



ITS secondary structure reconstruction to resolve taxonomy and phylogeny of the *Betula* L. genus

Andrii S. Tarieiev¹, Oliver Gailing^{1,2} and Konstantin V. Krutovsky^{1,2,3,4,5}

¹ Department of Forest Genetics and Forest Tree Breeding, Georg-August University of Göttingen, Göttingen, Germany

² Center for Integrated Breeding Research, Georg-August University of Göttingen, Göttingen, Germany

³ Laboratory of Forest Genomics, Genome Research and Education Center, Institute of Fundamental Biology and Biotechnology, Siberian Federal University, Krasnoyarsk, Russia

⁴ Laboratory of Population Genetics, N.I. Vavilov Institute of General Genetics, Russian Academy of Sciences, Moscow, Russia

⁵ Department of Ecosystem Science and Management, Texas A&M University, College Station, TX, United States of America

ABSTRACT

The taxonomy and phylogeny of the *Betula* L. genus remain unresolved and are very difficult to assess due to several factors, especially because of frequent hybridization among different species. In the current study, we used nucleotide sequences of two internal transcribed spacer regions (ITS1 and ITS2), which are commonly used as phylogenetic markers. In addition to their nucleotide variation we reconstructed their secondary structure and used it to resolve phylogenetic relationships of some birch species. We explored whether consideration of secondary structure in phylogenetic analyses based on neighbor-joining, maximum parsimony, maximum likelihood, and Bayesian inference methods would help us obtain more solid support of the reconstructed phylogenetic trees. The results were not unambiguous. There were only a few clades with higher support when secondary structure was included into analysis. The phylogenetic trees generated using different methods were mostly in agreement with each other. However, the resolving power of these markers is still insufficient to reliably discriminate some closely related species. To achieve this aim more reliably there is a need for application of modern genomic approaches in combination with traditional ones.

Subjects Computational Biology, Evolutionary Studies, Genetics, Plant Science, Taxonomy

Keywords *Betula*, Taxonomy, Phylogeny, ITS, Secondary structure, Phylogenetics

INTRODUCTION

Birches (*Betula* L. genus) are common pioneer trees and shrubs of temperate and boreal zones in the Northern Hemisphere. Trees classified within this genus are highly diverse, especially in the Old World. Birches occupy habitats predominantly in cool and moist regions, including peatlands, stream banks, lakeshores, cool, damp woods, and moist slopes in cool coves. Most of them represent organisms adapted to low temperatures and able to grow under different extreme conditions (bogs, high altitude mountains, sands,

Submitted 16 October 2020

Accepted 12 January 2021

Published 23 March 2021

Corresponding author

Konstantin V. Krutovsky,
konstantin.krutovsky@forst.uni-goettingen.de

Academic editor

Gabriele Casazza

Additional Information and
Declarations can be found on
page 16

DOI 10.7717/peerj.10889

© Copyright
2021 Tarieiev et al.

Distributed under
Creative Commons CC-BY 4.0

OPEN ACCESS

subarctic zone, etc.) and on very poor soils, frequently as the first colonizing pioneer woody plants (Ashburner & McAllister, 2013).

The taxonomy and exact phylogenetic relations in the *Betula* L. genus are still unresolved and very difficult to assess because of several reasons. Number of taxa and their range vary greatly in different studies or sources from ~30 (De Jong, 1993) to ~120 (World Checklist of Selected Plant Families; Ashburner & McAllister, 2013) species. Birches comprise taxonomically a notoriously complex group because of their high morphological variation, phenotypic plasticity, and frequent hybridization (especially introgressive one) (Johnsson, 1945; Johnsson, 1949; Brown, Kennedy & Williams, 1982; Anamthawat-Jonsson & Tomasson, 1999; Anamthawat-Jonsson & Thorsson, 2003; Thorsson, Salmela & Anamthawat-Jonsson, 2001; Thorsson et al., 2010; Palmé et al., 2004; Bona, Petrova & Jadwiszczak, 2018).

Earlier classifications were mostly based on morphological traits (Regel, 1865; Schneider, 1915; Schneider, 1916; De Jong, 1993; Walters, 1993; Chen, Manchester & Sun, 1999; Skvortsov, 2002; Tzvelev, 2002; Ashburner & McAllister, 2013). The internal classification of the genus (subgenera, sections, and subsections) was changed several times, but even till now it is still mostly based on morphological traits. However, they are often insufficient to discriminate different birch species because of their high phenotypic plasticity and frequent hybridization, which blurs boundaries between different taxa. Only a few birch taxa were studied using different molecular genetic markers, such as *rbcL* and ITS sequences (Chen, Manchester & Sun, 1999), nuclear *ADH* and chloroplast *matK* sequences (Jarvinen et al., 2004), nuclear ribosomal DNA sequences (Li, Shoupa & Chen, 2005), nuclear nitrate reductase DNA sequences (Li, Shoupa & Chen, 2007), other multiple chloroplast and nuclear regions (Singevar et al., 2020) and AFLP markers (Schenk et al., 2008), but sampling sizes were quite low in most studies. In addition, there was a recent study of the genetic structure and hybridization of some birch species across Eurasia that provided new data regarding the hybridization patterns and the origin of *B. pubescens* (Tsuda et al., 2017). Some birch taxa were recently reassessed with propositions to lower taxonomic ranks using also internal transcribed spacers (ITS) as molecular genetic markers (e.g., Kuneš et al., 2019; Tarieiev et al., 2019; Jadwiszczak et al., 2020; Linda et al., 2020).

Genome-wide markers inferred by restriction site associated DNA sequencing (RAD-seq) were used to study the phylogeny of 27 diploid and 31 polyploid birch taxa (Wang et al., 2021) and specifically *Costatae* section with description of a new species, *Betula buggsii* N. Wang (Wang et al., 2020). Complete reference genome sequence assemblies are also available for three birch species, *B. nana* L. (Wang et al., 2013), *B. pendula* Roth (Salojärvi et al., 2017) and *B. platyphylla* (Chen et al., 2021) respectively.

ITS are among the most common phylogenetic markers in plants including also several studies on birch genus (Chen, Manchester & Sun, 1999; Li, Shoupa & Chen, 2005; Wang et al., 2016; Kuneš et al., 2019; Tarieiev et al., 2019). There is a large number of ITS sequences deposited and freely available in the NCBI GenBank, ITS2, and EMBL databases. The most complex phylogenetic reconstruction of ~60 birch taxa (233 sequences in total) was done relatively recently using ITS sequences as phylogenetic markers (Wang et al., 2016). However, this phylogenetic study was unable to discriminate some closely related and hybrid taxa. The objective of our study was to test whether inclusion of secondary structure

of ITS would improve phylogenetic analysis. In our study we inferred secondary structure of both ITS1 and ITS2.

MATERIALS & METHODS

ITS markers

We analyzed 223 complete nucleotide sequences of ITS1 that represented 73 birch taxa* (57 species, including two species with hybrid origin, 6 subspecies, 9 varieties, and one form) and 263 sequences of ITS2 that represented 78 birch taxa* (60 species, including two with hybrid origin, one interspecific hybrid, 6 subspecies, 10 varieties, and one form) retrieved from the NCBI GenBank and compared for secondary structure reconstruction with the ITS2 Database (for ITS2 only; [Ankenbrand et al., 2015](#)) (see [Table S1](#)).

In addition, we sequenced several ITS sequences representing the two endemic birch species from Ukraine *B. borysthena* Klokov ([Klokov, 1946](#)) and *B. klokovii* Zaverucha ([Zaverucha, 1964](#)), dark-barked birches *B. pubescens* var. *sibakademica* Baranov (Kuzeneva) (\equiv *B. pubescens* f. *sibakademica* (Baranov) [Tarieiev et al., 2019](#)), ([Baranov, 1924](#); [Kuzeneva, 1936](#); [Tarieiev et al., 2019](#)) *B. kotulae* Zaverucha (\equiv *B. pendula* f. *obscura* (Kotula) [Tarieiev et al., 2019](#)), ([Zaverucha, 1964](#); [Tarieiev et al., 2019](#)) *B. atrata* Domin ([Domin, 1927](#)) and carpathian birch *B. pubescens* ssp. *carpatica* from Poland. We used specimens originating from natural habitats of Ukraine (especially endemic species from type localities (locus classicus) and type specimens listed in [Olshansky et al., 2016](#)), the Botanical Garden of the Georg-August University of Göttingen, and other herbaria—GOET, KW, LE and the former herbarium of Taras Shevchenko National University of Kyiv (see [Table S2](#)).

Birch taxa are listed in [Tables S1](#) and [S2](#) according to the NCBI taxonomy with more detailed explanation provided in [Data S1](#). However, the NCBI taxonomic ranks do not always reflect the current changes in taxonomy and, therefore, could be wrong or outdated in some cases.

ITS1 and ITS2 from ribosomal gene clusters were amplified and sequenced in nine birch species either as a single fragment that included partial 18S, ITS1, 5.8S, ITS2, and partial 28S using universal PCR primers ITS5 and ITS4 ([White et al., 1990](#)) or in parts using ITS5 and ITS2 primer pairs modified for birch species to amplify the ITS1 region and ITS3 and ITS4 primer pairs to amplify the ITS2 region, respectively. PCR products were sequenced on an ABI 3130x Genetic Analyzer (Applied Biosystems, Thermo Fisher Scientific Inc., Waltham, MA, USA). Using BioEdit software ([Hall, 1999](#)) these sequences were checked visually for quality and aligned together with 223 and 263 complete sequences of ITS1 and ITS2 retrieved from the NCBI GenBank and ITS2 databases and presented in [Tables S1](#) and [S2](#), respectively. Secondary structure ([Mai & Coleman, 1997](#); [Coleman, 2000](#); [Coleman, 2003](#); [Coleman, 2007](#); [Coleman, 2009](#); [Müller et al., 2007](#); [Wolf et al., 2013](#)) was inferred for all analyzed sequences using mfold web server ([Zuker, 2003](#)). Obtained structures were checked for correctness (e.g., correct position of conservative motives, U-U mismatch, and 5.8S-28S rRNA hybrid stem formation were manually inspected for ITS2) and Gibbs energy value. Structures were compared with models proposed in the ITS2-Database ([Koetschan et al., 2010](#); [Koetschan et al., 2012](#)) and aligned automatically

with manual adjustment afterwards using the 4SALE program (Seibel *et al.*, 2006; Seibel *et al.*, 2008). Alignments are presented in Data S2 and S3. Obtained structures were visualized using VARNA v3-93 (Darty, Denise & Ponty, 2009), and then their features were colored using Inkscape v. 0.92.3 (Free Software Foundation, Inc.; <https://inkscape.org>).

Phylogenetic relationships were reconstructed using Bayesian inference implemented in MrBayes 3.2.6. (Huelsenbeck & Ronquist, 2001; Ronquist & Huelsenbeck, 2003; Ronquist *et al.*, 2012) on the same sequence set with exclusion of identical sequences. Secondary structure was inferred by applying two different models: doublet model for paired regions and 4 by 4 with GTR +I for unpaired ones, respectively. The analysis was performed in four chains with sampling frequency 100 and different number of iterations –20,000,000, 50,000,000, and 100,000,000. Burnin was performed first separately in each run by monitoring the standard error value, convergence was considered to be reached when the standard error decreased below 0.01. The obtained final phylogenetic tree was visualised using FigTree (<http://tree.bio.ed.ac.uk/software/figtree>) and TreeGraph2 (Stöver & Müller, 2010).

To analyse the impact of rRNA secondary structure on the topology of the phylogenetic tree and support of the clades the following four methods of phylogenetic analysis were performed: (1) Bayesian analysis using an identical number of iterations and 4 by 4 GTR+I+G model suggested by the ModelTest in PAUP v.4.0a (<http://paup.phylosolutions.com>; Swofford, 2003) for all the regions, (2) neighbor-joining (NJ) clustering in ProfDist (Friedrich *et al.*, 2005; Wolf *et al.*, 2008) using GTR model, 1000 bootstrapping iterations and rate matrix Q for ITS2 in case of considering secondary structure or rate matrix Q for *Chlorophyta* in case of not considering secondary structure, (3) maximum parsimony (MP) analysis using PAUP v.4.0a with GTR+I+G model, 1000 bootstrap iterations, and 12-letter coding of secondary structure performed using 4SALE, 4) maximum likelihood (ML) analysis using MEGA X without considering secondary structure and R script provided with 4SALE package (<http://4sale.bioapps.biozentrum.uni-wuerzburg.de/mlseqstr.html>) together with 12-letter coding to consider the structure. It should be noted that 12-letter coding does not include IUPAC codes for ambiguous nucleotides, and, therefore, some information is lost, especially in case of considering hCBCs.

Pairwise topology of phylogenetic trees was adjusted manually for better comparison by merging and rearranging some clusters with statistically insignificant support, and respective drawings were generated using Inkscape v. 0.92.3 (Free Software Foundation, Inc.; <https://inkscape.org>).

RESULTS

Secondary structure reconstruction

Structural variation was observed for different groups of birch taxa. The consensus structural models comprising all variants were reconstructed for the ITS1 and ITS2 markers (Fig. 1).

The ITS1 structure consists of five helices: for main (H1, H2, H3, and H4) plus one additional (Ha). Additional helix Ha was present in four structural variants (Fig. 2): the A

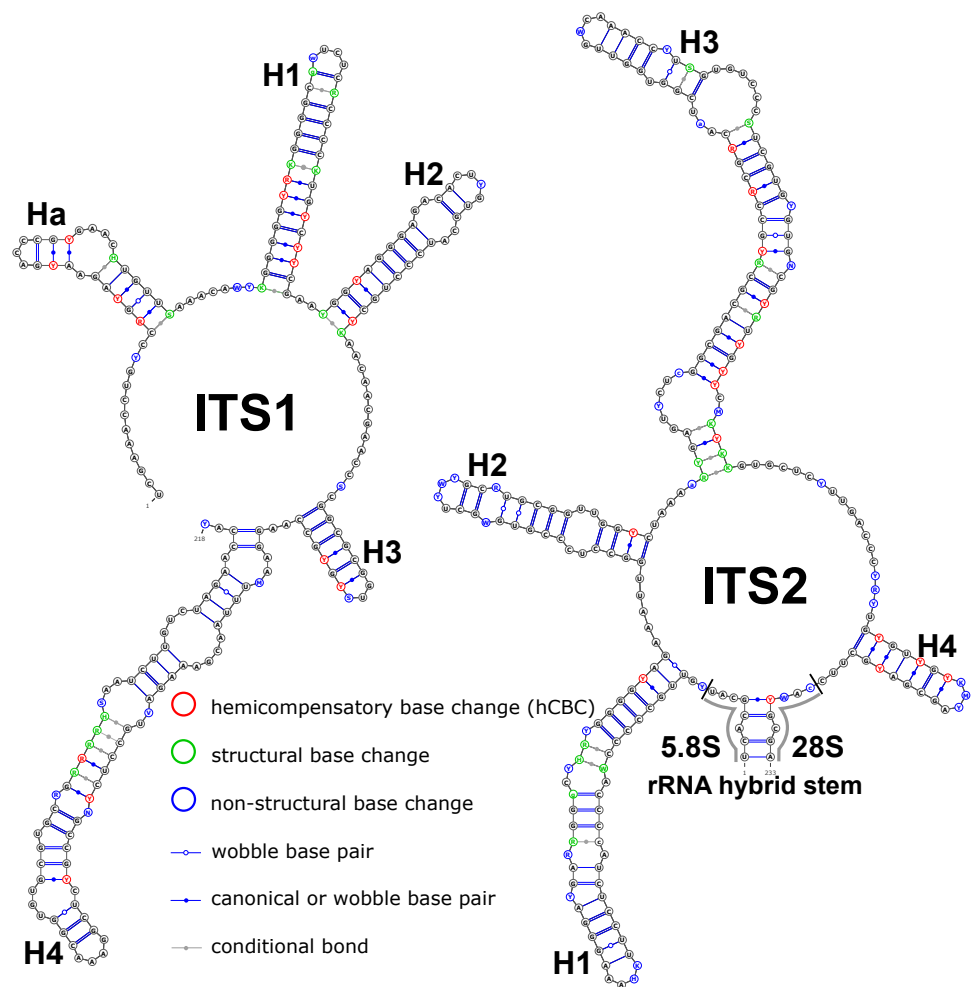


Figure 1 Consensus models of the secondary structure comprising all variants of ITS1 and ITS2 markers for the entire *Betula L.* genus. Indels are depicted by small letters.

Full-size DOI: 10.7717/peerj.10889/fig-1

variant was common in most samples (Fig. 2A), B was specific for *B. globispica* (AB243883.1, KT308904.1, KT308905.1, AY761111.1), *B. chinensis* (KT308917.1, KT308918.1), *B. fargesii* (KT308906) from section *Asperae*, subsection *Chinenses*, *B. glandulosa* (AY761110.1) from section *Apterocaryon*, *B. lenta* (AY761115.1, AY352330.1), *B. medwediewii* (AY761120.1), and *B. uber* (AY352334) from section *Lentae* (Fig. 2B), C was specific for *B. alnoides* (AY763114.1) and *B. luminifera* (KT308944.1) from the section *Acuminatae* (Fig. 2C), and D occurred only in two haplotypes of *B. apoiensis* (AB243908.1 and AB243912.1) (Fig. 2D).

The H1 helix was the most variable in ITS1 and represented by seven structural variants (Fig. 3). However, the last two ones (F and G, respectively) could be reduced to the first and second one (A and B) depending on the nucleotide in ambiguous sites N and R. The H2 helix was represented by two structural variants A and B (Fig. 4). B was specific only for *B. insignis* (KT308927.1, KT308928.1, KT308929.1). The H3 helix was the same for all analyzed sequences and without any structural variation. Four structural variants were

ITS1 additional helix Ha

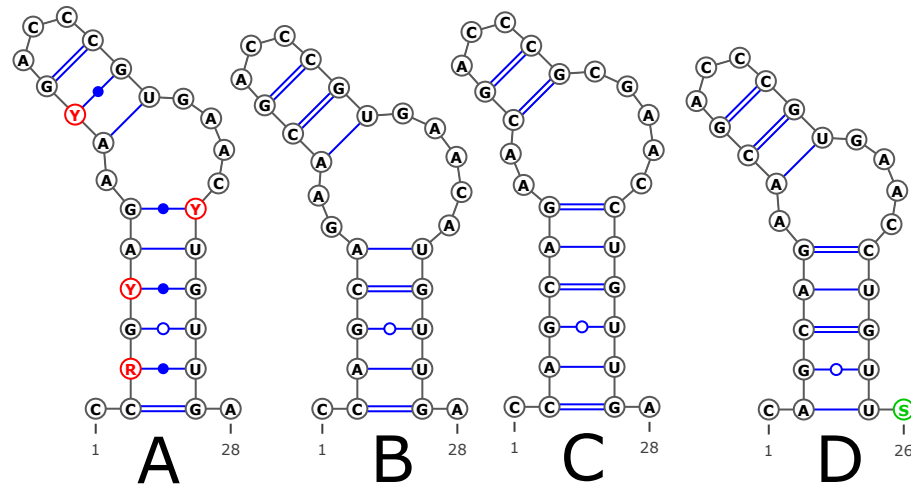


Figure 2 Structural variants of additional helix Ha.

Full-size DOI: [10.7717/peerj.10889/fig-2](https://doi.org/10.7717/peerj.10889/fig-2)

ITS1 Helix 1

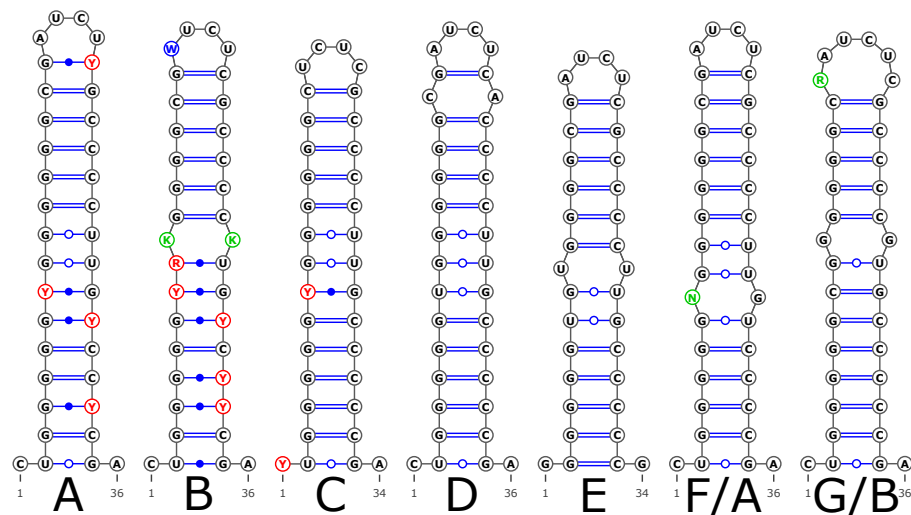


Figure 3 Structural variants of helix 1 (H1) in ITS1.

Full-size DOI: [10.7717/peerj.10889/fig-3](https://doi.org/10.7717/peerj.10889/fig-3)

observed for the H4 helix (Fig. 5). The B variant was species-specific for *B. nigra* (see detailed information about distribution of other variants in Data S4).

The ITS2 marker had no structural variation for H2 and H4 helices. Five structural variants were observed for the H1 helix (Fig. 6) (see also details in Data S3). The helix 3 (H3) was the most variable with 10 structural variants in total (Fig. 7). Further details

ITS1

Helix 2

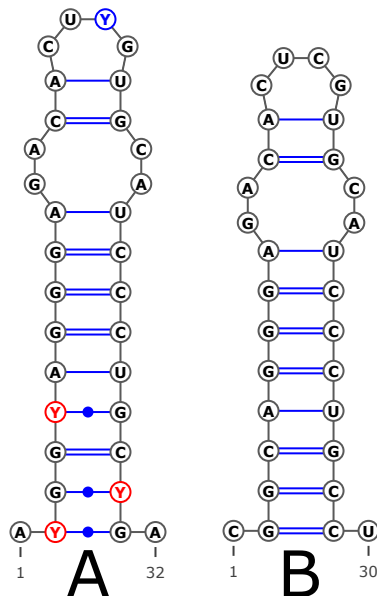


Figure 4 Structural variants of helix 2 (H2) in ITS1.

Full-size  DOI: [10.7717/peerj.10889/fig-4](https://doi.org/10.7717/peerj.10889/fig-4)

regarding distribution of different variants among analyzed sequences are presented in [Data S5](#).

There were no compensatory base changes (CBCs) detected in both structural models of markers across all samples. However, there were several hemicompensatory base changes (hCBCs): in total, 16 hCBCs sites were detected in ITS1 and 15 in ITS2, respectively (more detailed information is in [Data S6](#) and [S7](#)). The exact distribution of those hCBCs is depicted in [Figs. 1–7](#) in red color.

Neighbor-joining (NJ) phylogenetic analysis

NJ phylogenetic tree with consideration of the secondary ITS structure showed higher statistical support of several clusters and more detailed clustering compared to the NJ trees without such consideration (42 vs. 35 supported clusters, respectively, considering 50% as a cut-off for bootstrap support; [Fig. 8](#), [Figs. S1–S2](#)). In addition, 14 clusters that were supported in both trees had higher bootstrap values when the secondary structure was considered. However, the resolution power remained insufficient to delineate some of closely related species and hybrids. In general, clustering was mainly in agreement with genus taxonomy, but not always.

ITS1 Helix 4

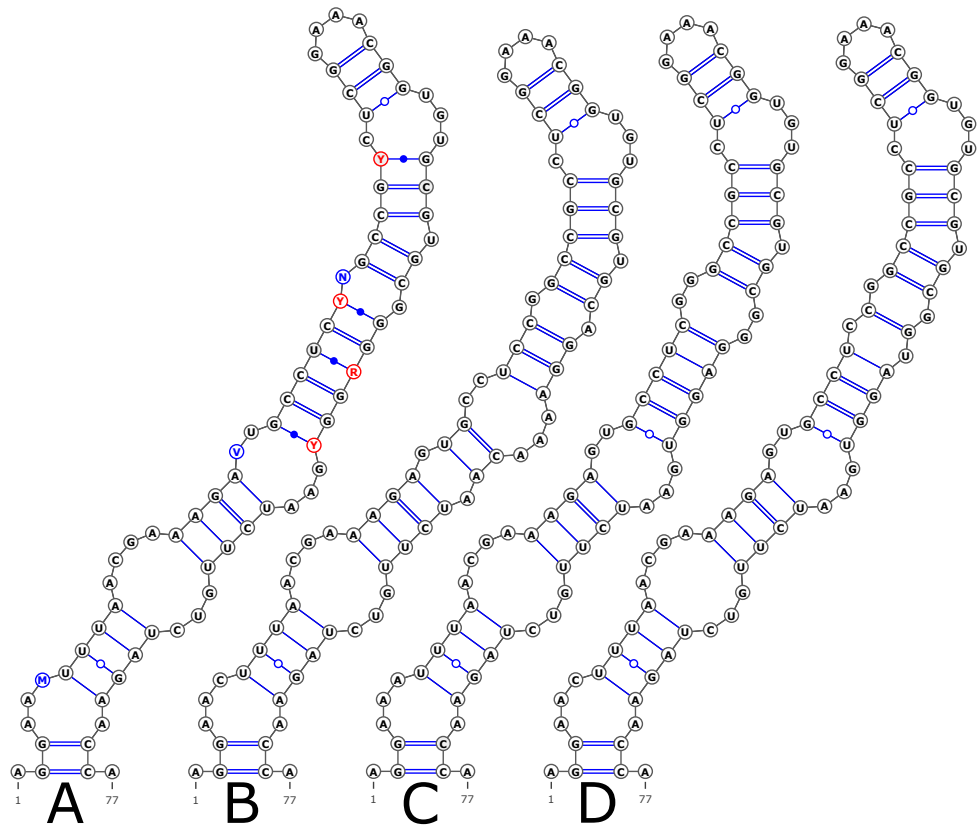


Figure 5 Structural variants of helix 4 (H4) in ITS1.

Full-size  DOI: 10.7717/peerj.10889/fig-5

ITS2 Helix 1

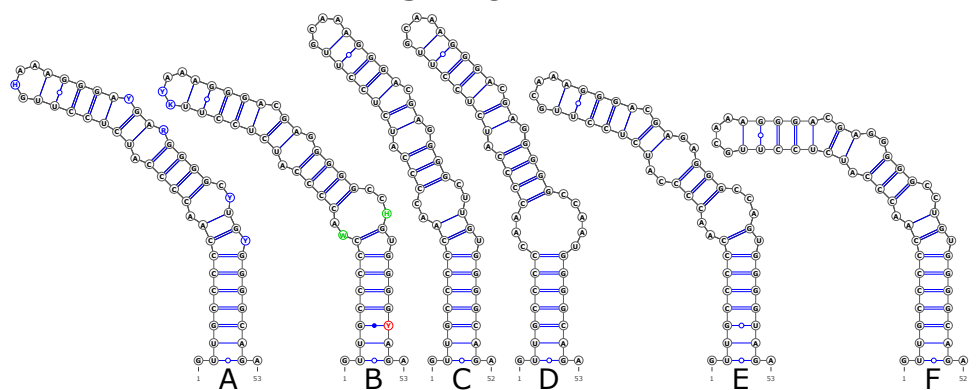


Figure 6 Structural variants of helix 1 (H1) in ITS2.

Full-size  DOI: 10.7717/peerj.10889/fig-6

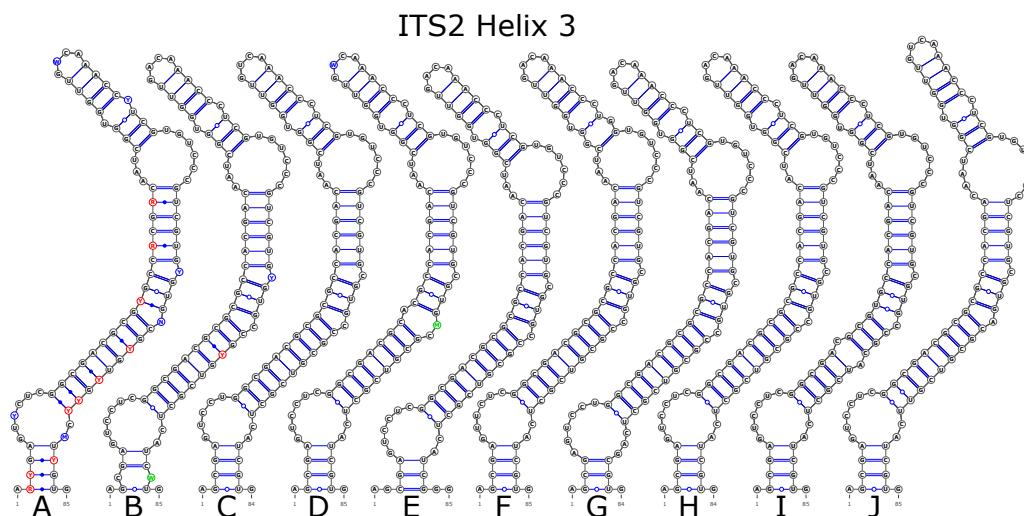


Figure 7 Structural variants of helix 3 (H3) in ITS2.

Full-size  DOI: [10.7717/peerj.10889/fig-7](https://doi.org/10.7717/peerj.10889/fig-7)

Maximum parsimony (MP) phylogenetic analysis

MP phylogenetic trees demonstrated the same pattern as the NJ trees: there were more supported clusters when the secondary structure was considered compared to the MP tree without such consideration: 27 vs. 18, respectively (Fig. 9, Figs. S3–S4). In addition, twelve clusters that were supported in both trees demonstrated higher bootstrap values when the secondary structure was considered. However, the resolution power of the MP analysis was lower than in the NJ analysis.

Maximum likelihood (ML) phylogenetic analysis

MP phylogenetic trees demonstrated 34 supported clusters when secondary structure was considered, but only 28 ones when traditional analysis without consideration of secondary structure was used (Fig. 10, Figs. S5–S6). In addition, twelve clusters that were supported in both trees demonstrated higher bootstrap values when the secondary structure was considered.

Bayesian phylogenetic analysis

Bayesian phylogeny reconstructions showed some changes in phylogenetic trees; both changes in the topology and support values (posterior probability) were observed when the secondary structure of ITS was considered. Clusters supported with posterior probability values >95% are highlighted in red in Fig. 11, Figs. S7–S11.

Several Bayesian trees were constructed using different number of iterations (20,000,000, 50,000,000, and 100,000,000). At 20,000,000 iterations the tree with consideration of the secondary structure had greater statistical support values for several clades (Figs. S10, S11) compared to the tree without consideration of the secondary structure (Fig. S9). However, with increasing iteration number this difference vanished, and for 100,000,000 iterations the situation is opposite –the tree without structure had greater statistical support (Fig. 11, Figs. S7, S8). At the same time, the stability of the tree in terms of topology, number

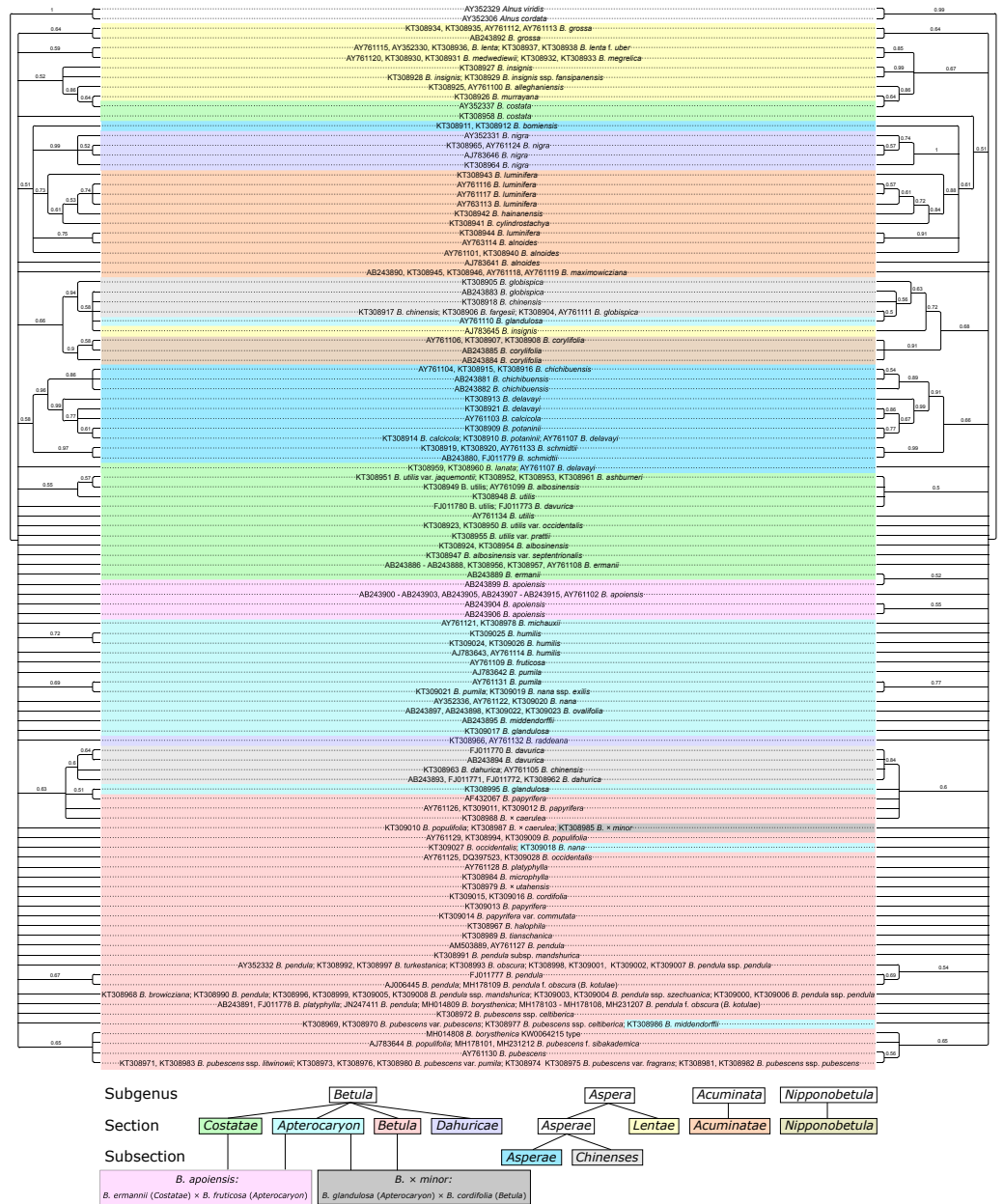


Figure 8 Two neighbor-joining (NJ) phylogenetic consensus trees based on 1,000 iterations without (left) and with (right) consideration of ITS secondary structure generated using ProfDistS with GTR model and rate matrix Q. They are presented also separately in Figs. S1–S2, respectively. Clusters with bootstrap value less than 0.5 (50%) were collapsed. Subgenera, sections, and subsections are marked by different colors according to *Ashburner & McAllister (2013)*. Identical taxa are connected by lines.

Full-size DOI: 10.7717/peerj.10889/fig-8

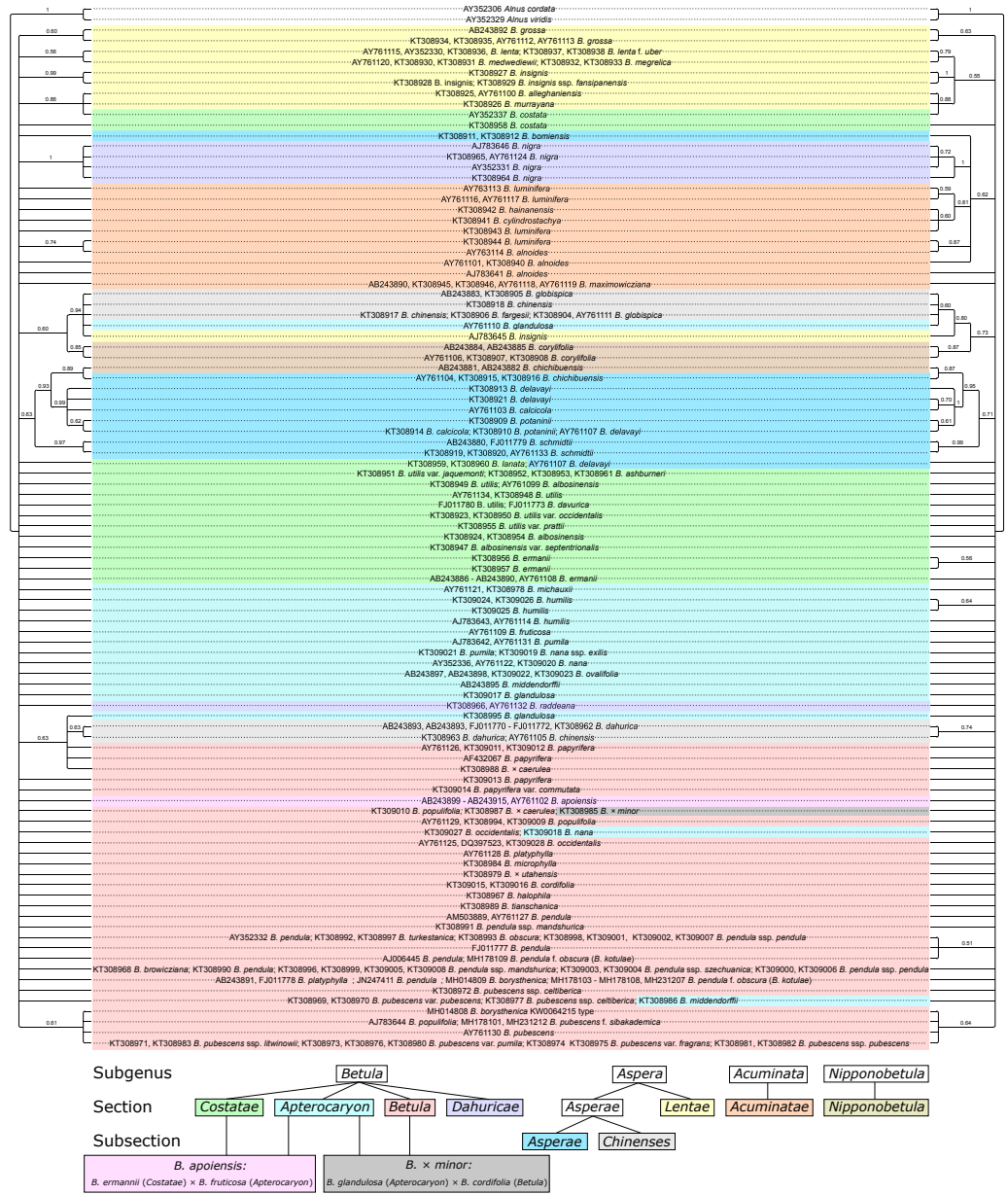


Figure 9 Two maximum parsimony (MP) phylogenetic consensus trees based on 1000 iterations and generated using PAUP without (left) and with (right) consideration of ITS secondary structure recorded as 12-letter code. They are presented also separately in Figs. S3–S4, respectively. Clusters with bootstrap value less than 0.5 (50%) were collapsed. Subgenera, sections, and subsections are marked by different colors according to *Ashburner & McAllister (2013)*. Identical taxa are connected by lines.

Full-size DOI: 10.7717/peerj.10889/fig-9

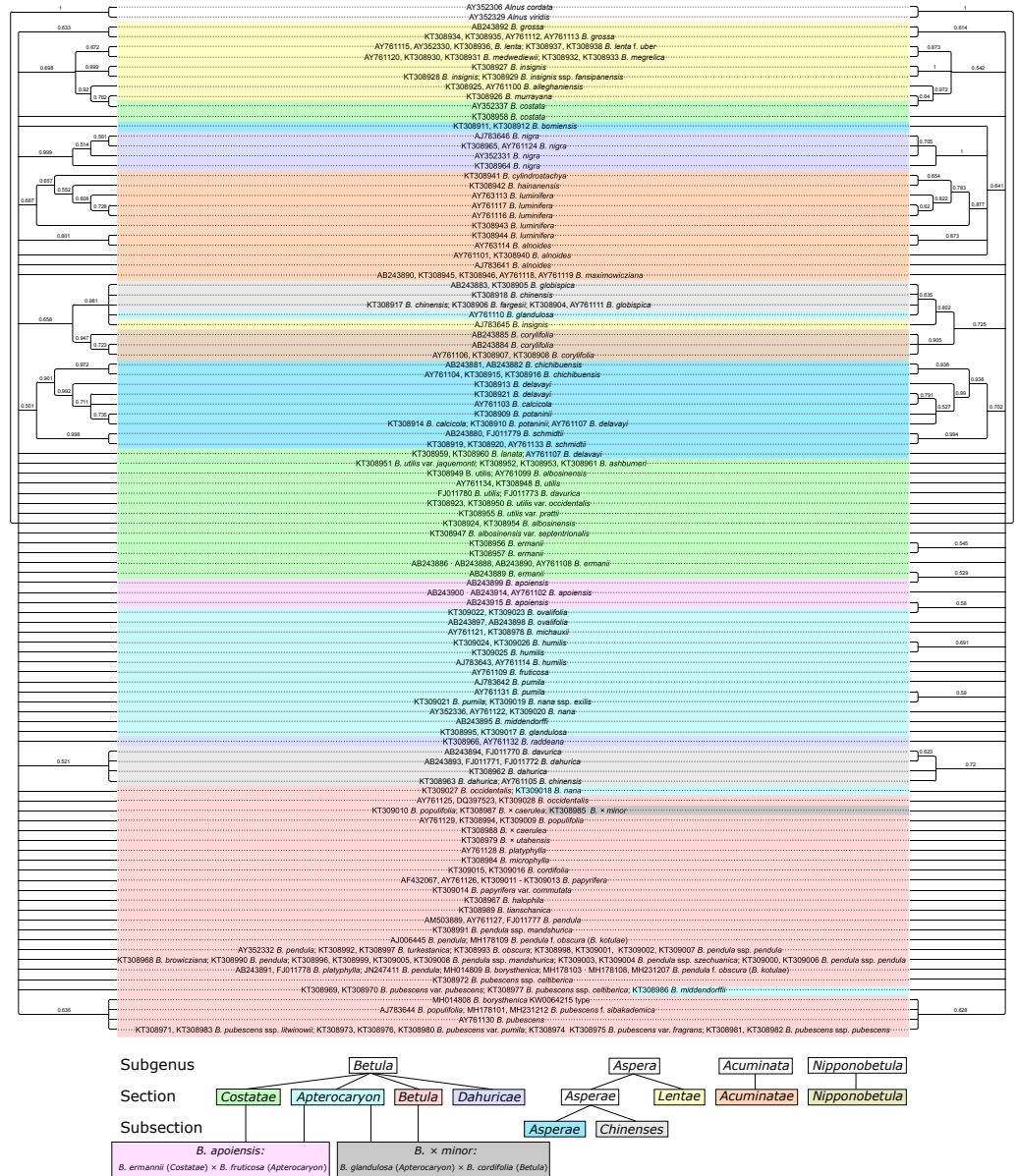


Figure 10 Two maximum likelihood (ML) phylogenetic consensus trees based on 1,000 iterations generated without consideration of ITS secondary structure using MEGA X (left) and with consideration of ITS secondary structure recorded as 12-letter code using 4SALE and phangorn script in R package (right). They are presented also separately in Figs. S5–S6, respectively. Clusters with bootstrap value less than 0.5 (50%) were collapsed. Subgenera, sections, and subsections are marked according to Ashburner & McAllister (2013). Identical taxa are connected by lines.

Full-size DOI: 10.7717/peerj.10889/fig-10

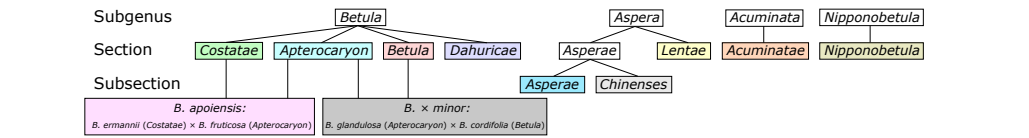
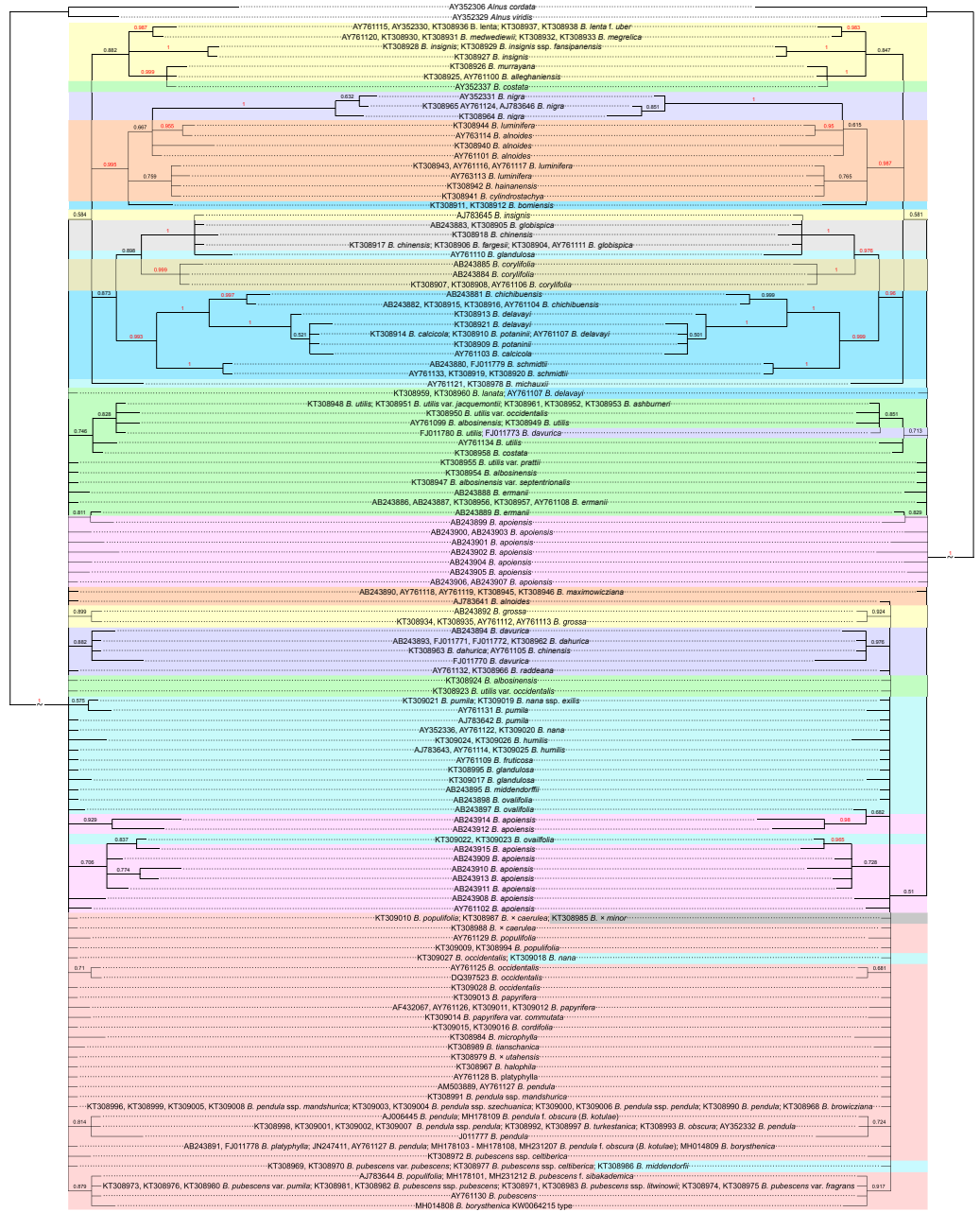


Figure 11 Two Bayesian phylogenetic trees based on 10000000 iterations including 10% burnin and 4 by 4 model for all regions without consideration of ITS secondary structure (left) and doublet model for paired regions, but 4 by 4 with GTR + I for unpaired ones with consideration of ITS secondary structure (right). They are presented also separately in Figs. S7–S8, respectively. Supported clades (> 0.95) are highlighted in red. Subgenera, sections, and subsections are marked according to *Ashburner & McAllister (2013)*. Identical taxa are connected by lines.

Full-size DOI: 10.7717/peerj.10889/fig-11

of clusters and support values was greater for the trees with structure. However, despite the fact that consideration of the secondary structure increased the stability of clusters, many closely related taxa inside the genus were still not resolved. Even subgenera, sections, and subsections (marked in Fig. 11, Figs. S7–S11 according to *Ashburner & McAllister (2013)* using the same symbols as in *Wang et al. (2016)*) could not be clearly distinguished since there are some cases where taxa from different sections and subsections have totally identical ITS sequences.

DISCUSSION

Reconstructed secondary structure of ITS1 and ITS2 reveals a common pattern for the whole genus *Betula* L. The secondary structure for both of these spacers is more conservative than their sequences, and some common patterns are present across very large groups (*Coleman, 2007*). At the same time, secondary structure reconstruction reveals several types of nucleotide base changes, such as structural, non-structural, hemicompensatory (hCBC) and compensatory (CBC) changes, and, therefore, they can considerably enhance phylogenetic studies (*Keller et al., 2010*). The approach is described for a wide range of land plants, algae and animals (*Coleman, 2009*). In case of birch, structural, non-structural and hemicompensatory (hCBC) changes were detected, but compensatory base changes (CBCs) were not found. This was expected because CBCs are usually observed for fully reproductively isolated species (*Coleman, 2000; Coleman, 2009; Müller et al., 2007*), and, therefore, they could be used in delineation of biological species according to Mayr's biological species concept (*Mayr, 1942*). Meanwhile, birches have very weak reproductive barriers with common hybridization events, and a number of taxa have hybrid origin (*Ashburner & McAllister, 2013; Anamthawat-Jonsson & Thorsson, 2003; Anamthawat-Jonsson & Tomasson, 1999; Brown, Kennedy & Williams, 1982; Johnsson, 1945; Johnsson, 1949; Thorsson, Salmela & Anamthawat-Jonsson, 2001; Thorsson et al., 2010; Palmé et al., 2004; Thomson et al., 2015; Wang et al., 2014; Zohren et al., 2016*). Therefore, the biological species concept is hardly applicable to many birch species. There are several other species concepts that could better accommodate this genus, but there is currently no strict consensus on this. Genetic clustering is a common approach to delineate birch species, but it could be spurious in case of species with high gene flow that can easily hybridize (*Tsuda et al., 2017*).

Some birch taxa, even from different subgenera, sections, and subsections, have identical ITS sequences and, therefore, cannot be resolved using ITS markers. Since most of the sequences were retrieved from the GenBank, it is not always possible to verify the exact origin of the samples and correct species identification. The errors could occur due to morphological misidentification, and it is often impossible to check the accuracy of morphological identification of those specimens. However, identity of ITS sequences in species that have very distinctive morphology could be real, especially if it was confirmed by sequence data obtained by different independent research groups for the same species. It is likely that birch is still in the process of speciation, and reproductive barriers are not established yet or weak isolation is beneficial for survival in extreme environments.

This reproductive feature is likely under positive selection, which could also explain the relatively low variation of ITS sequences within diverse groups. There are circa 15 known hybrids between different subgenera, sections, and subsections (Ashburner & McAllister, 2013). However, most of them were not included in this study. The exceptions are *B. × minor* (Tuck.) Fernald, which is a hybrid between *B. cordifolia* Regel ($2n = 2x = 28$, sect. *Betula*) and *B. glandulosa* Michx ($2n = 2x = 28$, sect. *Apterocaryon*), and *B. apoiensis* which is likely a hybrid between *B. ermannii* Cham. ($2n = 4x = 56$, sect. *Costatae*) and *B. fruticosa* Pallas ($2n = 4x = 56$, sect. *Apterocaryon*) (Ashburner & McAllister, 2013). This could explain a variety of different haplotypes of *B. apoiensis* with different clustering in the dendrograms. It should also be noted that those taxonomic categories such as subgenera, sections and subsections within birch genus were based only on morphological traits and do not necessarily reflect genetic differences. Multiple copies of ITS regions of different origin could be present in allopolyploids and interspecific hybrids, but only one of them could be amplified by PCR, which can also result in spurious phylogenetic trees. Recent genome-wide RAD-seq studies (Wang et al., 2020; Wang et al., 2021) solved some specific problems in birch taxonomy and phylogeny, especially clarifying parentage of some tetraploid species, enhancing resolution for diploid birch species with proposing changes on the level of sections and subsections and describing the new species *B. buggsii* that is morphologically similar to *B. utilis* ssp. *albosinensis*. However, some problems especially with polyploids with high ploidy levels remain. Therefore, we believe that further genome wide analyses can help to resolve the phylogenetic and systematic uncertainties in genus *Betula* even better.

CONCLUSIONS

The general models of secondary structure of the ITS1 and ITS2 sequences across the whole genus *Betula* L. were presented for the first time in this study. Species-specific structural variants were detected for *B. insignis* and *B. nigra*. In combination with sequence variation it provides higher confidence in clustering, but only when the number of iterations is relatively low or moderate. With increasing the number of iterations the number of supported clusters and support values for the tree considering secondary structure remain more or less the same, while for the tree without considering secondary structure the number of supported clusters and support values tend to increase. In case of very closely related and hybrid taxa, ITS markers did not provide sufficient discrimination, as well as even some subsections and sections could not be clearly separated. In general, our data were in agreement with the relatively recently published phylogenetic study of the entire genus *Betula* L. based on the nucleotide variation of the ITS markers (Wang et al., 2016), but structural variation can additionally be used to distinguish or support some taxa.

ACKNOWLEDGEMENTS

We thank Prof. Dr. Sergei L. Mosyakin and Dr. Nataliia Shiyan for herbarium type material of *B. borysthenica*, *B. klokovii*, and *B. kotulae* from KW herbarium, Institute of Botany, National Academy of Sciences of Ukraine; Prof. Dr. Dmitry V. Geltman and Dr. Maryna

Legchenko for authentic material of *B. pubescens* var. *sibakademica* from LE herbarium, Botanical Institute of Russian Academy of Sciences (BIN RAS); Dr. Dmytro Iakushenko for the plant material of *B. pubescens* ssp. *carpatica* collected in Poland; Dr. Oksana Abduloieva for herbarium material from the (former) herbarium of Taras Shevchenko University of Kyiv; Dr. Thomas Pröschold for valuable advice and help with Bayesian inference analysis; Dr. Matthias Wolf for useful advice concerning ProfDist; Oleksandra Dolynska for technical assistance during experiments.

ADDITIONAL INFORMATION AND DECLARATIONS

Funding

The research was funded by an EMA2ALRAKIS II internship, a Rufford Small Grant Foundation grant, and a DAAD (German Academic Exchange Service) PhD scholarship for Andrii Tarieiev. Support was also provided by the German Research Foundation (DFG) and the Open Access Publication Funds of the University of Göttingen. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Grant Disclosures

The following grant information was disclosed by the authors:
EMA2 ALRAKIS II internship, Rufford Small Grant Foundation grant.
DAAD (German Academic Exchange Service).

Competing Interests

The authors declare there are no competing interests.

Author Contributions

- Andrii S. Tarieiev conceived and designed the experiments, performed the experiments, analyzed the data, prepared figures and/or tables, authored or reviewed drafts of the paper, and approved the final draft.
- Oliver Gailing analyzed the data, authored or reviewed drafts of the paper, and approved the final draft.
- Konstantin V. Krutovsky conceived and designed the experiments, analyzed the data, authored or reviewed drafts of the paper, and approved the final draft.

Data Availability

The following information was supplied regarding data availability:
Raw data is available in the [Supplemental Files](#).

Supplemental Information

Supplemental information for this article can be found online at <http://dx.doi.org/10.7717/peerj.10889#supplemental-information>.

REFERENCES

- Anamthawat-Jónsson K, Thórsson AT. 2003. Natural hybridization in birch: triploid hybrids between *Betula nana* and *B. pubescens*. *Plant Cell Tissue and Organ Culture* 75:99–107 DOI 10.1023/A:1025063123552.
- Anamthawat-Jónsson K, Tomasson T. 1999. High frequency of triploid birch hybrid by *Betula nana* seed parent. *Hereditas* 130:191–193.
- Ankenbrand MJ, Keller A, Wolf M, Schultz J, Förster F. 2015. ITS2 database V: Twice as much. *Molecular Biology and Evolution* 32(11):3030–3032 DOI 10.1093/molbev/msv174.
- Ashburner K, McAllister HA. 2013. *The Genus Betula: a taxonomic revision of birches*. Kew: Kew publishing, Royal Botanical Gardens.
- Baranov VI. 1924. Notes about birches from forest steppe of Western Siberia. *Reports of Western-Siberian Department of Russian Geographical Society* 4:47–54.
- Bona A, Petrova G, Jadwiszczak KA. 2018. Unfavourable habitat conditions can facilitate hybridisation between the endangered *Betula humilis* and its widespread relatives *B. pendula* and *B. pubescens*. *Plant Ecology & Diversity* 11(3):295–306.
- Brown IR, Kennedy D, Williams DA. 1982. The occurrence of natural hybrids between *Betula pendula* Roth and *B. pubescens*. *Ehrh. Watsonia* 14:133–145.
- Chen S, Wang Y, Yu L, Zheng T, Wang S, Yue Z, Jiang J, Kumari S, Zheng C, Tang H, Li J, Li Y, Chen J, Zhang W, Kuang H, Robertson JS, Zhao PX, Li H, Shu S, Yordanov YS, Huang H, Goodstein DM, Gai Y, Qi Q, Min JM, Xu CY, Wang SB, Qu G-Z, Paterson AH, Sankoff D, Wei H, Liu G, Yang C. 2021. Genome sequence and evolution of *Betula platyphylla*. *Horticulture Research* 8:37 DOI 10.1038/s41438-021-00481-7.
- Chen ZD, Manchester SR, Sun HY. 1999. Phylogeny and evolution of the Betulaceae as inferred from DNA sequence, morphology, and paleobotany. *American Journal of Botany* 86:1168–1181 DOI 10.2307/2656981.
- Coleman AW. 2000. The significance of a coincidence between evolutionary landmarks found in mating affinity and a DNA sequence. *Protist* 151:1–9 DOI 10.1078/1434-4610-00002.
- Coleman AW. 2003. ITS2 is a double-edged tool for eukaryote evolutionary comparisons. *Trends in Genetics* 19:370–375 DOI 10.1016/S0168-9525(03)00118-5.
- Coleman AW. 2007. Pan-eukaryote ITS2 homologies revealed by RNA secondary structure. *Nucleic Acids Research* 35:3322–3329 DOI 10.1093/nar/gkm233.
- Coleman AW. 2009. Is there a molecular key to the level of biological species in eukaryotes? A DNA guide. *Molecular Phylogenetics and Evolution* 50:197–203 DOI 10.1016/j.ympev.2008.10.008.
- Darty K, Denise A, Ponty Y. 2009. VARNA: interactive drawing and editing of the RNA secondary structure. *Bioinformatics* 25(15):1974–1975 DOI 10.1093/bioinformatics/btp250.
- De Jong PC. 1993. An introduction to *Betula*: its morphology, evolution, classification and distribution, with a survey of recent work. In: Hunt D, ed. *Proceedings of the*

- IDS Betula symposium*, 199. Richmond, UK: International Dendrology Society, Richmond, UK, 2–4.
- Domin K. 1927.** *Betula atrata* Domin, nová bříza z Českomoravské vysočiny. *Rozpravy České akademie věd a umění* **36(52)**:1–10.
- Friedrich J, Dandekar T, Wolf M, Müller T. 2005.** ProfDist: a tool for the construction of large phylogenetic trees based on profile distances. *Bioinformatics* **21(9)**:2108–2109 DOI [10.1093/bioinformatics/bti289](https://doi.org/10.1093/bioinformatics/bti289).
- Hall TA. 1999.** BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series* **41**:95–98.
- Huelsenbeck JP, Ronquist F. 2001.** MRBAYES: Bayesian inference of phylogeny. *Bioinformatics* **17**:754–755 DOI [10.1093/bioinformatics/17.8.754](https://doi.org/10.1093/bioinformatics/17.8.754).
- Jadwiszczak KA, Vetchinnikova LV, Bona A, Tyburski Ł, Kuznetsova TJ, Isidorov VA. 2020.** Analyses of molecular markers and leaf morphology of two rare birches, *Betula obscura* and *B. pendula* var. *carelica*. *Annals of Forest Research* **63(2)**:121–113 DOI [10.15287/afr.2020.1973](https://doi.org/10.15287/afr.2020.1973).
- Järvinen P, Palmé A, Orlando Morales L, Lännenpää M, Keinänen M, Sapanen T, Lascoux M. 2004.** Phylogenetic relationships of *Betula* species (Betulaceae) based on nuclear ADH and chloroplast matK sequences. *American Journal of Botany* **91(11)**:1834–1845 DOI [10.3732/ajb.91.11.1834](https://doi.org/10.3732/ajb.91.11.1834).
- Johnsson H. 1945.** Interspecific hybridization within the genus *Betula*. *Hereditas* **31**:163–176.
- Johnsson H. 1949.** Studies on birch hybrids. *Hereditas* **35(2)**:115–135.
- Keller A, Förster F, Müller T, Dandekar T, Schultz J, Wolf M. 2010.** Including RNA secondary structures improves accuracy and robustness in reconstruction of phylogenetic trees. *Biology Direct* **5**:4 DOI [10.1186/1745-6150-5-4](https://doi.org/10.1186/1745-6150-5-4).
- Klokov MV. 1946.** New materials to knowledge Ukrainian flora. *Botanical Journal AS Ukrainian SSR* **3(1–2)**:17–18.
- Koetschan C, Förster F, Keller A, Schleicher T, Ruderisch B, Schwarz R, Müller T, Wolf M, Schultz J. 2010.** The ITS2 Database III—sequences and structures for phylogeny. *Nucleic Acids Research* **38**:D275–D279 DOI [10.1093/nar/gkp966](https://doi.org/10.1093/nar/gkp966).
- Koetschan C, Hackl T, Müller T, Wolf M, Schultz J, Förster F. 2012.** ITS2 database IV: interactive taxon sampling for internal transcribed spacer 2 based phylogenies. *Molecular Phylogenetics and Evolution* **63**:585–588 DOI [10.1016/j.ympev.2012.01.026](https://doi.org/10.1016/j.ympev.2012.01.026).
- Kuneš I, Linda R, Fér T, Karlík P, Baláš M, Ešnerova J, Vítámvás J, Bílý J, Urfus T. 2019.** Is *Betula carpatica* genetically distinctive? A morphometric, cytometric and molecular study of birches in the Bohemian Massif with a focus on Carpathian birch. *PLOS ONE* **14(10)**:e0224387 DOI [10.1371/journal.pone.0224387](https://doi.org/10.1371/journal.pone.0224387).
- Kuzeneva IO. 1936.** *Betuleae* Döll. In: Komarov VL, ed. *Flora of the USSR*. vol. 5. Mosqua-Leningrad: Editio Academiae Scientiarum URSS 296.
- Li J, Shoupa S, Chen Z. 2005.** Phylogenetics of *Betula* (Betulaceae) inferred from sequences of nuclear ribosomal DNA. *Rhodora* **107(929)**:69–86.

- Li J, Shoupa S, Chen Z. 2007. Phylogenetic relationships of diploid species of *Betula* (Betulaceae) inferred from DNA sequences of nuclear nitrate reductase. *Systematic Botany* 32(2):357–365 DOI 10.1600/036364407781179699.
- Linda R, Kuneš I, Baláš M. 2020. Genetic and morphometric variability between populations of *Betula Ůoycoviensis* from Poland and Czechia: A revised view of the taxonomic treatment of the Ojców birch. *PLoS ONE*. e0243310 DOI 10.1371/journal.pone.0243310.
- Mai JC, Coleman AW. 1997. The internal transcribed spacer 2 exhibits a common secondary structure in green algae and flowering plants. *Journal of Molecular Evolution* 44:258–271 DOI 10.1007/PL00006143.
- Mayr E. 1942. *Systematics and the origin of species*. New York: Columbia Univ. Press.
- Müller T, Philippi N, Dandekar T, Schultz J, Wolf M. 2007. Distinguishing species. *RNA* 13:1469–1472 DOI 10.1261/rna.617107.
- Olshanskyi IG, Shyian NM, Tarieiev AS. 2016. Types of *Betula* (Betulaceae) names described from Ukraine, kept in the National Herbarium of Ukraine (KW). *Ukrainian Botanical Journal* 73(4):401–403 DOI 10.15407/ukrbotj73.04.401.
- Palmé AE, Su Q, Palsson S, Lascoux M. 2004. Extensive sharing of chloroplast haplotypes among European birches indicates hybridization among *Betula pendula*, *B. pubescens* and *B. nana*. *Molecular Ecology* 13:167–178 DOI 10.1046/j.1365-294X.2003.02034.x.
- Regel E. 1865. Bemerkungen über die Gattungen *Betula* und *Alnus* nebst Beschreibung einiger neuer Arten. *Bulletin de la Société impériale des naturalistes de Moscou* 38(2):388–434.
- Ronquist F, Huelsenbeck JP. 2003. MRBAYES 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19:1572–1574 DOI 10.1093/bioinformatics/btg180.
- Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Höhna S, Larget B, Liu L, Suchard MA, Huelsenbeck JP. 2012. MRBAYES 3.2: Efficient Bayesian phylogenetic inference and model selection across a large model space. *Systematic Biology* 61:539–542 DOI 10.1093/sysbio/sys029.
- Salojärvi J, Smolander OP, Nieminen K, Rajaraman S, Safronov O, Safdari P, Lamminmäki A, Immanen J, Lan T, Tanskanen J, Rastas P, Amiryousefi A, Jayaprakash B, Kammonen JI, Hagqvist R, Eswaran G, Ahonen VH, Serra JA, Asiegbo FO, de Dios Barajas-Lopez J, Blande D, Blokhina O, Blomster T, Broholm S, Brosché M, Cui F, Dardick C, Ehonen SE, Elomaa P, Escamez S, Fagerstedt KV, Fujii H, Gauthier A, Gollan PJ, Halimaa P, Heino PI, Himanen K, Hollender C, Kangasjärvi S, Kauppinen L, Kelleher CT, Kontunen-Soppela S, Koskinen JP, Kovalchuk A, Kärenlampi SO, Kärkönen AK, Lim K-J, Leppälä J, Macpherson L, Mikola J, Mouhu K, Mähönen AP, Niinemets Ů, Oksanen E, Overmyer K, Palva ET, Pazouki L, Pennanen V, Puhakainen T, Poczai P, Possen BJHM, Punkkinen M, Rahikainen MM, Rousi M, Ruonala R, Van der Schoot C, Shapiguzov A, Sierla M, Sipilä TP, Sutela S, Teeri TH, Tervahauta AI, Vaattovaara A, Vahala J, Vetchinnikova L, Welling A, Wrzaczek M, Xu E, Paulin LG, Schulman AH, Lascoux M, Albert VA, Auvinen P, Helariutta Y, Kangasjärvi J. 2017. Genome sequencing and population

- genomic analyses provide insights into the adaptive landscape of silver birch. *Nature Genetics* **49**:904–912 DOI [10.1038/ng.3862](https://doi.org/10.1038/ng.3862).
- Schenk MF, Thienpont C-N, Koopman WJM, Gilissen LJWJ, Smulders MJM. 2008.** Phylogenetic relationships in *Betula* (Betulaceae) based on AFLP markers. *Tree Genetics and Genomes* **4**(4):911–924 DOI [10.1007/s11295-008-0162-0](https://doi.org/10.1007/s11295-008-0162-0).
- Schneider CK. 1915.** Bemerkungen zur Systematik der Gattung *Betula*. *Österreichische botanische Zeitschrift* **65**(10–12):305–312 DOI [10.1007/BF01660999](https://doi.org/10.1007/BF01660999).
- Schneider CK. 1916.** Betulaceae. In: Sargent CS, ed. *Plantae Wilsonianae*. vol. 2. Cambridge: The University Press, 423–508.
- Seibel PN, Müller T, Dandekar T, Schultz J, Wolf M. 2006.** 4SALE - A tool for synchronous RNA sequence and secondary structure alignment and editing. *BMC Bioinformatics* **7**:498 DOI [10.1186/1471-2105-7-498](https://doi.org/10.1186/1471-2105-7-498).
- Seibel PN, Müller T, Dandekar T, Wolf M. 2008.** Synchronous visual analysis and editing of RNA sequence and secondary structure alignments using 4SALE. *BMC Research Notes* **1**:91 DOI [10.1186/1756-0500-1-91](https://doi.org/10.1186/1756-0500-1-91).
- Singevar K, Moschner CR, Hartung E, Fladung M. 2020.** Species determination and phylogenetic relationships of the genus *Betula* inferred from multiple chloroplast and nuclear regions reveal the high methyl salicylate-producing ability of the ancestor. *Trees* **34**:1131–1146 DOI [10.1007/s00468-020-01984-x](https://doi.org/10.1007/s00468-020-01984-x).
- Skvortsov AK. 2002.** A new system of the genus *Betula* L. –the birch. *Bulletin of Moscow Society of Naturalists* **107**:73–76.
- Stöver BC, Müller KF. 2010.** TreeGraph 2: combining and visualizing evidence from different phylogenetic analyses. *BMC Bioinformatics* **11**:7 DOI [10.1186/1471-2105-11-7](https://doi.org/10.1186/1471-2105-11-7).
- Swofford DL. 2003.** PAUP*. *Phylogenetic analysis using parsimony (*and other methods)*. Version 4. Sunderland: Sinauer Associates.
- Tarieiev AS, Olshanskyi I, Gailing O, Krutovsky KV. 2019.** Taxonomy of dark- and white-barked birches related to *Betula pendula* Roth and *B. pubescens* Ehrh. (Betulaceae) in Ukraine based on both morphological traits and DNA markers. *Botanical Journal of the Linnean Society* **191**(1):142–154 DOI [10.1093/botlinnean/boz031](https://doi.org/10.1093/botlinnean/boz031).
- Thomson AM, Dick CW, Pascoini AL, Dayanandan S. 2015.** Despite introgressive hybridization, North American birches (*Betula* spp.) maintain strong differentiation at nuclear microsatellite loci. *Tree Genetics and Genomes* **11**:101 DOI [10.1007/s11295-015-0922-6](https://doi.org/10.1007/s11295-015-0922-6).
- Thórsson ÆTh, Pálsson S, Lascoux M, Anamthawat-Jónsson K. 2010.** Introgression and phylogeography of *Betula nana* (diploid), *B. pubescens* (tetraploid) and their triploid hybrids in Iceland inferred from cpDNA haplotype variation. *Journal of Biogeography* **37**:2098–2110 DOI [10.1111/j.1365-2699.2010.02353.x](https://doi.org/10.1111/j.1365-2699.2010.02353.x).
- Thórsson ÆTh, Salmela E, Anamthawat-Jónsson K. 2001.** Morphological, cytological, and molecular evidence for introgressive hybridisation in birch. *Journal of Heredity* **92**:404–408 DOI [10.1093/jhered/92.5.404](https://doi.org/10.1093/jhered/92.5.404).
- Tsuda Y, Semerikov V, Sebastiani F, Vendramin GG, Lascoux M. 2017.** Multispecies genetic structure and hybridization in the *Betula* genus across Eurasia. *Molecular Ecology* **26**:589–605 DOI [10.1111/mec.13885](https://doi.org/10.1111/mec.13885).

- Tzvelev NN. 2002.** De generibus *Betula* L. et *Alnus* Mill (Betulaceae) in Europa Orientali. *Novitas Systematicae Plantarum Vascularum* **34**:47–73.
- Walters SM. 1993.** *Betula* L. In: Tutin TG, Heywood VH, Burges NA, Moore DM, Valentine DH, Walters SM, Webb DA, eds. *Flora Europaea 2nd edition Volume 1 Psilotaceae to Platanaceae*. Cambridge University Press, 57–58.
- Wang N, Borrell JS, Bodles WJ, Kuttapitiya A, Nichols RA, Buggs RJ. 2014.** Molecular footprints of the Holocene retreat of dwarf birch in Britain. *Molecular Ecology* **23**:2771–2782 DOI [10.1111/mec.12768](https://doi.org/10.1111/mec.12768).
- Wang L, Ding J, Borrell JS, McAllister HA, Wang F, Liu L, Wang N. 2020.** Molecular and morphological analyses clarify species delimitation and reveal a new *Betula* species in section *Costatae*. *bioRxiv*. DOI [10.1101/2020.10.29.361519](https://doi.org/10.1101/2020.10.29.361519).
- Wang N, Kelly LJ, McAllister HA, Zohren J, Buggs RJA. 2021.** Resolving phylogeny and polyploid parentage using genus-wide genome-wide sequence data from birch trees. *Molecular Phylogenetics and Evolution*. 107126 DOI [10.1016/j.ympev.2021.107126](https://doi.org/10.1016/j.ympev.2021.107126).
- Wang N, McAllister HA, Bartlett PR, Buggs RG. 2016.** Molecular phylogeny and genome size evolution of the genus *Betula* (Betulaceae). *Annals of Botany* **117**(6):1023–1035 DOI [10.1093/aob/mcw048](https://doi.org/10.1093/aob/mcw048).
- Wang N, Thomson M, Bodles WJ, Crawford RM, Hunt HV, Featherstone AW, Pellicer J, Buggs RJ. 2013.** Genome sequence of dwarf birch (*Betula nana*) and cross-species RAD markers. *Molecular Ecology* **22**:3098–3111 DOI [10.1111/mec.12131](https://doi.org/10.1111/mec.12131).
- White TJ, Bruns T, Lee S, Taylor J. 1990.** Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. In: Innis MA, Gelfand DH, Sninsky JJ, White TJ, eds. *PCR Protocols: a guide to methods and applications*. New York: Academic Press, 315–322.
- Wolf M, Chen S, Song J, Ankenbrand M, Müller T. 2013.** Compensatory base changes in ITS2 secondary structures correlate with the biological species concept despite intragenomic variability in ITS2 sequences - a proof of concept. *PLOS ONE* **8**(6):e66726 DOI [10.1371/journal.pone.0066726](https://doi.org/10.1371/journal.pone.0066726).
- Wolf M, Ruderisch B, Dandekar T, Müller T. 2008.** ProfdistS: (Profile-) Distance based phylogeny on sequence-structure alignments. *Bioinformatics* **24**:2401–2402 DOI [10.1093/bioinformatics/btn453](https://doi.org/10.1093/bioinformatics/btn453).
- Zaverucha BV. 1964.** New and rare birch species of Ukrainian flora. *Ukrainian Botanical Journal* **21**(5):78–86.
- Zohren J, Wang N, Kardailsky I, Borrell JS, Joecker A, Nichols RA, Buggs RJA. 2016.** Unidirectional diploidtetraploid introgression among British birch trees with shifting ranges shown by RAD markers. *Molecular Ecology* **25**(11):2413–2426 DOI [10.1111/mec.13644](https://doi.org/10.1111/mec.13644).
- Zuker M. 2003.** Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Research* **31**(13):3406–3415 DOI [10.1093/nar/gkg595](https://doi.org/10.1093/nar/gkg595).