of the examined representatives.

5 Conclusions

Our study revealed huge diversity in plastome interspecies variability and the number of species-specific mutations within Poales. The efficacy of the molecular species identification method in the analyzed data set generally was associated with the level of plastome genetic diversity, however, cases deviating from this rule have also been reported. It was observed that a particularly large number of species without species-specific mutations belong to the evolutionarily youngest genera within the Pooideae and Panicoideae, although they have also been noted in other parts of the Poales phylogenetic tree. It has also been shown that in Poales the genera whose representatives occur only in open areas are characterized by a higher level of plastome variability than genera occupying forest habitats. Polyploidization and hybridization of species seem to be of particular importance for reducing the effectiveness of super-barcoding, although the scale of this phenomenon is not yet well studied.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material. Further inquiries can be directed to the corresponding author.

Author contributions

KK: Conceptualization, Formal analysis, Investigation, Methodology, Supervision, Writing – original draft. LP: Formal analysis, Visualization, Writing – original draft, Methodology. MM:

Visualization, Writing – review & editing. JS: Conceptualization, Formal analysis, Investigation, Methodology, Supervision, Visualization, Writing – review & editing.

Funding

The author(s) declare that no financial support was received for the research, authorship, and/or publication of this article.

Acknowledgments

We would like to especially thank Professor Jacek Nowakowski for his invaluable contribution to the statistical analysis of data.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

- Abdullah, A., Mehmood, F., Rahim, A., Heidari, P., Ahmed, I., and Pocza, P. (2021). Comparative plastome analysis of *Blumea*, with implications for genome evolution and phylogeny of Asteroideae. *Ecol. Evol.* 11, 7810–7826. doi: 10.1002/ece3.7614
- Amiryousefi, A., Hyvönen, J., and Pocza, P. (2018). IRscope: an online program to visualize the junction sites of chloroplast genomes. *Bioinformatics* 34, 3030–3031. doi: 10.1093/bioinformatics/bty220
- AusGrass2 Grasses of Australia (2014). Available at: <https://ausgrass2.myspecies.info/> (Accessed May 15, 2023).
- Baikhmetov, E., Nowak, A., Gudkova, P. D., and Nobis, M. (2020). Morphological and genome-wide evidence for natural hybridisation within the genus *Stipa* (Poaceae). *Sci. Rep.* 10, 1–14. doi: 10.1038/s41598-020-70582-1
- Bennetzen, J. (2007). Patterns in grass genome evolution. *Curr. Opin. Plant Biol.* 10, 176–181. doi: 10.1016/j.pbi.2007.01.010
- Bernhardt, N., Brassac, J., Kilian, B., and Blattner, F. R. (2017). Dated tribe-wide whole chloroplast genome phylogeny indicates recurrent hybridizations within Triticeae. *BMC Evol. Biol.* 17, 141. doi: 10.1186/s12862-017-0989-9
- CBOL Plant Working Group1, Hollingsworth, P. M., Forrest, L. L., Spouge, J. L., Hajibabaei, M., Ratnasingham, S., et al. (2009). A DNA barcode for land plants. *Proc. Natl. Acad. Sci. U.S.A.* 106, 12794–12797. doi: 10.1073/pnas.0905845106
- Charif, D., and Lobry, J. R. (2007). “SeqinR 1.0-2: A contributed package to the R project for statistical computing devoted to biological sequences retrieval and analysis,” in *Structural Approaches to Sequence Evolution Biological and Medical Physics, Biomedical Engineering*. Eds. U. Bastolla, M. Porto, H. E. Roman and M. Vendruscolo (Berlin, Heidelberg: Springer Berlin Heidelberg), 207–232. doi: 10.1007/978-3-540-35306-5_10
- Chen, Y., Zhong, H., Zhu, Y., Huang, Y., Wu, S., Liu, Z., et al. (2020). Plastome structure and adaptive evolution of *Calanthe* s.l. species. *PeerJ* 8, e10051. doi: 10.7717/peerj.10051
- Coastal Plain Plants (2021) (University of North Carolina Herbarium). Available at: <https://www.herbarium.unc.edu/coastal/> (Accessed May 14, 2023).
- Cowan, R. S., and Fay, M. F. (2012). “Challenges in the DNA barcoding of plant material,” in *Plant DNA Fingerprinting and Barcoding Methods in Molecular Biology*. Eds. N. J. Sucher, J. R. Hennell and M. C. Carles (Totowa, NJ: Humana Press), 23–33. doi: 10.1007/978-1-61779-609-8_3
- Dong, W., Xu, C., Li, C., Sun, J., Zuo, Y., Shi, S., et al. (2015). ycf1, the most promising plastid DNA barcode of land plants. *Sci. Rep.* 5, 8348. doi: 10.1038/srep08348
- Ferri, G., Corradini, B., Ferrari, F., Santunione, A. L., Palazzoli, F., and Alu', M. (2015). Forensic botany II, DNA barcode for land plants: Which markers after the international agreement? *Forensic Sci. International: Genet.* 15, 131–136. doi: 10.1016/j.fsigen.2014.10.005
- Florabase—the Western Australian Flora (1998). Available at: <https://florabase.dpaw.wa.gov.au/> (Accessed 18 May 2023).
- Ford, C. S., Ayres, K. L., Toomey, N., Haider, N., Van Alphen Stahl, J., Kelly, L. J., et al. (2009). Selection of candidate coding DNA barcoding regions for use on land plants. *Botanical J. Linn. Soc.* 159, 1–11. doi: 10.1111/j.1095-8339.2008.00938.x
- Fu, C.-N., Wu, C.-S., Ye, L.-J., Mo, Z.-Q., Liu, J., Chang, Y.-W., et al. (2019). Prevalence of isomeric plastomes and effectiveness of plastome super-barcodes in yews (*Taxus*) worldwide. *Sci. Rep.* 9, 2773. doi: 10.1038/s41598-019-39161-x
- Gallaher, T. J., Peterson, P. M., Soreng, R. J., Zuloaga, F. O., Li, D., Clark, L. G., et al. (2022). Grasses through space and time: An overview of the biogeographical and macroevolutionary history of Poaceae. *J. Systematics Evol.* 60, 522–569. doi: 10.1111/jse.12857
- Gao, C., Ren, X., Mason, A. S., Liu, H., Xiao, M., Li, J., et al. (2014). Horizontal gene transfer in plants. *Funct. Integr. Genomics* 14, 23–29. doi: 10.1007/s10142-013-0345-0
- Grass Base - The Online World Grass Flora (2014). Available at: <http://www.kew.org/data/grasses-db/sppindex.html/> (Accessed May 14, 2023).
- Gu, Z., Gu, L., Eils, R., Schlesner, M., and Brors, B. (2014). *circIze* implements and enhances circular visualization in R. *Bioinformatics* 30, 2811–2812. doi: 10.1093/bioinformatics/btu393
- Guisinger, M. M., Chumley, T. W., Kuehl, J. V., Boore, J. L., and Jansen, R. K. (2010). Implications of the plastid genome sequence of *Typha* (Typhaceae, poales) for understanding genome evolution in Poaceae. *J. Mol. Evol.* 70, 149–166. doi: 10.1007/s00239-009-9317-3
- Guisinger, M. M., Kuehl, J. V., Boore, J. L., and Jansen, R. K. (2011). Extreme reconfiguration of plastid genomes in the angiosperm family geraniaceae: rearrangements, repeats, and codon usage. *Mol. Biol. Evol.* 28, 583–600. doi: 10.1093/molbev/msq229
- Guo, X., Shi, N., Xie, P., Zhang, G., Liu, H., and Ji, Y. (2022). Plastome sequencing for accurate and effective authentication of *Polygonatum kingianum* (Asparagaceae). *Ind. Crops Products* 184, 115056. doi: 10.1016/j.indcrop.2022.115056
- Hand, M. L., Spangenberg, G. C., Forster, J. W., and Cogan, N. O. I. (2013). Plastome sequence determination and comparative analysis for members of the *Lolium - Festuca* grass species complex. *G3 Genes|Genomes|Genetics* 3, 607–616. doi: 10.1534/g3.112.005264
- Hebert, P. D. N., Cywinska, A., Ball, S. L., and deWaard, J. R. (2003). Biological identifications through DNA barcodes. *Proc. R. Soc. Lond. B* 270, 313–321. doi: 10.1098/rspb.2002.2218
- Henriquez, C. L., Arias, T., Pires, J. C., Croat, T. B., and Schaal, B. A. (2014). Phylogenomics of the plant family Araceae. *Mol. Phylogenet. Evol.* 75, 91–102. doi: 10.1016/j.ympev.2014.02.017

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2023.1275377/full#supplementary-material>

SUPPLEMENTARY FIGURE 1

The phylogenetic Poales tree built using the Maximum Likelihood method implemented in the IQ-TREE 2 from consensus sequences calculated for the analyzed plant genera. The bootstrap values are given above branches. Genus names given in red indicate taxa with the highest share of species with no MDCs discussed in the text.

- Hodkinson, T., Salamin, N., and Chase, M. (2007). Large trees, supertrees, and diversification of the grass family. *aliso* 23, 248–258. doi: 10.5642/aliso.20072301.19
- Hollingsworth, P. M., Graham, S. W., and Little, D. P. (2011). Choosing and using a plant DNA barcode. *PLoS One* 6, e19254. doi: 10.1371/journal.pone.0019254
- Huang, B., Ruess, H., Liang, Q., Colleoni, C., and Spooner, D. M. (2019). Analyses of 202 plastid genomes elucidate the phylogeny of *Solanum* section *Petota*. *Sci. Rep.* 9, 4454. doi: 10.1038/s41598-019-40790-5
- Jansen, R. K., Cai, Z., Raubeson, L. A., Daniell, H., dePamphilis, C. W., Leebens-Mack, J., et al. (2007). Analysis of 81 genes from 64 plastid genomes resolves relationships in angiosperms and identifies genome-scale evolutionary patterns. *Proc. Natl. Acad. Sci. U.S.A.* 104, 19369–19374. doi: 10.1073/pnas.0709121104
- Jansen, R. K., and Ruhlman, T. A. (2012). “Plastid genomes of seed plants,” in *Genomics of Chloroplasts and Mitochondria Advances in Photosynthesis and Respiration*. Eds. R. Bock and V. Knoop (Dordrecht: Springer Netherlands), 103–126. doi: 10.1007/978-94-007-2920-9_5
- Jenkins, G. (1986). Synaptonemal complex formation in hybrids of *Lolium temulentum* x *Lolium perenne* (L.): III. Tetraploid. *Chromosoma* 93, 413–419. doi: 10.1007/BF00285823
- Ji, Y., Liu, C., Yang, J., Jin, L., Yang, Z., and Yang, J.-B. (2020). Ultra-barcoding discovers a cryptic species in Paris yunnanensis (Melanthiaceae), a medicinally important plant. *Front. Plant Sci.* 11. doi: 10.3389/fpls.2020.00411
- Jiang, Y., Zhu, C., Wang, S., Wang, F., and Sun, Z. (2023). Identification of three cultivated varieties of *Scutellaria baicalensis* using the complete chloroplast genome as a super-barcode. *Sci. Rep.* 13, 5602. doi: 10.1038/s41598-023-32493-9
- Jiao, L., Lu, Y., He, T., Li, J., and Yin, Y. (2019). A strategy for developing high-resolution DNA barcodes for species discrimination of wood specimens using the complete chloroplast genome of three *Pterocarpus* species. *Planta* 250, 95–104. doi: 10.1007/s00425-019-03150-1
- Katoh, K., and Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* 30, 772–780. doi: 10.1093/molbev/mst010
- Kole, C. (2011). *Wild crop relatives: genomic and breeding resources* (Berlin, Heidelberg: Springer).
- Krawczyk, K., Myszczyński, K., Nobis, M., and Sawicki, J. (2022). Insights into adaptive evolution of plastomes in *Stipa* L. (Poaceae). *BMC Plant Biol.* 22, 525. doi: 10.1186/s12870-022-03923-z
- Krawczyk, K., Nobis, M., Myszczyński, K., Klichowska, E., and Sawicki, J. (2018). Plastid super-barcodes as a tool for species discrimination in feather grasses (Poaceae: *Stipa*). *Sci. Rep.* 8, 1924. doi: 10.1038/s41598-018-20399-w
- Kress, W. J. (2017). Plant DNA barcodes: Applications today and in the future: Plant DNA barcode applications. *Jnl Systematics Evol.* 55, 291–307. doi: 10.1111/jse.12254
- Kress, W. J., Wurdack, K. J., Zimmer, E. A., Weigt, L. A., and Janzen, D. H. (2005). Use of DNA barcodes to identify flowering plants. *Proc. Natl. Acad. Sci. U.S.A.* 102, 8369–8374. doi: 10.1073/pnas.0503123102
- Lady Bird Johnson Wildflower Center (2023). Available at: <http://www.wildflower.org/> (Accessed May 20, 2023).
- Lahaye, R., Savolainen, V., Duthoit, S., Maurin, O., and van der Bank, M. (2008). A test of psbK-psbI and atpF-atpH as potential plant DNA barcodes using the flora of the Kruger National Park (South Africa) as a model system. *Nat. Prec.* doi: 10.1038/npre.2008.1896.1
- Li, J., Tang, J., Zeng, S., Han, F., Yuan, J., and Yu, J. (2021). Comparative plastid genomics of four *Pilea* (Urticaceae) species: insight into interspecific plastid genome diversity in *Pilea*. *BMC Plant Biology* 21 (1), 1–13. doi: 10.1186/s12870-020-02793-7
- Li, X., Yang, Y., Henry, R. J., Rossetto, M., Wang, Y., and Chen, S. (2015). Plant DNA barcoding: from gene to genome: Plant identification using DNA barcodes. *Biol. Rev.* 90, 157–166. doi: 10.1111/brv.12104
- Lin, X. C., Lou, Y. F., Liu, J., Peng, J. S., Liao, G. L., and Fang, W. (2010). Crossbreeding of *Phyllostachys* species (Poaceae) and identification of their hybrids using ISSR markers. *Genet. Mol. Res.* 9, 1398–1404. doi: 10.4238/vol9-3gmr855
- Liu, Q., Li, X., Li, M., Xu, W., Schwarzacher, T., and Heslop-Harrison, J. S. (2020). Comparative chloroplast genome analyses of *Avena*: insights into evolutionary dynamics and phylogeny. *BMC Plant Biol.* 20, 406. doi: 10.1186/s12870-020-02621-y
- Liu, Z.-F., Ma, H., Ci, X.-Q., Li, L., Song, Y., Liu, B., et al. (2021). Can plastid genome sequencing be used for species identification in Lauraceae? *Botanical J. Linn. Soc.* 197, 1–14. doi: 10.1093/botlinn/boab018
- Merckelbach, L. M., and Borges, L. M. S. (2020). Make every species count: fastachar software for rapid determination of molecular diagnostic characters to describe species. *Mol. Ecol. Resour.* 20, 1761–1768. doi: 10.1111/1755-0998.13222
- Minh, B. Q., Schmidt, H. A., Chernomor, O., Schrempf, D., Woodhams, M. D., Von Haeseler, A., et al. (2020). IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol. Biol. Evol.* 37, 1530–1534. doi: 10.1093/molbev/msaa015
- Morales-Briones, D. F., Liston, A., and Tank, D. C. (2018). Phylogenomic analyses reveal a deep history of hybridization and polyploidy in the Neotropical genus *Lachemilla* (Rosaceae). *New Phytol.* 218, 1668–1684. doi: 10.1111/nph.15099
- Native Plant Trust (2023). Available at: <https://www.nativeplanttrust.org/> (Accessed May 17, 2023).
- Nobis, M., Gudkova, P. D., Baiakhmetov, E., Žabicka, J., Krawczyk, K., and Sawicki, J. (2019). Hybridisation, introgression events and cryptic speciation in *Stipa* (Poaceae): A case study of the *Stipa heptapotamica* hybrid-complex. *Perspect. Plant Ecology Evol. Systematics* 39, 125457. doi: 10.1016/j.ppees.2019.05.001
- Nowak, A., Nobis, M., Nowak, S., Nobis, A., Wróbel, A., Świercz, S., et al. (2020). *Illustrated flora of Tajikistan and adjacent areas* (Warsaw: Polish Academy of Sciences).
- Palmer, J. D. (1985). Chloroplast DNA and molecular phylogeny. *Bioessays* 2, 263–267. doi: 10.1002/bies.950020607
- Parks, M., Cronn, R., and Liston, A. (2009). Increasing phylogenetic resolution at low taxonomic levels using massively parallel sequencing of chloroplast genomes. *BMC Biol.* 7, 84. doi: 10.1186/1741-7007-7-84
- Pauksztó, L., Górski, P., Krawczyk, K., Maździarz, M., Szczecińska, M., Ślipiko, M., et al. (2023). The organellar genomes of Pellidae (Marchantiophyta): the evidence of cryptic speciation, conflicting phylogenies and extraordinary reduction of mitogenomes in simple thalloid liverwort lineage. *Sci. Rep.* 13, 8303. doi: 10.1038/s41598-023-35269-3
- Petersen, G., Seberg, O., Yde, M., and Berthelsen, K. (2006). Phylogenetic relationships of *Triticum* and *Aegilops* and evidence for the origin of the A, B, and D genomes of common wheat (*Triticum aestivum*). *Mol. Phylogenet. Evol.* 39, 70–82. doi: 10.1016/j.ympev.2006.01.023
- Piot, A., Hackel, J., Christin, P.-A., and Besnard, G. (2018). One-third of the plastid genome evolved under positive selection in PACMAD grasses. *Planta* 247, 255–266. doi: 10.1007/s00425-017-2781-x
- Qiu, Y., Hirsch, C. D., Yang, Y., and Watkins, E. (2019). Towards improved molecular identification tools in fine fescue (*Festuca* L., Poaceae) turfgrasses: nuclear genome size, ploidy, and chloroplast genome sequencing. *Front. Genet.* 10. doi: 10.3389/fgene.2019.01223
- Ramsey, A. J., and Mandel, J. R. (2019). “When one genome is not enough: organellar heteroplasmy in plants,” in *Annual Plant Reviews online*. Ed. J. A. Roberts (Wiley), 619–658. doi: 10.1002/9781119312994.apr0616
- Revell, L. J. (2012). phytools: an R package for phylogenetic comparative biology (and other things): *phytools: R package. Methods Ecol. Evol.* 3, 217–223. doi: 10.1111/j.2041-210X.2011.00169.x.calypogeia
- Rice, A., Glick, L., Abadi, S., Einhorn, M., Kopelman, N. M., Salman-Minkov, A., et al. (2015). The Chromosome Counts Database (CCDB)—a community resource of plant chromosome numbers. *New Phytol.* 206, 19–26.
- Rieseberg, L. H., and Soltis, D. E. (1991). *Phylogenetic consequences of cytoplasmic gene flow in plants* (Zurich Switzerland: Evolutionary Trends in Plants).
- Saarela, J. M., Wysocki, W. P., Barrett, C. F., Soreng, R. J., Davis, J. I., Clark, L. G., et al. (2015). Plastid phylogenomics of the cool-season grass subfamily: clarification of relationships among early-diverging tribes. *AoB Plants* 7, plv046. doi: 10.1093/aobpla/plv046
- Saddhe, A. A., and Kumar, K. (2017). DNA barcoding of plants: Selection of core markers for taxonomic groups. *Plant Sci. Today* 5, 9–13. doi: 10.14719/pst.2018.5.1.356
- Sarkar, I. N., Planet, P. J., and Desalle, R. (2008). caos software for use in character-based DNA barcoding. *Mol. Ecol. Resour.* 8, 1256–1259. doi: 10.1111/j.1755-0998.2008.02235.x
- Sawicki, J., Bączkiewicz, A., Buczkowska, K., Górski, P., Krawczyk, K., Mizia, P., et al. (2020). The increase of simple sequence repeats during diversification of marchantiidae, an early land plant lineage, leads to the first known expansion of inverted repeats in the evolutionarily-stable structure of liverwort plastomes. *Genes* 11, 299. doi: 10.3390/genes11030299
- Sawicki, J., Krawczyk, K., Ślipiko, M., Szandar, K., and Szczecińska, M. (2021). Comparative analysis of *Apopellia endiviifolia* plastomes reveals a strikingly high level of differentiation between its terrestrial and water form. *Diversity* 13, 674. doi: 10.3390/d13120674
- Sayers, E. W., Bolton, E. E., Brister, J. R., Canese, K., Chan, J., Comeau, D. C., et al. (2022). Database resources of the national center for biotechnology information. *Nucleic Acids Res.* 50, D20–D26. doi: 10.1093/nar/gkab112
- Ślipiko, M., Myszczyński, K., Buczkowska, K., Bączkiewicz, A., Szczecińska, M., and Sawicki, J. (2020). Molecular delimitation of European leafy liverworts of the genus *Calypogeia* based on plastid super-barcodes. *BMC Plant Biol.* 20, 243. doi: 10.1186/s12870-020-02435-y
- Soltis, P., and Soltis, D. (1991). Genetic variation in endemic and widespread plant species. *aliso* 13, 215–223. doi: 10.5642/aliso.19911301.10
- Soltis, P. S., and Soltis, D. E. (2016). Ancient WGD events as drivers of key innovations in angiosperms. *Curr. Opin. Plant Biol.* 30, 159–165. doi: 10.1016/j.cpb.2016.03.015
- Soreng, R. J., Peterson, P. M., Romaschenko, K., Davidse, G., Teisher, J. K., Clark, L. G., et al. (2017). A worldwide phylogenetic classification of the Poaceae (Gramineae) II: An update and a comparison of two 2015 classifications: Phylogenetic classification of the grasses II. *Jnl Systematics Evol.* 55, 259–290. doi: 10.1111/jse.12262
- Stegemann, S., and Bock, R. (2009). Exchange of genetic material between cells in plant tissue grafts. *Science* 324, 649–651. doi: 10.1126/science.1170397
- Stegemann, S., Keuthe, M., Greiner, S., and Bock, R. (2012). Horizontal transfer of chloroplast genomes between plant species. *Proc. Natl. Acad. Sci. U.S.A.* 109, 2434–2438. doi: 10.1073/pnas.1114076109

- Su, H.-J., Liang, S., and Nickrent, D. L. (2021). Plastome variation and phylogeny of *Taxillus* (Loranthaceae). *PLoS One* 16, e0256345. doi: 10.1371/journal.pone.0256345
- Szczecińska, M., and Sawicki, J. (2015). Genomic resources of three *Pulsatilla* species reveal evolutionary hotspots, species-specific sites and variable plastid structure in the family ranunculaceae. *IJMS* 16, 22258–22279. doi: 10.3390/ijms160922258
- Taberlet, P., Coissac, E., Pompanon, F., Gielly, L., Miquel, C., Valentini, A., et al. (2007). Power and limitations of the chloroplast trnL (UAA) intron for plant DNA barcoding. *Nucleic Acids Res.* 35, e14–e14. doi: 10.1093/nar/gkl938
- The PLANTS Database, USDA and NRCS (2023). Available at: <http://plants.usda.gov/> (Accessed May 17, 2023).
- Tomaszewska, P., Vorontsova, M. S., Renvoize, S. A., Ficinski, S. Z., Tohme, J., Schwarzacher, T., et al. (2023). Complex polyploid and hybrid species in an apomictic and sexual tropical forage grass group: genomic composition and evolution in *Urochloa* (*Brachiaria*) species. *Ann. Bot.* 131, 87–108. doi: 10.1093/aob/mcab147
- Useful Tropical Plants Database (2014). Available at: <https://tropical.theferns.info/> (Accessed May 17, 2023).
- Wang, J., Fu, C.-N., Mo, Z.-Q., Möller, M., Yang, J.-B., Zhang, Z.-R., et al. (2022). Testing the complete plastome for species discrimination, cryptic species discovery and phylogenetic resolution in *Cephalotaxus* (Cephalotaxaceae). *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.768810
- World Flora Online (2023). Available at: <http://www.worldfloraonline.org/> (Accessed May 15, 2023).
- Wu, L., Wu, M., Cui, N., Xiang, L., Li, Y., Li, X., et al. (2021). Plant super-barcode: a case study on genome-based identification for closely related species of *Fritillaria*. *Chin. Med.* 16, 52. doi: 10.1186/s13020-021-00460-z
- Wu, H., Yang, J.-B., Liu, J.-X., Li, D.-Z., and Ma, P.-F. (2022). Organelle phylogenomics and extensive conflicting phylogenetic signals in the monocot order poales. *Front. Plant Sci.* 12. doi: 10.3389/fpls.2021.824672
- Wysocki, W. P., Clark, L. G., Attigala, L., Ruiz-Sanchez, E., and Duvall, M. R. (2015). Evolution of the bamboos (Bambusoideae; Poaceae): a full plastome phylogenomic analysis. *BMC Evol. Biol.* 15, 50. doi: 10.1186/s12862-015-0321-5
- Xu, S., Dai, Z., Guo, P., Fu, X., Liu, S., Zhou, L., et al. (2021). ggtreeExtra: compact visualization of richly annotated phylogenetic data. *Mol. Biol. Evol.* 38, 4039–4042. doi: 10.1093/molbev/msab166
- Xu, C., Dong, W., Li, W., Lu, Y., Xie, X., Jin, X., et al. (2017). Comparative analysis of six *lagerstroemia* complete chloroplast genomes. *Front. Plant Sci.* 8. doi: 10.3389/fpls.2017.00015
- Xu, J.-H., Liu, Q., Hu, W., Wang, T., Xue, Q., and Messing, J. (2015). Dynamics of chloroplast genomes in green plants. *Genomics* 106, 221–231. doi: 10.1016/j.ygeno.2015.07.004
- Yu, G., Smith, D. K., Zhu, H., Guan, Y., and Lam, T. T. (2017). ggtree: an R package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods Ecol. Evol.* 8, 28–36. doi: 10.1111/2041-210X.12628
- Zhang, W., Sun, Y., Liu, J., Xu, C., Zou, X., Chen, X., et al. (2021). DNA barcoding of *Oryza*: conventional, specific, and super barcodes. *Plant Mol. Biol.* 105, 215–228. doi: 10.1007/s11103-020-01054-3
- Zhang, L., Zhu, X., Zhao, Y., Guo, J., Zhang, T., Huang, W., et al. (2022). Phylotranscriptomics resolves the phylogeny of pooideae and uncovers factors for their adaptive evolution. *Mol. Biol. Evol.* 39, msac026. doi: 10.1093/molbev/msac026