Christian-Albrechts-Universität zu Kiel
Promotionsordnung Agrar- u. Ernährungswiss. Fakultät 2020

Aus dem Institut für Landwirtschaftliche Verfahrenstechnik
der Christian-Albrechts-Universität zu Kiel

Phylogenetic relationships, marker analysis, and investigation of genes mediating high and low methyl salicylate biosynthesis in different birch species (*Betula* L., Betulaceae)

Dissertation zur Erlangung des Doktorgrades der Agrar- und
Ernährungswissenschaftlichen Fakultät
der Christian-Albrechts-Universität zu Kiel

vorgelegt von

M. Sc. Kiran Singewar

aus Nanded, Maharshtra, Indien

Kiel, 2020

Dekan: Prof. Dr. Karl H. Mühling

1. Berichterstatter: Prof. Dr. Eberhart Hartung

2. Berichterstatter: PD. Dr. Matthias Fladung

Tag der mündlichen Prüfung: 17th February 2021

Table of Contents

1. General introduction 1
1.1 Birch- its importance, morphology, and genetic architecture 1
1.2 Birch- taxonomy and phylogenetic analysis 2
1.3 Methyl salicylate (MeSA) in plants 4
1.4 Research issues and objectives 7
1.5 Scientific approach 9
2. 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 10
2.1 Key message 10
2.2 Abstract ..... 11(1131)
2.3 Introduction ..... 11(1131)
2.4 Materials and methods ..... 13(1133)
2.5 Results ..... 15(1135)
2.6 Discussion ..... 18(1138)
2.7 Conclusion ..... 23(1143)
2.8 Future perspectives ..... 24(1144)
2.9 Acknowledgment ..... 24(1144)
2.10 References ..... 24(1144)
3. Identification and analysis of key genes involved in methyl salicylate biosynthesis in different birch species 27
3.1 Key message 27
3.2 Abstract ..... 28(1/24)
3.3 Introduction ..... 28(1/24)
3.4 Materials and methods ..... 30(3/24)
3.5 Results ..... 32(5/24)
3.6 Discussion ..... 40(13/24)
3.7 Acknowledgment ..... 46(19/24)
3.8 References ..... 47(20/24)

4. Genome-wide bioinformatics and phylogenetic analysis revealed putative substrate specificities of SABATH and MES family members in silver birch (*Betula pendula*) 52
4.1 Key message 52
4.2 Abstract 53
4.3 Introduction 54
4.4 Materials and methods 57
4.5 Results 59
4.6 Discussion 74
4.7 Conclusion 79
4.8 Acknowledgment 80
4.9 Supplementary material 80
4.10 References 81
5. General discussion 85
5.1 Phylogenetic investigations in different birch species revealed the high MeSA producing ability of the ancestor 85
5.2 Candidate gene, marker, and tissue-specific expression analysis in different birches 86
5.3 Evolutionary relationships and putative substrate specificities of SABATH and MES enzyme family members of silver birch 87
6. Conclusions 89
7. Summary 90
8. Zusammenfassung 92
9. References 95
10. Declaration of co-authorship 99
11. Acknowledgements 103

1 General introduction

1.1 Birch- its importance, morphology, and genetic architecture

The Northern Hemisphere incorporates around 30% of the global forest area and contains highly dense trees that maintain the global forest ecosystem (Crowther et al., 2015; Gauthier et al., 2015). The most common genera distributed in the region is *Alnus* (Alders), *Betula* (Birches), *Carpinus* (Hornbeams), *Corylus* (Iron-wood), *Ostrya* (Hazel), and *Ostryopsis* (Hazel-hornbeam) all belonging to the *Betulaceae* family (Chen et al., 1999; Rastogi et al., 2015; Christenhusz et al., 2016). The family is distributed in Europe, the Caucasus, and Asia (Chen, 1994; Shaw et al., 2015).

The genus *Betula* is the largest group of dominant trees and shrubs of the *Betulaceae* family, found in the Northern Hemisphere (Rastogi et al., 2015; Shaw et al., 2015). They diverged from the sister genera of the *Betulaceae* family around 75 million years ago (Bina et al., 2016). The number of *Betula* species in the genus extends from 30 to 120 (Furlow, 1990; Koropachinskii, 2013) and some new species have also been described lately (Zeng et al., 2008; McAllister et al., 2011). Birches are considered ecologically significant plants that play an important role in forestry and horticulture (Ashburner et al., 2013). They are fertile seed producers, light-demanding pioneer species, able to grow within a short time, and the foremost species, even after forest fire, to show juvenile growth (Fischer et al., 2002).

The potential of renewable energy and biomass production from birch with economic benefits has been successful in Europe (Hynynen et al., 2009). Additionally, the breeding of birch for the rapid development and income-generating commercial plantations has been a traditional activity in many European countries (Stener et al., 2005; Gailis et al., 2020). Several species of the genus *Betula,* including *Betula pendula* and *Betula pubescens*, are widely populated from the Mediterranean to central Siberia in Europe. Some of the species are considered endangered in the IUCN Red List (Ashburner et al., 2013; Shaw et al., 2015).

The average height of birches is 25 meters, while their lifespan lies between 60 to 150 years. Some birches have been recorded to live up to 300 years as well (Ashburner et al., 2013). Birches are usually regenerated by seeds that are very small, uncountable, and able to be dispersed by wind. The species is monoecious, thus male and female catkins are found at the same tree (Viherä-Aarnio et al., 2008; Vakkari, 2009). Male catkins appear on the trees at the end of

summer, which can be seen during winter, while female catkins survive through the winter by sheltering in the buds, and in the spring they become noticeable (Vakkari, 2009).

Species of the genus *Betula* are polyploid, ranging from diploid to dodecaphonic, and the chromosome number varies from 2n = 2x = 28 to 2n = 12x = 168 (Ashburner et al., 2013). The *de novo* reference genome of European silver birch (Figure 1A) is the only species within the genus *Betula* that is publicly available on the genome browser (https://genomevolution.org/coge/) (Salojärvi et al., 2017). It is a diploid organism with 28 numbers of chromosomes. The genome size of silver birch is about 440 megabase pairs and genome assembly with syntenic alignment suggested that the species has not been through the whole genome duplication evolutionary event (Salojärvi et al., 2017). A survey has been conducted using keywords, '*Betula* genome' and '*Betula* transcriptome' at the PubMed database of NCBI (https://pubmed.ncbi.nlm.nih.gov/) (Table 1). Information about the genome sequencing of the subgenus *Aspera* is missing in the literature.

Betula lenta, (sweet birch) is one of the example species of birch (Figure 1B) belonging to the subgenus *Aspera*, native to the northeastern United States ranging from southern Maine west to southernmost Ontario and from the southern Appalachian Mountains to northern Georgia. It is a diploid and deciduous organism with 28 numbers of chromosomes. It is also known as black and cherry birch and was the only source of wintergreen oil. Extensive harvesting of wintergreen oil from sweet birch made it endangered until the 1950s-60s (Leak, 1965). A large gap has been observed about the genetics of sweet birch (Zoladeski, 2013).

1.2 Birch- taxonomy and phylogenetic analysis

Plant taxonomy includes description, naming, and classification while systematics provides the evolutionary relationship of the plant species. Both branches of biology help in the monitoring and conservation of biodiversity (Hinchliff et al., 2015). Nonetheless, species of the genus *Betula* have always been difficult to describe and classify due to the continuous hybridization and introgression which result in misidentification of *Betula* species (Winkler, 1904; Furlow, 1990; Thórsson et al., 2010). Although birch species vary in their bark color and shape (Li et al., 2005; Ashburner et al., 2013) and their bark contains dissimilar biochemical compounds (Williams et al., 2001), these features have not been enough to facilitate the interspecies relationships of birches (Li et al., 2005; Schenk et al., 2008). These different attempts of analysis with various

criteria resulted rather in several taxonomical classifications of the genus *Betula* (Regel, 1866; Winkler, 1904; De Jong, 1993; Skvortsov, 2002).

Table 1: Detailed information on the whole genome or transcriptome sequences of birch species available on the PubMed database of NCBI: Names, available genomic information, subgenus, and ploidy levels of *Betula* species. The ploidy levels and taxonomic positions were allocated according to Wang et al. (2016) and Ashburner et al., (2013), respectively.

Species	Genomic information	Subgenus	n	References
B. platyphylla	Transcriptome	*Betula*	2n	Mu et al., (2012)
B. nana	*De novo* genome	*Betula*	2n	Wang et al., (2013)
B. pendula	Transcriptome	*Betula*	2n	Lin et al., (2013)
B. platyphylla	Transcriptome	*Betula*	2n	Wang et al., (2014)
B. platyphylla	Transcriptome	*Betula*	2n	Su et al., (2014)
B. platyphylla	Transcriptome	*Betula*	2n	Huang et al., (2015)
B. pendula	*De novo* genome	*Betula*	2n	Salojärvi et al., (2017)
B. papyrifera	Transcriptome	*Betula*	4n	Theriault et al., (2017)
B. luminifera	Transcriptome	*Acuminata*	2n	Cai et al., (2018)
B. platyphylla	Chloroplast genome	*Betula*	2n	Wang et al., (2018)
B. pendula	Transcriptome	*Betula*	2n	Alonso-Serra et al., (2019)

According to Regel (1866), the genus *Betula* holds two subgenera including *Eubetula* with six sections, and *Alnaster*, consisting of only a single section, *Acuminatae*. Further, Winkler (1904) divided the genus into two sections, *Betulaster* and *Eubetula*.

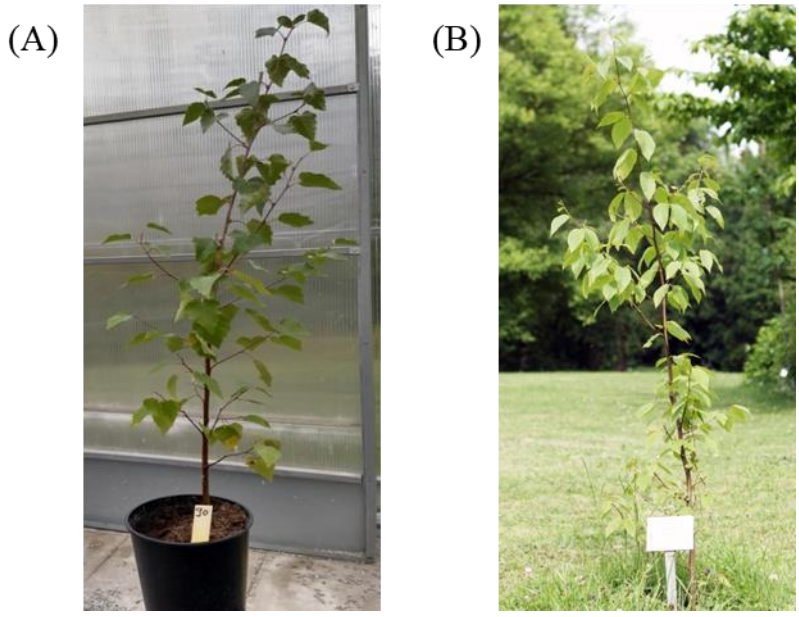

Figure 1: Two-year young birch plants. (A) The European birch, *Betula pendula*, is also known as silver birch. (B) The North American birch, *Betula lenta*. It is commonly known as sweet birch.

Section *Eubetula,* additionally categorized into three subsections, *Costatae*, *Nanae,* and *Albae*. Winkler (1904), also suggested reconsidering the classification of subsections, *Costatae* (Winkler, 1904). In the most recent classification, the genus *Betula* is designated into four subgenera and eight sections by Ashburner et al. (2013).

Besides morphological classifications, various endeavors have been made to produce systematics of the genus *Betula* using molecular genetics approaches (Jianhua et al., 2005; Li et al., 2005; Bina et al., 2016; Wang et al., 2016). Many studies have utilized only a single gene (Li et al., 2005; Wang et al., 2016), the amplified fragment length polymorphism (AFLP) markers (Schenk et al., 2008), and also a blend of single chloroplast and nuclear genes (Järvinen et al., 2004; Bina et al., 2016) to determine the relationships of *Betula* species.

Moreover, some studies claim that the methyl salicylate (MeSA) is the major constituent of essential oil extracted from many birches (Başer et al., 2007; Ashburner et al., 2013; Pal et al., 2015), however, systematics based on this important trait have not been described yet.

1.3 Methyl salicylate (MeSA) in plants

MeSA is widely synthesized by many plant species for long-distance mobile signaling in the systemic acquired resistance (SAR) which is an inducible defense mechanism, activated in response to pathogen attack (Shulaev et al., 1997; Park et al., 2007). Salicylic acid (SA) converted to MeSA by one of the members of SABATH enzyme family, salicylic acid methyltransferase (SAMT) (Ross et al., 1999; D'Auria et al., 2003). The SABATH family is a group of S-adenosyl-L-methionine (SAM)-dependent methyltransferases (SAM-MTs) representing an associated group of *O*-methyltransferases (OMTs) (D'Auria et al., 2003) in plants. The name SABATH is designated based on the first three characterized members in this family; SAMT, benzoic acid carboxyl methyltransferase (BAMT), and Theobromine synthase (D'Auria et al., 2003). Further, MeSA reverted to SA by one of the member enzymes from methyl esterases (MES) family, salicylic acid-binding protein 2 (SABP2) that having strong esterase activity (Kumar et al., 2003) (Figure 2). Demethylation of the resulting methyl esters of SA (MeSA), is catalyzed by the members of the MES enzyme family, affiliated to the α/β hydrolase superfamily (Nardini et al., 1999).

Besides SA, the plant SABATH and MES family members participate in various plant developmental processes through methylation and demethylation of many hormones, signaling

molecules, and floral scent metabolites, including jasmonic acid (JA), and indol-3 acetic acid (IAA) (Yang et al., 2008; Han et al., 2017).

MeSA is one of the herbivore-induced plant volatile (HIPV) compounds (Gadino AN. et al., 2012). Plants produce MeSA amid herbivore attacks to entice the natural antagonist of the herbivores (Cindy et al., 2004; Mallinger et al., 2011). The plant development process emits an abundant quantity of volatiles including MeSA into the environment. The volatile compounds are stored in specialized glandular cells and can be released constitutively under stress and optimal conditions (Gatehouse, 2002; Joó et al., 2011). Conscientious analysis of SA-dependent MeSA production has been performed on different agricultural plants like *Oryza sativa* (Chern et al., 2005), *Arabidopsis thaliana* (Clarke et al., 2000; Wittek et al., 2015), and *Nicotiana tabacum* (Yan et al., 2014).

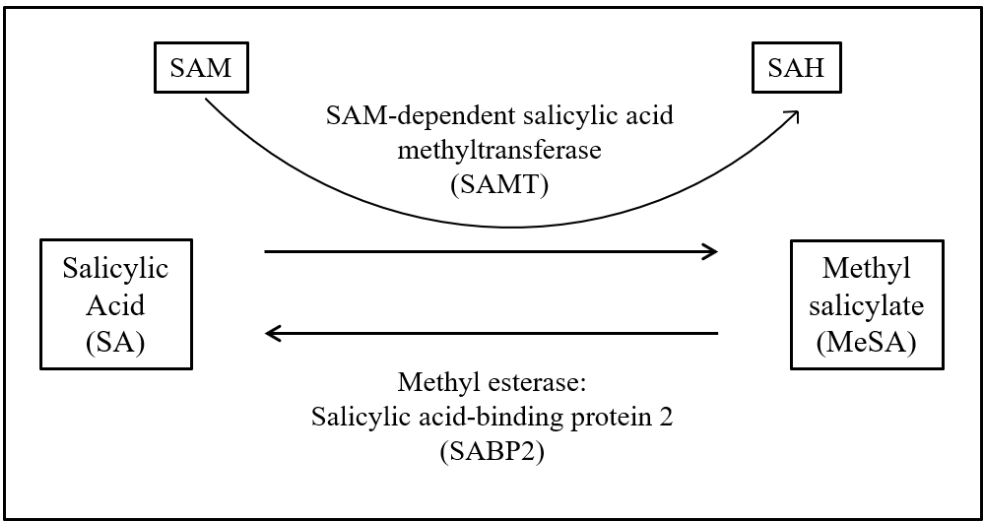

Figure 2: Salicylic acid produces MeSA using one of the enzyme members from SABATH* family, salicylic acid carboxyl methyltransferase (SAMT) which are S-adenosyl-L-methionine (SAM) dependent enzymes. SAM donates a methyl group to complete the methylation reaction and modifies itself to S-adenosyl homocysteine (SAH). Synthesized MeSA is converted back through one of the enzyme members from the methylesterases family, Salicylic acid-binding protein 2 (SABP2). *SABATH: SAMT, benzoic acid carboxyl methyltransferase (BAMT), and Theobromine synthase.

Extensive studies suggest that almost all plant species produce MeSA as a signal in response to pathogen attack in SAR, (Shulaev et al., 1997; Liu et al., 2011; Chen et al., 2019; Shine et al., 2019). Thus, we designated these species as "low-MeSA-producers" since the species synthesize a limited amount of MeSA when necessary (Singewar et al., 2020a). Unlike, species that produce MeSA not just in SAR, while have constitutively elevated quantity are referred to as "high MeSA-producers" (Singewar et al., 2020a). To the best of knowledge, studies about molecular genetics behind the constitutive MeSA production in plants are limited (Singewar et al., 2020a, 2020b) compared to MeSA as a stress-induced signaling molecule.

Various plants constantly produce MeSA which is adequate for the specific fragrance to be well distinguishable to the human nose including many species of the genus *Gaultheria*, *Polygala* (*Polygalaceae*), and *Betula* (specifically the subgenus *Aspera*, section *Lentae*) (Tyler, 1981; Ribnicky et al., 2003; Başer et al., 2007; Pizzolatti et al., 2009; Ashburner et al., 2013; Pal et al., 2015; Singewar et al., 2020a). The fragrance of MeSA emerges from all parts of the various plant species, specifically from bark and young twigs following different scent intensities (Hopkins 1986; Lopes-Lutz et al. 2010; Liu et al., 2011). For example, *P. bicolor* and *P. villosa* carry a weak aroma while strong fragrance has been observed in *P. cecropiifolia* and *P. guianensis* (Gaglioti 2015). The chemical analysis of *P. paniculata* and *P. cyparissias* roots showed MeSA is the main constituent of total extracted oils with 89.1% and 97.8% respectively (Pizzolatti et al. 2009).

Although many species from the genus *Gaultheria* and *Polygala* are rich sources of MeSA, these are the small flowering plants growing between 5 to 50 centimeters high (Ribnicky et al., 2003; Pizzolatti et al., 2009). In contrast, the birches are the long-living and ecologically beneficial forest tree species. According to analytical examination, MeSA is the major constituent of essential oil, extracted from the leaf and twigs of many birches, (Başer et al., 2007; Pal et al., 2015). MeSA has been extensively used for the treatment of skin diseases, infections, rheumatism, and urinary disorders due to its anti-inflammatory properties (Clark, 1999; Menković et al., 2011; Shikov et al., 2011; Anderson et al., 2017).

Even though MeSA can be extracted from plants, at present, it is mainly produced synthetically. The contrived MeSA has been used as an ingredient in fragrance, flavoring, and as a cosmetic ointment in various food, beverages, and beauty products, respectively (NCBI; Lapczynski et al., 2007). From its original extraction procedure through traditional plants, MeSA has now become a synthetic, mass-produced, and pharmaceutically significant ingredient (Anderson et al., 2017).

Nevertheless, information about the content of the pharmacologically important MeSA in birch species is limited and can seldom be found in the research articles (Başer et al., 2007; Liu et al. 2013; Pal et al., 2015; Singewar et al., 2020a).

Considering the pharmacological importance of MeSA, birch represents a relevant target for the development of natural MeSA production strategies. Knowledge about the genetic framework of the available birches is required for the selection and breeding of potentially high MeSA-producing birch trees (Moose et al., 2008; Rasmussen, 2020). The current thesis encourages the commercial production of birches for the timber in short-rotation coppices (SRC) (Walle et al., 2007) along with pharmacological benefits through extracting MeSA from the bark of the birches. The SRC and extracted MeSA would be highly advantageous for forest farming as well as for bio-economy.

1.4 Research issues and objectives

Several attempts have been made to classify the genus *Betula* according to its subgenus and sections (Regel, 1866; Winkler, 1904; Ashburner et al., 2013; Wang et al., 2016). Additionally, the molecular phylogenetic analysis has also been utilized to determine relationships of birch species. However, the distribution and evolution of the high MeSA-producing trait in the genus *Betula* are largely unknown.

Also, the use of natural MeSA is restricted since the production varies significantly in different birch species and the limited knowledge of molecular genetics. The distribution of the natural product in certain species, tissues, and organs requires an improvement in molecular and genetic knowledge. This knowledge is essential to developing organic medicine.

The main objectives of the current thesis are to:

- Analyses the evolutionary relationships of birch species from the perspective of their high MeSA producing ability,
- Examine potential molecular markers and investigate the genes involved in the high MeSA production, and finally,
- Investigate the SABATH and MES family members at gene and enzyme levels that mediate not only MeSA but also JA and IAA biosynthesis.

These main objectives create the general structure of the content and logical succession of the research topics and hypotheses dealt with in chapters 2 to 4.

In Chapter 2, various chloroplast and nuclear genomic regions should be critically analyzed and fragrance analysis has to be conducted to address the following questions:

(1) Is the olfactory analysis of the MeSA fragrance consistent with the previous monographic report (Ashburner et al., 2013);

(2) Are the sequences of multiple chloroplasts and nuclear genomic regions capable of distinguishing between and within *Betula* groups of high and low MeSA productivity;

(3) Are the phylogenies based on chloroplast and nuclear sequences are compatible?

In Chapter 3, information on the two most prominent genes involved to maintain the biosynthesis of MeSA in birches has to be gathered; in targeting on:

(1) The intra- and inter-specific comparative bioinformatics analysis of candidate genes in eight different low and high MeSA-producing *Betula* species;

(2) The sequence variation analysis and marker development in candidate genes associated with MeSA production and

(3) The relative expression analysis of two candidate genes that mediate MeSA biosynthesis in high and low MeSA-producing *Betula* species.

In Chapter 4, the main aim is to:

(1) Identify and characterize the members of *SABATH* and *MES* gene families in *B. pendula;*

(2) Facilitate our understanding of the putative substrate specifications of SABATH and MES enzyme members; and

(3) Provide useful bioinformatics information for the selection of respective *SABATH* and *MES* candidate genes involved in the methylation and demethylation of SA, JA, and IAA in *B. pendula*.

1.5 Scientific approach

The flow chart in figure 3 reflects the general scientific approach of the current thesis. The initial objective was to determine and analyze the phylogenetic relationships of high and low MeSA-producing *Beula* species. Further, identification of markers and expression of the candidate genes were examined that mediate MeSA biosynthesis in birches. In the third section, the SABATH and MES family members at gene and enzyme levels were studied using bioinformatics and phylogenetic analysis. The members of SABATH and MES families are responsible for methylation and demethylation of SA, JA, and IAA, respectively. Finally, the flow chart in figure 3 indicates the application and future perspective of the current research project.

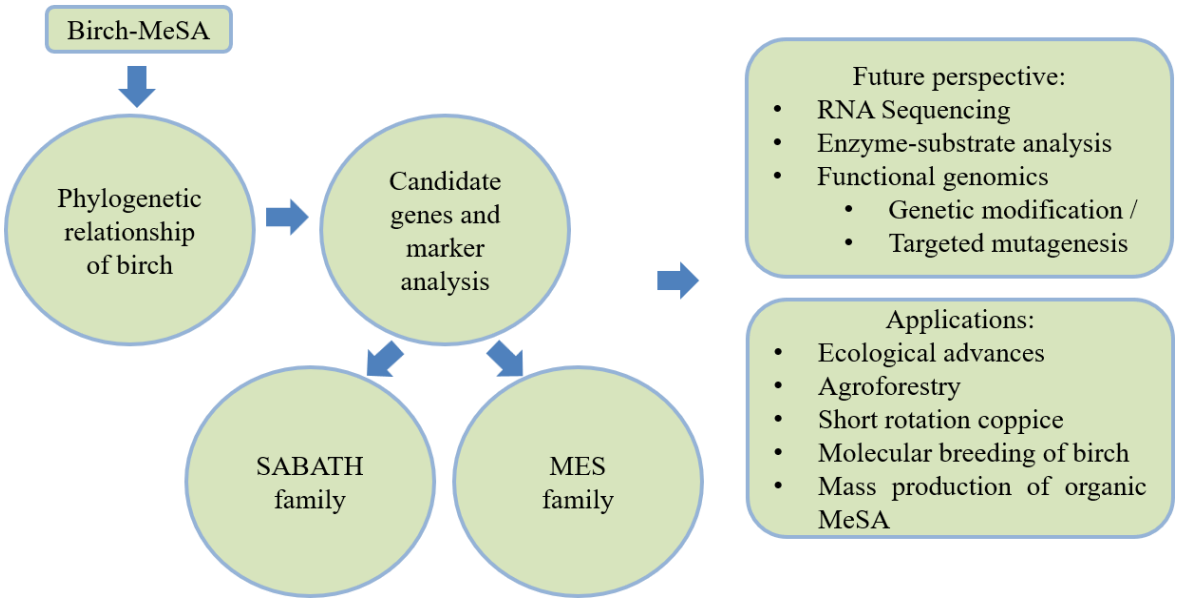

Figure 3: Scientific approach. First, the phylogenetic relationship analysis of the high and low MeSA producing birches has to be determined. Further, studies of candidate genes and marker analysis, SABATH, and MES family members at gene and enzyme levels also have to be conducted. Vital information obtained could be applied for the molecular breeding and mass production technology of the organic MeSA. In the future, whole transcriptome analysis and functional genomics, including genetic modification or targeted mutagenesis will be conducted to decide the function of a gene.

2 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

(Published in *Trees- Structure and Function*; 34, 1131–1146, May 2020)

2.1 Key message

The investigation has provided evolutionary information on the distribution of the high MeSA-producing trait in the different species of the genus *Betula*. The investigation helped to select species for further molecular genetic analyses. Moreover, the study helped to remove misclassified species from the analysis.

Trees (2020) 34:1131–1146
https://doi.org/10.1007/s00468-020-01984-x

ORIGINAL ARTICLE

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

Kiran Singewar[1,2] · Christian R. Moschner[1] · Eberhard Hartung[1] · Matthias Fladung[2]

Received: 11 July 2019 / Accepted: 29 April 2020 / Published online: 16 May 2020

Abstract

Key message **The investigation provides initial knowledge on the distribution and evolution of the high and low methyl salicylate-producing trait in the *Betula* genus. Mislabelled birch species could be identified and removed.**

Abstract The genus *Betula* is the largest group of ecologically and economically dominant perennial woody plants in subalpine forests. The taxonomy of *Betula* is complex due to an extensive history of hybridization and periodic introgression events among the species. Although almost all land plants including birches produce methyl salicylate (MeSA) as a signaling molecule and in response to stress ("low MeSA producer"), some birch species produce high amounts of MeSA in the leaves and bark ("high MeSA producer"). Unfortunately, the evolution of high levels of MeSA production in the genus *Betula* remains unclear. The salicylic acid-binding protein 2 (*SABP2*) and salicylic acid methyltransferase (*SAMT*) genes involved in MeSA biosynthesis were incorporated into this study to examine the interspecific relationship of high and low MeSA-producing birches. Additionally, eight chloroplast and three nuclear regions were included to evaluate their potential application in species determination. The analysis resulted in 25 and 61 nucleotide variations, respectively, which allowed for a visualization of the genetic architecture in the 18 *Betula* species investigated. The high MeSA-producing *B. lenta, B. grossa,* and *B. alleghaniensis* formed the basal clade in the phylogenetic analysis, thus revealing their ancestral status, and the network analysis postulates that the diploid *B. lenta* is one of the ancestors of the genus *Betula*. The results indicate that the ability to produce high levels of MeSA that were initially present in the genus has been lost several times during its evolution. Placing species of the subgenus *Acuminata* alongside the subgenus *Betula*, together with a fragrance analysis, questions their ability to produce high levels of MeSA.

Keywords *Betula* · Chloroplast · Methyl salicylate · Molecular marker · Nuclear · Phylogeny

Introduction

Birches (*Betula* L., Betulaceae) are ecologically important trees and shrubs that are widely distributed throughout the Northern Hemisphere (Furlow 1990). They diverged from other genera of the Betulaceae family around 75 million years ago (Bina et al. 2016). The genus *Betula* plays a vital role in landscape structure, forestry, breeding for biomass production and horticulture (Repola 2008; Ashburner and McAllister 2013; Smith et al. 2014), and some *Betula* species are listed as endangered in the International Union for Conservation of Nature's (IUCN) red list of threatened species (Shaw et al. 2014).

Plant taxonomy helps in the monitoring and conservation of biodiversity (Hinchliff et al. 2015). However, hybridization, introgression, and misidentification of *Betula* species

Communicated by J. Hormaza.

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s00468-020-01984-x) contains supplementary material, which is available to authorized users.

✉ Matthias Fladung
matthias.fladung@thuenen.de

[1] Institute of Agricultural Process Engineering, Christian-Albrechts University of Kiel, Max-Eyth-Strasse 6, 24118 Kiel, Germany

[2] Thünen Institute of Forest Genetics, Sieker Landstrasse 2, 22927 Grosshansdorf, Germany

have led to difficulties in the systematics of this genus (Winkler 1904; Furlow 1990). Although birch species contain different biochemical compounds in their bark and buds (Williams and Arnold 2001) and vary in leaf shape and bark color (Li et al. 2005), these characteristics have not provided reliable interspecies relationships (Li et al. 2005; Schenk et al. 2008), and have resulted in several taxonomical classifications (Regel 1865; Winkler 1904; De Jong 1993; Skvortsov 2002). Regel (1865) acknowledged two subgenera: *Eubetula* consisting of six sections and *Alnaster* consisting of a single section, *Acuminatae*. Winkler (1904) classified the *Betula* genus extensively and further divided the genus into two main sections: *Betulaster* and *Eubetula*, with three subsections of *Eubetula* (*Costatae, Nanae,* and *Albae*). The same author suggested that the subsection *Costatae* required reconsideration (Winkler 1904). Ashburner and McAllister (2013) proposed the latest classification with four subgenera and eight sections, thereby substantially supporting the previous categorization (Skvortsov 2002).

Several studies have variously investigated the amplified fragment length polymorphism (AFLP) markers (Schenk et al. 2008), single genes (Li et al. 2007) and a combination of chloroplast barcoding and nuclear genes (Järvinen et al. 2004; Bina et al. 2016) to resolve the systematics of *Betula*. The official DNA barcoding regions *rbcL* and *matK* have been used to recognize different plant species due to their discrimination power, easy amplification, and good sequence quality (CBOL Plant Working Group 2009; Hollingsworth et al. 2011). However, the interspecific relationships of birch species remains unclear (Järvinen et al. 2004). The first (Li et al. 2005) and most recent (Wang et al. 2016) molecular phylogenetic analyses of the genus *Betula*, based on internal transcribed spacers (ITS), did not resolve all the ambiguities in the interspecific relationships of birches. These studies also advocated the inclusion of multiple nuclear and/or chloroplast regions to resolve these classification issues. The phylogenetic analysis by Wang et al. (2016), however, is congruent with the conclusions of Ashburner and McAllister (2013).

Apart from the classification debate, birches have an extensive history of medicinal use in the treatment of skin diseases, infections, rheumatism and urinary disorders (Menković et al. 2011; Shikov et al. 2011; Angmo et al. 2012; Rastogi et al. 2015). Analytical evidence suggests that terpenoids and methyl salicylate (MeSA) are the major components of birch essential oil (Can Baser et al. 2007; Pal et al. 2015). Terpenoids play a key role in plant defense (Singh and Sharma 2015) and their medicinal value is the subject of current clinical trials (Yin et al. 2017).

On the other hand, MeSA is a volatile compound produced by almost all land plants, including birches, as a signaling molecule and in response to stress (Forouhar et al. 2005) (termed here "low MeSA producers"). Some birch species produce high amounts of MeSA ("high MeSA producers") in the leaves and bark and are used in many drugs for the treatment of muscle and joint pain (Williams and Arnold 2001; Dadákova et al. 2010). However, information about the content of the pharmacologically important MeSA in birch species is almost absent in the literature. Therefore, improved knowledge of the genetic framework of birches is required for selection and breeding purposes (Baum 2008).

Salicylic acid methyltransferase (*SAMT*) is a member of the *O*-methyltransferase gene family that catalyzes the methylation of salicylic acid (SA), while salicylic acid-binding protein 2 (*SABP2*) reverts MeSA to SA (Ross et al. 1999; Forouhar et al. 2005) by demethylation (Du and Klessig 1997). Therefore, *SAMT* and *SABP2* are the putative vital genes for the variation in MeSA production. Thus, they are included in the study together with the internal transcribed spacer (ITS) (White et al. 1990), nitrate reductase (*NIA*) (Li et al. 2007) and alcohol dehydrogenase (*ADH*) (Järvinen et al. 2004) gene regions. Additionally, eight chloroplast coding and intergenic spacers, including *rbcL*, *matK*, *trnH-psbA*, *psbK-psbI*, *trnQ-trnS*, *matK-trnK*, *trnC-petN*, and *rpoC2-rpoC1* (Schroeder et al. 2012; Schroeder and Fladung 2015), were also used to investigate their discrimination potential.

According to the monograph by Ashburner and McAllister (2013), *B. alleghaniensis, B. alnoides, B. globispica, B. grossa, B. lenta, B. luminifera, B. maximowicziana* and *B. medwediewii* are classified as high MeSA producers, whereas *B. costata, B. davurica, B. ermani, B. nana, B. papyrifera, B. pendula, B. platyphylla, B. pubenscens, B. tianchanica* and *B. utilis* are considered as low MeSA producers. Considering the uneven distribution of high MeSA content in the genus *Betula*, the purpose of the study was to uncover the interspecific relationships of high and low MeSA-producing birch species. Since data of the MeSA content in birches is only available in Ashburner and McAllister (2013), we applied an olfactory analysis by scratching and then smelling the bark to obtain an indication of high and low MeSA producers among birch species.

This study aims to answer the following questions: (1) Is the olfactory analysis of the MeSA fragrance consistent with the previous monographic report (Ashburner and McAllister 2013)? (2) Are the sequences of multiple chloroplasts and nuclear genomic regions capable of distinguishing between and within *Betula* groups of high and low MeSA productivity? (3) Are the phylogenies based on chloroplast and nuclear sequences compatible?

To this end, each chloroplast and the nuclear genomic region has been critically analyzed to detect nucleotide variations and to examine the discrimination potential. Although the taxonomy of *Betula* is complex, this investigation has helped in the design of further state-of-the-art experiments.

Materials and methods

Sample collection

Seeds from eight *Betula* species (*B. alleghaniensis, B. maximowicziana, B. medwediewii, B. grossa, B. costata, B. lenta, B. alnoides* and *B. globispica*) were collected from different botanical gardens in Germany (Frankfurt, Giessen, Leipzig, Marburg and Tharandt), Great Britain (Liverpool) and Lithuania (Alytus). Samples of *B. utilis* were collected from a private supplier (Kiel, Germany). In addition, the Northwest German Forestry Research Institute provided frozen seeds of *B. alleghaniensis* (Table 1). Seeds were germinated in soil with pH 6.5 and 10–30% humidity in a natural environment without any fertilizer in a polyhouse at the Institute of Agricultural Process Engineering, Kiel University, Germany. Plantlet cultivation was carried out with the required dose of fertilizers in a glasshouse without any artificial light at the Thünen Institute of Forest Genetics, Grosshansdorf, Germany.

In addition, leaf or twig samples from nine more species (*B. pubescens, B. pendula, B. papyrifera, B. luminifera, B. ermanii, B. platyphylla, B. nana, B. tianschanica,* and *B. davurica*) were collected from botanical gardens in Canada (Montreal), Germany (Grosshansdorf, Göttingen and Munich) and Great Britain (Cambridge). The geographical distribution and place of origin are described in Table 1.

Table 1 Detailed information on the species used for phylogenetic analyses: names of the species, places of sample origin, geographical distribution, ploidy levels (according to Wang et al. (2016)) and taxonomic positions according to Ashburner and McAllister (2013)

Species name	Place of sample origin	Distribution	2n	Subgenus	Section
B. alleghaniensis	BG Tharandt, Germany	North America	6n	*Aspera*	*Lentae*
B. alleghaniensis	NW-FVA[a], Germany	North America	6n	*Aspera*	*Lentae*
B. alleghaniensis	BG Giessen, Germany	North America	6n	*Aspera*	*Lentae*
B. maximowicziana	BG Giessen, Germany	Japan	2n	*Acuminata*	*Acuminata*
B. maximowicziana	BG Tharandt, Germany	Japan	2n	*Acuminata*	*Acuminata*
B. maximowicziana	BG Leipzig, Germany	Japan	2n	*Acuminata*	*Acuminata*
B. maximowicziana	BG Grosshansdorf, Germany	Japan	2n	*Acuminata*	*Acuminata*
B. medwediewii	BG Giessen, Germany	Caucasus Mountains	10n	*Aspera*	*Lentae*
B. medwediewii	BG Liverpool, Great Britain	Caucasus Mountains	10n	*Aspera*	*Lentae*
B. medwediewii Regel	BG Gosshansdorf, Germany	Caucasus Mountains	10n	*Aspera*	*Lentae*
B. grossa	BG Tharandt, Germany	Japan	12n	*Aspera*	*Lentae*
B. utilis	Kiel (private), Germany	Himalayas	4n	*Betula*	*Costatae*
B. costata	OMC Seeds, Lithuania	Central China and Korea	2n	*Betula*	*Costatae*
B. pubescens	BG Grosshansdorf, Germany	Europe and North Asia	4n	*Betula*	*Betula*
B. pendula	BG Grosshansdorf, Germany	Europe and East Asia	2n	*Betula*	*Betula*
B. papyrifera	BG Grosshansdorf, Germany	North America	4n	*Betula*	*Betula*
B. luminifera	BG Grosshansdorf, Germany	China	2n	*Acuminata*	*Acuminatae*
B. ermanii	BG Grosshansdorf, Germany	China, Korea, and Japan	4n	*Betula*	*Costatae*
B. platyphylla	BG Grosshansdorf, Germany	Japan, China, Siberia	2n	*Betula*	*Betula*
B. lenta	BG Giessen, Germany	North America	2n	*Aspera*	*Lentae*
B. lenta	BG Tharandt, Germany	North America	2n	*Aspera*	*Lentae*
B. lenta	BG Marburg, Germany	North America	2n	*Aspera*	*Lentae*
B. alnoides	BG Tharandt, Germany	India, Bhutan, Nepal, China	2n	*Acuminata*	*Acuminatae*
B. globispica	BG Frankfurt, Germany	Japan	10n	*Aspera*	*Asperae*
B. nana	BG Munich, Germany	Arctic region	2n	*Betula*	*Apterocaryon*
B. nana	BG Cambridge, England	Arctic region	2n	*Betula*	*Apterocaryon*
B. nana	BG Montreal, Canada	Arctic region	2n	*Betula*	*Apterocaryon*
B. tianschanica	BG Göttingen, Germany	Kazakhstan	4n	*Betula*	*Betula*
B.davurica	BG Göttingen, Germany	China, Japan, Korea	8n	*Betula*	*Dahuricae*

The ploidy conditions and taxonomical positions (subgenus and sections) of species are defined according to Wang et al. (2016) and Ashburner and McAllister (2013), respectively

[a]Northwest German Forestry Research Institute

BG Botanical Garden

Alnus incana and *Carpinus betulus*, as outgroup species, were sampled from BG Grosshansdorf.

Olfactory analysis of MeSA fragrance

To maintain consistency with previous studies, we conducted the olfactory analysis of MeSA fragrance by scratching the bark of young plants and twigs of the trees. Volunteers from the institute participated in the analysis. All volunteers were provided with bark from the young plants and twig samples from the different birches for scratching and sniffing. According to MeSA fragrance intensity, the samples were unanimously categorized as high, intermediate and low MeSA-producing *Betula* species (Supplementary Table 1).

Selection of chloroplast and nuclear regions and PCR primer design

Functions of the selected chloroplast and nuclear regions were defined according to previous reference studies (Supplementary Table 2). Previously described and newly designed primer combinations of all chloroplast and nuclear regions were used in the current study (Supplementary Table 2). Birch-specific primers were newly designed to amplify the *SABP2, SAMT* and *NIA* genes as well as the ITS regions with the help of the *B. pendula* genome (Salojärvi et al. 2017).

DNA extraction, PCR amplification and sequencing

Total DNA was extracted from the leaves of the plants (and the twigs of *B. tianschanica,* and *B. davurica*) according to the CTAB protocol of Dumolin et al. (1995). DNA extraction in *B. nana, B. medwediewii, B. alleghaniensis,* and *B. lenta* was difficult due to the presence of high levels of polysaccharides. Thus a pre-washing buffer (1.6 mL ice-cold TNE buffer: 200 mM Tris–HCl, 250 mM NaCl, 50 mM EDTA; Wang et al. 2013) was used to extract DNA of good quality.

The following PCR cycling conditions were used for all regions except *ADH*: 95 °C for 3 min, 40 cycles at 95 °C for 10 s, annealing with respective temperature for 30 s and 72 °C for 30 s. The amplification of *ADH* was carried out using the PCR reaction described by Järvinen et al. (2004). All PCR reactions were performed in a SensoQuest thermocycler (Göttingen, Germany). *B. papyrifera, B. alnoides, B. globispica* and *B. pubescens* individuals displayed two fragments after amplification with *ADH* primers. Therefore, gel purification of the ~ 1060 bp fragment of these species was carried out using a QIAquick Gel Extraction Kit. PCR products were confirmed on 1% agarose gel stained with Roti®-GelStain (Carl Roth, Karlsruhe, Germany). The StarSEQ (Mainz, Germany) service was used for sequencing.

Sequence examination and data validation

The electropherograms of each sequence were visually inspected. Upstream and downstream regions were trimmed. In addition, all sequences were aligned and screened for the presence of nucleotide substitutions using the SeqMan Pro15 program from the DNASTAR Lasergene bioinformatics software suite (Madison, Wisconsin USA). The sequences generated were validated with the NCBI gene database, if available. Verifying the reliability of the newly obtained sequences was an important step in avoiding any sequence misclassification.

Phylogenetic analysis

All sequences generated from the 29 birch individuals and the outgroup species were aligned with the bioinformatics computer program ClustalW and nexus files were created using MEGA7 (Kumar et al. 2016). Phylogenetic trees were inferred using ML (maximum likelihood) analysis with the substitution model JC69 + I for heterozygosity in the program PAUP* 4.0a (build 163) (https://paup.phylosolutions.com/). All characters, including indels and extended codes from the International Union of Pure and Applied Chemistry (IUPAC) (Johnson et al. 2010), were equally weighted and their status was kept unordered. Further, a bootstrap analysis of 100 replicates was performed to evaluate the individual clades with random and simple sequence additions (Maddison 1991). Generated trees were visualized using FigTree v1.4.3 (https://tree.bio.ed.ac.uk/software/figtree/) program. Furthermore, a phylogenetic network analysis was performed using the median-joining algorithm with the program package Network 5 (Bandelt et al. 1999).

Analyzed sequences of nuclear and chloroplast regions were treated as an individual-specific single-locus. Three different data sets (chloroplast, nuclear and a combined "league data set") were created to evaluate the congruence between chloroplast and nuclear DNA. In addition, another phylogeny was constructed by deleting the possible ambiguous regions from the aligned league data set sequences (Supplementary Fig. 1). During alignment, sequences from fast-evolving genes and different species may produce potential gaps in the alignment defined as "ambiguous regions". The presence of these regions may mean that the assumptions are inaccurate, while deleting them might result in the loss of vital information (Lutzoni et al. 2000).

The outgroup species, *A. incana* and *C. betulus*, were amplified with all the chloroplast primers. However, *NIA* and *ADH* primers were only able to amplify *A. incana*. Therefore, *C. betulus* was excluded from the nuclear phylogenetic

analysis. Several heterozygous sites were observed in nuclear regions and treated according to the IUPAC nucleotide extended codes (https://www.bioinformatics.org/sms/iupac.html). The pairwise distance between *Betula* and the outgroup species was summarized using a function in MEGA7 based on the league data.

Results

Olfactory analysis of MeSA fragrance

To observe the consistency of high and low MeSA producers with monographic evidence (Ashburner and McAllister 2013), we conducted an olfactory analysis of MeSA fragrance by scratching the bark of young plants and twigs from the trees. The MeSA fragrance comparison analysis resulted in high (*B. grossa*, *B. lenta* and *B. alleghaniensis*), intermediate (*B. maximowicziana, B. medwediewii*) and low (all others) MeSA producers. Notably, the analysis showed that *B. costata* (subgenus *Betula*, section *Costatae*) produced the MeSA fragrance, while *B. luminifera*, *B. alnoides* (subgenus *Acuminata*, section *Acuminatae*) and *B. globispica* (subgenus *Aspera*, section *Asperae*) did not produce any MeSA that could unanimously be recognized by the human nose (Supplementary Table 1).

Sequence analysis

In the present study, 400 new *Betula* sequences were submitted to the NCBI GenBank database (Supplementary Table 3). Annotations and descriptions were assigned according to previous records available in the GenBank.

Polymorphic sites were found in chloroplast regions excluding the official barcode *rbcL.* The sequencing results indicated that the intergenic spacers *trnH-psbA*, *psbK-psbI*, *trnC-petN* and *trnQ-trnS* retained a higher degree of polymorphism than the coding regions *matK*, *matK-trnK* and *rpoC2-rpoC1*. The chloroplast intergenic spacer *trnH-psbA* had the highest number of substitutions, whereas the coding region *rpoC2-rpoC1* showed an average number of nucleotide variations (Table 2).

The exon regions of the nuclear gene *SABP2* (exon-3) and *SAMT* (exon-2) revealed a total number of 50 and 68 polymorphic sites, and 21 and 12 substitutions, respectively. The *NIA* and *ADH* sequences had a total number of 22 and 84 variable sites including six and 16 substitutions, respectively. In the ITS region, only 12 sites appeared to show variation (Table 3).

In addition, intraspecific sequence analysis of chloroplast regions was also conducted and this showed variations in *B. maximowicziana, B. medwediewii, B. lenta* and *B. nana* (Supplementary Table 4). Sequence alignments of *B. alleghaniensis_10* sampled from the arboretum of the Thünen Institute, Grosshansdorf (BG), differed from the other *B. alleghaniensis* individuals. Therefore, we considered this individual as mislabelled and removed it from the study.

Phylogenetic analysis

Chloroplast regions

The aligned sequences of eight chloroplast regions of 29 individuals belonging to 18 birch species and two outgroup species contained 4587 characters, of which 204 were variable and 81 were parsimony-informative. A maximum likelihood phylogenetic tree was constructed (Fig. 1). According to our fragrance analysis (Supplementary Table 1), all species which produce low MeSA, except for *B. maximowicziana* and *B. medwediewii*, which produce an intermediate amount of MeSA, are clustered in clade I. Moreover, *B. luminifera, B. alnoides* and *B. globispica*, which are classified as high MeSA according to Ashburner and McAllister (2013) but which failed the olfactory analysis (Supplementary Table 1), are also clustered in clade I. On the other hand, the high MeSA producers are clustered together in clade II, except for *B. costata, B. utilis* and *B. papyrifera*.

Nuclear regions

The aligned sequences of five nuclear regions of 29 individuals belonging to 18 birch species and one outgroup species contained 2,818 characters, of which 377 were variable and 133 were parsimony-informative. *B. papyrifera* and *B. utilis*, which clustered with the high MeSA producers *B. lenta, B. grossa* and *B. alleghaniensis* in the chloroplast phylogeny (Fig. 1), congregated with most of the species from the subgenus *Betula* and *Acuminata* in clade I in the nuclear maximum likelihood phylogeny (Fig. 2). Therefore, clade II contains all the high MeSA producer species and only one low MeSA producer, *B. costata.*

League data set (combined chloroplast and nuclear sequence data)

The aligned chloroplast and nuclear sequences of all birch individuals were used to construct the concatenated data phylogeny (Fig. 3) which revealed a similar topology to the nuclear phylogeny (Fig. 2), with one exception: *B. medwediewii* displayed a discrete lineage (clade II). The pairwise distance between *Betula* and the outgroup species was summarized using a function in MEGA7 based on the league data (Supplementary Table 5).

Table 2 Characteristics of the variable sites in eight cpDNA regions in all the *Betula* species studied: *B. alleganienisis* (*ale*), *B. maximowicziana* (*max*), *B. medwiediewi* (*med*), *B. grossa* (*gro*), *B. utilis* (*util*), *B. costata* (*cos*), *B. pubescens* (*pub*), *B. pendula* (*pen*), *B. papyrifera* (*pap*), *B. luminifera* (*lum*), *B. ermanii* (*erm*), *B. platyphylla* (*play*), *B. lenta* (*len*), *B. alnoides* (*aln*), *B. globispica* (*glo*), *B. nana* (*nan*), *B. tianschanica* (*ta*), *B. davurica* (*dav*)

Chloroplast regions	Fragment length[a]	No. of var.[b]	No. of S.[c]	Position of S	*ale*	*max*	*med*	*gro*	*util*	*cos*	*pub*	*pen*	*pap*	*lum*	*erm*	*play*	*len*	*aln*	*glo*	*nan*	*tia*	*dav*	%[d] poly
rbcL	488	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
matK	809	3	3	230	A	C	C	C	C	A	C	C	A	C	C	C	C	C	C	C	C	C	0.42
				432	C	T	T	T	T	C	T	T	C	T	T	T	T	T	C	T	T	T	
				434	A	A	A	A	A	A	A	C	A	C	A	A	A	A	A	A	A	C	
trnH-psbA	481	12	6	57	C	T	T	C	C	C	T	T	C	T	T	T	C	T	T	T	T	T	3.44
				93	G	G	G	G	G	G	G	G	G	G	G	G	T	G	G	G	T	T	
				98	G	T	T	G	G	G	T	T	G	G	T	T	T	T	T	T	T	T	
				162	A	A	A	G	G	A	A	A	A	A	A	A	A	A	A	A	A	A	
				183	T	G	G	T	T	T	G	G	T	G	G	G	T	G	G	G	G	G	
				188	C	C	C	T	C	C	C	C	C	C	C	C	C	C	C	C	G	G	
psbK-psbI	789	6	3	154	G	G	G	T	G	G	G	G	G	G	G	G	G	G	G	G	G	G	1.31
				218	T	T	T	C	C	T	T	T	T	T	T	T	T	T	T	T	T	T	
				279	T	T	T	C	C	T	T	T	T	T	T	T	T	T	T	T	T	T	
trnQ-trnS	577	9	3	216	T	T	T	T	C	T	T	T	T	T	T	T	T	T	T	T	T	T	1.9
				256	C	T	T	T	T	C	T	T	C	T	T	T	T	T	T	T	T	T	
				269	A	A	A	A	G	A	A	A	A	A	A	A	A	A	A	A	A	A	
matK-trnK	740	9	4	164	C	C	C	T	C	C	C	C	C	C	C	C	C	C	C	C	C	C	1.22
				506	C	C	C	C	A	C	C	C	C	C	C	C	C	C	C	C	C	C	
				566	G	T	T	G	G	G	T	T	G	T	T	T	T	G	T	T	T	T	
				716	T	T	T	T	A	T	T	T	T	T	T	T	T	T	T	T	T	T	
rpoC2-rpoC1	506	2	1	218	G	A	A	A	A	G	A	A	G	A	A	A	A	A	A	A	A	A	0.50
trnC-petN	958	6	5	302	A	G	G	G	G	A	G	G	A	G	G	G	G	G	G	G	G	G	0.72
				355	T	T	T	C	T	T	T	T	T	T	T	T	T	T	T	T	T	T	
				453	G	G	G	A	G	G	G	G	G	G	G	G	G	G	G	G	G	G	
				476	G	G	G	G	T	G	G	G	G	G	G	G	G	G	G	G	G	G	
				538	T	T	T	T	G	T	T	T	T	T	T	T	T	T	T	T	T	T	

The "–" indicates no variation

S. Substitutions

[a]Fragment length obtained after trimming

[b]Total number of nucleotide variations present in all selected individuals

[c]Detected number of substitutions (S)

[4]Percentage variable sites calculated as the total number of polymorphisms in relation to the sequence length

Table 3 Characteristics of the variable sites in five nuclear DNA regions in five *Betula* species: *B. alleghaniensis* (*ale*), *B. maximowicziana* (*max*), *B. medwediewii* (*med*), *B lenta* (*len*) *and B. nana* (*nan*)

Nuclear regions	Fragment length[a]	No. of var.[b]	No. of S.[c]	Position of S	*ale*	*max*	*med*	*len*	*nan*	% poly.[d]
ITS	597	12	5	182	G	A	A	A	A	2.01
				203	A	G	A	A	G	
				252	T	A	T	T	A	
				552	G	A	A	A	A	
				571	G	A	G	G	A	
NIA	451	22	6	165	T	G	G	T	G	4.87
				243	A	T	T	A	T	
				333	C	C	C	C	G	
				365	C	G	G	G	G	
				393	G	G	G	T	G	
				414	T	G	G	T	G	
ADH	915	84	16	62	C	C/T	T	C	C	9.18
				77	C/T	T	T	T	T	
				80	G	G	A	G	G	
				102	T	C/T	C	C/T	C	
				118	A/G	A/T	A	T	T	
				213	A	C	C	C	C	
				245	C	A	C	C	C	
				292	C	C/T	T	T	T	
				325	A	G	G	G	G	
				364	G	G	T	T/G	T/G	
				713	C	T	T	T	T	
				735	G	T	G	G	G	
				736	A	A	A	C	A	
				775	A	A	G	A	A	
				811	T	C	T	T	T	
				899	Y	C	C	C	C	
SABP2	400	50	21	44	T	T	T	T	C/T	12.50
				72	A	A	G	A	G	
				76	A	A	C	A	C	
				79	T	C	C	C	C	
				83	G/T	T	T	T	T	
				88	G/C	C	C	C	C	
				94	T	T	T	T	C/T	
				98	T	T	T	T	C	
				100	C/T	C	C	C	C	
				127	A/C	A	A	A	A	
				148	A/G	C	C	C	A/T	
				151	A	G	G	G	G	
				160	A/C	A	A	A	A	
				234	G/T	G	G	G	G	
				244	A	A	A	A	A/G	
				294	A	G	G	G	G	
				299	T	T	T	C	T	
				302	A	T	T	A	T	
				307	A/G	G	G	G	G	
				313	A/G	G	G	G	G	
				347	G	G	G	G	A/G	

Table 3 (continued)

Nuclear regions	Fragment length[a]	No. of var.[b]	No. of S.[c]	Position of S	*ale*	*max*	*med*	*len*	*nan*	% poly.[d]
SAMT	383	68	12	62	C	A/G	C	C	C	17.75
				72	G/T	A/G	A	T	A	
				94	G	G	G	G	G/T	
				118	A	G	G	A	G	
				126	A/T	T	T	A/T	T	
				132	G	A	A	A/G	A	
				191	C	A	A	C	A	
				259	A/C	A	A	C	A	
				265	A/G	A	A	A/G	A	
				333	T	C	C	T	C	
				338	A	C	C	C	A	
				363	T	C	C	T	T	

Substitutions detection was carried out only in the species which have more than one individual in the study

S. Substitutions

[a]Fragment length obtained after sequence analysis and trimming

[b]Total number of nucleotide variations present in all selected individuals

[c]Detected number of substitutions (S)

[d]Percentage variable sites calculated as the total number of polymorphisms in relation to the sequence length

Silencing of ambiguous regions

A phylogenetic tree was reconstructed by deleting the possible ambiguous regions in the league data set to examine the reliability of the generated data (Supplementary Fig. 1) and then compared with the league data set (Fig. 3). Constructing a phylogeny excluding any ambiguous regions did not result in any marked differences (Supplementary Fig. 1). Therefore, it can be concluded that alignment artifacts do not have any substantial effect on the existing phylogenetic analyses.

Network analysis

According to the requirements, only the variable sites of the league data set were used to create a FASTA file and perform the network analysis (Bandelt et al. 1999). The network revealed four clades (Fig. 4) similar to the nuclear phylogeny (Fig. 2). Clade I, II, III, and IV included both high and low MeSA-producing birches with complete heterozygosity between the birches. The network analysis clearly shows the ancestral node formed by *B. lenta* that carries a high MeSA-producing ability.

Discussion

The aim of the present study was to examine the differences in the ability of the multiple chloroplast and nuclear regions to determine the phylogenetic relationship between high and low MeSA-producing birches.

The relevance of the investigated markers

A total number of thirteen regions, including the chloroplast and nuclear regions were investigated for nucleotide variation in selected birches. Among the eight chloroplast regions, the intergenic spacer *trnaH-psbA* delivered the highest number of variable sites, which is consistent with previous studies (Bina et al. 2016; Schroeder et al. 2017). The often-recommended official barcoding region *matK* is more variable in genera like *Lycium* (Fukuda et al. 2001) and indicated low variation in the genus *Betula*, which is in agreement with previous studies (Järvinen et al. 2004; Cräutlein et al. 2011).

Although the official barcoding region *rbcL* varies in *Populus* (Schroeder et al. 2012; Schroeder and Fladung 2015) and complements the *trnH-psbA* sequences (Kress and Erickson 2007), no variants were found in the present study. This observation agrees with the previous analysis that the gene content of the chloroplast genomes is preserved in comparison with the nuclear genome, resulting in low variation (Kato et al. 2000; Järvinen et al. 2004; Palme et al. 2004; Tang et al. 2004). The current study suggests using the multi-locus chloroplast combinations *trnH-psbA*, *matK-trnK,* and *trnC-petN* in agreement with previous results (Fazekas et al. 2008; Hollingsworth et al. 2009; Levin et al. 2009) since the official barcoding and the coding regions of the chloroplast are highly conserved with less polymorphic sites.

In agreement with Järvinen et al. (2004), the low-copy nuclear *ADH* gene region displayed the highest number of variables in this study, most probably since the amplified *ADH* sequences include few exons and few introns. The

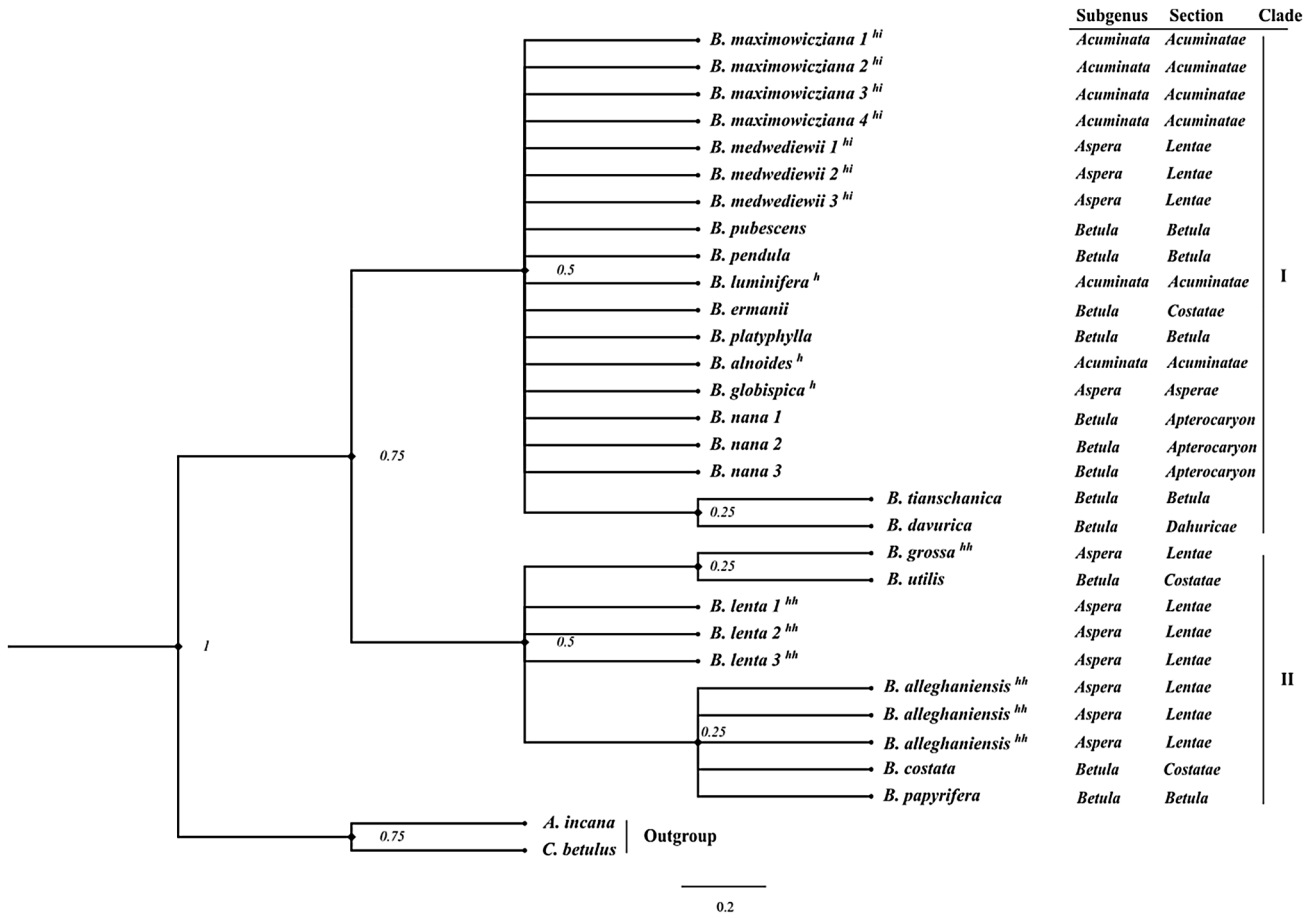

Fig. 1 Maximum likelihood (ML) phylogenetic tree based on eight cpDNA regions of different *Betula* species, using the substitution model JC69 + I for heterozygosity in the program PAUP* 4.0a. The *Betula* taxa with "h" produce a high amount of MeSA according to Ashburner and McAllister (2013), while those with "hh" produce a high amount of MeSA according to the olfactory fragrance analysis conducted in this study. The species with "hi" represent "h" high (according to Ashburner and McAllister (2013)) and "i" intermediate (according to the olfactory analysis, respectively)

universal ITS primers are also able to amplify fungal DNA and additional steps are needed to confirm the resulting sequences (Jobes et al. 1997). To bypass this step, *Betula*-specific ITS primers were designed (Supplementary Table 1) using *B. pendula* as the reference genome (Salojärvi et al. 2017). The new primers allowed for an easy amplification of the nuclear ribosomal ITS regions (ITS1 and ITS2). According to previous analysis (Li et al. 2007), *NIA* is a useful marker and helps in the inference of *Betula* phylogeny. However, *NIA* demonstrates low genetic variation. The reason could be that the new *NIA* PCR primers, designed on the first exon of the gene in the current study, might reduce the degree of polymorphism.

The *SABP2* and *SAMT* gene regions are involved in the biogenesis of MeSA and serve as a signaling molecule in the systemic acquired resistance (SAR) mechanism of plants (Shah et al. 2014). Considering their organized interaction with numerous molecules, *SABP2* and *SAMT* genes displayed a higher percentage of polymorphism (17.75% and 12.50%, respectively) in comparison with all analyzed regions. The current analysis indicates that slow-evolving uniparental chloroplast and fast-evolving bi-parental nuclear regions are more efficient in discrimination of the low and high MeSA birch producers than those gene sets studied independently.

Phylogenetic relationship of birches

The phylogenetic analysis of the genus *Betula* showed that the MeSA producing ability is not monophyletic but distributed among species from different subgenera. The phylogeny based on chloroplast regions (Fig. 1) revealed the relationship between three North American species, *B. alleghaniensis, B. lenta,* the high MeSA-producers and *B. papyrifera* located in one clade. Although we used eight different chloroplast regions, the results based on the birch

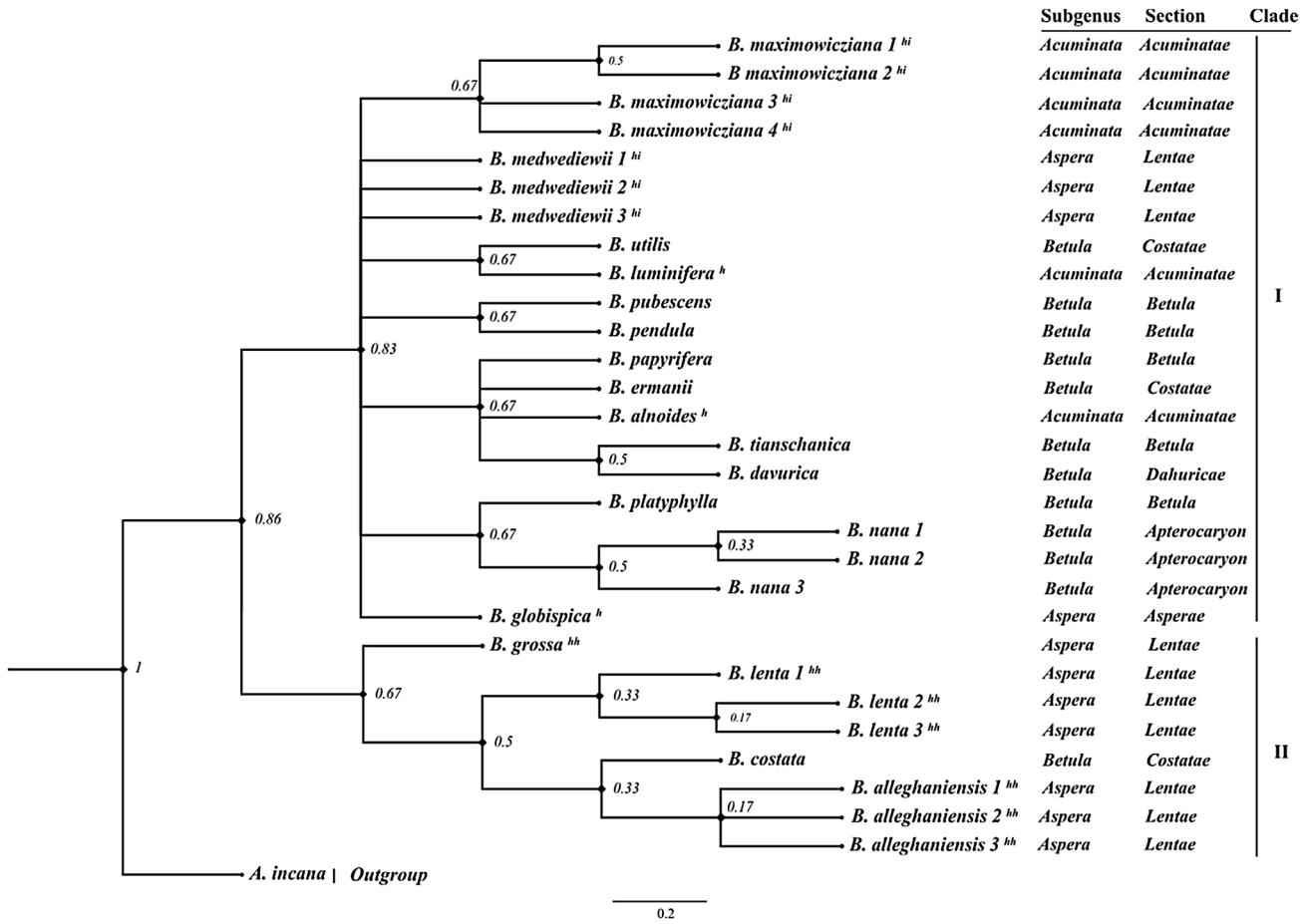

Fig. 2 Maximum likelihood (ML) phylogenetic tree based on five nuclear regions of different *Betula* species, using the substitution model JC69 + I for heterozygosity in the program PAUP* 4.0a. The *Betula* taxa with "h" produce a high amount of MeSA according to Ashburner and McAllister (2013), while those with "hh" produce a high amount of MeSA according to the olfactory fragrance analysis conducted in this study. The species with "hi" represent "h" high (according to Ashburner and McAllister (2013)) and "i" intermediate (according to the olfactory analysis, respectively)

species considered in Järvinen et al. (2004) and Bina et al. (2016) are consistent based on just the chloroplast either *matK* or *trnH-psbA* regions. However, *B. grossa*, *B. costata* and *B. utilis* were not studied in Järvinen et al. (2004) and *B. grossa* not in Bina et al. (2016). In our study, *B. utilis* is closely grouped with *B. grossa*, while it formed a separate lineage in the phylogeny based on *trnH-psbA* region (Bina et al. 2016).

In addition to the chloroplast regions, five fast-evolving nuclear regions were also used to generate robust data for phylogenetic analysis. In the nuclear phylogeny (Fig. 2), *B. papyrifera* and *B. utilis* formed a branch with the species from the subgenus *Betula*, which agrees with previous ITS phylogenies (Bina et al. 2016; Wang et al. 2016). Individuals in the nuclear phylogeny (Fig. 2) were more revealing when compared with the chloroplast region (Fig. 1), which might be due to a high number of variations in the nuclear regions.

The combined analysis (Fig. 3) was influenced by phylogenetic signals from nuclear sequences and chloroplast regions which contributed enough information to improve resolution within the clades. It could also explain why the league data set (Fig. 3) resembles that of the nuclear phylogeny (Fig. 2). It is possible that most variable sequences dominate the moderate data set in a concatenated analysis, thus resulting in the topology (Gontcharov et al. 2004). Species from the subgenus *Aspera* (*B. alleghaniensis, B. grossa* and *B. lenta*) formed a separate clade, whereas species from the subgenus *Betula* were congregated in another clade (Figs. 1, 2 and 3). Section-wise distribution of species from the subgenus *Betula* remains complex in all the data sets (Figs. 1, 2 and 3), which supports previous analysis (Li et al. 2005; Bina et al. 2016; Wang et al. 2016).

Due to the morphological similarities with *B. lenta* and *B. alleghaniensis*, the hexaploid *B. grossa* is placed in the *Lentae* section by Ashburner and McAllister (2013), agreeing

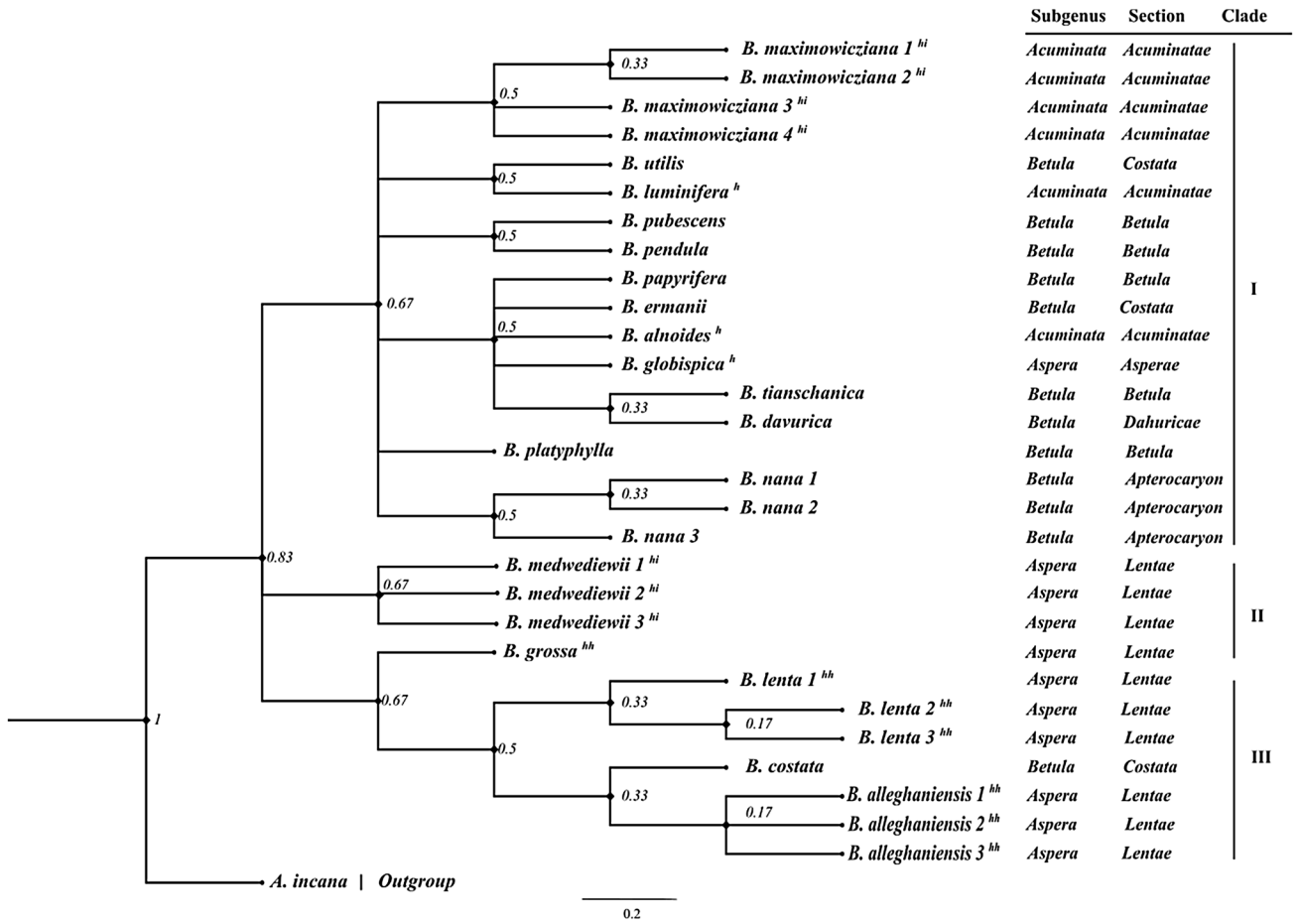

Fig. 3 Maximum likelihood (ML) phylogenetic tree based on the league data set of different *Betula* species, using the substitution model JC69 + I for heterozygosity in the program PAUP* 4.0a. The *Betula* taxa with "h" produce a high amount of MeSA according to Ashburner and McAllister (2013), while those with "hh" produce a high amount of MeSA according to the olfactory fragrance analysis conducted in this study. The species with "hi" represents, "h" high (according to Ashburner and McAllister (2013)) and "i" intermediate (according to the olfactory analysis, respectively)

with our phylogenetic analysis (Figs. 1, 2 and 3). However, this result disagrees with previous results (Nagamitsu et al. 2006; Schenk et al. 2008; Wang et al. 2016), where it is clustered with species from the subgenus *Betula*. The reason for this incongruity might be the multiple loci used and may also indicate that one of the parents of *B. grossa* belongs to the subgenus *Betula* (Wang et al. 2016). The possible hybridization with a morphologically similar species of section *Lentae* lead to the formation of an allopolyploid, resulting in it being thus placed (Ashburner and McAllister 2013; Wang et al. 2016).

Ashburner and McAllister (2013) placed *B. globispica* in a list of high MeSA producer *Betula* species that is incongruous with our olfactory analysis (Supplementary Table 1) where it produced no MeSA fragrance. Also, in the phylogenetic analysis (Figs. 1, 2, 3 and Supplementary Fig. 1), it clustered with species from the subgenus *Betula*, which is consistent with previous results (Bina et al. 2016). The reason for this might be that *B. globispica* hybridized with a species from the *Betula* section and thus appeared with the species from the *Betula* section (Wang et al. 2016).

According to the olfactory analysis (Supplementary Table 1), species from the subgenus *Acuminata* are either intermediate or low MeSA producers and always clustered with species from the subgenus *Betula*. They did not form a distinct clade in our analysis (Figs. 1, 2 and 3 and Supplementary Fig. 1). The close relationship between the diploid *B. maximowicziana* and subgenus *Betula* clearly supports the previous AFLP and ITS studies (Schenk et al. 2008; Wang et al. 2016). The olfactory analysis partially agrees with the morphological distribution (Ashburner and McAllister 2013) since the fragrance analysis could not detect MeSA in the diploid *B. luminifera* and tetraploid *B. alnoides*, suggesting further analytical examination. Additionally, it is experimentally proven that no post-zygotic barriers exist between *B. maximowicziana* and *B. pendula* subspecies *mandshurica.*

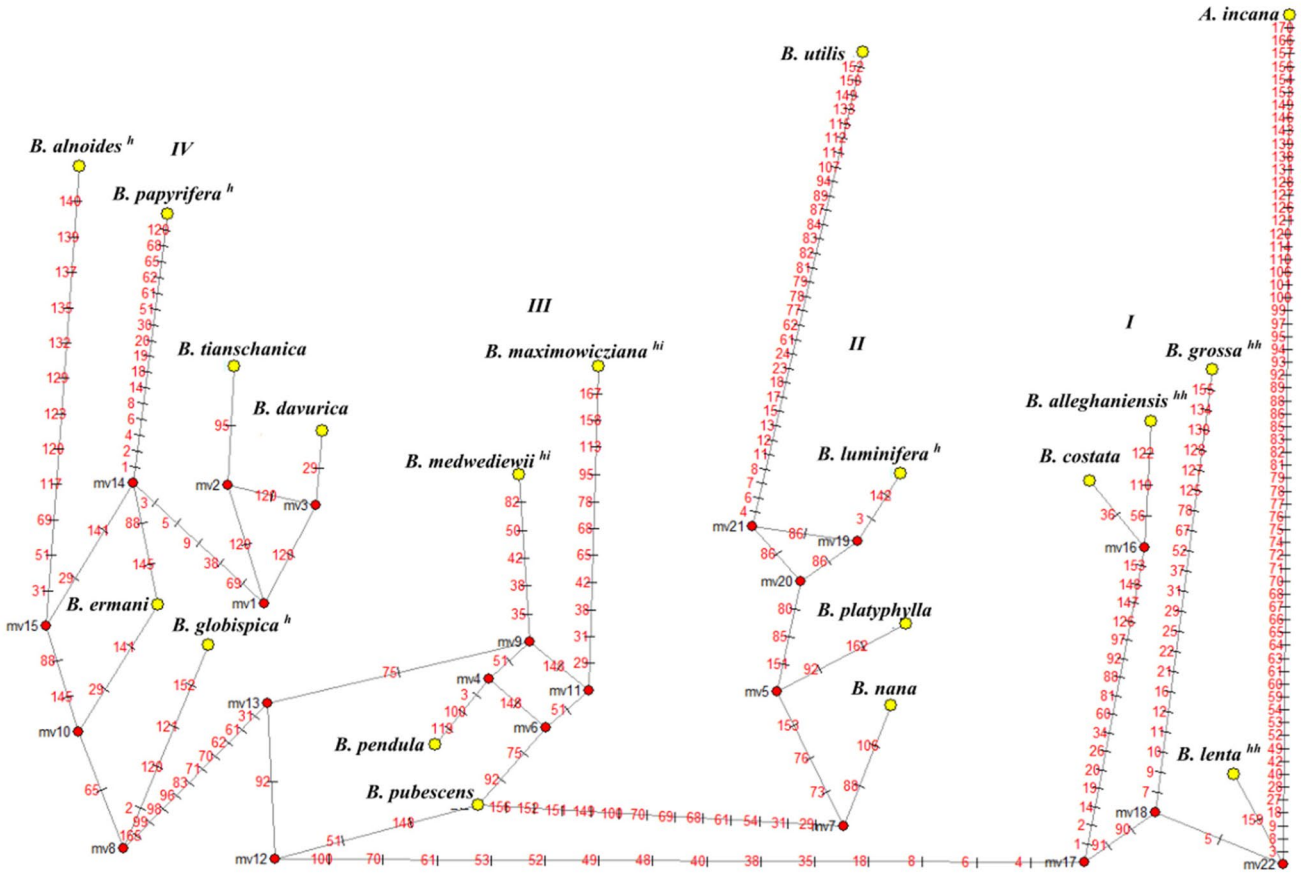

Fig. 4 The evolutionary network analysis of the *Betula* individuals based on variable sites of eight cpDNA and five nuclear loci. Red points (mv) and numbers on the lineages refer to ancestral nodes and mutation events, respectively. The *Betula* taxa with "h" produce a high amount of MeSA according to Ashburner and McAllister (2013), while those with "hh" produce a high amount of MeSA according to the olfactory fragrance analysis conducted in this study. The species with "hi" represent "h" high (according to Ashburner and McAllister (2013)) and "i" intermediate (according to the olfactory analysis, respectively)

Thus, fertile hybrids can be formed in crosses between two species (Johnsson 1945). Such evidence of hybridization could explain the intermediate or low production of MeSA in the subgenus *Acuminata.*

The decaploid *B. medwediewii* always clustered with different species of the subgenus *Betula* (Figs. 1, 2, 3 and Supplementary Fig. 1), in contrast to previous results where it clustered with *B. lenta* and *B. alleghaniensis* (Li et al. 2005; Bina et al. 2016; Wang et al. 2016). All three *B. medwediewii* accessions were from different botanical gardens but each displayed similar clustering, making misidentification unlikely. The unexpected placement of *B. medwediewii* in a clade with species of the subgenus *Betula* may indicate that one of the progenitors of this polyploid belongs to the subgenus *Betula.* This possibility is likely since in the league data set (Fig. 3), *B. medwediewii* occupied its own clade. In addition, it is an intermediate MeSA producer species (Supplementary Table 1) suggesting its authenticity as well as providing a valid reason for Ashburner and McAllister (2013) placing it in the subgenus *Aspera*.

The current analysis led to the identification of a misclassified *B. alleghaniensis_10* individual from the arboretum of the Thünen Institute in Grosshansdorf, Germany (Supplementary Fig. 2). Both phylogenies (Fig. 3 and Supplementary Fig. 2) were constructed using the league data set. However, considerable differences were found, suggesting that the misclassification of a single individual could affect the complete phylogenetic tree. Such conflicting results and differing opinions are common scenarios in the genus *Betula*, and could be due to the continuous process of hybridization and introgression (Furlow 1990; Järvinen et al. 2004).

In addition, grouping of the diploid *B. costata* (subgenus, *Betula*; section, *Costatae*) with species from the section *Lentae* (Figs. 1, 2 and 3 and Supplementary Fig. 1) is consistent with former studies (Li et al. 2005; Bina et al. 2016). However, *B. costata* individuals used in the present study passed the olfactory analysis (Supplementary Table 1) producing considerable MeSA fragrance, thus contradicting the earlier report (Ashburner and McAllister 2013). However,

the possibility of hybridization with one of the high MeSA-producing *Betula* species from East Asia and the problem of misclassification needs to be considered. Nevertheless, an analytical examination of more individuals will be required for stronger conclusions regarding the taxonomical and high MeSA-producing status of *B. costata.*

The use of *SAMT* and *SABP2* was considered worthwhile for the phylogenetic analysis since these genes are involved in the biosynthesis of MeSA (Shah et al. 2014). As far as we know, this is the first study where *SABP2* and *SAMT* are used in the phylogenetic analysis of *Betula.* The *SAMT* gene was previously investigated for the Solanaceae family (Talline et al. 2005) and provided advantages and disadvantages in the analysis. In the current analysis, although *SAMT* and *SABP2* gene regions displayed higher variations (Table 3), they did not confirm a profound phylogenetic relationship between the birches.

An assessment of the sequencing data set generated is a vital operation in systematics for characterizing intrinsic features. The alignment of sequences generated from diverse genomic regions and species may produce ambiguous regions resulting in incongruent assumptions. In addition, if these regions are ignored, they may threaten the correct distribution of the species in the phylogeny (Lutzoni et al. 2000). Nevertheless, differences between the chloroplast and nuclear phylogenies have been reported in previous studies (Semerikov et al. 2003; Järvinen et al. 2004). This also occurred, to a certain extent, in the current analysis, suggesting that multigene analyses may not fully resolve all internal branches (Karol et al. 2001). Considering the disagreement between the data sets, a phylogenetic tree was constructed which ignored ambiguous regions in the alignment (Supplementary Fig. 1). In this case, the phylogeny displayed similar topologies to the nuclear phylogeny (Fig. 3) with better resolution through the branch length. It was concluded that alignment artifacts do not have any substantial effect on the phylogenetic analyses.

In the chloroplast (Fig. 1), nuclear (Fig. 2) and league data (Fig. 3) phylogenies, an exclusive clustering was observed between *B. grossa*, *B. alleghaniensis* and *B. lenta*, which produce high MeSA levels in the bark (Supplementary Table 1; Ashburner and McAllister 2013). The constant phylogenetic positioning of these species suggests their evolutionary relationships. It is possible that one of the species could be among the ancestors of selected *Betula* species. Although, the biogeographic reconstruction analysis by Bina et al. (2016) suggests that the genus *Betula* originated in East Asia. Bina et al. (2016) and De Jong (1993) proposed that the two subgenera of *Betulenta* (including *B. grossa*, *B. alleghaniensis,* and *B. lenta*) and *Betulaster* contain the oldest species, followed by the ancestral reconstruction that advocated the migration of the species from North America to Europe (Bina et al. 2016). Even the occurrence of disjunction events recognized within the genus *Betula* has been reported between eastern Asia and North America (*B. costata* and *B. alleghaniensis*) (Li et al. 2005). Following this disjunction, Wang et al. (2016) also hypothesized that the common ancestor of the *Betula* species might have been constantly scattered over the Northern Hemisphere. Bina et al. (2016) also support the occurrence of vicariance events in different regions, including the East Asian, North American and European lineages.

Considering parallel migration, hybridization and introgression events in the genus *Betula* and the network analysis (Fig. 4), we speculate that the diploid *B. lenta*, (subgenera; *Betulenta*) which has the ability to produce high MeSA, is one of the ancestors of the genus *Betula*. To construct the network analysis (Fig. 4), we utilized a substantial number of nuclear and chloroplast regions. Our network analysis (Fig. 4) results are consistent with previous studies where *B. lenta* also formed an ancestor node (Bina et al. 2016), along with the high MeSA producer *B. uber* (Ashburner and McAllister 2013).

We also speculate that in the process of evolution, the ancestrally high MeSA-producing trait was passed on to some younger *Betula* species, whereas some species lost their MeSA-producing ability, suggesting divergence in the evolutionary events among the species (Hughes 2012; Bina et al. 2016). Our analysis supports some of the previous studies; traits which are not maintained through natural selection may become less functional over time and follow a relaxed evolutionary selection (Reich et al. 2003). If some traits have no active function in the genome, they could disappear in future generations (Lahti et al. 2009).

Conclusions

The present study provides initial knowledge on the distribution of the high and low MeSA-producing trait in the *Betula* genera. The network analysis suggests that *B. lenta* is one of the ancestors and that the MeSA-producing trait has been lost several times in the evolution of younger *Betula* species. The chloroplast and nuclear phylogenies reveal homogenous topologies suggesting compatibility between organelle and nuclear results. This is noteworthy, considering that even a small segment of a nuclear DNA sequence produces high-resolution levels and provides more parsimony information to the chloroplast DNA. The data generated could be important for the Barcode of Life project and will be used to develop rapid molecular test methods to reduce the cost of sequencing for the breeding and selection of birches.

Comprehensive sampling together with genome-wide studies could resolve the questionable phylogenetic position of some *Betula* species. Additionally, an analytical examination can help to make precise decisions about the taxonomical positions of high and low MeSA-producing *Betula* species.

Future perspectives

The study successfully provides strong evidence about the evolutionary state of the ancestral species to establish a visible and reliable basis for further next-generation sequencing experiments. The *SABP2* and *SAMT* genes serve as candidate genes for MeSA biosynthesis. The original idea to include these genes in the phylogenetic analysis was to correlate putative nucleotide substitutions with MeSA production. On the basis of the current study, different low and high MeSA-producing *Betula* species will be selected for bark and leaf RNA sequencing analyses.

Acknowledgements Open Access funding provided by Projekt DEAL. We are highly grateful to the botanical gardens and Northwest German Forestry Research Institute (listed in Table 1) for providing *Betula* seed samples upon our request for scientific research purposes only without any conflict of interest. We thank PD Dr. Birgit Kersten for assisting with the submission of data to the NCBI GenBank, to Dr. Niels Müller for consultation in sequencing and the manuscript review (both Thünen Institute of Forest Genetics, Grosshansdorf, Germany), and to all colleagues at the Thünen Institute of Forest Genetics, Grosshansdorf, Germany and the Institut für Landwirtschaftliche Verfahrenstechnik Kiel for their support, including Katrin Groppe for laboratory assistance, and Dr. Hilke Schröder, Dr. Jutta Buschbom and Khira Deecke for helpful discussions on phylogenic analyses. Thanks are due to Dina Führmann (Thünen Institute, Braunschweig, Germany) and Nicola Wilton (Language Boutique®, Hamburg, Germany) for English language editing.

Author contribution statement MF is the first supervisor of this work and contributed to the experimental outline, data validation and correcting of the manuscript. Project leader CRM is the second supervisor and contributed to sample collection and writing the project proposal. EH provided guidance throughout the experiments and was also responsible for English language editing. Lab work, writing of the paper, data collection and analysis was completed by KS as a part of his Ph.D. work.

Funding This project was funded by the "Energiewende und Umweltinnovationen" state program for the economy; the Ministry of Energy Transition, Agriculture, Environment and Rural Areas of Schleswig–Holstein, Germany.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

Angmo K, Adhikari BS, Rawat GS (2012) Changing aspects of traditional healthcare system in western Ladakh, India. J Ethnopharmacol. https://doi.org/10.1016/j.jep.2012.07.017

Ashburner K, McAllister HA (2013) The genus Betula: a taxonomic revision of birches. Kew Publishing, London

Bandelt HJ, Peter F, Arne R (1999) Median-joining networks for inferring intraspecific phylogenies. Mol Biol Evol. https://doi.org/10.1093/oxfordjournals.molbev.a026036

Baser KHC, Demirci B (2007) Studies on Betula essential oils. ARKIVOC. https://doi.org/10.3998/ark.5550190.0008.730

Baum D (2008) Trait evolution on a phylogenetic tree: relatedness, similarity, and the myth of evolutionary advancement; Nature Education. https://www.nature.com/scitable/topicpage/trait-evolution-on-a-phylogenetic-tree-relatedness-41936/. Accessed 15 Jun 2019

Bina H, Yousefzadeh H, Ali SS et al (2016) Phylogenetic relationships, molecular taxonomy, biogeography of Betula, with emphasis on phylogenetic position of Iranian populations. Tree Genet Genomes. https://doi.org/10.1007/s11295-016-1037-4

CBOL Plant Working Group, Hollingsworth PM, Forrest L et al (2009) DNA barcode for land plants. Proc Natl Acad Sci. https://doi.org/10.1073/pnas.0905845106

Cräutlein M, Korpelainen H, Pietiläinen M (2011) DNA barcoding: a tool for improved taxon identification and detection of species diversity. Biodivers Conserv. https://doi.org/10.1007/s10531-010-9964-0

Dadákova E, Vrchotová N, Tříska J (2010) Content of selected biologically active compounds in tea infusions of widely used European medicinal plants. J Agrobiol. https://doi.org/10.2478/s10146-009-0003-2

De Jong PC (1993) An introduction to Betula: Its morphology, evolution, classification and distribution with a survey of recent work, Proceedings of the IDS Betula symposium, 2–4 October 1992. International Dendrology Society, Richmond

Du H, Klessig DF (1997) Identification of a soluble, high-affinity salicylic acid-binding protein in Tobacco. Plant Physiol. https://doi.org/10.1104/pp.113.4.1319

Dumolin S, Demesure B, Petit RJ (1995) Inheritance of chloroplast and mitochondrial genomes in pedunculate oak investigated with an efficient PCR method. Theor Appl Genet. https://doi.org/10.1007/BF00220937

Fazekas AJ, Burgess KS, Kesanakurti PR, Graham SW et al (2008) Multiple multilocus DNA barcodes from the plastid genome discriminate plant species equally well. PLoS ONE. https://doi.org/10.1371/journal.pone.0002802

Ferris C, King RA, Gray AJ (1997) Molecular evidence for the maternal parentage in the hybrid origin of Spartina anglica C.E. Hubbard. Mol Ecol. https://doi.org/10.1046/j.1365-294X.1997.00165.x

Forouhar F, Yang Y, Kumar D, Chen Y et al (2005) Structural and biochemical studies identify tobacco SABP2 as a methyl

salicylate esterase and implicate it in plant innate immunity. PNAS. https://doi.org/10.1073/pnas.0409227102
Fukuda T, Yokoyama J, Ohashi H (2001) Phylogeny and biogeography of the genus Lycium (Solanaceae): inferences from chloroplast DNA sequences. Mol Phylogenet Evol. https://doi.org/10.1006/mpev.2001.0921
Furlow JJ (1990) The genera of Betulaceae in the southeastern United States. J Arnold Arboretum 71:1–67. https://biostor.org/reference/61852
Gontcharov AA, Marin B, Melkonian M (2004) Are combined analyses better than single gene phylogenies? A case study using SSU rDNA and rbcL sequence comparisons in the Zygnematophyceae (Streptophyta). Mol Biol Evol. https://doi.org/10.1093/molbev/msh052
Hinchliff CE, Smith SA, Allman JF, Burleigh JG et al (2015) Synthesis of phylogeny and taxonomy into a comprehensive tree of life. Proc Natl Acad Sci USA. https://doi.org/10.1073/pnas.1423041112
Hollingsworth ML, Andra CA, Forrest LL et al (2009) Selecting barcoding loci for plants: evaluation of seven candidate loci with species-level sampling in three divergent groups of land plants. Mol Ecol Resour. https://doi.org/10.1111/j.1755-0998.2008.02439.x
Hollingsworth PM, Graham SW, Little DP (2011) Choosing and using a plant DNA barcode. PLoS ONE. https://doi.org/10.1371/journal.pone.0019254
Hughes AL (2012) Evolution of adaptive phenotypic traits without positive Darwinian selection. Heredity. https://doi.org/10.1038/hdy.2011.97
Järvinen P, Palmé A, Morales LO et al (2004) Phylogenetic relationships of Betula species (Betulaceae) based on nuclear ADH and chloroplast matK sequences. Am J Bot. https://doi.org/10.3732/ajb.91.11.1834
Jobes DV, Thien LB (1997) A conserved Motif in the 5.8S Ribosomal RNA (rRNA) gene is a useful diagnostic marker for plant internal transcribed spacer (ITS) sequences. Mol Biol Report. https://doi.org/10.1023/A:1007462330699
Johnson AD (2010) An extended IUPAC nomenclature code for polymorphic nucleic acids. Bioinformatics. https://doi.org/10.1093/bioinformatics/btq098
Johnsson H (1945) Interspecific Hybridization within the genus Betula. Hereditas. https://doi.org/10.1111/j.1601-5223.1945.tb02752.x
Karol KG, McCourt RM, Cimino MT, Delwiche CF (2001) The closest living relatives of land plants. Science. https://doi.org/10.1126/science.1065156
Kato T, Kaneko T, Sato S, Nakamura Y, Tabata S (2000) Complete structure of the chloroplast genome of a Legume, Lotus japonicus. DNA Res 5:6. https://doi.org/10.1093/dnares/7.6.323
Kress WJ, Erickson DL (2007) A two-locus global DNA barcode for land plants: the coding rbcL gene complements the non-coding trnH-psbA spacer region. PLoS ONE. https://doi.org/10.1371/journal.pone.0000508
Kumar S, Stecher G, Tamura K (2016) MEGA7: molecular evolutionary genetics analysis version 7.0 for Bigger Datasets. Mol Biol Evol. https://doi.org/10.1093/molbev/msw054
Lahti DC, Johnson NA, Ajie BC, Otto SP et al (2009) Relaxed selection in the wild. Trends Ecol Evol. https://doi.org/10.1016/j.tree.2009.03.010
Levin RA, Blanton J, Miller JS (2009) Phylogenetic utility of nuclear nitrate reductase: a multi-locus comparison of nuclear and chloroplast sequence data for inference of relationships among American Lycieae (Solanaceae). Mol Phylogenet Evol. https://doi.org/10.1016/j.ympev.2008.12.005
Li JH, Shoup S, Chen ZD (2005) Phylogenetics of Betula (Betulaceae) inferred from sequences of nuclear ribosomal DNA. Rhodora. https://doi.org/10.3119/04-14.1
Li JH, Shoup S, Chen ZD (2007) Phylogenetic relationships of diploid species of Betula (Betulaceae) inferred from DNA sequences of nuclear nitrate reductase. Syst Bot. https://doi.org/10.1600/036364407781179699
Lutzoni F, Wagner P, Reeb V, Zoller S (2000) Integrating ambiguously aligned regions of DNA sequences in phylogenetic analyses without violating positional homology. Syst Biol. https://doi.org/10.1080/106351500750049743
Maddison D (1991) The discovery and importance of multiple Islands of most-parsimonious trees. Syst Biol. https://doi.org/10.1093/sysbio/40.3.315
Martins TR, Barkman TJ (2005) Reconstruction of solanaceae phylogeny using the nuclear gene SAMT. Syst Bot. https://doi.org/10.1600/0363644054223675
Menković N, Savikin K, Tasić S, Zdunić G, Stesević D et al (2011) Ethnobotanical study on traditional uses of wild medicinal plants in Prokletije Mountains (Montenegro). J Ethnopharmacol. https://doi.org/10.1016/j.jep.2010.09.008
Nagamitsu T, Kawahara T, Kanazashi A (2006) Endemic dwarf birch *Betula apoiensis* (Betulaceae) is a hybrid that originated from *Betula ermanii* and *Betula ovalifolia*. Plant Species Biol. https://doi.org/10.1111/j.1442-1984.2006.00147.x
Pal M, Mishra T, Kumar A et al (2015) Chemical constituents and antimicrobial potential of essential oil from *Betula utilis* growing in high altitude of Himalaya (India). J Essent Oil Bear Plants. https://doi.org/10.1080/0972060X.2015.1036569
Palme 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. Mol Ecol. https://doi.org/10.1046/j.1365-294X.2003.02034.x
Rastogi S, Pandey MM, Kumar Singh Rawat A (2015) Medicinal plants of the genus Betula-traditional uses and a phytochemical-pharmacological review. J Ethnopharmacol. https://doi.org/10.1016/j.jep.2014.11.010
Regel E (1865) Bemerkungen uber die Gattungen Betula und Alnus nebst Beschreibung einiger neuer Arten. Bull Soc Nat (Moscou) 38:388–434
Reich PB, Wright IJ, Cavender-Bares JM, Craine J et al (2003) The evolution of plant functional variation: traits, spectra, and strategies. Int J Plant Sci. https://doi.org/10.1086/374368
Repola J (2008) Biomass equations for birch in Finland. Silva Fennica. https://doi.org/10.14214/sf.236
Ross JR, Nam KH, D'Auria JC, Pichersky E (1999) S-Adenosyl-L-methionine: salicylic acid carboxyl methyltransferase, an enzyme involved in floral scent production and plant defense, represents a new class of plant methyltransferases1. Arch Biochem Biophys. https://doi.org/10.1006/abbi.1999.1255
Salojärvi J, Smolander O, Nieminen K, Rajaraman S, Safronov O et al (2017) Genome sequencing and population genomic analyses provide insights into the adaptive landscape of silver birch. Nat Genet. https://doi.org/10.1038/ng.3862
Schenk MF, Thienpont CN, Koopman WJM et al (2008) Phylogenetic relationships in Betula (Betulaceae) based on AFLP markers. Tree Genet Genom. https://doi.org/10.1007/s11295-008-0162-0
Schroeder H, Hoeltken AM, Fladung M (2012) Differentiation of Populus species using chloroplast single nucleotide polymorphism (SNP) markers-essential for comprehensible and reliable poplar breeding. Plant Biol. https://doi.org/10.1111/j.1438-8677.2011.00502.x
Schroeder H, Fladung M (2015) Differentiation of Populus species by chloroplast SNP markers for barcoding and breeding approaches. iForest Biogeosci Forestry. https://doi.org/10.3832/ifor1326-007

Schroeder H, Kersten B, Fladung M (2017) Development of multiplexed marker sets to identify the most relevant poplar species for breeding. Forests. https://doi.org/10.3390/f8120492

Semerikov VL, Yhang H, Sun M, Lascoux M (2003) Conflicting phylogenies of Larix (Pinaceae) based on cytoplasmic and nuclear DNA. Mol Phylogenet Evol. https://doi.org/10.1016/S1055-7903(02)00447-5

Shah J, Chaturvedi R, Chowdhury Z, Venables B, Petros RA (2014) Signaling by small metabolites in systemic acquired resistance. Plant J Cell Mol Biol. https://doi.org/10.1111/tpj.12464

Shaw K, Stritch L, Rivers M, Roy S, Becky W, Govaerts R (2014) The Red List of Betulaceae. BGCI. Richmond. https://www.bgci.org/resources/bgci-tools-and-resources/the-red-list-of-betulaceae/. Accessed 15 Oct 2019

Shikov AN, Djachuk GI, Sergeev DV et al (2011) Birch bark extract as therapy for chronic hepatitis C-a pilot study. Phytomedicine. https://doi.org/10.1016/j.phymed.2011.01.021

Singh B, Sharma RA (2015) Plant terpenes: defense responses, phylogenetic analysis, regulation and clinical applications. 3 Biotech. https://doi.org/10.1007/s13205-014-0220-2

Skvortsov AK (2002) A new system of the genus *Betula* L.-the birch. Bull Mosc Soc Nat 107:73–76

Small RL (2004) Phylogeny of Hibiscus sect. Muenchhusia (Malvaceae) Based on chloroplast rpL16 and ndhF, and nuclear ITS and GBSSI Sequences. Syst Bot. https://doi.org/10.1600/036364404774195575

Small RL, Cronn RC, Wendel JF (2004) Use of nuclear genes for phylogeny reconstruction in plants. Aust Syst Bot. https://doi.org/10.1071/sb03015

Smith A, Granhus A, Astrup R et al (2014) Functions for estimating above ground biomass of birch in Norway. Scand J For Res. https://doi.org/10.1080/02827581.2014.951389

Tang J, Xia H, Cao M et al (2004) A comparison of rice chloroplast genomes. Plant Physiol. https://doi.org/10.1104/pp.103.031245

Wang N, McAllister HA, Bartlett PR et al (2016) Molecular phylogeny and genome size evolution of the genus Betula (Betulaceae). Ann Bot. https://doi.org/10.1093/aob/mcw048

Wang N, Thomson M, Bodles WJ et al (2013) Genome sequence of dwarf birch (Betula nana) and cross-species RAD markers. Mol Ecol. https://doi.org/10.1111/mec.12131

White TJ, Bruns T, Lee S, Taylor T (1990) Amplification and direct sequencing of fungal ribosomal RNA Genes for phylogenetics. PCR protocols: a guide to methods and applications. Academic Press, New York, pp 315–322

Williams J Jr, Arnold M (2001) Sources of genetic structure in the woody perennial *Betula occidentalis*. Int J Plant Sci. https://doi.org/10.1086/321921

Winkler H (1904) Betulaceae. In: Engler A (ed) Das Pflanzenreich Heft, 19 (IV.61), 1–49. W. Engelmann, Leipzig

Yin J, Li X, Zhan Y et al (2017) Cloning and expression of BpMYC4 and BpbHLH9 genes and the role of BpbHLH9 in triterpenoid synthesis in birch. BMC Plant Biol. https://doi.org/10.1186/s12870-017-1150-z

3 Identification and analysis of key genes involved in methyl salicylate biosynthesis in different birch species

(Published in *PLOS ONE*; 15(10): e0240246, October 2020)

3.1 Key message

Expression and bioinformatics analyses of the two candidate genes in different high and low methyl salicylate producing birch species aided to select the species for further whole-genome and transcriptome sequencing. This initial knowledge was essential to understand the genetic architecture of the candidate genes which will be helpful in the follow-up projects. The analysis was important to understand the belonging of the candidate genes to different families, as described in the fourth chapter of the thesis.

PLOS ONE

RESEARCH ARTICLE

Identification and analysis of key genes involved in methyl salicylate biosynthesis in different birch species

Kiran Singewar[1,2]*, **Christian R. Moschner**[1], **Eberhard Hartung**[1], **Matthias Fladung**[2]*

1 Institute of Agricultural Process Engineering, Christian-Albrechts University of Kiel, Kiel, Schleswig-Holstein, Germany, **2** Thünen Institute of Forest Genetics, Grosshansdorf, Schleswig-Holstein, Germany

* kiran.singewar@thuenen.de (KS); matthias.fladung@thuenen.de (MF)

Citation: Singewar K, Moschner CR, Hartung E, Fladung M (2020) Identification and analysis of key genes involved in methyl salicylate biosynthesis in different birch species. PLoS ONE 15(10): e0240246. https://doi.org/10.1371/journal.pone.0240246

Editor: Heping Cao, USDA-ARS Southeast Area, UNITED STATES

Received: June 19, 2020

Accepted: September 22, 2020

Published: October 8, 2020

Data Availability Statement: All relevant data is available within the manuscript and the Supporting Information file. Also, all the sequence accession numbers obtained from the NCBI gene bank are mentioned in the Supporting Information file.

Funding: Thank you very much to the funding agency; "Energiewende und Umweltinnovationen" state program for the economy; Ministry for Energy Transition, Agriculture, Environment, Nature and Digitalization of Schleswig–Holstein, Germany. (Project number: LPW-L/1.2/24). The

Abstract

Species of the perennial woody plant genus *Betula* dominate subalpine forests and play a significant role in preserving biological diversity. In addition to their conventional benefits, birches synthesize a wide range of secondary metabolites having pharmacological significance. Methyl salicylate (MeSA) is one of these naturally occurring compounds constitutively produced by different birch species. MeSA is therapeutically important in human medicine for muscle injuries and joint pain. However, MeSA is now mainly produced synthetically due to a lack of information relating to MeSA biosynthesis and regulation. In this study, we performed a comprehensive bioinformatics analysis of two candidate genes mediating MeSA biosynthesis, *SALICYLIC ACID METHYLTRANSFERASE* (*SAMT*) and *SALICYLIC ACID-BINDING PROTEIN* 2 (*SABP2*), of high (*B. lenta*, *B. alleghaniensis*, *B. medwediewii*, and *B. grossa*) and low (*B. pendula*, *B. utilis*, *B. alnoides*, and *B. nana*) MeSA-producing birch species. Phylogenetic analyses of *SAMT* and *SABP2* genes and homologous genes from other plant species confirmed their evolutionary relationships. Multiple sequence alignments of the amino acid revealed the occurrence of important residues for substrate specificity in SAMT and SABP2. The analysis of *cis* elements in different birches indicated a functional multiplicity of *SAMT* and *SABP2* and provided insights into the regulation of both genes. We successfully developed six prominent single nucleotide substitution markers that were validated with 38 additional birch individuals to differentiate high and low MeSA-producing birch species. Relative tissue-specific expression analysis of *SAMT* in leaf and bark tissue of two high and two low MeSA-synthesizing birches revealed a high expression in the bark of both high MeSA-synthesizing birches. In contrast, *SABP2* expression in tissues revealed indifferent levels of expression between species belonging to the two groups. The comparative expression and bioinformatics analyses provided vital information that could be used to apply plant genetic engineering technology in the mass production of organic MeSA.

Introduction

Methyl salicylate (MeSA) is a volatile compound, widespread in many plant species, that has been extensively studied as a long-distance mobile signaling molecule in systemic acquired

Thünen -Institute provided basic office and lab equipment to perform this study, we thank for the cooperation. We acknowledge financial support by Land Schleswig-Holstein within the funding programme Open Access Publikationsfonds.

Competing interests: The authors have declared that no competing interests exist.

resistance (SAR) [1]. SAR is an inducible defense mechanism, activated in response to pathogen attack [2]. Salicylic acid (SA) produces MeSA by the action of salicylic acid methyltransferase (SAMT) utilizing S-adenosyl-L-methionine (SAM) as a cofactor, which is the most widely used methyl donor for enzymatic methyl transfer reactions [3]. The SAMT enzyme accumulates at the site of infection [4] since the systematic collection of MeSA in the infected tissue is required for the successful functioning of SAR [5]. In contrast, the salicylic acid-binding protein 2 (SABP2) with strong esterase activity catalyzes the reaction of MeSA to SA synthesis [2, 6]. The overexpression of *SABP2* or silencing of *SAMT* reduces the accumulation of MeSA in the infected plant leaves, resulting in SAR depletion [7], indicating that *SAMT* and *SABP2* synchronize MeSA levels in plants [8]. High concentrations of SA, arising from the shikimate acid pathway, might be toxic to plants [9] and thus could be one reason for MeSA production [10].

Many studies have shown that plants produce MeSA following herbivore attacks in order to attract the natural enemies of these herbivores [11, 12]. The term for these compounds is herbivore-induced plant volatile (HIPV) [13]. Arthropods, the natural enemies of herbivores, are attracted by HIPVs and in this way increase biological control activity [14]. According to Coppola et al., [15], aphids behave differently on plants treated with MeSA compared to untreated plants. Aphids rapidly abandon MeSA-treated plants, confirming its direct impact on their dispersal. Considerable amounts of MeSA are emitted from plant vegetation into the environment [16]. Before emission, the volatile compound is stored in specialized glandular cells or organelles and can be released constitutively under stress, but also under optimal conditions [17]. Rigorous studies on SA-dependent MeSA production have been conducted on a variety of annual plant species, including rice [18], *Arabidopsis* [19, 20] and tobacco [21]. However, to the best of our knowledge, studies of constitutive MeSA production in plants are mostly absent in the literature.

The importance of MeSA as a natural phytochemical compound in therapeutics was acknowledged very early on in human history [22]. For the American and Canadian indigenous communities, plants were the main source of MeSA as a substance to reduce pain [22, 23]. The leaves and bark of birches were used as a basis for herbal infusions for the treatment of fever and gastrointestinal ailments [24, 25]. Commonly, MeSA is used as an essential oil and fragrance [26] and possesses an anti-inflammatory effect that has been used for pain relief and in many medicinal products for muscle and joint pain and rheumatic conditions [27]. From its original extraction through traditional herbal procedures, MeSA is now a mass-produced, synthetic and pharmaceutically significant substance [23]. The use of natural MeSA is limited due to an insufficient supply of the natural raw product, and the constructive utilization of natural plant resources is often critical. The limited distribution of the natural product in certain species, tissues and organs requires an improvement in the molecular and phytochemical knowledge relating to the product that is vital to the development of herbal medicine [28].

Several species of the genus *Betula* (specifically the subgenus *Aspera*, section *Lentae*) are among the examples that constitutively produce high levels of MeSA [22, 29, 30]. A chemical analysis of birch extract (bark and leaf) revealed a concentration of between 49–99.8% of MeSA in the extracted essential oil constituents [31–33]. Birches belong to one of the vital angiosperm genera that support and benefit thousands of living organisms, as well as maintain the ecosystem in boreal forests [30, 34, 35]. Apart from their medicinal and traditional uses, many species of this genus have a long history of difficulties in their classification [36–38]. Considering taxonomical issues, the classification of birches using molecular markers for high and low MeSA production is mostly absent in the literature.

New pharmaceutically significant and naturally existing substances, together with their therapeutic and regenerative features, are being constantly investigated. Therefore, bioinformatics and the expression analysis of genes involved in MeSA biosynthesis will be crucial for

cloning and functional analysis studies. It is possible to regulate MeSA biosynthesis through genetic engineering or genome editing technologies.

In this study, we attempted to answer the question of whether, if any, there are common variations in the genes that contribute to an increased MeSA content in some *Betula* species. Thus, we targeted (1) the intra and interspecific comparative bioinformatics analysis of candidate genes in different low and high MeSA-producing *Betula* species, (2) the sequence variation analysis and marker development in candidate genes associated with MeSA production and (3) the relative expression analysis of two candidate genes that mediate MeSA biosynthesis in high and low MeSA-producing *Betula* species from two different subgenera.

To achieve this, we analyzed *SAMT* and *SABP2* candidate genes that mediate MeSA biosynthesis in eight *Betula* species. The sequencing analysis of *SAMT* and *SABP2* genes revealed putative nucleotide variation associated with high and low MeSA production in birches. The tissue-specific expression analyses of the candidate genes showed differential expression in the *Betula* species.

Materials and methods

Plant material

The seeds of eight *Betula* species (*B. alleghaniensis*, *B. alnoides*, *B. lenta*, *B. grossa*, *B. medwediewii*, *B. pendula* and *B. utilis*) were collected from different botanical gardens in Germany. Specific permissions for the sample collection were obtained from the authority responsible for respective botanical gardens. Seed germination was carried out in standard soil with 10–30% humidity and pH 6.5 in a natural environment without any fertilizer in a polyhouse at the Institute of Agricultural Process Engineering, Kiel University, Germany. Plantlet cultivation was implemented with the required dose of fertilizers in a glasshouse under natural daylight conditions at the Thünen-Institute of Forest Genetics, Grosshansdorf, Germany. The birch species previously confirmed through barcoding and phylogenetic analysis [29] were selected for this study (Table 1).

MeSA-producing ability was classified according to monographic descriptions [22, 30–32] and own analytical evidence, i.e. olfactory analysis following the scratching of the bark [29]: *B. alleghaniensis*, *B. medwediewii*, *B. grossa*, and *B. lenta* (subgenus: *Aspera*) were classified as high MeSA producers, while *B. pendula*, *B. alnoides*, *B. uitilis*, and *B. nana* (subgenus: *Betula*) were classified as low MeSA producers.

Table 1. Details of the species used for *SAMT* and *SABP2* candidate gene analyses: Names of the species, place of sample origin, geographical distribution, ploidy levels and taxonomic positions were allocated according to Wang et al. (2016) and Ashburner and McAllister (2013).

Species Name	Individuals*	Physical origin	Distribution	2n	Subgenus	Section
B. alleghaniensis	6	BG Tharandt, Germany	North America	6n	*Aspera*	*Lentae*
B. lenta	6	BG Giessen, Germany	North America	2n	*Aspera*	*Lentae*
B. medwediewii	4	BG Tharandt, Germany	Japan	2n	*Aspera*	*Lentae*
B. grossa	4	BG Tharandt, Germany	Japan	12n	*Aspera*	*Lentae*
B. alnoides	4	BG Tharandt, Germany	India, Bhutan, Nepal	2n	*Acuminata*	*Acuminatae*
B. pendula	6	BG Grosshansdorf, Germany	Europe and East Asia	2n	*Betula*	*Betula*
B. utilis	4	Kiel (private), Germany	Himalayas	4n	*Betula*	*Costatae*
B. nana	4	BG Cambridge, England	Arctic region	2n	*Betula*	*Apterocaryon*

*Selected number of individuals per species.

https://doi.org/10.1371/journal.pone.0240246.t001

Identification of candidate genes and retrieval of sequences

The previously functionally characterized protein sequence of *Clarkia breweri* SAMT (accession number: AF133053) [4] and *Nicotiana tabacum* SABP2 (accession number: AY485932) [7] were used as queries in the silver birch (*B*. *pendula*) genome database (https://genomevolution.org/coge/CoGeBlast.pl) to search for homologous sequences using the BlastP algorithm [39]. The silver birch candidate genes encoding SAMT and SABP2 proteins were selected according to the highest-scoring pair and E-value after the Blast search [39] (S1 Table in S1 File). Additionally, the phylogenetic closeness of the *B*. *pendula* candidate genes with the references *CbSAMT* (S1 Fig in S1 File) and *NtSABP2* (S2 Fig in S1 File), were also considered as selection criteria. Thirty-two SAMT and thirteen SABP2 previously identified protein sequences from different plant species were retrieved from the NCBI (https://www.ncbi.nlm.nih.gov/) and PopGenIE (http://popgenie.org/) gene databases for phylogenetic analysis (S2 Table in S1 File).

DNA extraction, amplification and sequence analysis

Total DNA was extracted from the leaves of the plants according to the CTAB protocol [40]. DNA extraction of *B*. *medwediewii*, *B*. *alleghaniensis*, and *B*. *lenta* was difficult due to the presence of a high level of polysaccharides, therefore a pre-washing buffer [41] (1.6 ml ice-cold TNE buffer: 200 mM Tris-HCl, 250 mM NaCl, 50 mM EDTA) was used to extract a sufficient quality of DNA.

Specific primers were designed (S3 Table in S1 File) based on the respective *B*. *pendula* gene sequences [42] as references for amplifying the exon and promoter regions of the *SABP2* and *SAMT* genes of the different birch species (Table 1) using polymerase chain reaction (PCR). The following cycling conditions were used for the PCR amplification: 95˚C for 3 min, 40 cycles at 95˚C for 10 s, 60˚C for 30 s and 72˚C for 30 s. All PCR reactions were performed in a SensoQuest thermocycler (Göttingen, Germany). The PCR products were confirmed on 1% agarose gel stained with Roti®GelStain (Carl Roth, Germany). The StarSEQ (Mainz, Germany) service was used for sequencing.

Electropherograms of each sequence were visually inspected. All sequences were aligned and screened for the presence of polymorphism using the SeqManPro 15 program from the DNASTAR Lasergene bioinformatics software suite (Madison, Wisconsin USA).

Multiple sequence alignment and phylogenetic analysis

The retrieved protein sequences from the NCBI (https://www.ncbi.nlm.nih.gov/) and PopGenIE (http://popgenie.org/) gene databases were aligned using ClustalW with default parameters and maximum likelihood (ML) trees were constructed using MEGA X [43] with a bootstrap value of 1,000 replicates. The MES *Beauveria bassiana* (PMB68924.1) and SAMT *Aspergillus niger* (NT166520) protein sequences were used as the outgroup species in the phylogenetic analysis. Additionally, two more (S4A and S4B Fig in S1 File) phylogenies were constructed using eight birch species to analyze intraspecific relationships. The DNA sequences from the birch *SAMT* and *SABP2* were translated into amino acid sequences using the ExPASy translation tool (https://web.expasy.org/translate/).

Gene structure, conserved domain, motif and promoter analysis

The intron/exon organization of the *SAMT* and *SABP2* genes of *B*. *pendula* were predicted based on the respective genomic and coding DNA sequences retrieved from the available *B*. *pendula* genome (https://genomevolution.org/coge/CoGeBlast.pl). The conserved domains were analyzed using the motif online search tool (https://www.genome.jp/tools/motif/). The

conserved motifs in the *Betula* proteins were identified using the MEME online tool (http://meme-suite.org/tools/meme) with the following parameters: maximum number of motifs, 11; minimum motif width, 6 and maximum motif width, 60. The promoter region of the *SAMT* and*SABP2* genes was examined in *B. pendula* and other *Betula* species under investigation using the option "search for care" in the PlantCARE database [44].

RNA extraction and RT-qPCR

The leaf and bark tissues of three-year-old plants were harvested on the morning of 8 July 2019 from the four different *Betula* species, including *B. alleghaniensis*, *B. lenta*, *B. pendula* and *B. utilis* (three biological replicates per species). Total RNA was extracted from the leaf and bark of the plants using the Spectrum Plant Total RNA Kit manufactured by Sigma-Aldrich. Extracted RNA was treated with Invitroge TURBO DNA- free Kit (Thermofisher Scientific, Dreieich, Germany) according to the manufacturer's instructions to remove residual DNA before the next steps. RNA samples were selected based on the rRNA band intensities (28S/18S = 2:1) with a Nanodrop spectrometer (Thermo Scientific, Germany), with optical density values A260 nm/A280 nm between 1.8–2.0 absorption ratio, and A260 nm/A230 nm higher than 1.8 absorption ratio. The first-strand of cDNA was synthesized using GoScrip Reverse Transcription Mix, Oligo(dT) (Promega, Germany). The reverse transcription reaction included 10 μl RNA, 4 μl reaction buffer, 2 μl GoScript Enzyme, and nuclease-free water to a final volume of 20 μl. The reaction conditions were as follows: 25˚C for 5 min followed by 43˚C for 60 min and 70˚C for 15 min. The reverse transcription product was diluted 10-fold and used as the template for quantitative real-time PCR (Bio-Rad, Munich, Germany) and at least three replicates were performed for each gene. Primers for qPCR were designed for the first and fifth exon regions of *SABP2* and *SAMT* genes, respectively (S3 Table in S1 File). The relative expression levels of the genes were calculated using 'delta Ct' (ΔCt) methods and evaluated in the Bio-Rad CFX Manager software. The *ubiquitin* (S4 Table in S1 File) and *actin* (accession number: FJ410442) as the housekeeping genes.

Results

The *Betula* species in Table 1, previously confirmed through barcoding and phylogenetic analyses [29], were used for *SAMT* and *SABP2* candidate gene identification and a comparative analysis.

Identification of *SAMT* and *SABP2* candidate genes

The birch *SAMT* and *SABP2* candidate genes revealing the highest level of sequence similarities in *C. breweri (SAMT)* and *N. tabacum (SABP2)* were chosen for the phylogenetic analyses. For *SAMT*, three hits with similar E-values appeared after a tBlastn search [Bpev01.c0161.g0056.m0001 (*BpSAMT2*; E-value: 4E-47), Bpev01.c0161.g0057.m0001 (*BpSAMT3;* E-value: 9E-46) and Bpev01.c0425.g0055 (*BpSAMT;* E-value: 1E-45)] (S1 Table in S1 File). The phylogenetic tree (S1 Fig in S1 File) clearly shows that Bpev01.c0425.g0055 (*BpSAMT*) clustered closest to the reference *SAMT* gene from *C. breweri*, and was thus selected as the candidate gene for the analyses in this study. The coverage percentage between *CbSAMT* and *BpSAMT* was 61.8%, with 40.1% identity in the tBlastn search. A similar strategy was followed to identify putative *B. pendula SABP2* candidate genes. The three hits [Bpev01.c0161.g0056.m0001 (*BpSABP2*; E-value: 7E-46) Bpev01.c0161.g0057.m0001 (*BpSABP2-2*; E-value: 6E-36) and Bpev01.c0425.g0055.m0001 (*BpSABP2-3*; E-value: 6E-35)] with the lowest E-values were selected to construct a phylogenetic tree (S1 Fig in S1 File). The BpSABP2 (Bpev01.c0015.

g0219) protein clustered closest to the reference NtSABP2, showed a 98% coverage and a 64.7% identity in the tBlastn search (S2 Fig in S1 File).

The *BpSAMT* gene is localized on chromosome IX and contains five exons and four introns, while *BpSABP2* is located on chromosome V and carries three exons and two introns. Using the *B. pendula* reference genes, specific primers (S3 Table in S1 File) were designed for *SAMT* and *SABP2* to amplify homologous regions in other *Betula* species, *B. alleghaniensis* (Ba), *B. lenta* (Bl), *B. medwediewii* (Bm), *B. grossa* (Bg), *B. utilis* (Bu), *B. alnoides* (Bal) and *B. nana* (Bn).

Comparative analysis of SAMT and SABP2 protein sequences

Functionally characterized SAMT and SABP2 reference protein sequences from different species that showed enzymatic activity toward SA and MeSA, respectively, were compared with *B. pendula* SAMT and SABP2 proteins (S5 Table in S1 File) using the BLASTp option in the NCBI gene database.

BpSAMT displayed 55% coverage and 63.7% identity to the first functionally characterized CbSAMT (AF133053), and 50% coverage and 39.9% identity to AtBSMT1 (AT3G11480). A SABATH gene from *P. trichocarpa*, *PtSABATH4*, showed the highest activity towards SA, displaying 51% coverage and 54.37% identity to BpSAMT. BpSAMT is 95.5% (100%), 99.6% (100%), 96.9% (100%), 97.5% (100%), 93.5% (100%), 95.9% (100%) and 97.53% (99%) identical (coverage) to BaSAMT, BalSAMT, BlSAMT, BmSAMT, BuSAMT, BnSAMT and BgSAMT protein sequences, respectively.

Likewise, a comparative analysis of the BpSABP2 protein sequence with previously characterized genes from NtSABP2 showed 98% coverage (64.73% identity), while the *P. trichocarpa* PtSABP2-1 (Potri.007G037700) and PtSABP2-1 (Eugene3.00070971) showed 9% and 4% coverage and 28.6% and 40.9% identity, respectively, to BpSABP2. The BpSABP2 showed 53.5% identity and 96% coverage to the AtMES9 protein sequence that showed the highest enzymatic activity to MeSA in *A. thaliana*. BpSABP2 is 93.2% (100%), 96.2% (100%), 95.1% (100%), 98.1% (100%), 94.7% (100%), 97.2% (100%) and 93.5% (100%) identical (coverage) to BaSABP2, BalSABP2, BlSABP2, BmSABP2, BuSABP2, BnSABP2 and BgSABP2 protein sequences, respectively.

Gene structure, sequence variation analysis and marker development

The structure of *SAMT* and *SABP2* genes in the different birch species was predicted on the basis of the exon/intron organization of the homologous *B. pendula* genes (Fig 1). Based on these predictions, exon regions were amplified from four high (*B. alleganienisis* (*al*), *B. lenta* (*len*), *B. grossa* (*bg*) *B. medwediewii* (*med*)) and four low (*B. pendula* (*pen*), *B. utilis* (*util*), *B. nana* (nan) *B. alnoides* (*aln*)) MeSA-producing *Betula* species using various primer

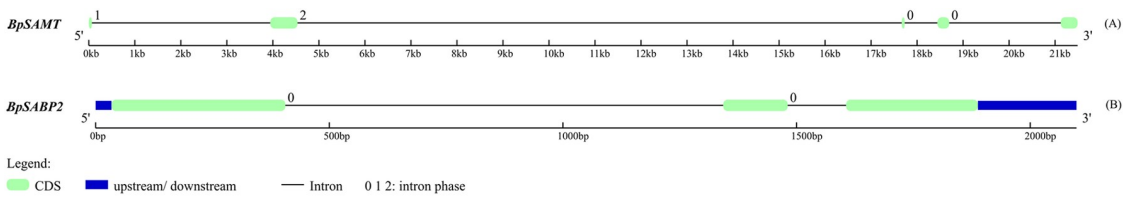

Fig 1. Structural features of *B. pendula SAMT* and *SABP2* genes. Exons are represented by round-corner rectangles, while the line between two exons represents an intron. Intron phases are represented by the numbers above the line. The intron phases are likely to assist in exon shuffling, recombination fusion, and protein domain exchange [45, 46].

https://doi.org/10.1371/journal.pone.0240246.g001

combinations to determine the nucleotide architecture of the *SABP2* and *SAMT* coding sequences (S3 Table in S1 File). Low sequence length variation in coding regions was observed within the eight *Betula* species (S4 Table in S1 File), ranging from 1,344 bp to 1,348 bp for *SAMT* and as 792 bp for *SABP2*.

A comparative analysis among the eight birch species displayed a considerable amount of nucleotide polymorphism within the genes. Here, putative group-specific single nucleotide polymorphisms (SNPs) were considered for marker analysis (Table 2). Groups were allocated on the basis of their ability to produce high and low levels of MeSA. In total, 38 individuals from different *Betula* species originating from different botanical gardens (S6 Table in S1 File) were analyzed for *SAMT* and *SABP2* gene sequence variation by designing specific primers (S3 Table in S1 File).

In total, six prominent group-specific SNPs were discovered to differentiate between high (highlighted in yellow) and low MeSA-producing birch species (Table 2). The occurrence of heterozygous nucleotide sites was frequently observed in all high MeSA-producing species, while no heterozygous positions were observed in low MeSA-producing birches. Four of the six SNP positions revealed heterozygous sites within the group of high MeSA producers, with exception of *B. medwediewii* (med). In the latter species, all six SNP positions revealed heterozygous sites.

High nucleotide variations were also detected for the *SAMT* gene, however, no group-specific nucleotide substitution that is putatively associated with high or low MeSA-producing ability was observed.

Functional domain and conserved motif analysis

The functional domain of the *Betula* SAMT amino acid sequences were analyzed and compared with the respective reference protein sequences. The Pfam (https://www.genome.jp/) domain search revealed that the methyltransferase 7 domain was conserved in all the SAMT protein sequences included in the study. All *Betula* SAMT only displayed domains described as SAM-dependent carboxyl methyltransferase (S7 Table in S1 File). Multiple sequence alignment of SAMT amino acid sequences was constructed using *Betula* and the respective reference amino acid sequences for structural analysis. The alignment of *B. lenta* (BlSAMT), *B. alleghaniensis* (BaSAMT), *B. grossa* (BgSAMT), *B. medwediewii* (BmSAMT), *B. pendula* (BpSAMT), *B. utilis* (BuSAMT), *B. nana* (BnSAMT), *B. alnoides* (BalSAMT) together with CbSAMT shown to contain the conserved domain of methyltransferase (Fig 2). In addition, we detected the occurrence of a previously defined [47] SAM-binding motif within the aligned sequences. Further, we observed the positions of residues involved in the substrate-binding

Table 2. Nucleotide characteristics and validation of SNPs discovered in *SABP2* gene regions to determine the eight high and low MeSA-producing *Betula* species. High MeSA (yellow): *B. alleganienisis* (ale), *B. lenta* (len), *B. grossa* (bg), *B. medwediewii* (med); low MeSA: *B. pendula* (pen), *B. utilis* (uti), *B. nana* (nan), *B. alnoides* (aln). Group-specific SNPs are shown in blue.

Nuclear region	Total number of SNP	Position	ale (6*)	len (6*)	bg (4*)	med (4*)	pen (6*)	uti (4*)	nan (4*)	aln (4*)
SABP2	6	160bp	S (6)	C (4)/S (2)	S (4)	S (4)	G (6)	G (4)	G (4)	G (4)
		189bp	R (6)	A (5) / R (1)	R (4)	R (3) / G (1)	G (6)	G (4)	G (4)	G (4)
		262bp	T (6)	T (6)	T (4)	T (2), W (1), A (1)	A (6)	A (4)	A (4)	A (4)
		298bp	G (6)	G (6)	G (3) / A (1)	G (2) / R (1) / A (1)	A (6)	A (4)	A (4)	A (4)
		304bp	G (5) / K (1)	K (6)	G (4)	K (3) / T (1)	T (6)	T (4)	T (4)	T (4)
		336bp	A (6)	A (6)	A (4)	R (4)	G (6)	G (4)	G (4)	G (4)

*Total number of species used to validate the discovered polymorphic sites.

https://doi.org/10.1371/journal.pone.0240246.t002

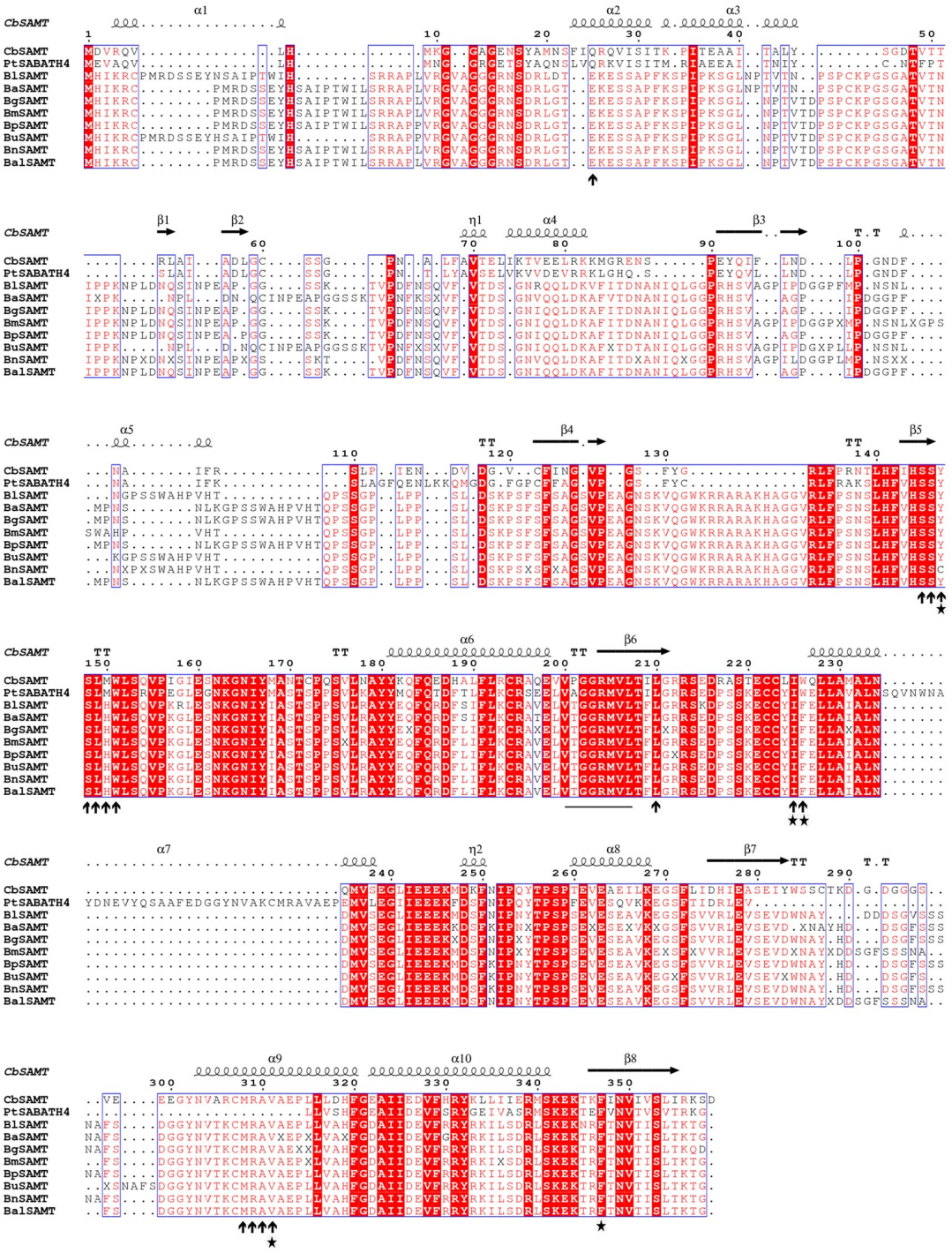

Fig 2. Multiple sequence alignment of SAMT amino acid sequences constructed from CbSAMT, *C. breweri* SAMT [4], PtSABATH4, *P. trichocarpa* SABATH [50] and eight *Betula* species, including *B. lenta* (BlSAMT), *B. alleghaniensis* (BaSAMT), *B. grossa* (BgSAMT), *B. medwediewii* (BmSAMT), *B. pendula* (BpSAMT), *B. utilis* (BuSAMT), *B. nana* (BnSAMT) and *B. alnoides* (BalSAMT). The blue frames represent the conserved residues, the white characters in red boxes suggest strict identity and the red characters in white boxes specify similarity. All the amino acid sequences carry the conserved domain of methyltransfer including the SAM-binding motif that had been previously defined, highlighted with green line [47]. The positions of residues involved in the SA substrate binding, identified from the three-dimensional structure [48], are indicated by arrows, while the residues indicated with a star are the aromatic moiety of the substrate and important for substrate selectivity [49]. The figure was prepared with ESPript [51].

https://doi.org/10.1371/journal.pone.0240246.g002

site that was identified from the three dimensional structure [48] and the aromatic moiety of the substrate that is important for substrate selectivity of SAMT proteins [49] (Fig 2).

Similarly, another multiple amino acid sequence alignment (Fig 3), including *B. lenta* (BlSABP2), *B. alleghaniensis* (BaSABP2), *B. grossa* (BgSABP2), *B. medwediewii* (BmSABP2), *B. pendula* (BpSABP2), *B. utilis* (BuSABP2), *B. nana* (BnSABP2), *B. alnoides* (BalSABP2) and the reference NtSABP2, was constructed. The α/β hydrolase-6 domain was conserved in all tested SABP2 protein sequences. The *Betula* SABP2 protein sequences displayed multiple domains, including α/β hydrolase-1, α/β hydrolase-4 and Lipase-3. We observed the three conserved amino acids form the catalytic triad that is commonly found in the hydrolase domain and the residues that contact SA [52] (Fig 3).

The MEME online tool was used to identify the conserved motifs and/or differences in protein structure among the *Betula* SAMT and SABP2 amino acid sequences. In total, 11 and five equally shared conserved motifs were observed in all the *Betula* SAMT and SABP2 amino acid sequences, respectively (S3A and S3B Fig in S1 File).

Phylogenetic analysis and functional prediction

To ascertain the evolutionary relationship of the *Betula* SAMT and SABP2 with the SAMT and SABP2 members of other plant species which have been functionally characterized (S2

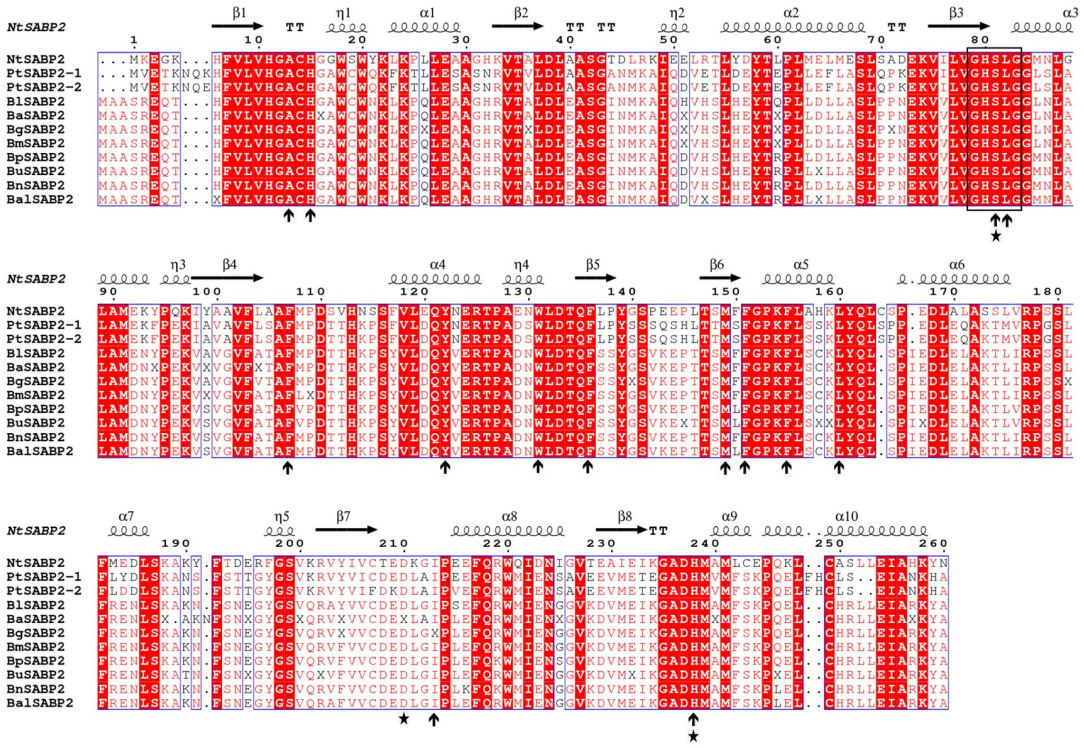

Fig 3. Multiple sequence alignment of SABP2 amino acid sequences constructed from *N. tobacco* SABP2 (*NtSABP2*) [7], *P. trichocarpa* SABP2 (*PtSABP2-1*, *PtSABP2-2*) [53] and eight *Betula* species, including *B. lenta* (BlSAMT), *B. alleghaniensis* (BaSAMT), *B. grossa* (BgSAMT), *B. medwediewii* (BmSAMT), *B. pendula* (BpSAMT), *B. utilis* (BuSAMT), *B. nana* (BnSAMT) and *B. alnoides* (BalSAMT). The blue frames represent the conserved residues, the white characters in red boxes suggest strict identity and the red characters in white boxes specify similarity. The lipase signature sequence of SABP2 is displayed with black frame. The three conserved amino acids forming a catalytic triad, S81, D210 and H238, commonly found in the hydrolase domain, are indicated with a star and are conserved in *Betula* SABP2 [7], while residues that contact to SA are indicated by arrows [5]. The figure was prepared with ESPript [51].

https://doi.org/10.1371/journal.pone.0240246.g003

Table in S1 File), a maximum likelihood phylogenetic tree with 1,000 bootstrap values was constructed using the amino acid sequences of *B. lenta* (BlSAMT and BlSABP2), *B. alleghaniensis* (BaSAMT and BaSABP2), *B. grossa* (BgSAMT and BgSABP2), *B. medwediewii* (BmSAMT and BmSABP2), *B. pendula* (BpSAMT and BpSABP2), *B. utilis* (BuSAMT and BuSABP2), *B. nana* (BnSAMT and BnSABP2) and *B. alnoides* (BalSAMT and BalSABP2). The possible substrate specificity of *Betula* SAMT and SABP2 proteins was determined on the basis of the phylogenetic clustering genes in the same subgroup, and might share a similar function.

According to the phylogenetic tree (Fig 4), the SAMT proteins were divided into two groups (Group A and B). All the *Betula* SAMT was clustered together in Group A with the SAMT from *P. trichocarpa* (PtSABATH4) and *C. breweri* (CbSAMT). In addition, the *Betula* SAMT was clustered in a subgroup of Group A having bootstrap values of 98 and accompanying the SAMT from other species, suggesting that *Betula* SAMT most probably shares a similar function. The SAMT with a higher homology infers the function of the unknown *Betula* SAMT according to the clustering of the phylogenetic tree. These SAMT proteins all use SA as a substrate that synthesizes the volatile ester MeSA. It should be noted that *Arabidopsis* BSMT (AT3G11480), which uses both SA and benzoic acid (BA) as a substrate, is not clustered together with *Betula* SAMT.

Similarly, the SABP2 phylogenetic tree (Fig 5) was constructed using the *Betula* SABP2 protein with other known SABP2 proteins from different plant species (S2 Table in S1 File). *Betula* SABP2 clustered in Group A together with the functionally characterized SABP2 from *P. trichocarpa* (PtSABP2-1 and PtSABP2-1) with a bootstrap value 88 for the clade, suggesting a possible functional similarity. The *Arabidopsis* MESs (AtMES1, 2, 4, 7 and 9) and NtSABP2 also clustered in Group A.

The candidate and reference proteins formed a clade in the phylogenetic tree (Figs 4 and 5) and were also included in the sequence alignment analysis (Figs 2 and 3).

The intraspecific evolutionary relationship of *SAMT* and *SABP2* in eight high and low MeSA-producing *Betula* species were also analyzed by constructing two phylogenetic trees using the maximum likelihood method in MEGA X software [43]. The exon regions of the *SAMT* and *SABP2* genes were sequenced for all the *Betula* species and converted into the amino acid sequences. Both the phylogenetic trees revealed two clades differentiating the high and low MeSA-producing birch species (S4A and S4B Fig in S1 File).

Expression analysis of *SAMT* and *SABP2* in different birch species and tissues

To detect the prior tissue-specific expression of *SAMT* and *SABP2* genes in *Betula*, we analyzed the expression of *SAMT* and *SABP2* in the leaf and bark of two high MeSA (*B. lenta* and *B. alleghaniensis*) and two low MeSA (*B. utilis* and *B. pendula*) producers using quantitative real-time RT-PCR (Fig 6). Altogether, *SAMT* and *SABP2* genes revealed differential expression patterns in the two tissues analyzed from high and low MeSA producers. The *B. alleghaniensis SAMT* (*BaSAMT*) displayed high expression in the bark as well as in the leaf (Fig 6A and 6B) tissues, while *B. lenta SAMT* (*BlSAMT*) was highly expressed only in the bark (Fig 6A and 6B). Both *B. utilis SAMT* (*BuSAMT*) and *B. pendula SAMT* (*BpSAMT*) had a low expression in both bark and leaf tissue.

For *SABP2*, *B. alleghaniensis* (*BaSABP2*) and *B. lenta* (*BpSABP2*) showed no expression in the bark and low expression in leaves (Fig 6C and 6D). On the other hand, *B. utilis SABP2* (*BuSABP2*) showed a low expression in the bark and a high expression in the leaf, while *B. pendula SABP2* (*BpSABP2*) was highly expressed only in the bark.

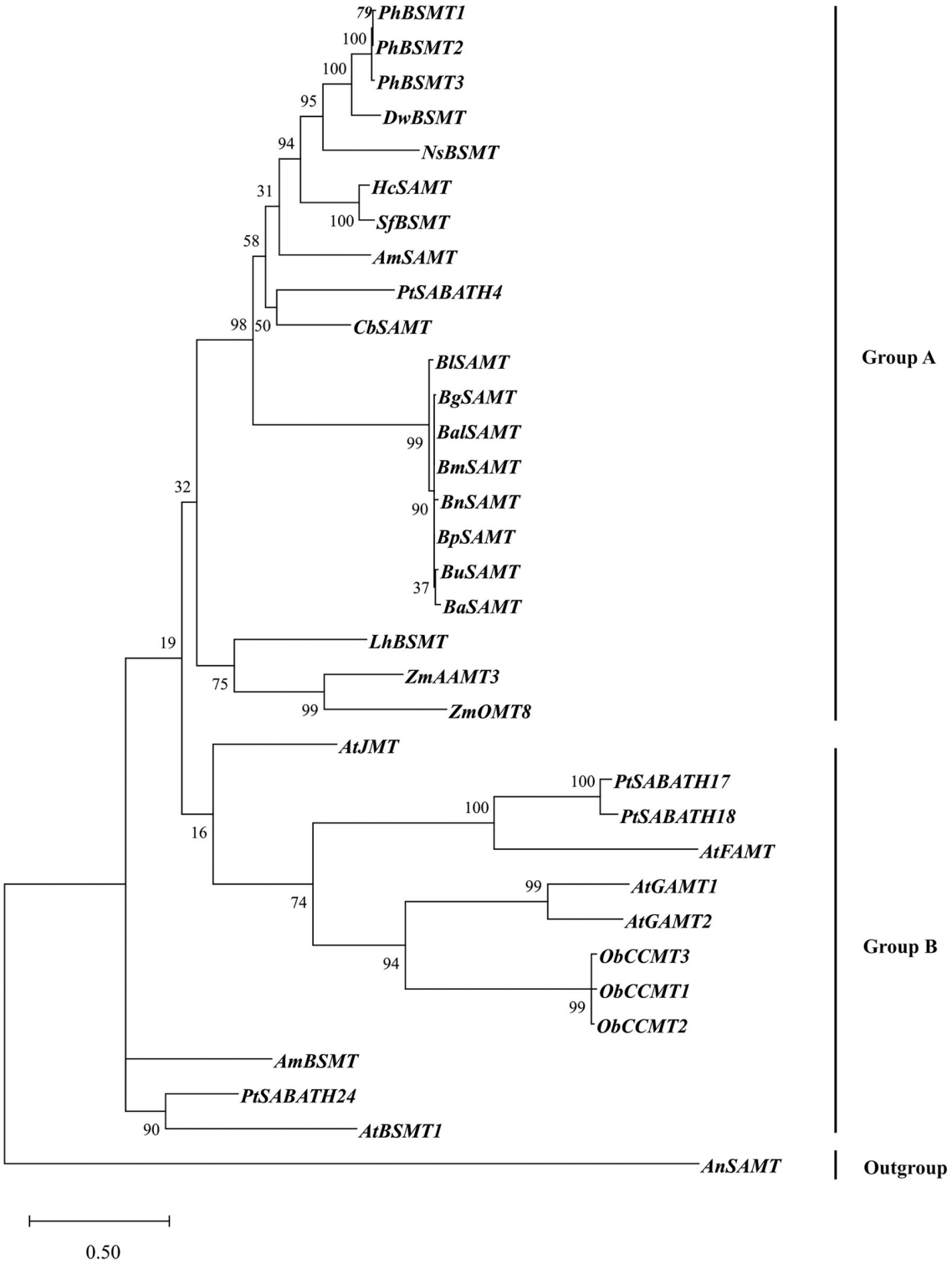

Fig 4. Evolutionary relationship of *Betula* SAMT proteins: The phylogenetic tree was constructed using amino acid sequences of *B. lenta* (Bl SAMT), *B. alleghaniensis* (Ba SAMT), *B. grossa* (Bg SAMT), *B. medwediewii* (Bm SAMT), *B. pendula* (Bp SAMT), *B. utilis* (Bu SAMT), *B. nana* (Bn SAMT) and *B. alnoides* (Bal SAMT) species with 26 functionally characterized SAMT from other species (S2 Table in S1 File). A total number of 34 SAMT amino acid sequences were used in the maximum likelihood method in the MEGA7 software [43]. A SAMT from *Aspergillus niger* (NT166520) was used as an outgroup species. The numbers at the nodes indicate bootstrap values calculated with 1,000 replicates. Branches are drawn to scale with the bar indicating 0.50 substitutions per site.

https://doi.org/10.1371/journal.pone.0240246.g004

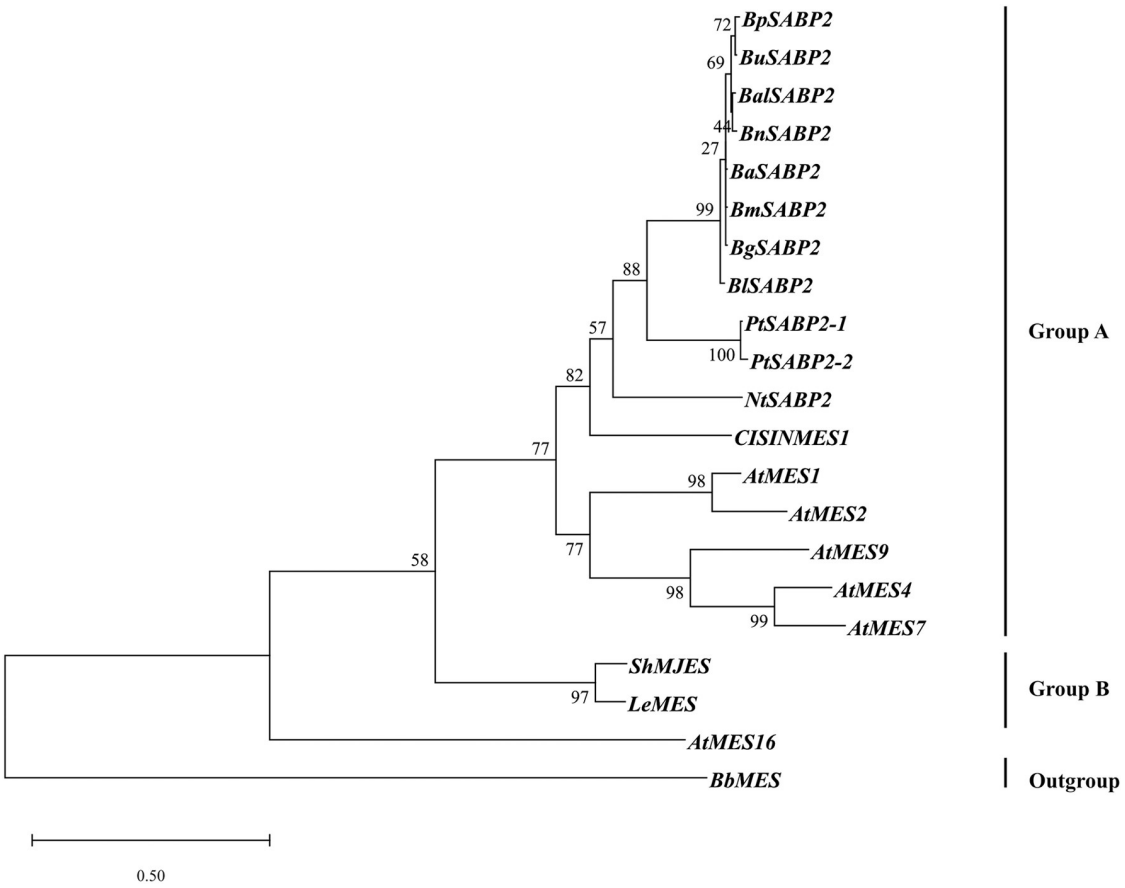

Fig 5. Evolutionary relationship of *Betula* SABP2 proteins: The phylogenetic tree was constructed using amino acid sequences of *B. lenta* (BlSABP2), *B. alleghaniensis* (Ba SABP2), *B. grossa* (Bg SABP2), *B. medwediewii* (Bm SABP2), *B. pendula* (Bp SABP2), *B. utilis* (Bu SABP2), *B. nana* (Bn SABP2) and *B. alnoides* (Bal SABP2) species with 13 functionally characterized SABP2/MES from other species (S2 Table in S1 File). A total number of 21 SABP2/MES amino acid sequences were used in the maximum likelihood method in the MEGA7 software [43]. A MES from *Beauveria bassiana* (PMB68924.1) was used as an outgroup species. The numbers at the nodes indicate bootstrap values calculated with 1,000 replicates. Branches are drawn to scale with the bar indicating 0.50 substitutions per site.

https://doi.org/10.1371/journal.pone.0240246.g005

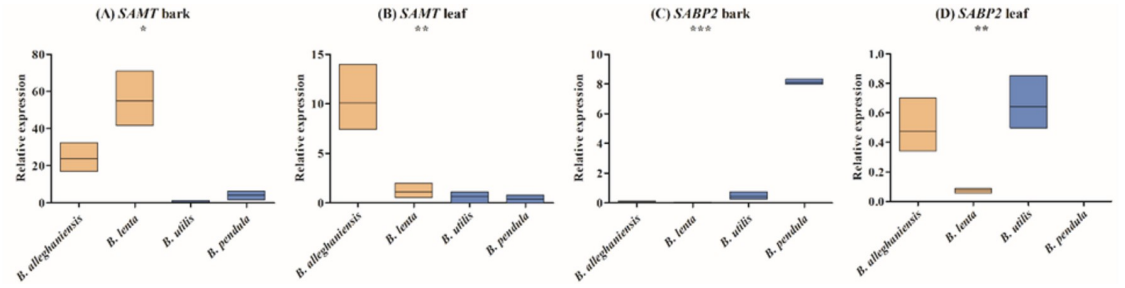

Fig 6. Tissue-specific expression of the *SAMT* and *SABP2* genes in two high (*B. alleghaniensis* and *B. lenta*; gray-tan columns) and two low (*B. utilis* and *B. pendula*; blue columns) MeSA-producing birch species. The expression of candidate genes was assessed by qRT-PCR. The y-axis indicates the relative expression level, while the x-axis indicates the different tissue of the different species. Three biological and three technical replicates were used. *Actin* and *ubiquitin* reference genes were used to normalize the expression.

https://doi.org/10.1371/journal.pone.0240246.g006

Betula SAMT and *SABP2* gene promoter analysis

The sequences obtained from the PCR-amplified promoter regions of the *SABP2* and *SAMT* genes from the low MeSA-producing *B*. *pendula* (pen), *B*. *utilis* (uti), *B*. *nana* (nan), *B*. *alnoides* (aln), and the high MeSA-producing *B*. *alleganienisis* (ale), *B*. *lenta* (len), *B*. *grossa* (bg), *B*. *medwediewii* (med) were submitted to the NCBI database (S8 Table in S1 File). All sequences were analyzed using PlantCARE [44] to identify putative *cis* elements (S9 and S10 Tables in S1 File). We obtained *BpSAMT* and *BpSABP2* gene promoter regions from the *B*. *pendula* genome sequence [42]. The length of the *BpSABP2* and *BpSAMT* promoters were 1,050 bp and 1,003 bp, respectively. The results indicated that the promoter regions contain multiple eukaryotic *cis*-acting elements, including TATA and CAAT boxes. In the *BpSABP2* promoter sequence, four abscisic acid response elements (ABRE) were found at positions bp 74+, 191-, 938- and 939+; three Box4 parts of conserved DNA module elements were located at positions bp 143+, 801- and 581-; one Sp1 at position bp863- and three G-Box light-responsive elements were located at positions bp 73-, 983- and 191+. Two light-responsive GATA-motifs were localized at positions bp 454- and 766+; two elements involved in circadian control were located at positions bp 964+ and 973+ and one auxin-responsive element (TGA-element) was found at position bp 62+. The *BpSAMT* contained the plant light-responsive elements (GTGGC-motifs) at position bp 170; a chs-CMA2a at position bp 75- and 246-, and one auxin responsive AuxRR-core element was located at position bp 714-.

Through the primer walking approach, about 600 bp and 700 bp of the *SAMT* and *SABP2* promoter regions were also successfully obtained for the eight other *Betula* species (except for *B*. *grossa SABP2*) for comparative analysis (S8 Table in S1 File). The presence of different *cis* elements, together with their frequencies in the *SAMT* and *SABP2* gene promoter regions, was evaluated in seven birch species (Figs 7 and 8, respectively). The fragment length of *SAMT* and *SABP2* promoters varied between 603–628 and 636–770 base pairs, respectively (S9 and S10 Tables in S1 File). A comparative analysis of the *cis* regulatory elements revealed considerable differences in the frequencies between the high and low MeSA producers *B*. *lenta* and *B*. *pendula*, respectively. The *B*. *lenta SAMT* promoter region showed two TATA boxes, while all other species contains only one TATA box. Different numbers of TATA boxes were observed in the *SABP2* gene promoter regions of *B*. *pendula* and *B*. *lenta*.

Discussion

Intensive studies have been conducted to reveal the role of MeSA in plant immunity and the signaling cascades of the SAR mechanism in plants [3, 54, 55]. SAMT, which belongs to the SAM-dependent methyltransferases and is commonly found in plants, forms MeSA by the methylation of SA [3]. MeSA reverts to SA by SABP2 under strong esterase activity [2, 6]. Almost all plants produce MeSA as a long-distance mobile signal in stress defense and as an SAR mechanism [1], including the birch species from the subgenus *Betula* (*B*. *utilis*, *B*. *pendula*, *B*. *nana* and *B*. *alnoides*), denoted "low MeSA producers" [29]. However, in addition to its role in immunity and the SAR signaling cascade, MeSA is produced constitutively in some winter green shrubs and birch species, and is therefore believed to be an important plant constituent [29, 30]. Constitutive MeSA production is expressed by a natural sweet and strong scent and has medicinal and pharmaceutical significance [31–33]. *Betula* species that constitutively produce MeSA are called "high MeSA producers." They belong to the subgenus *Aspera* and include *B*. *alleghaniensis*, *B*. *lenta*, *B*. *grossa* and *B*. *medwediewii* [29, 30].

It has been confirmed that SAMT and SABP2 enzymes mediate MeSA biosynthesis in many plant species [19, 52–54, 56, 57]. In particular, both enzymes have been functionally characterized and well-studied in different plant species, including *N*. *tabacum*, *C*. *breweri*, *P*.

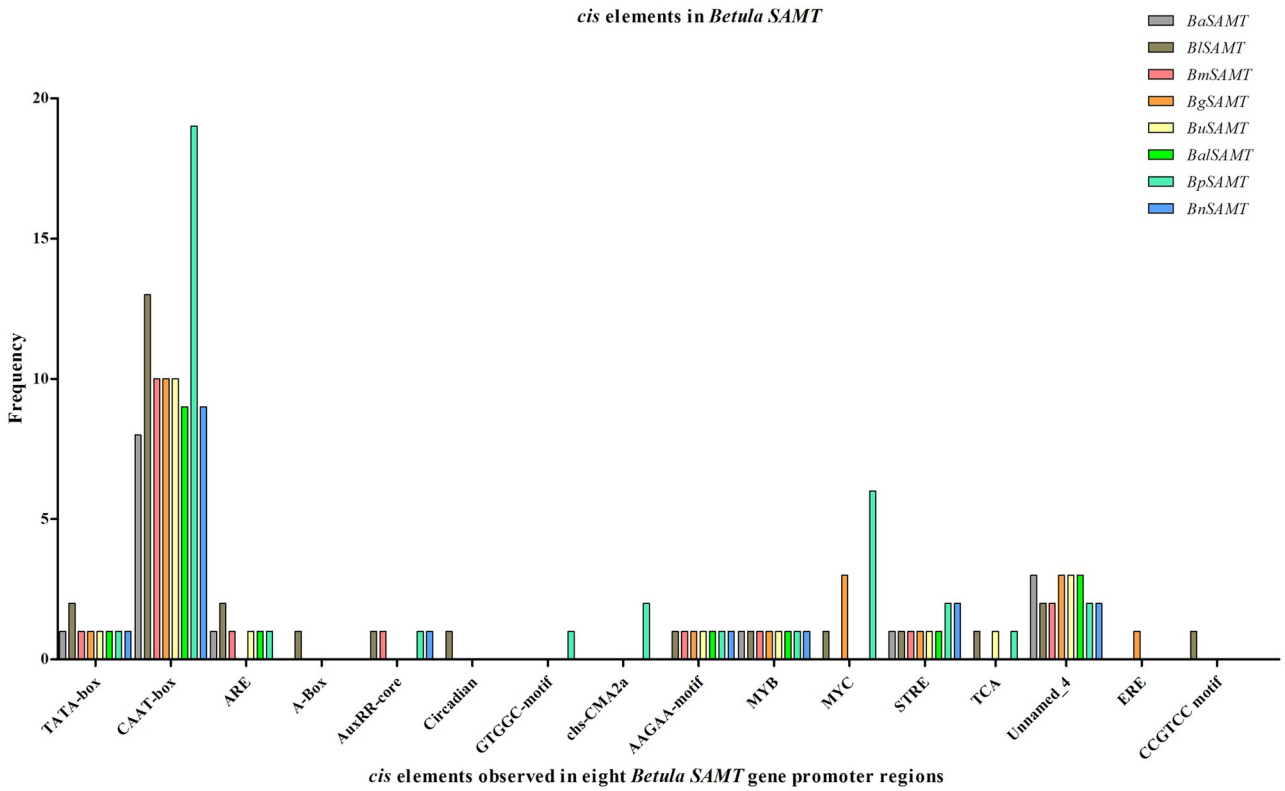

Fig 7. Frequencies of identified *cis* elements using the PlantCARE database [44] in the promoter regions of *SAMT* genes of four high (*B. lenta* (Bl), *B. alleghaniensis* (Ba), *B. grossa* (Bg) and *B. medwediewii* (Bm)) and four low (*B. pendula* (Bp), *B. utilis* (Bu), *B. nana* (Bn) and *B. alnoides* (Bal)) MeSA-producing birch species.

https://doi.org/10.1371/journal.pone.0240246.g007

trichocarpa, *V. vinifera*, and *A. thaliana* [4, 50, 52, 54]. However, very little is known about the genetic architecture of SAMT and SABP2 in the ecologically important tree species of the genus *Betula*.

In this study, we identified two candidate genes, *SAMT* and *SABP2*, in low and high MeSA-producing birch species using protein sequences of previously functionally characterized from *C. breweri* SAMT [4] and *N. tabacum* SABP2 [7] as references. Also, to the best of our knowledge, this is the first study that has identified putative high and low MeSA-specific nucleotides in the *SABP2* gene that could be used to develop molecular markers to differentiate high and low MeSA-producing *Betula* species.

Marker development and validation

Although hybridization, introgression and misidentification have often been obstacles in the systematics of the genus *Betula* [58, 59], the characteristics of leaf shape, bark color, and varying chemical composition of the bark and leaves have successfully been applied to classify the majority of birch species [36–38]. Until now, the classification of some birch species is still under discussion, with several taxonomical classifications still in existence for the genus *Betula* [30, 58, 60, 61]. The use of molecular markers based on, e.g. microsatellites (SSR) and chloroplast regions, could help resolve systematics issues and this technique has already been widely introduced into plant systematics [62–64]. Next-generation sequencing technologies have

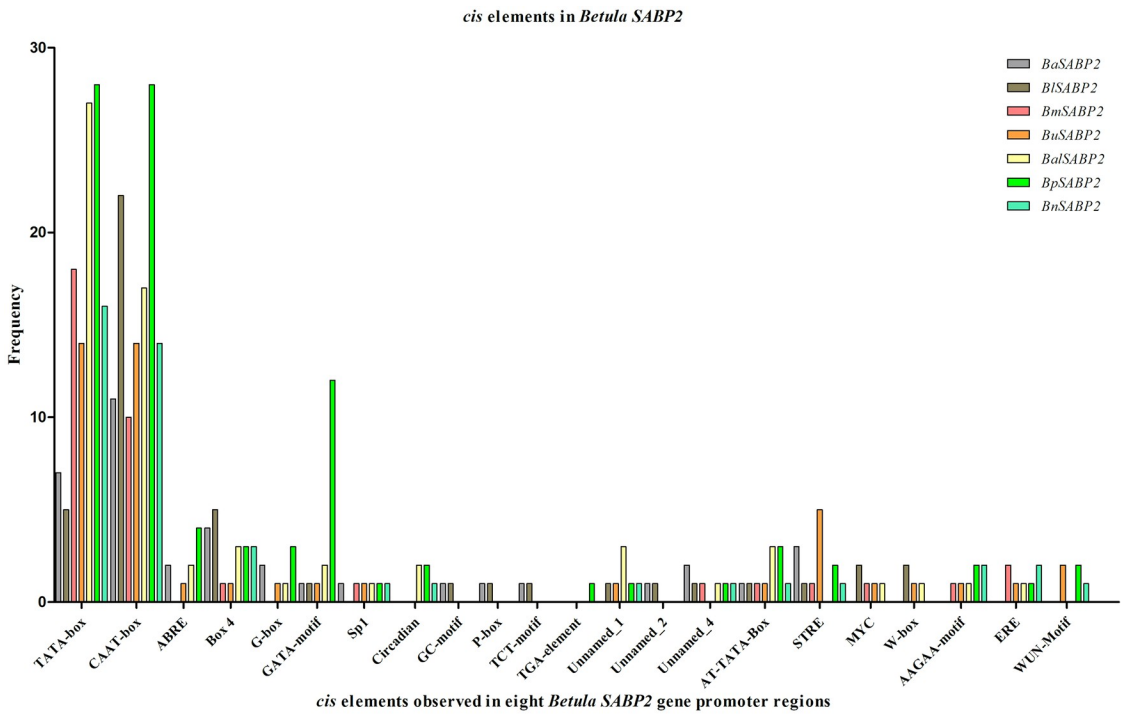

Fig 8. Frequencies of identified *cis* elements using the PlantCARE database [44] in the promoter regions of *SABP2* genes of three high (*B. lenta* (Bl), *B. alleghaniensis* (Ba) and *B. medwediewii* (Bm)) and four low (*B. pendula* (Bp), *B. utilis* (Bu), *B. nana* (Bn) and *B. alnoides* (Bal)) MeSA-producing birch species. Unfortunately, the promoter region of *B. grossa* (Bg) could not be amplified.

https://doi.org/10.1371/journal.pone.0240246.g008

recently made it possible to conduct whole genome sequencing, allowing the generation of a large number of genome-wide markers [65–67]. Population genetics studies of the silver birch have identified genetic variations in genes that are associated with local adaptations to different environmental conditions [42].

For the genus *Betula*, low and high MeSA-synthesis ability could be an additional criterion for species systematics in this genus [29, 30]. Therefore, we defined the *SAMT* and *SABP2* candidate genes involved in the MeSA biosynthesis of eight *Betula* species, including *B. lenta*, *B. alleghaniensis*, *B grossa*, *B. medwediewii* (high MeSA-producing), *B. utilis*, *B. pendula*, *B. alnoides* and *B. nana* (low MeSA-producing).

Various studies have advocated the importance of *SABP2* and *SAMT* genes in plant developmental stages and signaling cascades [3, 7, 50, 52, 68]. Considering their interaction with numerous molecules, *SABP2* and *SAMT* displayed a higher percentage of polymorphism compared to previously analyzed genomic regions in *Betula* [64, 69, 70]. The relative analysis is in agreement with our earlier study where small segments of both genes were investigated [29].

Although significant nucleotide variation was observed between all *Betula* species analyzed, only low and high MeSA-specific nucleotides were considered. We discovered six specific positions on the *SABP2* gene that could be associated with high MeSA production in birches (Table 2). Validation of the nucleotide substitutions (SNPs) was performed with 38 additional birch individuals belonging to eight different birch species from different botanical gardens (S6 Table in S1 File), indicating that species-specific nucleotide substitutions are associated with high MeSA production. To the best of our knowledge, this is the first study that has attempted to identify high and low MeSA-specific nucleotides in the *SABP2* gene in different

Betula species that could be used to develop SNP markers associated with low and high MeSA content. Unfortunately, in the *SAMT* gene, no high or low MeSA-specific nucleotides could be detected. However, the SNPs identified in the *SABP2* gene need to be validated in additional low and high MeSA-producing birch species and by including more individuals.

The decaploid *B*. *medwediewii* (subgenus, *Aspera*; section, *Lentae*) exhibited considerable heterozygous nucleotides at all six SNP positions in the *SABP2* gene (Table 2). Ashburner et. al., (2013) revealed *B*. *medwediewii* as a high MeSA producer, while the olfactory fragrance analysis unanimously categorized *B*. *medwediewii* as an intermediate MeSA producer [29]. In addition to the presence of substantial heterozygous SNPs in the *SABP2* gene and intermediate MeSA production, the clustering of *B*. *medwediewii* with the species of the subgenus *Betula* [29] supports the idea that during the evolution of this species, one of the parents belonged to the subgenus *Betula*. Its partial MeSA-producing ability could be a rational motivation for Ashburner et al., (2013) allocating this species to the subgenus *Aspera*.

Comprehensive bioinformatics analysis

The genetic architecture of the silver birch has been recently enhanced due to the available genome. In our study, we used different bioinformatics tools, including sequencing, gene structure analysis, multiple sequence alignment, domain characterization, conserved motifs, promoter analysis and phylogenetic relationships analysis. The aim was to collect vital information on the different high and low MeSA-producing birch species for biotechnological purposes, including functional analysis, molecular breeding and the commercial use of natural medicinal products.

All the *Betula SAMT* candidate genes from eight different birch species in the study showed the presence of a methyltransferase 7 domain (Methyltransf_7; S7 Table in S1 File) and a conserved motif III that possess SAM-binding sites described previously (Joshi et al., 1998). The occurrence of the motif III in 56 different plant species suggests it plays a major role in the binding of the SAM-dependent *O*-methyltransferases to their specific substrate, which also includes SAMT that catalyzes SA into MeSA [47, 71]. The crystallography analysis of the *C*. *breweri* SAMT protein and the substrate SA complex possesses active sites responsible for the selection of SA that were also characterized in *Betula* SAMT [48], suggesting its role in MeSA biosynthesis (Fig 2).

Amino acid sequences of *Betula* and *C*. *breweri* SAMT proteins revealed only three mismatches: the *Betula* SAMT has histidine, phenylalanine and tyrosine at positions 150, 209 and 224, rather than methionine, isoleucine and leucine, respectively, in the *C*. *breweri* SAMT [48] (Fig 2). In total, 14 SA binding residues were identified in the *Betula* SAMT, compared to 16 in *C*. *breweri* [48].

The alignment of the *Betula* SAMT (Fig 2) and previously functionally characterized members of SABATH family suggests that *Betula* SAMT probably methylates both SA and the structurally similar substrate BA [54]. It has also been experimentally proven that members of the SABATH methyltransferase family catalyze multiple substrates with different K_m values [8, 50, 54, 55]. Additionally, it has been suggested that a single amino acid substitution might play a critical role in the specificity of SAMT/BSMT with SA and BA [72]. Further, the detailed study by Han et al., (2017) on *P*. *trichocarpa* revealed the evolutionary substrate specificity of the members of the methyltransferase family, including SAMT, can be achieved by changes in amino acid sequences and that alterations in a single amino acid might result in a divergence in substrate specificity [50]. Despite the cited study, the actual mechanism behind the substrate specificity of the SAMT enzyme is still unclear. However, through structural analyses, it has been suggested that the size and shape of the active sites may play an important role in the

differentiation of individual substrates [73]. The *Betula* SAMT protein alignment also revealed the presence of hydrophobic and aromatic-rich residues of the carboxyl bearing substrate-binding pockets that were previously observed in the detailed study of *A*. *thaliana* indole-3-acetic acid methyltransferase (AtIAMT) and CbSAMT [4, 49] (Fig 2).

Likewise, the amino acid sequence alignment of SABP2 from eight different birch species revealed the presence of a catalytic α/β hydrolase domain (Abhydrolase_6; S7 Table in S1 File) conserved in the SABP2 family which is in agreement with *A*. *thaliana* SABP2/MES enzymes [74]. The signature amino acid sequence, conserved in the *N*. *tobacco* SABP2 [5], was recognized in all eight *Betula* SABP2 species (Fig 3). All eight *Betula* SABP2 displayed the conserved catalytic triad found in the hydrolase domain that was proved in the protein profiling of *N*. *tobacco* SABP2 [7]. The conserved catalytic triad is in agreement with previous analyses conducted with *Arabidopsis* and the grapevine [52, 74]. Moreover, the 14 residues observed in *Betula* SABP2 that contact to SA were consistent with a previous structural study of tobacco SABP2 [5].

The phylogenetic tree revealed that SAMT from the investigated *Betula* species cluster together with the first functionally characterized *C*. *breweri* SAMT [4] and *P*. *trichocarpa* PtSABATH4 (Fig 4). It is noteworthy that the eight *Betula* and *Populus* SAMT (PtSABATH4) sequences clustered more closely to known SAMTs from *Antirrhinum majus* and *Hoya carnosa* flowers. Additionally, *Stephanotis floribunda*, *Nicotiana suaveolens*, *Datura wrightii* and *Petunia hybrid* BSMT (benzoic acid/salicylic acid methyltransferase) were also clustered in the same clade. The clustering of SAMT and BSMT might have occurred since the purified SAMT enzymes from *C*. *breweri*, and *S*. *floribunda* are able to methylate both SA and BA with higher and lower affinity, respectively [4, 75, 76]. In addition to the SAMT enzymes in Group A, the methyltransferases with different substrate specificity clustered in a paraphyletic Group B containing, for example, *Arabidopsis* jasmonic acid carboxyl methyltransferase (AtJMT) (Fig 4). It was hypothesized that JMT and SAMT/BSMT might have evolved from the indole-3-acetic acid carboxyl methyltransferase (IAMTs) [49].

The *SABP2* phylogeny, with functionally characterized genes from other species, showed *SABP2-1* and *SABP2-2* from *P*. *trichocarpa* clustered together with 94 bootstrap values (Fig 5). The two copies of the *SABP2* gene in *P*. *trichocarpa* were most probably the result of genome duplication events [77], while no signs of duplication events were observed in *B*. *pendula*, resulting in only one copy of *SABP2* in the investigated *Betula* species [42]. Both *PtSABP2-1* and *PtSABP2-2* genes showed explicit esterase activity to MeSA that produced salicylic acid [53]. Since the *Betula* species and *P*. *trichocarpa* are both woody plants and *Betula* SABP2 and PtSABP2 occur in one clade, we can predict that the *Betula* SABP2 functions similarly to PtSABP2. Although the *Betula* candidate genes showed low coverage and identity (S5 Table in S1 File) to the most closely related species, *P*. *trichocarpa*, still they formed a single clade. Therefore, we also recommend using phylogenetic analysis as a candidate gene selection criterion.

Additionally, NtSABP2, PtSABP2 and all eight *Betula* SABP2 displayed the three conserved amino acids forming a catalytic triad. We therefore hypothesize that *Betula* SABP2 catalyzes MeSA with its esterase activity (Fig 3). A comparative analysis of identity (coverage) of *Betula* SABP2 with functionally characterized Arabidopsis AtMES1, AtMES2, AtMES4, AtMES7 and AtMES9 protein sequences resulted in 58% (100%), 53% (100%), 54% (97%), 52% (97%), 54% (96%), respectively. All *Arabidopsis* AtMES1, 2, 4, 7 and 9 showed esterase activity towards SA [9] and the phylogenetic tree also suggests their evolutionary closeness with *Betula* SABP2 (Fig 5). It has been shown that AtMES proteins are responsible for the hydrolysis of other methyl esters, suggesting that almost of all these proteins are able to utilize multiple substrates with different enzymatic activity [74].

Studies of the eukaryotic promoter have shown that gene transcription activity is controlled by multiple cis and trans-acting elements [78]. Detailed studies of these elements were obtained

from diverse experiments, including deletion, element relocation and mutagenesis analysis [79]. Considering the importance of cis elements, we successfully amplified the promoter regions of all the *Betula SAMT* and *SABP2* genes used for analyzing the functions of regulatory elements. We conducted a detailed comparative analysis between high (*B. alleghaniensis*, *B. lenta*, *B. grossa* and *B. medwedweii*) and low (*B. pendula*, *B. utilis*, *B. alnoides* and *B. nana*) MeSA producers. The *Betula SAMT* and *SABP2* gene promoters contain a variety of common elements, including the TATA and CAAT boxes (Figs 7 and 8). The promoter region of *B. lenta SAMT* showed two TATA boxes, while the *SAMT* of all the other *Betula* had only one. In the case of *SABP2* promoters, the low MeSA producers, *B. utilis*, *B. nana*, *B. alnoides* and *B. pendula*, displayed a higher number of TATA boxes compared to the high MeSA producers *B. lenta* and *B. alleghaniensis* (Figs 7 and 8). The occurrence of additional transcription starting sites suggests a higher likelihood of relevant expression, since *B. pendula* is a low MeSA-producing birch. The only functionally known *cis*-acting element involved in the circadian rhythm was observed in the promoter regions of both genes and the collective analysis indicated that both *SAMT* and *SABP2* might be induced by the plant hormones [5, 80, 81].

Expression analysis

In order to detect the possible tissue-specific expression of *SAMT* and *SABP2* genes in high (*B. lenta* and *B. alleghaniensis*) and low (*B. pendula* and *B. utilis*) MeSA-producing *Betula* species, we analyzed the expression of *Betula SAMT* and *SABP2* genes in the bark and leaves of three-year-old plants.

The expression of *SAMT* was higher in the bark of *B. alleghaniensis* and *B. lenta* than in *B. pendula* and *B. utilis* (Fig 6A), suggesting its importance in the bark of high MeSA producers. The high expression of the *SAMT* gene in *B. alleghaniensis* and *B. lenta* species also reveals its significance in high MeSA production, since MeSA could be extracted in abundance from the stems of these plants [29, 31].

The characterized activity of *SAMT* in *C. breweri*, *S. floribunda* and snapdragons showed that the enzyme can methylate both SA and BA at different K_m values [4, 75, 76]. In *P. trichocarpa*, PtSABATH4 showed a higher enzymatic activity towards SA than BA, and a higher expression in all tissues studied when compared to other family members of SABATH [50]. Likewise, under normal growth conditions, the *A. thaliana BSMT1 (AtBSMT1)* and *A. lyrata BSMT1 (AlBSMT1)* genes showed considerable expression in leaves. In contrast, the AtBSMT1 protein showed higher enzymatic activity towards SA than BA, while the AlBSMT1 protein had a lower affinity for SA than BA [54]. The studies showed diversions within the substrate specificity of SA/BAMT proteins and collectively suggest that the *Betula SAMT* candidate gene could putatively also catalyze both SA and BA with divergent K_m values. This hypothesis also supports the deduced amino acid sequence of *Betula* SAMT, aligned with *C. breweri* and *Populus* SAMT (Fig 2), where amino acid shifts were observed. Considering the already published expression analysis of *O-methyltransferase* genes in poplar [82], *Arabidopsis* [83], citrus [84] and the results in this paper, we hypothesize that SAMT, which methylates SA to form MeSA, is highly expressed in the bark of the high MeSA producers *B. lenta* and *B. alleghaniensis*, resulting in the constitutive production of MeSA. Our hypothesis is in agreement with a previous SAMT analysis conducted in *C. breweri* flowers in order to characterize the molecules responsible for scent production [4].

For the first time, the SABP2 enzyme was identified in tobacco [85] and was shown to be a MeSA esterase and an important protein that is required for SAR development [5, 7]. In addition, the members of MES/SABP2 family have been isolated and characterized in many other plant species, including the grapevine [52], the potato [86], citrus [80] and poplars [53].

SABP2 is one of the many crucial elements of the SA signaling cascade that was identified by conducting intensive biochemical and molecular genetics studies in different plant species [53, 80, 87]. The bioinformatics sequence analysis of the *Arabidopsis* genome revealed 20 genes coding for proteins with relatively high sequence similarities to the tobacco *SABP2* [5, 74]. This suggests that methylesterases are involved in the hydrolysis of MeSA [74, 88].

The expression of *SABP2*, which converts MeSA into SA, was higher in the bark of the low MeSA-producing species *B*. *utilis* and *B*. *pendula* than in *B*. *alleghaniensis* and *B*. *lenta* (Fig 6C). The results suggest an abundance of SABP2 in *B*. *utilis* and *B*. *pendula*, and thus a higher affinity for reverting MeSA to SA. Likewise, in poplars, *SABP2-1* and *SABP2-2* showed the highest and a moderate level of expression in leaves and bark, respectively, while the expression of *PtSABP2-2* was found to be low in leaves under "normal" growing conditions [53].

To the best of our knowledge, this is the first study that has attempted to gather information about the genes involved in the biosynthesis of MeSA in birches. Detailed bioinformatics studies and expression analysis have led to the identification of candidate genes in eight species of the genus *Betula* that mediate MeSA biosynthesis. The results obtained in this study will be beneficial for further functional and enzymatic substrate specificity analysis of the *SAMT* and *SABP2* genes. In addition, this is the first attempt to identify high and low MeSA-specific nucleotides which can be used to develop SNP markers associated with low and high MeSA content for molecular breeding purposes.

Supporting information

S1 File.
(DOCX)

Acknowledgments

We are extremely grateful to the botanical gardens and Northwest German Forestry Research Institute (listed in Table 1) for providing the requested *Betula* seed samples for scientific research purposes only without any conflict of interest. Thanks also to PD Dr. Birgit Kersten for assisting with the submission of data to the NCBI GenBank, to Dr. Niels Müller for consultation in sequencing and manuscript review (both from the Thünen Institute of Forest Genetics, Grosshansdorf, Germany), and to all the colleagues at the Thünen Institute of Forest Genetics, Grosshansdorf, Germany and the Institut für Landwirtschaftliche Verfahrenstechnik, Kiel for their support, including Katrin Groppe, Susanne Jelkmann and Vivian Kuhlenkamp for laboratory assistance and Dr. Hilke Schröder, Ana Montalvao, Dr. Hans Hönicka and Khira Lettkemann for helpful discussions on methyl salicylate biosynthesis. Thanks are due to Nicola Wilton (Language Boutique®, Hamburg, Germany) for English language editing.

Author Contributions

Data curation: Kiran Singewar.

Funding acquisition: Christian R. Moschner.

Investigation: Kiran Singewar, Matthias Fladung.

Methodology: Matthias Fladung.

Project administration: Christian R. Moschner.

Resources: Eberhard Hartung.

Software: Kiran Singewar.

Supervision: Matthias Fladung.

Validation: Matthias Fladung.

Visualization: Kiran Singewar.

Writing – original draft: Kiran Singewar.

Writing – review & editing: Eberhard Hartung, Matthias Fladung.

References

1. Shulaev V, Silverman P, Raskin I. Airborne signalling by methyl salicylate in plant pathogen resistance. Nature. 1997; 385, 718–721.
2. Park SW, Kaimoyo E, Kumar D, Mosher S, Klessig DF. Methyl salicylate is a critical mobile signal for plant systemic acquired resistance. Science. 2007; 318(5847):113–6. https://doi.org/10.1126/science.1147113 PMID: 17916738
3. D'Auria JC, Chen F, Pichersky E. Chapter eleven The SABATH family of MTS in Arabidopsis Thaliana and other plant species. In: Romeo JT, editor. Recent Advances in Phytochemistry. 37: Elsevier; 2003. p. 253–83.
4. Ross J, R., Kyoung Hee Nam, D'Auria John C., Pichersky E. S-Adenosyl-L-methionine:salicylic acid carboxyl methyltransferase, an enzyme involved in floral scent production and plant defense, represents a new class of plant methyltransferases. Archives of Biochemistry and Biophysics. 1999.
5. Forouhar F, Lee IS, Vujcic J, Vujcic S, Shen J, Vorobiev SM, et al. Structural and functional evidence for Bacillus subtilis PaiA as a novel N1-spermidine/spermine acetyltransferase. J Biol Chem. 2005; 280 (48):40328–36. https://doi.org/10.1074/jbc.M505332200 PMID: 16210326
6. Kumar D. Salicylic acid signaling in disease resistance. Plant Sci. 2014; 228:127–34. https://doi.org/10.1016/j.plantsci.2014.04.014 PMID: 25438793
7. Kumar D, Klessig Daniel F. High-affinity salicylic acid-binding protein 2 is required for plant innate immunity and has salicylic acid-stimulated lipase activity. Proceedings of the National Academy of Sciences. 2003; 100, 16101–16106.
8. Chaiprasongsuk M, Zhang C, Qian P, Chen X, Li G, Trigiano RN, et al. Biochemical characterization in Norway spruce (Picea abies) of SABATH methyltransferases that methylate phytohormones. Phytochemistry. 2018; 149:146–54. https://doi.org/10.1016/j.phytochem.2018.02.010 PMID: 29501924
9. Vlot C, A., D'Maris AD, Klessig Daniel F. Salicylic Acid, a Multifaceted Hormone to Combat Disease. Annual Review of Phytopathology. 2009; 47(1):177–206.
10. Manthe B, Schulz M, Schnabl H. Effects of salicylic acid on growth and stomatal movements ofVicia faba L.: Evidence for salicylic acid metabolization. Journal for Chemical Ecology. 1992; 18:1525.
11. Cindy E, M., Van Den Boom, Van Beek TA, Posthumus MA, De Groot A, Dicke M. Qualitative and Quantitative Variation Among Volatile Profiles Induced by Tetranychus urticae Feeding on Plants from Various Families. Journal of Chemical Ecology. 2004; 30(1):69–89. https://doi.org/10.1023/b:joec.0000013183.72915.99 PMID: 15074658
12. Mallinger RE, Hogg DB, Gratton C. Methyl Salicylate Attracts Natural Enemies and Reduces Populations of Soybean Aphids (Hemiptera: Aphididae) in Soybean Agroecosystems. Journal of Economic Entomology. 2011; 104(1):115–24. https://doi.org/10.1603/ec10253 PMID: 21404848
13. Gadino A, N., Walton Vaughn M. Lee J, C. Evaluation of methyl salicylate lures on populations of Typhlodromus pyri (Acari: Phytoseiidae) and other natural enemies in western Oregon vineyards,. Biological Control,. 2012; 63(1):48–55.
14. Fang T, Ye Y-YFJ-R. The effect of methyl salicylate on the induction of direct and indirect plant defense mechanisms in poplar (Populus × euramericana 'Nanlin 895'),. Journal of Plant Interactions. 2015; 10:1, 93–100,.
15. Coppola M, Manco E, Vitiello A, Di Lelio I, Giorgini M, Rao R, et al. Plant response to feeding aphids promotes aphid dispersal. Entomologia Experimentalis et Applicata. 2018; 166(5):386–94.
16. Joó É, Dewulf J, Amelynck C, Schoon N, Pokorska O, Šimpraga M, et al. Constitutive versus heat and biotic stress induced BVOC emissions in *Pseudotsuga menziesii*. Atmospheric Environment. 2011; 45 (22):3655–62.
17. Gatehouse JA. Plant resistance towards insect herbivores: a dynamic interaction. New Phytologist. 2002; 156(2):145–69.

18. Chern M, Fitzgerald HA, Canlas PE, Navarre DA, Ronald PC. Overexpression of a rice NPR1 homolog leads to constitutive activation of defense response and hypersensitivity to light. Mol Plant Microbe Interact. 2005; 18(6):511–20. https://doi.org/10.1094/MPMI-18-0511 PMID: 15986920

19. Clarke J, D., Volko Sigrid M., Heidi Ledford Frederick M. Ausubel XD. Roles of Salicylic Acid, Jasmonic Acid, and Ethylene in cpr-Induced Resistance in Arabidopsis. The Plant Cell. 2000; 12, 2175–2190. https://doi.org/10.1105/tpc.12.11.2175 PMID: 11090217

20. Wittek F, Kanawati B, Wenig M, Hoffmann T, Franz-Oberdorf K, Schwab W, et al. Folic acid induces salicylic acid-dependent immunity in Arabidopsis and enhances susceptibility to Alternaria brassicicola. Mol Plant Pathol. 2015; 16(6):616–22. https://doi.org/10.1111/mpp.12216 PMID: 25348251

21. Yan S, Dong X. Perception of the plant immune signal salicylic acid. Curr Opin Plant Biol. 2014; 20:64–8. https://doi.org/10.1016/j.pbi.2014.04.006 PMID: 24840293

22. Clark GSIV. An aroma-chemical profile. Methyl salicylate, or oil of wintergreen1999; 24(1): 5–11. Available from: https://www.perfumerflavorist.com/flavor/application/mint/Methyl-Salicylate-or-Oil-of-Wintergreen-369182341.html.

23. Anderson A, McConville A, Fanthorpe L, Davis J. Salicylate Poisoning Potential of Topical Pain Relief Agents: From Age Old Remedies to Engineered Smart Patches. Medicines (Basel). 2017; 4(3).

24. Applequist WL. Medicinal Plants in Folk Tradition: An Ethnobotany of Britain and Ireland by David E. Allen and Gabrielle Hatfield. Systematic Botany 2004; 29(4), 1021.

25. Angier B. Field Guide to Medicinal Plants. Mechanicsburg, PA, USA: Stackpole Books; 1978.

26. Wendy A. Medicinal Plants in Folk Tradition: An Ethnobotany of Britain and Ireland by David E. Allen and Gabrielle Hatfield. Systematic Botany. 2004; 29(4):1021-, 1.

27. Mason L, Moore RA, Edwards JE, McQuay HJ, Derry S, Wiffen PJ. Systematic review of efficacy of topical rubefacients containing salicylates for the treatment of acute and chronic pain. BMJ. 2004; 328 (7446):995. https://doi.org/10.1136/bmj.38040.607141.EE PMID: 15033879

28. Zhang M, Wang S, Yin J, Li C, Zhan Y, et al. Molecular cloning and promoter analysis of squalene synthase and squalene epoxidase genes from Betula platyphylla. Protoplasma. 2016; 253(5):1347–63. https://doi.org/10.1007/s00709-015-0893-3 PMID: 26464187

29. Singewar K, Moschner CR, Hartung E, Fladung M. 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. 2020.

30. Ashburner K, McAllister HA, Hague J, Brown A, Williams P, Williams M, et al. The Genus Betula: A Taxonomic Revision of Birches: Royal Botanic Gardens, Kew; 2013.

31. Demirci F, Demirci B., Baser K. H. C., Guven aK. The composition and antifungal bioassay of the essential oils of different Betula species growing in Turkey. Chemistry of Natural Compounds. 2000; Vol. 36, No. 2.

32. Dũng NX, Mõi LD, Leclercq PA. Constituents of the Bark Oil ofBetula alnoidesHam ex. D. Don from Vietnam. Journal of Essential Oil Research. 1995; 7(5):565–6.

33. Schnaubelt K. Medical Aromatherapy. North Atlantic Books; Berkley, CA, USA: Frog Ltd.; 1998.

34. Mery G, Katila P, Galloway G, Alfaro R, Kanninen M, Max L, et al. Forests and society: responding to global drivers of change. Vienna, Austria: IUFRO World Series; 2010.

35. Kuuluvainen TT, Bergeron Y, Coates KD. Restoration and Ecosystem-Based Management in the Circumboreal Forest: Background, Challenges, and Opportunities. In: (Ed.) IJS, editor. Restoration of temperate and boreal forests (2 ed, pp 251–270) Boca Raton, FL: CRC Press2015.

36. Williams JJH, Arnold ML. Sources of Genetic Structure in the Woody Perennial Betula occidentalis. International Journal of Plant Sciences. 2001; 162(5):1097–109.

37. Jianhua L, Suzanne S, Zhiduan C. Phylogenetics of Betula (Betulaceae) inferred from sequences of nuclear ribosomal DNA. Rhodora. 2005; 107(929):69–86.

38. Schenk MF, Thienpont C-N, Koopman WJM, Gilissen LJWJ, Smulders MJM. Phylogenetic relationships in Betula (Betulaceae) based on AFLP markers. Tree Genetics & Genomes. 2008; 4(4):911.

39. Altschul S, F., Warren Gish, Webb Miller, Eugene W. Myers DJL. Basic local alignment search tool. Journal of Molecular Biology. 1990; 215(3):403–10. https://doi.org/10.1016/S0022-2836(05)80360-2 PMID: 2231712

40. Dumolin S, Demesure B, Petit RJ. Inheritance of chloroplast and mitochondrial genomes in pedunculate oak investigated with an efficient PCR method. Theoretical and Applied Genetics. 1995; 91(8):1253–6. https://doi.org/10.1007/BF00220937 PMID: 24170054

41. Wang N, Thomson M, Bodles WJA, Crawford RMM, Hunt HV, Featherstone AW, et al. Genome sequence of dwarf birch (Betula nana) and cross-species RAD markers. Molecular Ecology. 2013; 22 (11):3098–111. https://doi.org/10.1111/mec.12131 PMID: 23167599

42. Salojärvi J, Smolander O, Nieminen Kea. Genome sequencing and population genomic analyses provide insights into the adaptive landscape of silver birch. Nature Genetics. 2017; 49, 904–912 https://doi.org/10.1038/ng.3862 PMID: 28481341

43. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. Molecular Biology and Evolution. 2018; 35(6):1547–9. https://doi.org/10.1093/molbev/msy096 PMID: 29722887

44. Lescot M, Déhais P, Thijs G, Marchal K, Moreau Y, Van de Peer Y, et al. PlantCARE, a database of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences. Nucleic Acids Research. 2002; 30(1):325–7. https://doi.org/10.1093/nar/30.1.325 PMID: 11752327

45. Gilbert W. The exon theory of genes. 1987(0091–7451 (Print)).

46. Patthy L. Intron-dependent evolution: Preferred types of exons and introns. FEBS Letters. 1987; 214 (1):1–7. https://doi.org/10.1016/0014-5793(87)80002-9 PMID: 3552723

47. Joshi CP, Chiang VL. Conserved sequence motifs in plant S-adenosyl-L-methionine-dependent methyltransferases. Plant Molecular Biology. 1998; 37(4):663–74. https://doi.org/10.1023/a:1006035210889 PMID: 9687070

48. Zubieta C, Ross JR, Koscheski P, Yang Y, Pichersky E, Noel JP. Structural Basis for Substrate Recognition in the Salicylic Acid Carboxyl Methyltransferase Family. The Plant Cell. 2003; 15(8):1704. https://doi.org/10.1105/tpc.014548 PMID: 12897246

49. Zhao N, Ferrer J-L, Ross J, Guan J, Yang Y, Pichersky E, et al. Structural, biochemical, and phylogenetic analyses suggest that indole-3-acetic acid methyltransferase is an evolutionarily ancient member of the SABATH family. Plant physiology. 2008; 146(2):455–67. https://doi.org/10.1104/pp.107.110049 PMID: 18162595

50. Han X-M, Yang Q, Liu Y-J, Yang Z-L, Wang X-R, Zeng Q-Y, et al. Evolution and Function of the Populus SABATH Family Reveal That a Single Amino Acid Change Results in a Substrate Switch. Plant and Cell Physiology. 2017; 59(2):392–403.

51. Gouet P, Courcelle E, Stuart DI, M√©toz F. ESPript: analysis of multiple sequence alignments in PostScript. Bioinformatics. 1999; 15(4):305–8. https://doi.org/10.1093/bioinformatics/15.4.305 PMID: 10320398

52. Zhao N, Lin H, Lan S, Jia Q, Chen X, Guo H, et al. VvMJE1 of the grapevine (Vitis vinifera) VvMES methylesterase family encodes for methyl jasmonate esterase and has a role in stress response. Plant Physiology and Biochemistry. 2016; 102:125–32. https://doi.org/10.1016/j.plaphy.2016.02.027 PMID: 26934101

53. Zhao N, Guan J, Forouhar F, Tschaplinski TJ, Cheng Z-M, Tong L, et al. Two poplar methyl salicylate esterases display comparable biochemical properties but divergent expression patterns. Phytochemistry. 2009; 70(1):32–9. https://doi.org/10.1016/j.phytochem.2008.11.014 PMID: 19136124

54. Chen F, D'Auria JC, Tholl D, Ross JR, Gershenzon J, Noel JP, et al. An Arabidopsis thaliana gene for methylsalicylate biosynthesis, identified by a biochemical genomics approach, has a role in defense. The Plant Journal. 2003; 36(5):577–88. https://doi.org/10.1046/j.1365-313x.2003.01902.x PMID: 14617060

55. Ament K, Krasikov V, Allmann S, Rep M, Takken FLW, Schuurink RC. Methyl salicylate production in tomato affects biotic interactions. The Plant Journal. 2010; 62(1):124–34. https://doi.org/10.1111/j.1365-313X.2010.04132.x PMID: 20059742

56. Effmert U, Saschenbrecker S, Ross J, Negre F, Fraser CM, Noel JP, et al. Floral benzenoid carboxyl methyltransferases: From in vitro to in planta function. Phytochemistry. 2005; 66(11):1211–30. https://doi.org/10.1016/j.phytochem.2005.03.031 PMID: 15946712

57. Fukami H, Asakura T, Hirano H, Abe K, Shimomura K, Yamakawa T. Salicylic Acid Carboxyl Methyltransferase Induced in Hairy Root Cultures of Atropa belladonna after Treatment with Exogeneously Added Salicylic Acid. Plant and Cell Physiology. 2002; 43(9):1054–8. https://doi.org/10.1093/pcp/pcf119 PMID: 12354924

58. Winkler H. Betulaceae. 1904.

59. Furlow JJ. The genera of Betulaceae in the southeastern United States. Journal of the Arnold Arboretum. 1990; 71:1–67.

60. Regel E. Bemerkungen über die Gattungen Betula und Alnus: nebst Beschreibung einiger neuer Arten: In der Buchdruckerei der Kaiserlichen Universität; 1866.

61. De Jong PC. An introduction to Betula: its morphology, evolution, classification and distribution, with a survey of recent work. Proceedings of the IDS Betula symposium, 2–4 October 1992. 1993.

62. Wang N, McAllister HA, Bartlett PR, Buggs RJA. Molecular phylogeny and genome size evolution of the genus Betula (Betulaceae). Annals of Botany. 2016; 117(6):1023–35. https://doi.org/10.1093/aob/mcw048 PMID: 27072644

63. Schroeder H, Kersten B, Fladung M. Development of Multiplexed Marker Sets to Identify the Most Relevant Poplar Species for Breeding. Forests. 2017; v. 8(no. 12).

64. Bina H, Yousefzadeh H, Ali SS, Esmailpour M. Phylogenetic relationships, molecular taxonomy, biogeography of Betula, with emphasis on phylogenetic position of Iranian populations. Tree Genetics & Genomes. 2016; 12(5):84.

65. Pinosio S, Giacomello S, Faivre-Rampant P, Taylor G, Jorge V, Le Paslier MC, et al. Characterization of the Poplar Pan-Genome by Genome-Wide Identification of Structural Variation. Molecular Biology and Evolution. 2016; 33(10):2706–19. https://doi.org/10.1093/molbev/msw161 PMID: 27499133

66. Zhang B, Zhu W, Diao S, Wu X, Lu J, Ding C, et al. The poplar pangenome provides insights into the evolutionary history of the genus. Communications Biology. 2019; 2(1):215.

67. Porth I, Klapšte J, Skyba O, Hannemann J, McKown AD, Guy RD, et al. Genome-wide association mapping for wood characteristics in Populus identifies an array of candidate single nucleotide polymorphisms. New Phytologist. 2013; 200(3):710–26. https://doi.org/10.1111/nph.12422 PMID: 23889164

68. Wang B, Wang S, Wang Z. Genome-Wide Comprehensive Analysis the Molecular Phylogenetic Evaluation and Tissue-Specific Expression of SABATH Gene Family in Salvia miltiorrhiza. Genes. 2017; 8 (12):365.

69. Järvinen P, Palmé A, Orlando Morales L, Lännenpää M, Keinänen M, Sopanen T, et al. Phylogenetic relationships of Betula species (Betulaceae) based on nuclear ADH and chloroplast matK sequences. American Journal of Botany. 2004; 91(11):1834–45. https://doi.org/10.3732/ajb.91.11.1834 PMID: 21652331

70. Li J, Shoup S, Chen Z. Phylogenetic Relationships of Diploid Species of Betula (Betulaceae) Inferred from DNA Sequences of Nuclear Nitrate Reductase. Systematic Botany. 2007; 32(2):357–65.

71. Vidgren J, Svensson LA, Liljas A. Crystal structure of catechol O-methyltransferase. Nature. 1994; 368 (6469):354–8. https://doi.org/10.1038/368354a0 PMID: 8127373

72. Barkman TJ, Martins TR, Sutton E, Stout JT. Positive Selection for Single Amino Acid Change Promotes Substrate Discrimination of a Plant Volatile-Producing Enzyme. Molecular Biology and Evolution. 2007; 24(6):1320–9. https://doi.org/10.1093/molbev/msm053 PMID: 17374877

73. Qian Ping ZNCFGUOH. Understanding Substrate Specificity of Related Plant Methylesterases(MESs) from Computational Investigations. Chemical Journal of Chinese Universities. 2015; 36(11):2283–91.

74. Yang Y, Xu R, Ma C-j, Vlot AC, Klessig DF, Pichersky E. Inactive Methyl Indole-3-Acetic Acid Ester Can Be Hydrolyzed and Activated by Several Esterases Belonging to the AtMES Esterase Family of Arabidopsis. Plant Physiology. 2008; 147(3):1034. https://doi.org/10.1104/pp.108.118224 PMID: 18467465

75. Negre F, Kolosova N, Knoll J, Kish CM, Dudareva N. Novel S-adenosyl-l-methionine:salicylic acid carboxyl methyltransferase, an enzyme responsible for biosynthesis of methyl salicylate and methyl benzoate, is not involved in floral scent production in snapdragon flowers. Archives of Biochemistry and Biophysics. 2002; 406(2):261–70. https://doi.org/10.1016/s0003-9861(02)00458-7 PMID: 12361714

76. Pott MB, Pichersky E, Piechulla B. Evening specific oscillations of scent emission, SAMT enzyme activity, and SAMT mRNA in flowers of Stephanotis floribunda. Journal of Plant Physiology. 2002; 159 (8):925–34.

77. Tuskan GA, DiFazio S, Jansson S, Bohlmann J, Grigoriev I, Hellsten U, et al. The Genome of Black Cottonwood, Populus trichocarpa (Torr. & Gray). Science. 2006; 313(5793):1596. https://doi.org/10.1126/science.1128691 PMID: 16973872

78. Fang RX, Nagy F, Sivasubramaniam S, Chua NH. Multiple cis regulatory elements for maximal expression of the cauliflower mosaic virus 35S promoter in transgenic plants. The Plant Cell. 1989; 1(1):141. https://doi.org/10.1105/tpc.1.1.141 PMID: 2535461

79. Farnham PJ. Insights from genomic profiling of transcription factors. Nature Reviews Genetics. 2009; 10(9):605–16. https://doi.org/10.1038/nrg2636 PMID: 19668247

80. Lima CCd Silva, Shimo HM, de Felício R, Mercaldi GF, Rocco SA, et al. Structure-function relationship of a citrus salicylate methylesterase and role of salicylic acid in citrus canker resistance. Scientific Reports. 2019; 9(1):3901. https://doi.org/10.1038/s41598-019-40552-3 PMID: 30846791

81. Zhao N, Yao J, Chaiprasongsuk M, Li G, Guan J, Tschaplinski TJ, et al. Molecular and biochemical characterization of the jasmonic acid methyltransferase gene from black cottonwood (Populus trichocarpa). Phytochemistry. 2013; 94:74–81. https://doi.org/10.1016/j.phytochem.2013.06.014 PMID: 23849543

82. Barakat A, Choi A, Yassin NBM, Park JS, Sun Z, Carlson JE. Comparative genomics and evolutionary analyses of the O-methyltransferase gene family in Populus. Gene. 2011; 479(1):37–46.

83. Seo HS, Song JT, Cheong JJ, Lee YH, Lee YW, Hwang I, et al. Jasmonic acid carboxyl methyltransferase: a key enzyme for jasmonate-regulated plant responses. Proc Natl Acad Sci U S A. 2001; 98 (8):4788–93. https://doi.org/10.1073/pnas.081557298 PMID: 11287667

84. Liu X, Luo Y, Wu H, Xi W, Yu J, Zhang Q, et al. Systematic analysis of O-methyltransferase gene family and identification of potential members involved in the formation of O-methylated flavonoids in Citrus. Gene. 2016; 575(2, Part 2):458–72. https://doi.org/10.1016/j.gene.2015.09.048 PMID: 26407870

85. Du H, Klessig DF. Identification of a Soluble, High-Affinity Salicylic Acid-Binding Protein in Tobacco. Plant Physiology. 1997; 113(4):1319. https://doi.org/10.1104/pp.113.4.1319 PMID: 12223676

86. Manosalva PM, Park S-W, Forouhar F, Tong L, Fry WE, Klessig DF. Methyl Esterase 1 (StMES1) Is Required for Systemic Acquired Resistance in Potato. Molecular Plant-Microbe Interactions®. 2010; 23 (9):1151–63.

87. Durrant WE, Dong X. SYSTEMIC ACQUIRED RESISTANCE. Annual Review of Phytopathology. 2004; 42(1):185–209.

88. Yang Y, Varbanova M, Ross J, Wang G, Cortes D, Fridman E, et al. Chapter Ten—Methylation and Demethylation of Plant Signaling Molecules. In: Romeo JT, editor. Recent Advances in Phytochemistry. 40: Elsevier; 2006. p. 253–70.

4 Genome-wide bioinformatics and phylogenetic analysis revealed putative substrate specificities of SABATH and MES family members in silver birch (*Betula pendula*)

(Submitted to *Silvae Genetica*, 29th October 2020)

4.1 Key message

The conducted study suggests the importance of *in silico* bioinformatics analysis before the wet lab experiments. The analysis provided a wide range of candidate genes that are important for different plant developmental processes including pathogen-induced signaling, regulation of growth hormones, and production of secondary metabolic aromatic compounds. All the candidate genes including vital regulatory elements will be subjected to genomic and transcriptomic analysis.

4.2 Abstract

Plant SABATH family members catalyze the methylation of many hormones, signaling molecules, and floral scent metabolites, including salicylic acid (SA), jasmonic acid (JA), and indol-3 acetic acid (IAA). Demethylation of resulting methyl esters executed by members of the MES family. Members of both families are significantly involved in plant developmental processes. Here, using different bioinformatics tools, we studied the evolutionary relationship and characterized the putative functions of the family members in silver birch (*Betula pendula*). It is a socio-ecologically important tree species, plays a vital role in reforestation. Ten and twelve members of the SABATH (BpSABATH1-10) and MES (BpMES1-12) family were identified in silver birch, respectively at the gene and enzyme levels. The *BpSABATH* and *BpMES* genes were distributed on seven of fourteen chromosomes, indicating the occurrence of moderate duplication events important for the expansion of both families. Phylogenetic clustering and gene ontology database suggest, BpSABATH8 is involved in the methylation of indole-3-acetic acid (IAA), while BpSABATH5, BpSABATH6, and BpSABATH7 methylate JA to methyl jasmonate (MeJA). BpSABATH9 was alone in the phylogenetic functional group 1 and prefers SA as a substrate to synthesize methyl salicylate (MeSA). Likewise, BpMES5 and BpMES12 are possibly involved in the demethylation of the methyl ester of IAA, while BpMES6, BpMES7, and BpMES8 responsible for the demethylation of MeJA. BpMES9 clustered with MES, prefer MeSA as a substrate. The current analysis helped to selected candidate genes that could be subjected to further molecular breeding of birch varieties adapted to biotic and abiotic stress conditions.

4.3 Introduction

Many plant metabolites, including salicylic acid (SA), jasmonic acid (JA), and indol-3 acetic acid (IAA), undergo methylation and demethylation in different environmental conditions (D'Auria et al., 2003, Han et al., 2017, Yang et al., 2008). The methylation of these metabolites is catalyzed by the members of SABATH enzyme family, a group of S-adenosyl-L-methionine (SAM)-dependent methyltransferases (SAM-MTs) representing an associated group of *O*-methyltransferases (OMTs) (D'Auria et al., 2003). The intra- and inter-specific comparative analysis showed high sequence similarities in the SABATH family members, though the individual members express different substrate specificities (Han et al., 2017, Yang et al., 2008).

The SAM: salicylic acid carboxyl MT (SAMT) and benzoic acid carboxyl MT (BAMT) from *Clarkia breweri* and *Antirrhinum majus*, respectively, were the first two enzymes isolated and characterized from the SABATH family (Dudareva et al., 2000, Ross et al., 1999). The cofactor SAM is the most widely used methyl donor for enzymatic methyl transfer (Joshi et al., 1998). The name SABATH was designated based on the first three identified and characterized genes (SAMT, BAMT, and Theobromine synthase). Although different members of the *SABATH* gene family showed high nucleotide sequence similarities with many plant species, their numbers vary considerably. In *Arabidopsis thaliana* (*AtSABATH*), and *Oryza sativa* (*OsSABATH*), a total number of 24, and 41 *SABATH* genes, respectively, were identified. The crystal structure of the *A. thaliana* IAA methyltransferase (AtIAMT) was determined and the *OsSABATH4* gene was identified as the most similar to *AtIAMT*. More than half of *OsSABATH* genes were expressed in leaves, roots, and stems representing their active participation in diverse molecular processes (Zhao et al., 2008). A variety of plant mechanisms are responsible for regulating the methylated and free forms of IAA (Delker et al., 2008, Teale et al., 2006). In *A. thaliana* and *Populus trichocarpa*, MeIAMT catalyzes the methylation of IAA (Zhao et al., 2008) involved in leaf development (Qin et al., 2005). The woody plant species *Picea abies* (PaSABATH1-10), *Picea glauca* (PgSABATH1-15), and *P. trichocarpa* (PtSABATH1-28) contain 10, 15, and 28 enzyme family members, respectively (Chaiprasongsuk et al., 2018, Han et al., 2017, Zhao et al., 2009). The enzymatic activity of the ten PaSABATHs was tested against IAA, SA, and JA phytohormones. The higher enzymatic activity with IAA and SA was shown by PaSABATH1 and PaSABATH2, respectively, while three PaSABATHs (4, 5, and 10) elected JA as a substrate (Chaiprasongsuk et al., 2018). Further, comprehensive evolutionary and biochemical functional

analysis disclosed the change in substrate specificity upon a shift in a single amino acid in the forward and reverse mutagenesis studies (Han et al., 2017). The finding indicates the fine-tuned regulation of the SABATH enzyme family members in woody plant species.

The demethylation of the resulting methyl esters of SA (MeSA), JA (MeJA), and IAA (MeIAA) is catalyzed by the members of the methylesterase (MES) enzymes, which is affiliated to the α/β hydrolase superfamily (Nardini et al., 1999). The first MES, salicylic-acid binding protein 2 (SABP2), was isolated from *N. tobacco* and was studied in the SA signaling pathway (Kumar et al., 2003). The amino acid sequence of NtSABP2 shares 77%, 46%, and 56% similarity with *P. trichocarpa* MeSA, *Solanum lycopersicum* MeJA, and *Rauvolfia serpentina* polyneuridine aldehyde esterase (PNAE), respectively (Dogru et al., 2000, Stuhlfelder et al., 2004, Yang et al., 2008, Zhao et al., 2009). In only two species, *A. thaliana* (AtMES) and the *V. vinifera* (VvMES), a total number of 20 and 15 members, respectively, of the *MES* gene family were identified (Yang et al., 2008, Zhao et al., 2016). Of the three members that showed enzymatic activity towards MeJA, the VvMES5 was 77% identical to *S. lycopersicum* MeJA at the protein level (Zhao et al., 2016). Further, the *VvMES5* denoted as *VvMJE1* and its differential expression was evaluated with heat, cold, and UV-B-treated *V. vinifera* plants. Upregulation in the expression of *VvMES1* upon cold and UV-B treatment was observed, suggesting its role in response to abiotic stresses. The active participation of MeJA in keeping fruits and vegetables fresh has also been demonstrated (Alvarez et al., 2015).

Biochemical analysis revealed, AtMES17 (At3g10870) vigorously catalyzes the hydrolysis of methyl ester IAA (MeIAA). However, the AtMES17 mutated with T-DNA insertions resulted in reduced sensitivity to MeIAA in comparison to wild-type roots of *A. thaliana* plants. In the same study, *A. thaliana* plants overexpressing AtMES17 showed induced activity to MeIAA and not to IAA. The study suggested, the MeIAA is an idle form, produced from free IAA and the exhibition of MeIAA is carried out by the free IAA due to hydrolysis by plant esterases (Yang et al., 2008). A recent study with *Citrus sinensis* has shown the participation of CsMES in the hydrolysis of MeSA into SA through molecular modeling. It could be demonstrated that the citrus canker caused by *Xanthomonas citri* is suppressed by SA and MeSA (Lima Silva et al., 2019). Accumulation of SA and CsMES occurred in the course of *X. aurantifolii* and *C. sinensis* interaction. The finding advocates the role of MeSA and SA in the pathogen-induced systemic acquired resistance (SAR) mechanism.

The role of SABATH and MES enzyme families in the synthesis of the hormones, signaling molecules, and floral scent metabolites necessary for plant development have been mostly studied in the model and crop plants (D'Auria et al., 2003, Yang et al., 2008). Thus, the information about these enzymes in the woody plant species is very limited. Considering the importance of forest trees for the ecosystem in the present area of climate change, it is advantageous to study the SABATH and MES enzymes in long-lived woody plant species. Here, *B. pendula* (silver birch) was selected for the *in-silico* analysis since it is one of the dominant species in the boreal forests of the Northern Hemisphere (Salojärvi et al., 2017). Silver birch is a commercially important tree species and plays a vital role in landscape structure, forestry, breeding for biomass production, and horticulture (Ashburner et al., 2013). The leaf color of the birch transforms to yellow-green in autumn and is usually green in the springtime and summer (Gang et al., 2019). The characteristics like short life cycle, rapid growth and plentiful production of seeds make birch a pioneer species that participate in the regeneration of forests after 'forest fires' (Fischer et al., 2002). Different species of the genus *Betula,* adapting to various climatic conditions, are distributed within the wide geographical region (Hemery et al., 2010, Hynynen et al., 2009). Birch is a wind-pollinated species, widely involved in cross-pollination (Atkinson 1992, Koski et al., 2005) and creating a large gene pool (Ranta et al., 2008). Thus, high genetic variability is maintained giving rise to tolerance formation and increasing the probability of survival in diverse environmental conditions (Araminienė et al., 2014, Aspelmeier et al., 2004). Birches create ideal living conditions for other tree species (Prévosto et al., 2004, Rosenvald et al., 2014), and thus, they significantly contribute to the recovery of forests after disturbances (Dubois et al., 2020). Silver birch plays a key role in maintaining the biodiversity of coniferous forests since the species coexists with other tree species (Hynynen et al., 2009).

The present study aims to (1) identify and characterize the *SABATH* and *MES* genes in *B. pendula;* (2) facilitate our understanding of the evolution and the putative substrate specifications of SABATH and MES enzyme members; and (3) provide useful bioinformatics information for the selection of appropriate candidate genes involved in the methylation and demethylation of SA, JA, and IAA in *B. pendula*.

To this end, we have successfully characterized the gene and enzyme members of the two SABATH and MES families in *B. pendula* by *in silico* analyses. The different bioinformatic

analyses were crucial and assisted in designing further state-of-art molecular and biochemical experiments to evaluate their functional role in *B. pendula*.

4.4 Materials and methods

4.4.1 Identification of *SABATH* and *MES* gene family members in *B. pendula*

The amino acid sequences of *C. breweri* SAMT (CbSAMT) and *N. tabacum* SABP2 (NtSABP2) were obtained from previous studies (Kumar et al., 2003, Ross et al., 1999). The amino acid sequences of the two genes were used as queries in a tBLASTn search of the *B. pendula* genome sequence (https://genomevolution.org/coge/GenomeInfo.pl?gid=35080). An E-value cutoff of 1^{-5} was applied to the homolog recognition and if the sequence satisfied $\leq$1e-10, it was selected as a candidate gene.

4.4.2 SABATH and MES amino acid sequence retrieval from different plant species

Only functionally characterized members from *A. thaliana* (AtSABATH), *P. trichocarpa* (PtSABATH), and *P. abies* (PaSABATH) (Chaiprasongsuk et al., 2018, D'Auria et al., 2003, Zhao et al., 2013), *A. thaliana* (AtMES), *V. vinifera* (VvMES) (Yang et al., 2008, Zhao et al., 2016), and other known members from different species were retrieved from the https://www.ncbi.nlm.nih.gov/ and http://popgenie.org/ for the comparative analysis (S Table 1).

4.4.3 Multiple sequence alignment and phylogenetic analysis

The retrieved amino acid sequences were aligned using the ClustalW program (Thompson et al., 1994) available in the MEGA X bioinformatics package (Kumar et al., 2018) with default parameters. The maximum likelihood (ML) phylogenetic trees of BpSABATH and BpMES with other known SABATH and MES enzymes were constructed using a bootstrap value of 1,000 replicates in the MEGA X bioinformatics tool (Kumar et al., 2018). *Aspergillus niger* SAMT (NT166520) and *Beauveri abassiana* MES (PMB68924.1) were used as outgroup species in the construction of the phylogenetic tree for substrate prediction.

4.4.4 Chromosomal localization of birch *SABATH* and *MES* genes

A physical map was constructed to confirm the chromosomal locations of the *SABATH* and *MES* genes. The karyoploteR (Gel et al., 2017) package was used to plot the chromosome map and to visualize the locations of *SABATH* genes on the *B. pendula* chromosomes.

Tandem duplications in the *BpSABATH* and *BpMES* gene family were determined when located within 100 kb neighboring regions and when a close phylogenetic relationship was formed among a group of genes at the same chromosome location (Kong et al., 2007).

4.4.5 Gene structure, conserved domain, gene ontology, and promoter analysis

Coding regions (CDSs) and genomic sequences were retrieved from the *B. pendula* genome (https://genomevolution.org/coge/GenomeInfo.pl?gid=35080) to analyze the intron/exon organization of *BpSABATH* and *BpMES* genes. Further, the sequences were submitted to the Gene Structure Display Server (http://gsds.cbi.pku.edu.cn/) to investigate the gene structure based on each of the CDSs and the corresponding genomic sequences.

The MEME online tool (http://meme-suite.org/) was utilized to identify the motifs present in the *BpSABATH* and *BpMES* genes. The following parameters were set: the maximum number of motifs, 11; minimum motif width, 6; maximum motif width, 60. Additionally, all predicted SABATH and MES gene family members in *B. pendula* were submitted to the Pfam database (El-Gebali et al., 2018) to confirm the conserved domains of all candidates (https://www.genome.jp/tools/motif/).

The theoretical isoelectric point (*pI*) and molecular weights (*Mw*) of the SABATH and MES enzymes in *B. pendula* were predicted using the 'Compute *pI*/*Mw* tool' on the ExPASy server (https://web.expasy.org/compute_pi/). Besides, the promoter regions of *BpSABATH* and *MES* genes were examined in the PlantCARE database (Lescot et al., 2002). DNA fragments of approximately 1,000 bp were retrieved from the 5'-untranslated region of the genes. Further, the raw sequences were subjected to the PlantCARE database and the option 'search for care' used to search for *Cis*-regulatory elements.

The *B. pendula* gene ontology browser available at the Hardwood Genome project (https://www.hardwoodgenomics.org/organism/Betula/pendula) was used to attribute the product of the *BpSABATH* and *BpMES* genes. Further, the protein structure homology-modeling of all the family members was carried out to validate the functional residues. The Swiss-Model, an automated server (https://swissmodel.expasy.org/), was used to build the protein models.

4.5 Results

4.5.1 Identification and comparative analysis of *SABATH* and *MES* genes

To identify the *SABATH* and *MES* gene family members in *B. pendula*, BLASTP analyses against the *B. pendula* genome were performed using amino acid sequences of CbSAMT for SABATH and NtSABP2 for MES as queries. A total number of 10 and 12 *SABATH* and *MES* most similar genes were obtained (Table 1). Protein sequences of both family members were subjected to Pfam analyses to confirm their protein domain.

Gene lengths of *BpSABATH* varied from 1,730 (*BpSABATH10*: Bpev01.c0800.g0038.m0001) to 21,489 bp (*BpSABATH9*: Bpev01.c0425.g0055.m0001). The lengths of the BpSABATH CDS and protein varied from 453 bp and 150 aa (BpSABATH1: Bpev01.c2345.g0001.m0001) to 1,548 bp and 515 aa (BpSABATH5: Bpev01.c0161.g0056.m0001), respectively (S Table 2).

Table 1: The SABATH and MES family members in *B. pendula*. A total number of 10 and 12 members were obtained, respectively. The gene locus, chromosome number, length of nucleotide sequence, protein, and CDS were determined using the *B. pendula* genome.

BpSABATH* and *BpMES	**Gene locus**	**Chromosome**	**The nucleotide sequence (bp)**	**CDS (bp)**	**Amino Acid (aa)**
BpSABATH1	Bpev01.c2345.g0001.m0001	Chr1	5,177	453	150
BpSABATH2	Bpev01.c1865.g0002.m0001	Chr1	2,113	858	285
BpSABATH3	Bpev01.c0759.g0006.m0002	Chr1	3,566	966	321
BpSABATH4	Bpev01.c0807.g0007.m0001	Chr3	9,210	954	317
BpSABATH5	Bpev01.c0161.g0056.m0001	Chr9	5,154	1,548	515
BpSABATH6	Bpev01.c0161.g0057.m0001	Chr9	2,693	1,125	374
BpSABATH7	Bpev01.c0161.g0058.m0001	Chr9	2,793	1,107	368
BpSABATH8	Bpev01.c0240.g0011.m0001	Chr12	3,226	1,029	342
BpSABATH9	Bpev01.c0425.g0055.m0001	Chr12	21,489	1,344	447
BpSABATH10	Bpev01.c0800.g0038.m0001	Chr13	1,730	1,020	340
BpMES1	Bpev01.c0449.g0051.m0001	Chr1	4,323	1,164	258
BpMES2	Bpev01.c0919.g0029.m0001	Chr1	15,239	996	214
BpMES3	Bpev01.c0135.g0098.m0001	Chr2	5,954	1,140	260
BpMES4	Bpev01.c0436.g0011.m0001	Chr3	1,764	738	167
BpMES5	Bpev01.c1072.g0010.m0001	Chr5	2,691	813	195
BpMES6	Bpev01.c0015.g0216.m0001	Chr5	1,014	765	170
BpMES7	Bpev01.c0015.g0217.m0001	Chr5	2,819	783	170
BpMES8	Bpev01.c0015.g0218.m0001	Chr5	3,644	780	172
BpMES9	Bpev01.c0015.g0219.m0001	Chr5	2,099	792	180
BpMES10	Bpev01.c0015.g0220.m0001	Chr5	5,425	561	121
BpMES11	Bpev01.c0015.g0221.m0001	Chr5	1,612	627	147
BpMES12	Bpev01.c0089.g0060.m0001	Chr13	6,019	777	179

The gene lengths of *BpMES* varied from 1,014 bp (*BpMES6*: Bpev01.c0015.g0216.m0001) to 15,239 bp (*BpMES2*: Bpev01.c0919.g0029.m0001). The lengths of the BpMES proteins and CDSs varied from 121 aa to 260 aa (BpMES10: Bpev01.c0015.g0220.m0001 and BpMES1: Bpev01.c0449.g0051.m0001) and 561 bp to 1,164 bp (BpMES1: Bpev01.c0449.g0051.m0001 and BpMES10Bpev01.c0015.g0220.m0001), respectively (S Table 2).

The molecular weights of the predicted BpSABATH enzymes ranged from 16.90 kDa (BpSABATH1) to 48.67 kDa (BpSABATH9), and the theoretical isoelectric points were predicted to range from 4.91 (BpSABATH1) to 8.75 (BpSABATH9) (S Table 3). Likewise, the molecular weights of BpMES enzymes ranged from 22.01 kDa (BpMES10) to 42.68 kDa (BpMES1), and the theoretical isoelectric points were predicted to range from 5.20 (BpMES6) to 9.41 (BpMES11) (S Table 3).

4.5.2 Chromosomal localization of *B. pendula SABATH* and *MES* genes

Analysis of the chromosomal locations showed that the ten and twelve *SABATH* and *MES* genes each mapped to only five chromosomes and were unevenly distributed throughout the genome (Figure 1). Of the 10 *BpSABATH* genes, three were located on chromosomes 1 (*BpSABATH1*, *BpSABATH2*, and *BpSABATH3*) and 9 (*BpSABATH5*, *BpSABATH6*, and *BpSABATH7*), with two genes on chromosome 12 (*BpSABATH8* and *BpSABATH9*). In contrast, the two genes *BpSABATH4* and *BpSABATH10* were located on chromosomes 3 and 13, respectively.

Of the 12 *BpMES* genes, only one gene each was located on chromosomes 2 (*BpMES3*), 3 (*BpMES4*), and 13 (*BpMES12*), with two genes located on chromosome 1 (*BpMES1* and *BpMES2*). In contrast, seven *BpMES* genes (*BpMES5*, *BpMES6*, *BpMES7*, *BpMES8*, *BpMES9*, *BpMES10*, and *BpMES11*) were located on chromosome 5.

Considering that duplication events are more likely to be customary in the gene family expansion (Moore et al., 2003), the possibilities of tandem duplications of the *BpSABATH* and *BpMES* genes were investigated (Figure 1). Of the 10 *BpSABATH* genes, five were found in two tandem repeats, including *BpSABATH1* with *BpSABATH2*, and *BpSABATH5* and *BpSABATH6* with *BpSABATH7*. Eight of the 12 *BpMES* genes were found in two tandem repeats, including *BpMES1* with *BpMES2*, and *BpMES5*, *BpMES6*, *BpMES7*, *BpMES8*, *BpMES9*, and *BpMES10* with *BpMES11*.

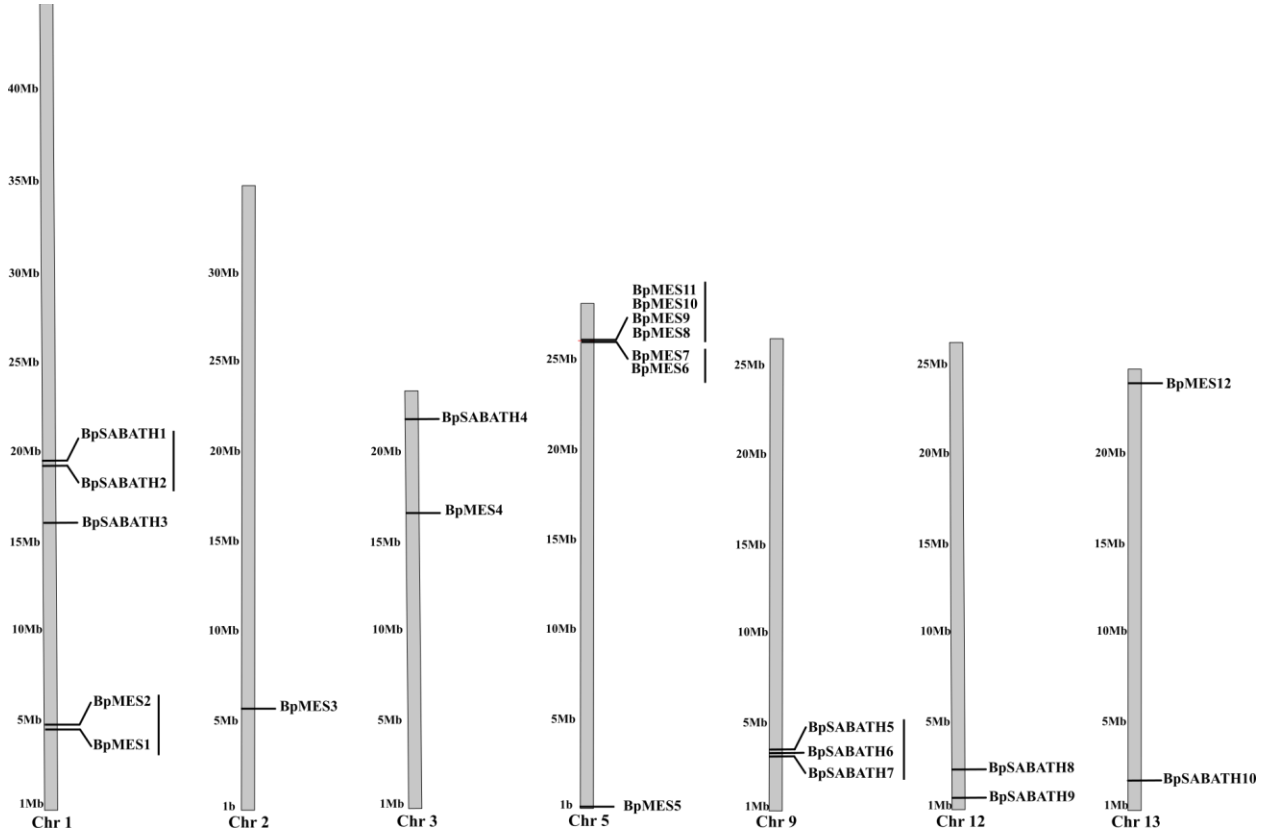

Figure 1: Localization of *SABATH* and *MES* genes on silver birch (*B. pendula*) chromosomes: Chromosomes 1 and 9 carry three *SABATH*s each, while chromosomes 3 and 13 carry only one *SABATH* gene. Chromosomes 2 and 5 carry one and seven *MES* genes, respectively. Chromosomes 1, 3, and 13 contain both *SABATH* and *MES* genes. The names of the chromosomes and their sizes (Mb) are indicated next to each chromosome and are based on the *B. pendula* genome. Tandemly duplicated genes are shown beside the black lines. No evidence of segmental duplication was identified in the *SABATH* gene family in the *B. pendula* genome. The karyoploteR package was used to plot the chromosome map.

4.5.3 Gene structure and intraspecies phylogenetic relationship analysis of *BpSABATH* and *BpMES* family members

The structural diversity of *BpSABATH* and *BpMES* genes was analyzed through their exon/intron organization. The phylogenetic tree to analyze the interspecies relationship between *BpSABATH* and *BpMES* genes was constructed using the maximum likelihood method (Figure 2A). All *BpSABATH* genes contain introns; no genes without introns were observed (Figure 2B). Tandem duplicated pairs (*BpSABATH1* with *BpSABATH2* and *BpSABATH5* and *BpSABATH6* with *BpSABATH7*) showed similar gene structures. Every exon of the gene was similar to its tandemly duplicated sister gene and also showed a similar size.

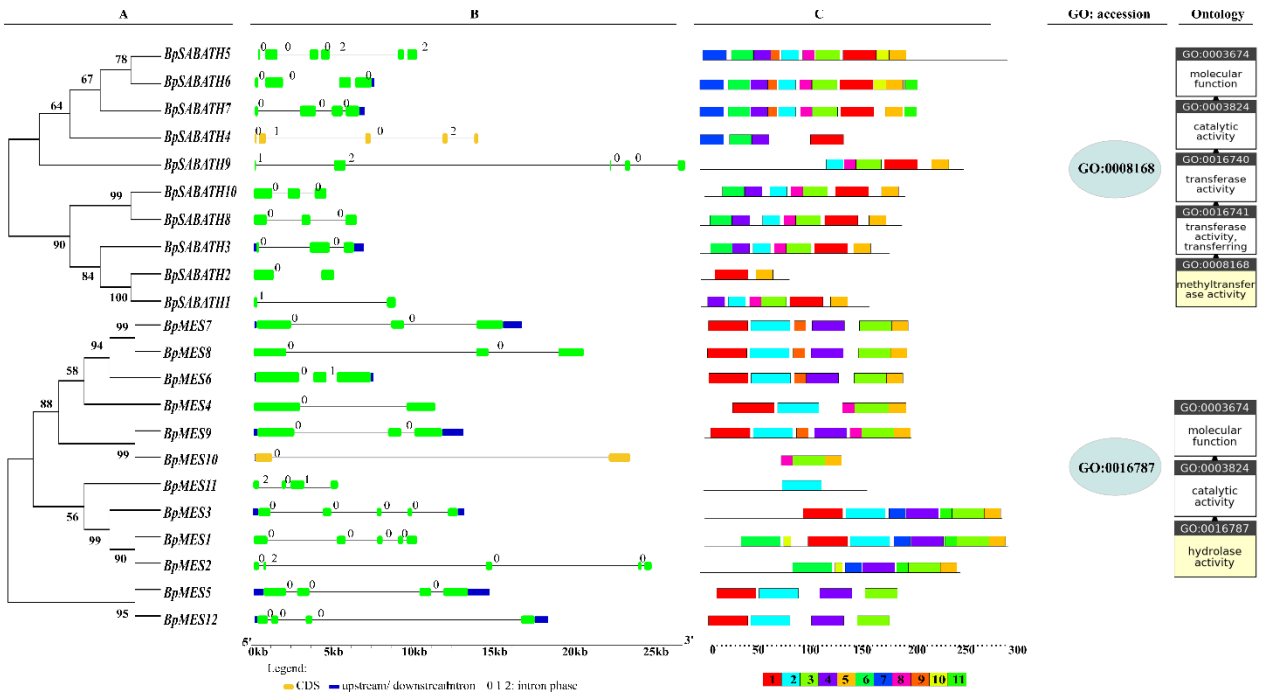

Figure 2: Intraspecies relationship, gene structure, motifs, and gene ontology of SABATH and MES families: (A) The phylogenetic tree was constructed using the maximum likelihood method in MEGA X software with 1,000 bootstrap replicates (Kumar et al., 2018) to analyze the interspecies relationship between *BpSABATH* and *BpMES* genes. (B) Structural features of the *SABATH* and *MES* genes in *B. pendula*. The coding regions (CDS) are indicated by green round-corner rectangles, while black lines between two exons represent the introns. Blue boxes indicate upstream/downstream UTRs. Intron phases are represented by the numbers above the intron (black line). Intron phases are likely to assist in exon shuffling, recombination fusion, and protein domain exchange (Gilbert 1987, Patthy 1987). (C) Schematic representation of the motifs in *B. pendula* SABATH and MES proteins. The lengths of the motifs can be estimated using the scale at the bottom of the figure. The names of the motifs are listed at the bottom of the figure. (D) The information about the gene ontology is also shown to confirm the putative functions of the candidate genes.

The *BpMES* gene structural analysis revealed that the number of exons varied from two to five. No genes lacking introns were observed, i.e., all genes contained introns (Figure 2B). Tandem duplicated gene pairs (*BpMES1* through *BpMES2*, and *BpMES6*, *BpMES7*, *BpMES8*, *BpMES9* through *BpMES10*) showed similar intron and exon structures, while the two tandem duplicates, *BpMES5* and *BpMES11* displayed a related intron/exon structure. Nevertheless, every exon of the gene was similar to its tandem duplicated sister gene and also showed a similar size.

The sequence length of introns was more variable in comparison with the conserved exon regions of their tandem duplicated sister genes. This analysis is consistent with the traditional theory that introns are more unstable than the conserved exon regions (Koonin 2006).

The symmetric exons represent the same splicing phase at both ends and an excess of symmetric exons and phase 0 introns are expected to accelerate protein domain exchange, exon shuffling,

and fusion in recombination (Gilbert 1987, Patthy 1987). According to the analyzed gene structures, the exons of six genes were symmetric with phase 0 introns and no exon was symmetrical with phase 1 and 2 introns. Of the 37 introns of the ten *BpSABATH* genes, 20 were phase 0, three were phase 1 and four were phase 2 (Figure 2B). Similarly, of the 31 introns of MES genes, 27 were phase 0, two were phase 1, and only two were phase 2 (Figure 2B).

The motif similarities and differences within *BpSABATH* genes were compared using the MEME online suite (Figure 2C). The *BpSABATH* gene family contains 11 distinct motifs (Figure 2C; S Figure 1A). Overall, the most closely related members of the family showed a similar motif organization (*BpSABATH5* and *BpSABATH6* with *BpSABATH7*, and *BpSABATH1* with *BpSABATH2*). Motifs 1 and 5 were shared by all the *BpSABATH*s, while motif 10 was present only in *BpSABATH5* and 6. Further, motif 9 was specific to the *BpSABATH5*, 6, and 7. Motifs 2, 3, and 4 were shared among all *BpSABATH*s, except for *BpSABATH1* and *BpSABATH9*. *BpSABATH1*, 2, 3, 8, and 10 lacked motifs 7, 9, 10, and 11 which were mainly distributed in the *BpSABATH*4, 5, 6, 7, and 9 presents in the N- or C- terminal. Similarly, the organizational variations of the motif in *BpMES* were also compared. Eleven different motifs were identified in the *BpMES* gene family and their logos were also extracted (S Figure 1B). The most closely related members within the family showed a similar motif organization (*BpMES6* and *BpMES7* with *BpMES8*, and *BpMES5* with *BpMES12*). Motif 3 was shared by all *BpMES*s, except for *BpMES11* which contained only one motif (motif 2) (Figure 2C), while motif 6 and 10 were present only in *BpMES1* and 2. Also, motif 9 was specific to *BpMES*5, 7, 8, and 9, while motif 7 was shared among *BpMES*1, 2, and 3 (Figure 2C). The putative functions of the candidate genes were confirmed by the information about the gene ontology (Figure 2D).

4.5.4 Multiple sequence alignment and homology modeling of BpSABATH and BpMES proteins

The protein sequence alignment of BpSABATH proteins with CbSAMT showed the presence of SAM/SAH-binding residues as well as the aromatic moiety of the substrate (Figure 3). A total number of six residues that actively participate in the SAM/SAH were observed at 22-Ser, 57-Asp, 98-Asp, 99-ASP, 129-Ser, and 130-Phe. Likewise, 16 residues involved in SA binding were detected at 25-Gln, 145-Ser, 146-Ser, 147-Tyr, 148- Ser, 149- Leu, 150-Met, and 151-Trp. Of the six SAM/SAH binding residues that have been studied in CbSAMT, BpSABATH4, BpSABATH5, BpSABATH6, BpSABATH7, BpSABATH8, and BpSABATH10 showed all six, while BpSABATH1 did not show any.

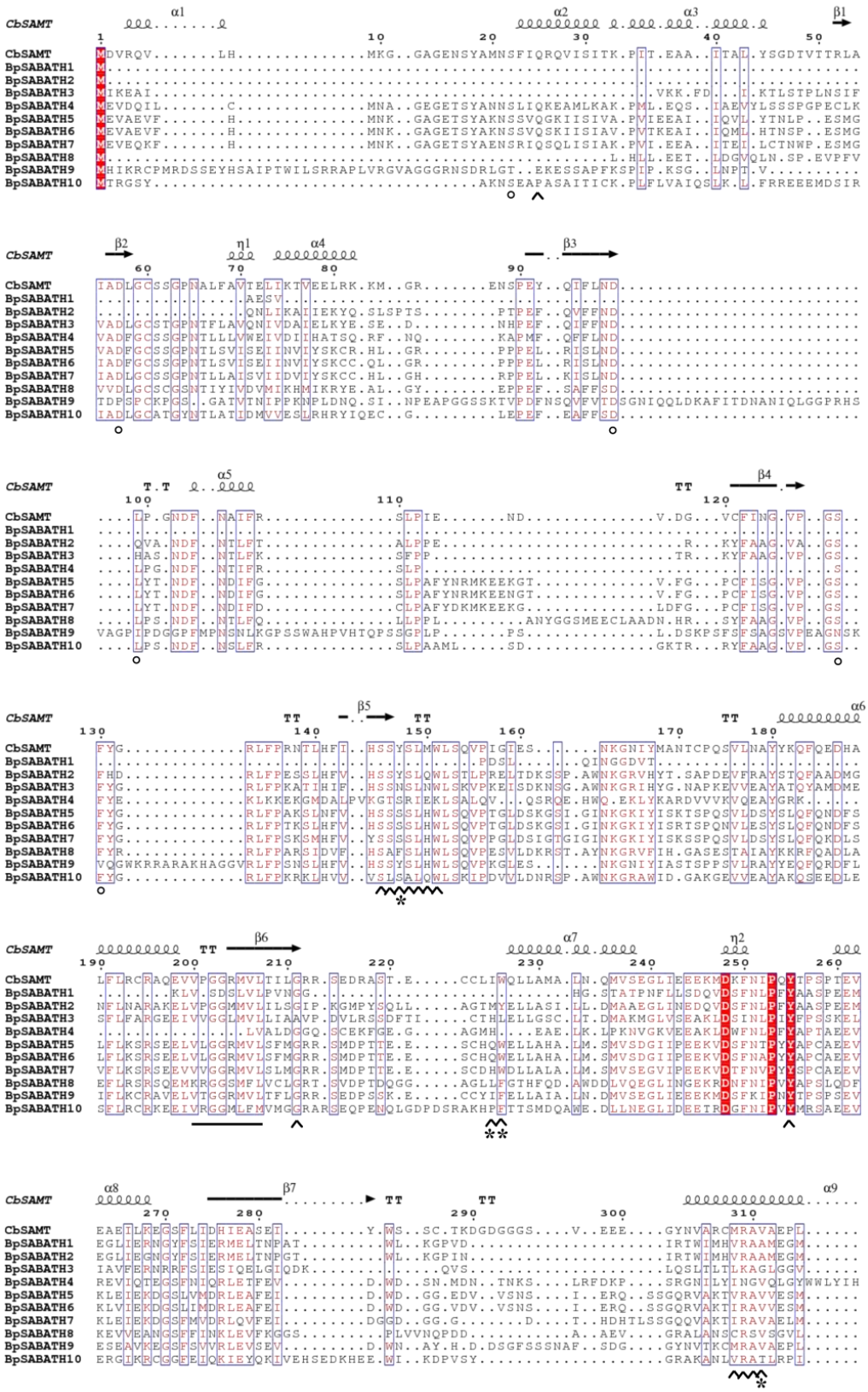
CbSAMT
BpSABATH1
BpSABATH2
BpSABATH3
BpSABATH4
BpSABATH5
BpSABATH6
BpSABATH7
BpSABATH8
BpSABATH9
BpSABATH10

Figure 3: Multiple sequence alignment of BpSABATH family members: Structure-based multiple sequence alignment of *B. pendula* SABATH protein family members and, for comparison, *CbSAMT* from *C. breweri* (Ross et al., 1999). Blue frames indicate conserved residues, white characters in red boxes indicate strict identity, and red characters in white boxes indicate similarity. The conserved domain of methyltransferase including the SAM-binding motif that had previously been defined is highlighted with a black line (Joshi et al., 1998). The secondary structure elements above the alignment are those of the CbSAMT protein whose structure has been previously experimentally determined and described (Zubieta et al., 2003). The positions of residues involved in the SA substrate-binding and SAM/SAH-binding residues, identified by the three-dimensional structures, are indicated by black arrows and circles, respectively (Zubieta et al., 2003). Residues indicated by an asterisk are the aromatic moiety of the substrate and are important for substrate selectivity identified in the previous study (Zhao et al., 2008). The figure was prepared with the help of ESPript (Gouet et al., 1999).

Further, BpSABATH2, BpSABATH3, and BpSABATH9 carried at least one of the SAM/SAH binding residues. The aromatic residue Val-311 was present in all the BpSABATH proteins except for BpSABATH1, BpSABATH2, BpSABATH3, and BpSABATH10 while BpSABATH2 only carried residue Tyr-147. BpSABATH9 contains the highest number of aromatic residues of all the BpSABATH and includes Tyr-147, Lue-210, Iso-225, and Phe-347, while residue Try-226 was present only in BpSABATH5, BpSABATH6, and BpSABATH7.

The protein sequence alignment of BpMES with NtSABP2 showed the presence of catalytic triad residues that were observed at 81-Ser, 210-Asp, and 238-His, a characteristic feature of the α/β hydrolase fold family, is conserved in 8 of these enzymes (Figure 4). In the BpMES1 and BpMES3, the conserved Ser in the catalytic triad is replaced by Asp, a substitution previously found in active α/β hydrolases in animals (Holmquist 2000, Yang et al., 2008), while, in BpMES10 it is replaced by Met (Figure 4). Similarly, 14 residues that contact SA were conserved at positions 13-Ala, 15-His, 81-Ser, 82-Leu, 107-Phe, 122-Tyr, 131-Trp, 136-Phe, 149-Met, 152-Phe, 155-Phe, 160-Leu, and 213-Ile (Figure 4). Homology modeling of both gene families was conducted to analyze the protein structure similarities as well as to visualize the functional residues within the amino acid sequences identified in multiple sequence alignment (S Table 4).

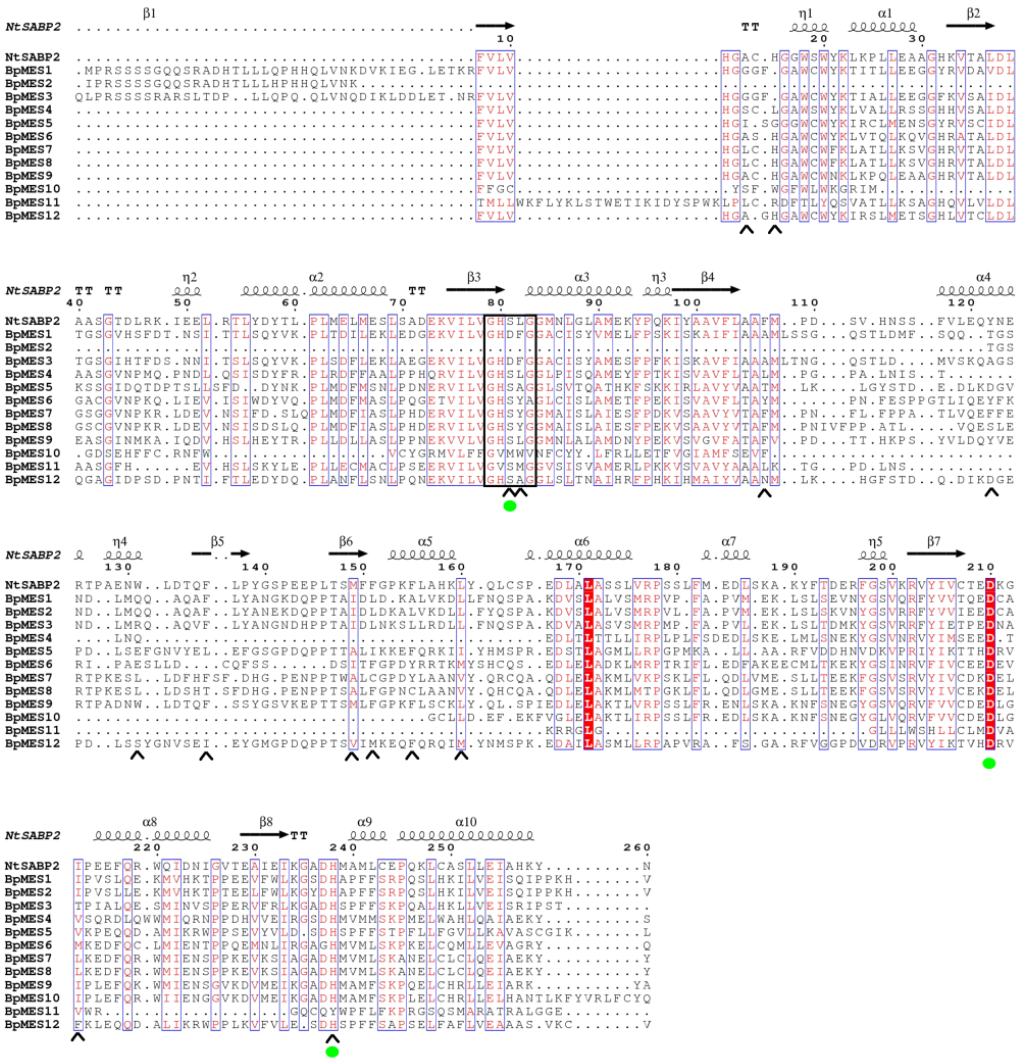

Figure 4: Multiple sequence alignment of BpMES family members: Multiple sequence alignment of *B. pendula* MES protein family members and, for comparison, NtSABP2 (Kumar et al., 2003). The blue frames represent the conserved residues, white characters in red boxes represent strict identity, and red characters in white boxes specify similarity. The lipase signature sequence of SABP2 is displayed with a black frame. The three conserved amino acids form a catalytic triad, S81, D210, and H238, commonly found in the hydrolase domain indicated by the green dot, conserved in BpMES (Kumar et al., 2003), while residues that contact to SA are indicated with an arrow (Forouhar et al., 2005). The figure was prepared with ESPript (Gouet et al., 1999).

4.5.5 Evolutionary relationships and putative substrate specificity of BpSABATH and BpMES enzyme family members

Phylogenetic clustering could preliminarily predict the functions of an unknown protein since grouped proteins in a clade showed similar gene structures and might possess similar functions (Kapteyn et al., 2007, Zhao et al., 2013). Besides, the proteins might be evolved from a recent common ancestor (Xie et al., 2014).

The probable role in plant and substrate specificity of BpSABATH proteins were determined based on a maximum likelihood phylogenetic tree (Figure 5) constructed using 10 BpSABATHs and 52 functionally characterized SABATHs from other plant species, including *A. thaliana*, *P. trichocarpa, P. abies,* and other respective species. Only functionally characterized SABATH members from the respective species were included to predict potential substrates of BpSABATH enzymes. The topology of the phylogenetic tree and distribution of the BpSABATHs, together with functionally characterized SABATHs (Figure 5), formed seven functional subgroups (functional subgroup 1 to 7).

Almost all functionally characterized SABATHs from functional subgroup 1 were observed to catalyze the conversion of SA and BA to SAMT and/or BSMT. Only one (BpSABATH9) of the 10 BpSABATHs joined functional group 1. The resulting products were subsequently involved in various biological functions. Most of the members from functional subgroup 2 were involved in catalyzing JA and IAA and accommodate a maximum of four BpSABATHs, suggesting their potential functions. BpSABATH3, BpSABATH10, and BpSABATH8 were clustered in functional groups 4, 6, and 7, respectively. Most of the functionally characterized members grouped in functional groups 4, 6, and 7 showed higher enzymatic activity towards FA, JA, and IAA, respectively. Considerably, functional groups 4 and 6 also catalyze VA, GA, and JA with low enzymatic activity. Functional groups 3 and 5 did not include any members of the BpSABATHs, while functional group 7 included equal numbers of SABATHs that utilize IAA, SA, and BA as substrates.

In the phylogenetic tree, two groups of paralogous genes (*BpSABATH1* and *BpSABATH2*; *BpSABATH5, BpSABATH6*, and *BpSABATH7*) were identified from the SABATH gene family in *B. pendula* (Figure 5). In addition, two pairs of orthologues, BpSABATH8 with Potri.001G359000 (enzymatic activity on indole-3-acetic acid), and BpSABATH4 with Potri.014G168200 (enzymatic activity on farnesoic acid) were identified in the SABATH gene family in *B. pendula* that most probably have the same function.

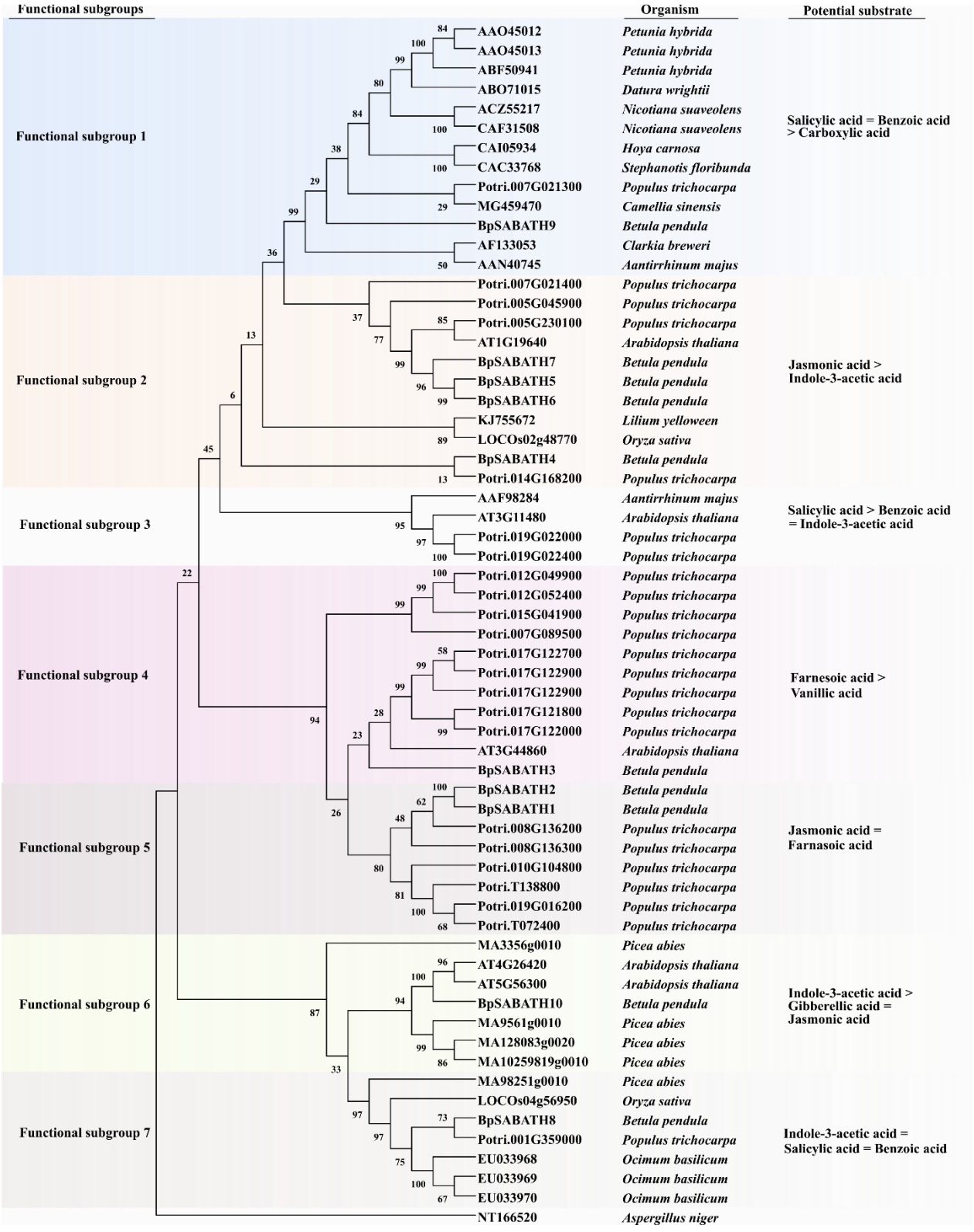
Functional subgroups
Organism
Potential substrate
Functional subgroup 1
Functional subgroup 2
Functional subgroup 3
Functional subgroup 4
Functional subgroup 5
Functional subgroup 6
Functional subgroup 7
AAO45012 Petunia hybrida
AAO45013 Petunia hybrida
ABF50941 Petunia hybrida
ABO71015 Datura wrightii
ACZ55217 Nicotiana suaveolens
CAF31508 Nicotiana suaveolens
CAI05934 Hoya carnosa
CAC33768 Stephanotis floribunda
Potri.007G021300 Populus trichocarpa
MG459470 Camellia sinensis
BpSABATH9 Betula pendula
AF133053 Clarkia breweri
AAN40745 Aantirrhinum majus
Salicylic acid = Benzoic acid > Carboxylic acid
Potri.007G021400 Populus trichocarpa
Potri.005G045900 Populus trichocarpa
Potri.005G230100 Populus trichocarpa
AT1G19640 Arabidopsis thaliana
BpSABATH7 Betula pendula
BpSABATH5 Betula pendula
BpSABATH6 Betula pendula
KJ755672 Lilium yelloween
LOCOs02g48770 Oryza sativa
BpSABATH4 Betula pendula
Potri.014G168200 Populus trichocarpa
Jasmonic acid > Indole-3-acetic acid
AAF98284 Aantirrhinum majus
AT3G11480 Arabidopsis thaliana
Potri.019G022000 Populus trichocarpa
Potri.019G022400 Populus trichocarpa
Salicylic acid > Benzoic acid = Indole-3-acetic acid
Potri.012G049900 Populus trichocarpa
Potri.012G052400 Populus trichocarpa
Potri.015G041900 Populus trichocarpa
Potri.007G089500 Populus trichocarpa
Potri.017G122700 Populus trichocarpa
Potri.017G122900 Populus trichocarpa
Potri.017G122900 Populus trichocarpa
Potri.017G121800 Populus trichocarpa
Potri.017G122000 Populus trichocarpa
AT3G44860 Arabidopsis thaliana
BpSABATH3 Betula pendula
Farnesoic acid > Vanillic acid
BpSABATH2 Betula pendula
BpSABATH1 Betula pendula
Potri.008G136200 Populus trichocarpa
Potri.008G136300 Populus trichocarpa
Potri.010G104800 Populus trichocarpa
Potri.T138800 Populus trichocarpa
Potri.019G016200 Populus trichocarpa
Potri.T072400 Populus trichocarpa
Jasmonic acid = Farnasoic acid
MA3356g0010 Picea abies
AT4G26420 Arabidopsis thaliana
AT5G56300 Arabidopsis thaliana
BpSABATH10 Betula pendula
MA9561g0010 Picea abies
MA128083g0020 Picea abies
MA10259819g0010 Picea abies
Indole-3-acetic acid > Gibberellic acid = Jasmonic acid
MA98251g0010 Picea abies
LOCOs04g56950 Oryza sativa
BpSABATH8 Betula pendula
Potri.001G359000 Populus trichocarpa
EU033968 Ocimum basilicum
EU033969 Ocimum basilicum
EU033970 Ocimum basilicum
Indole-3-acetic acid = Salicylic acid = Benzoic acid
NT166520 Aspergillus niger

Figure 5: Potential substrates of BpSABATH protein family members according to phylogenetic clustering: Only functionally characterized protein sequences were used for the phylogenetic tree construction. The functional relationship between SABATH protein family members from the respective species was considered according to the subgroups formed in the phylogenetic tree. The phylogenetic tree was constructed using BpSABATH and 52 functionally characterized members of SABATH from other species (S Table 1). A total number of 62 members of the SABATH family were used for the maximum likelihood method in MEGA 7 software (Kumar et al., 2018). A member of the SABATH family, SAMT from *Aspergillus niger* (NT166520), was used as an outgroup species. Numbers at nodes indicate bootstrap values calculated with 1,000 replicates.

The clustering patterns of the phylogenetic tree provide a preliminary understanding for predicting the functions of an unknown protein since proteins grouped in one clade showed similar functions (Kapteyn et al., 2007, Zhao et al., 2013) and the proteins might be evolved from a recent common ancestor (Xie et al., 2014).

Various members of the *MES* gene family were identified and their functional characterization has been described in numerous plants (Lima Silva et al., 2019, Vlot et al., 2008, Yang et al., 2008, Zhao et al., 2009, Zhao et al., 2016). Here, the putative role and substrate specificity of BpMES proteins were determined based on an ML phylogenetic tree (Figure 6) constructed using 23 functionally characterized MESs from other plant species, including *A. thaliana, P. trichocarpa,* and *V. vinifera*. Only *B. pendula* and functionally characterized MES protein members from the respective species were included to construct the phylogenetic tree and to predict the potential substrates of BpMES enzymes (Figure 6).

The topology of the phylogenetic tree and distribution of *B. pendula* along with functionally characterized MESs formed four functional subgroups (functional subgroups 1 to 4). All the functionally characterized MES and SABP2 genes from functional subgroup 1 were observed to catalyze the conversion of MeSA to SA which is subsequently involved in various biological functions. Two (BpMES9 and BpMES10) of the 12 BpMES were included in functional subgroup 1. Functional subgroup 2 contained seven AtMES that showed hydrolyze activity towards MeSA, PNPA, and MeIAA (Yang et al., 2008). In functional subgroup 3, all the functionally characterized MES members were involved in catalyzing MeJA and accommodated four BpMESs (BpMES4, BpMES6, BpMES7, and BpMES8), suggesting their potential functions.

B. pendula MES5, MES11, and MES12 formed a cluster in functional group 4. One of the functionally characterized AtMES17 (At3g10870) showed the highest and most specific activity

towards MeIAA, while AtMES17 (At4g16690) catalyzed MeIAA as well as PNPA and MeJA (Yang et al., 2008). Considering the clustering, BpMES5 and AtMES17 showed as orthologous genes (Figures 6) and share the same function in the plant genome. BpMES1, BpMES2, and BpMES3 were not clustered in any of the functional subgroups and gave no evidence about their putative function.

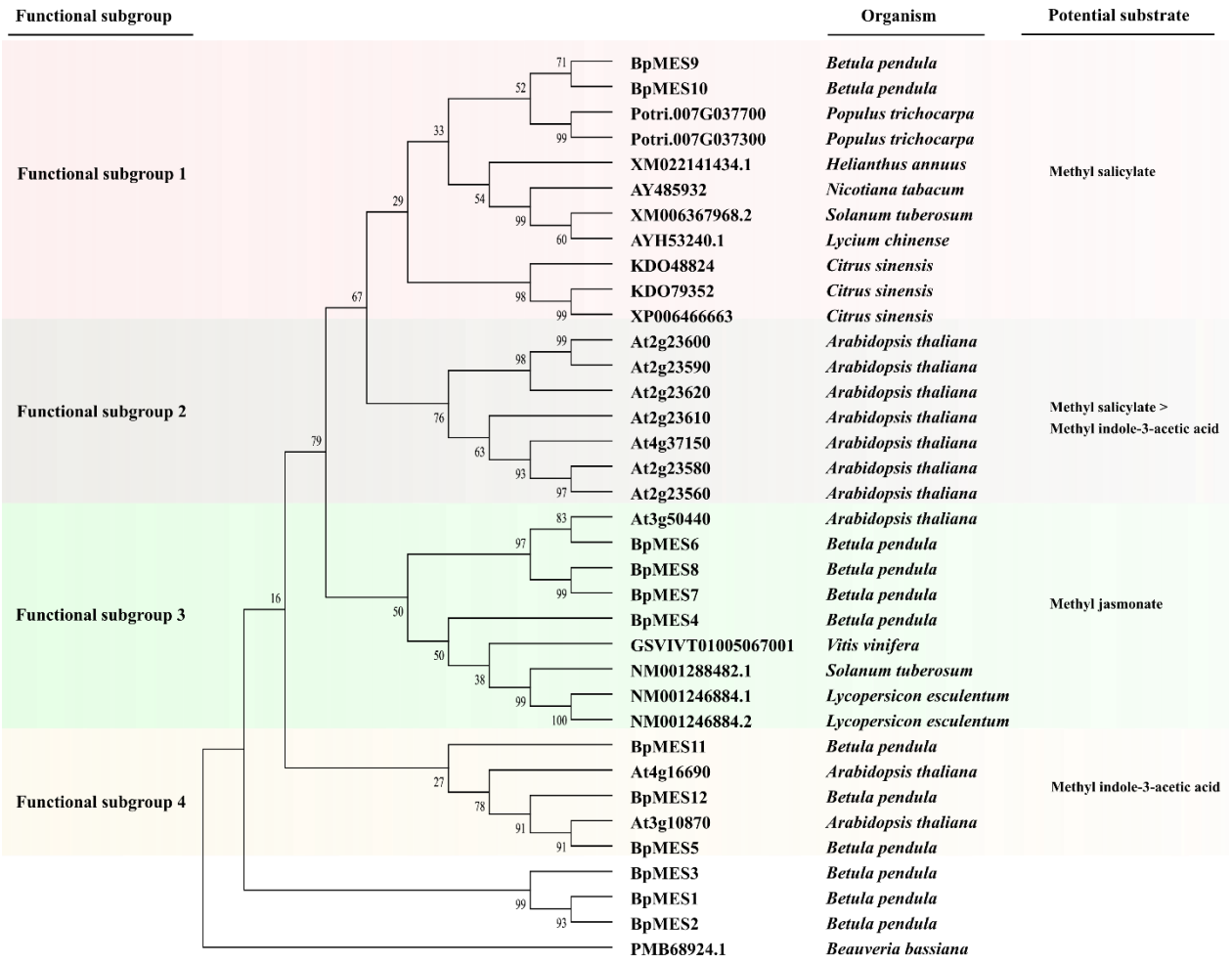

Figure 6: Potential substrates of BpMES protein family members according to phylogenetic clustering: Functionally characterized protein sequences were only included for the phylogenetic tree construction. The functional relationship between MES protein family members from the respective species was considered according to the functional subgroups formed in the phylogenetic tree. The phylogenetic tree was constructed using BpMES and 24 functionally characterized members of SABATH from other species (S Table 1). A total number of 36 members of the MES family were used to construct the maximum likelihood phylogenetic tree in MEGA 7 software (Kumar et al., 2018). A member of the MES family from *Beauveri abassiana* (PMB68924.1) was used as an outgroup species. Numbers at nodes indicate bootstrap values calculated with 1000 replicates.

In the phylogenetic tree, three groups of paralogous genes (*BpMES1, BpMES2*, and *BpMES3*; *BpMES7* and *BpMES8; BpMES9* and *BpMES10*) were identified from the MES gene family in *B. pendula* (Figure 6). Also, two pairs of orthologues, BpMES5 with AtMES17 (enzymatic activity towards MeIAA), and BpMES6 with At3g50440 (enzymatic activity towards MeJA) were identified in the MES gene family in *B. pendula* and most probably have the same function (Figure 6).

4.5.6 Promoter analysis of *BpSABATH* and *BpMES* gene family members

Promoter regions of the *BpSABATH* and *BpMES* genes were retrieved from the available *B. pendula* genome (Salojärvi et al., 2017). Retrieved promoter sequences (S Table 5) were analyzed using the PlantCARE database to identify the putative *cis*-elements (S Table 6A and 6B). The presence of different *cis*-elements along with their frequencies in *BpSABATH* and *BpMES* gene promoter regions were evaluated (Figure 7 and 8). In total, 33 *cis*-acting elements were observed in the 10 *BpSABATH* genes. The TATA and CAAT box *cis*-elements were abundant and present in all the *BpSABATH* genes. The auxin-responsive element AuxRR was present in *BpSABATH9*, while the GC motif in *BpSABATH*1 and *BpSABATH*2. The MeJA responsive elements, the CGTCA, and TGACG motifs were present only in *BpSABATH*1, *BpSABATH*2, *BpSABATH*5, *BpSABATH*6, *BpSABATH*7, and *BpSABATH*8. Only one SA responsive element, TCA, was present in *BpSABATH*1, *BpSABATH*2, *BpSABATH*3, *BpSABATH*4, *BpSABATH*9, and *BpSABATH*10.

In total, 66 different *cis*-acting elements were observed in the twelve *B. pendula* MES genes. The *cis*-elements, like the TATA and CAAT box, were considered abundant and were present in all the *Betula MES* genes. The *cis*-acting elements, the AACA motif, C box, and GCN4 motif, were present only on the promoter region of the *BpMES11* gene, while ARE box, CAT box, GCN4 motif, and RY element were present only on the promoter region of the *BpMES6* gene. Likewise, the GTGGC motif and LAMP elements were only present on the *BpMES12* promoter regions. The ACE, GAP box, LTR, TCCC element, and TC rich repeat elements were observed only on the promoter regions of *BpMES3, BpMES10, BpMES7, BpMES4*, and *BpMES5* genes, respectively.

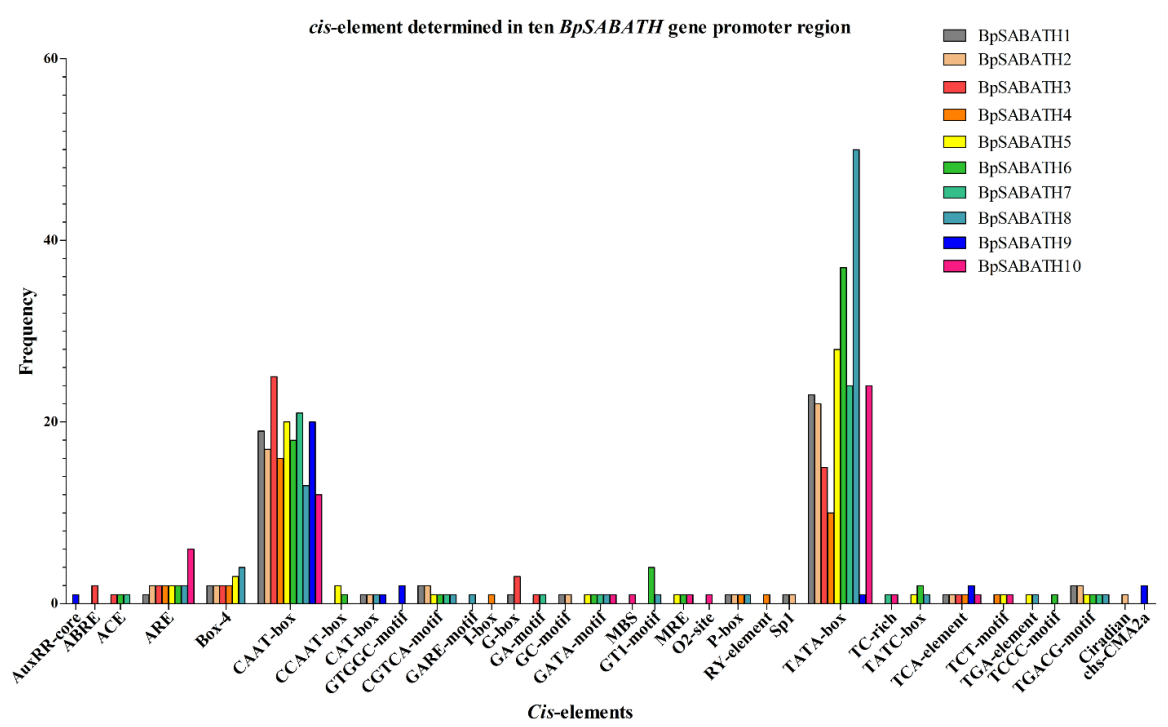

Figure 7: *BpSABATH* gene promoter analysis: Frequencies of identified *cis*-elements using the PlantCARE database (Lescot et al., 2002) in the promoter regions of 10 *BpSABATH* genes. Each *BpSABATH* is represented by a different color.

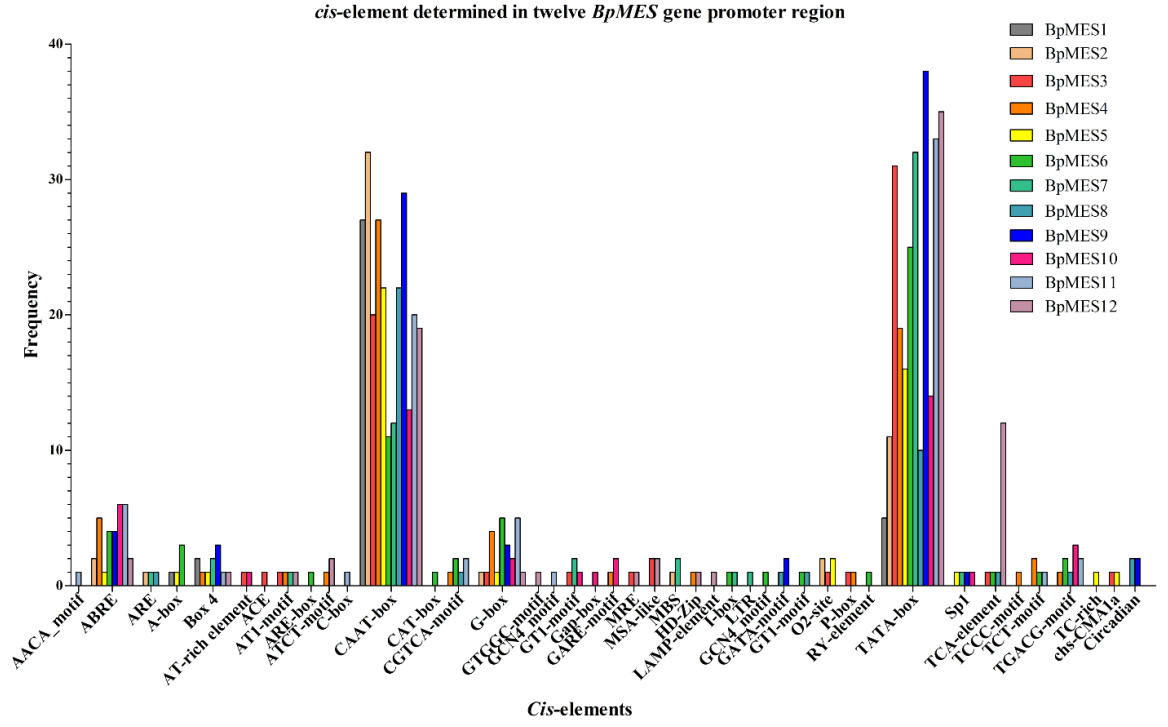

Figure 8: *BpMES* gene promoter analysis: Frequencies of identified *cis*-elements using the PlantCARE database (Lescot et al., 2002) in the promoter regions of 12 *BpMES* genes. Each *BpMES* is represented by a different color.

The MeJA responsive CGTCA motif was present on *BpMES4, BpMES6, BpMES8*, and *BpMES11*, and the TGACG motif was on the *BpMES4*, *BpMES6, BpMES8, BpMES10,* and *BpMES11* gene promoter regions, while the gibberellin-responsive GARE motif showed on *BpMES4* and *BpMES10*, and P-box on *BpMES3* and *BpMES4* (S Table 6B).

The abscisic acid-responsive ABRE was observed in *BpMES2*, *BpMES 4, BpMES5*, *BpMES6*, *BpMES9, BpMES10, BpMES11*, and *BpMES12* promoters, while the TGA auxin responsible element was on *BpMES1* and *BpMES9*. The TCA element involved in SA responses was present on *BpMES3*, *BpMES6, BpMES8*, and *BpMES12*, while TC-rich repeats involved in defense and stress responses were present only in *BpMES5*.

4.6 Discussion

Over the last few years, many plant genomes have been sequenced and this has contributed important information to plant improvement and development (Die et al., 2018). However, the genome sequencing of woody plant species is still rare (Tuskan et al., 2006, Velasco et al., 2010). In this study, we used protein sequences of SABATH and MES and performed BLAST analyses against the *B. pendula* genome to extract candidate genes belonging to both families (Salojärvi et al., 2017).

To the best of our knowledge, following *A. thaliana* and *V. vinifera*, *B. pendula* is the third plant species, and the first tree species, in which the complete *MES* gene family has been recognized and comprehensively studied. Although the genome size of *A. thaliana* (~135 Mbp) is, compared to *V. vinifera* (~500 Mbp) and *B. pendula* (~440 Mbp), very small, it contains 20 members of the *MES* gene family (Yang et al., 2008), while *V. vinifera* and *B. pendula* only contain 15 and 12 members of *MES*, respectively (Zhao et al., 2016). Compared to the *MES* gene family, the *SABATH* gene family has been studied intensively in many plant species (Ament et al., 2010, Chaiprasongsuk et al., 2018, Chen et al., 2003, D'Auria et al., 2003, Han et al., 2017). The highest and lowest number of *SABATH* gene family members have been recorded in the *O. sativa* (41) and *P. abies* (10), respectively (Chaiprasongsuk et al., 2018, Zhao et al., 2008).

4.6.1 Comparative bioinformatics analysis

Plant SABATHs catalyze the methylation of numerous secondary metabolites, play an important role in different biological mechanisms, including stress responses, development, and growth (Ament et al., 2010, Chen et al., 2003, Effmert et al., 2005).

The *B. pendula* 10 BpSABATH proteins were divided into two clades in the intraspecific SABATH members (Figure 2A). The BpMES proteins were divided into three distinct clades (Figure 2A). All the *BpSABATH* genes displayed a methyltransferase 7 domain (Methyltransf 7; S Table 7) and a conserved motif III that possessed SAM-binding sites (Joshi et al., 1998). The occurrence of motif III in 56 different plant species suggests it plays a major role in the binding of SAM-dependent *O*-methyltransferases to their specific substrate (Joshi et al., 1998, Vidgren et al., 1994). The crystallography analysis of the CbSAMT enzyme and the SA binding residues were also characterized in the BpSABATHs (Zubieta et al., 2003) (Figure 3). A total number of 3, 5, and 1 SAM/SAH binding residues were present in BpSABATH1, 2, and 9, while BpSABATH4-8 and 10 carried all six compared to the template CbSAMT amino acid sequence.

The BpSABATH enzyme alignment also revealed the presence of aromatic-rich residues of the carboxyl-bearing substrate-binding pockets that were previously observed in the detailed study of *A. thaliana* indole-3-acetic acid methyltransferase (AtIAMT) (Figure 3).

Except for BpMES2, all amino acid sequence alignments of BpMES enzyme family members displayed the presence of the catalytic α/β hydrolase domain (α/β hydrolase; S Table 7) conserved in the MES family, which is in agreement with *A. thaliana* MES enzymes (Yang et al., 2008). The signature amino acid sequence, conserved in the NtSABP2 proteins (Forouhar et al., 2005), was identified in all 12 BpMES members except for BpMES2 at positions 79-83 aa (Figure 4). Except for BpMES1-2 and BpMES10-11, all the BpMES enzymes displayed the conserved catalytic triad found in the hydrolase domain identified through the protein profiling of NtSABP2 protein (Kumar et al., 2003). The conserved catalytic triad was also observed in *A. thaliana* and *V. vinifera* MES enzymes (Yang et al., 2008, Zhao et al., 2016). Moreover, the observed 14 residues in BpMES enzymes that bind to SA were consistent with the previous structural study of NtSABP2 (Forouhar et al., 2005). The BpMES family members encode proteins sizing between 121 (BpMES8) and 258 (BpMES1) amino acids (S Table 2), unlike NtSABP2, LeMJE, and most of the AtMES proteins (Kumar et al., 2003, Stuhlfelder et al., 2004, Yang et al., 2008). BpMES1, BpMES2, BpMES3, and BpMES11 carry some extra 25–43 amino acids at the N- terminal (Figure 4). The extended N terminal may not contain a directing signal peptide, suggesting that these BpMES enzymes are situated in the cytosol, similar to other members of the family (Yang et al., 2008).

4.6.2 Substrate specificity of BpSABATH and BpMES members

Diverse *SABATH* and *MES* genes in numerous plants have been functionally described (Ament et al., 2010, Chaiprasongsuk et al., 2018, Chen et al., 2003, Han et al., 2017, Köllner et al., 2010). The members of a family from different species having similar functions will most probably be clustered together in the phylogenetic tree, suggesting the genes in the same clade might share a similar origin. It can be postulated that the SABATH and MES protein family members from different plant species with a higher similarity might have the same feature and function (Xie et al., 2014). Therefore, we can deduce the function of the new members of the SABATH and MES protein families in *B. pendula* according to their clustering pattern.

SA is one of the key molecules in several plants that are involved in plant development and many other mechanisms, including defense against various pathogens. When SA is methylated by one

of the SABATH enzymes to form MeSA, the volatile ester, it functions as a signaling molecule after the plant experiences an infection (Park et al., 2007, Vlot et al., 2008, Zubieta et al., 2003). In *P. trichocarpa*, four SABATHs showed enzymatic activity towards SA, and PtSABATH4 displayed the highest activity towards SA. Han et al., (2017) suggested that PtSABATH4 might be the only carrier of this salicylic acid carboxyl methyltransferase activity (Han et al., 2017). Notably, the SABATHs that utilized SA and/or BA above all other tested substrates formed one clade (Figure 5), confirming the potential function of BpSABATH9, since it is the only member of the BpSABATH clustered together. In *A. thaliana*, SABATH8 (At3g11480) is one of the members that converted SA to MeSA (Chen et al., 2003). It is unlikely that At3g11480 and the *O. sativa* SABATH3 (LOCs02g48770) would not cluster with the SABATHs that catalyze SAMT in flowering plants (Figure 5). The uneven distribution of SABATH enzymes that methylate SA supports the occurrence of multiple independent evolutionary events (Chaiprasongsuk et al., 2018) and displays the paraphyletic status of the SAMT gene in seed plants.

Studies have shown that *A. thaliana* IAMT (At5g55250) converts IAA into non-polar methyl-IAA (Qin et al., 2005). Of 28 PtSABATH enzymes in *P. trichocarpa*, six (PtSABATH2, 3, 12, 17, 21, and 24) had very low enzymatic activity towards IAA, while PtSABATH1 showed a 40.5-fold higher activity (Han et al., 2017). One *P. glauca* (PgSABATH1) (Zhao et al., 2009) and *P. abies* (PaSABATH1: PaIAMT) (Chaiprasongsuk et al., 2018) enzyme showed the highest enzymatic activity towards IAA, grouped in the functional subgroup 7 (Figure 5). Similarly, among 23 *O. sativa* SABATHs, only one OsSABATH4 (LOCs04g56950) gene showed catalytic activity towards IAA (Zhao et al., 2008). Likewise, a single BpSABATH (BpSABATH8) enzyme accommodated in the functional subgroup 7 (Figure 5), along with all the functionally characterized SABATHs from the different respective species, preferred IAA as a most favorable substrate and might actively take part in auxin homeostasis. The homology modeling also suggested a structural similarity to the protein accession, 3b5i (Indole-3-acetic Acid Methyltransferase) (S. Table 4). Previous studies suggested that the IAMT gene in the *P. abies*, *A. thaliana*, *O. sativa*, *P. trichocarpa*, and *B. pendula* are conserved (Chaiprasongsuk et al., 2018, Qin et al., 2005, Zhao et al., 2008) and might have evolved from a common ancestor of plants that produced seeds. The presented analysis supports the conclusion that IAMT might be the earliest member of SABATH family since a monophyletic clade has formed by the IAMTs from

different plant species, including *A. thaliana*, *O. sativa*, *P. trichocarpa*, and *P. glauca* (Zhao et al., 2009), which is consistent with our study (Figure 5).

Several studies have concluded that MeJA is involved in many diverse mechanisms, including defense, flowering, and seed germination (Cheong et al., 2003). However, very little is known about the function of MeJA in woody plants, other than that it promotes the synthesis of traumatic resin ducts in *P. abies* (Martin et al., 2003). The biochemical analysis of SABATH in *P. abies* revealed that three enzymes select JA as a favorable substrate (Chaiprasongsuk et al., 2018), while in *P. trichocarpa*, nine SABATHs showed enzymatic activity towards JA (Han et al., 2017). The involvement of multiple SABATHs in preferring JA was consistent since multiple (four of ten) BpSABATHs clustered together with the *P. trichocarpa* JAMT (Figure 5). This observation is in contrast to *A. thaliana*, where only one member of AtSABATH (At1G19640) was identified as having catalytic activity towards JA (Seo et al., 2001). At1G19640 and PtSABATH3 displayed the highest level of enzymatic activity towards JA and clustered together with four BpSABATHs (BpSABATH4, BpSABATH5, BpSABATH6, and BpSABATH7), while three *P. abies* JAMTs (PaJAMT1, PaJAMT2, and PaJAMT3) clustered in a different clade (Figure 5). Although the three PaJAMTs are most similar to each other, they are the result of the latest gene duplication and displayed a divergence in biochemical properties that indicates a functional divergence after gene duplication (Chaiprasongsuk et al., 2018). BpSABATH1 and BpSABATH3 were clustered with six *P. trichocarpa* SABATHs preferring JA as well as FA as a substrate. It has been shown that members of the SABATH enzyme family utilize multiple substrates at different levels of enzymatic activity (Han et al., 2017), suggesting their multifunctional ability to survive under diverse stress conditions.

The esterase activities of fifteen Arabidopsis MES enzymes with four methyl esters of phytohormones (IAA, SA, GA, and JA) were tested. Eleven *A. thaliana* MESs showed esterase activity with at least one of the three substrates IAA, SA, and JA, while no enzyme showed activity towards GA (Yang et al., 2008). The biochemical analysis suggests that the demethylation of methyl esters of phytohormones might play an important role in modulating the different biological functions (Westfall et al., 2013).

AtMES17 (At3g10870) was hypothesized to encode for MeIAA esterase since amino acid sequences were firmly related to LeMJE and NtSABP2. Additionally, it has been confirmed in vitro through biochemical analyses that AtMES17 encodes an esterase which specifically

hydrolyzes MeIAA to IAA, and it is presumed that this will also occur *in vivo* (Yang et al., 2008). The biochemical functional analysis suggests that AtMES17 is responsible for auxin homeostasis since MeIAA could transport through membranes to neighboring cells where it could be hydrolyzed back to the active auxin IAA by one of the MES members. The phylogenetic tree showed the orthologous relationship between AtMES17 and BpMES5, since they formed a close clade, suggesting that both encode an enzyme with a similar function (Figure 6).

BpMES4, BpMES6, BpMES7, and BpMES8 were clustered together with the functionally characterized MES enzymes (Figure 6) that more specifically hydrolyze MeJA than other tested substrates (Stuhlfelder et al., 2004, Zhao et al., 2016). The clustering of multiple BpMESs with LeMJE (NM001246884.1 and NM001246884.2) was similar to the VvMES study where three VvMES formed a clade with LeMJE (Zhao et al., 2016). In the previous analysis, evidence of an orthologous relationship between VvMES5 (GSVIVT01005067001: VvMJE1), LeMJE, and AtMJE (At3g50440) was not found, which is in contrast to this study. Here, BpMES6 formed a cluster with AtMJE, suggesting proof of an orthologous relationship (Figure 6). The orthologous relationship is contrary to the previous analysis where the *V. vinifera, S. lycopersicum*, and *A. thaliana* MJE displayed different lineages, supporting the hypothesis of independent evolution (Zhao et al., 2016). VvMJE1 from grapevine utilizes MeJA as a substrate since it showed a very high esterase activity towards MeJA in biochemical analyses (Zhao et al., 2016). Considering the phylogenetic clustering of BpMES4, BpMES6, BpMES7, and BpMES8 with previously biochemically characterized members of the MES family, we can predict that these BpMES also hydrolyze MeJA esters. Further biochemical analysis will be required to confirm which of these shows significant enzymatic activity towards MeJA.

BpMES9 and BpMES10 were the only two members clustered together with two functional methyl esterase *SABP2* genes from the *P. trichocarpa* (Zhao et al., 2009) (Figure 6). Further, members of MES, including *S. tuberosum, Lycium chinense*, *Helianthus annuus,* and *N. tobacum* (Kumar et al., 2003) from functional group 1, specifically utilized MeSA as substrate. Interestingly, the two MeSA esterases from *P. trichocarpa* possess similar functions, however, a likely difference in biological roles was observed (Zhao et al., 2009). Competitively, *B. pendula* have not been through intense evolutionary genome duplication events, possibly resulting in only one copy of *SABP2* (Salojärvi et al., 2017, Zhao et al., 2009). The amino acid sequence of BpMES10 showed one mismatch at the conserved catalytic triad in the multiple sequence

alignment, while BpMES9 carries all three conserved residues (Figure 4). Thus, we could speculate that BpMES9 is the most trusted candidate in catalyzing the demethylation of MeSA. The hypothesis also supports the previous study where BpMES9/BpSABP2 was comprehensively studied in different *Betula* species and was proposed as a candidate gene to revert MeSA to SA (Singewar et al., 2020a, Singewar et al., 2020b).

Substrate specificity of MES has been always a point of curiosity that has led to the biochemical analysis of different MESs with various possible substrates (Kumar et al., 2003, Yang et al., 2008, Zhao et al., 2009, Zhao et al., 2016, Zhao et al., 2013). This study has shown that, in addition to MeJA, the VvMJE1 also prefers MeSA, but only at a high concentration of the substrate (Zhao et al., 2016). Similarly, the MeSA esterase from both *N. tabacum* and *P. trichocarpa* showed the highest activity towards MeSA. The MeSA ester was also enzymatically active towards MeIAA and MeJA at only very high concentrations of substrate, which was considered to be physiologically immaterial (Forouhar et al., 2005, Zhao et al., 2009). Further, the AtMES enzymes showed enzymatic activity towards multiple substrates under the experimental setups, suggesting that possibly all MES enzymes use more than one substrate (Yang et al., 2008). These initial studies encourage a detailed examination using diverse MES members from different species, which will be beneficial for recognizing the evolution behind substrate specificity among the MES plant family.

4.7 Conclusion

To the best of our knowledge, this is the first study that has attempted to gather information about SABATH and MES family members at gene and enzyme levels in *B. pendula*. The vital comparative bioinformatics analysis revealed discrete patterns present in the SABATH and MES family members, involved in the biosynthesis of many hormones, signaling molecules, and floral scent. The identification and extensive *in silico* analysis of the BpSABATH and BpMES genes and enzymes revealed their putative functions and substrate specificities. The vital information will be an asset for further functional and enzymatic substrate specificity, respectively. Additionally, the advantageous information of candidate genes could be exploited for genetic modification or targeted mutagenesis and genotype building to decide the function of a gene. Further, trait-specific markers would be designed to breed birch varieties that are adapted to certain environments.

4.8 Acknowledgments

We are extremely grateful to PD Dr. Birgit Kersten for assisting with the Data retrieval, to Dr. Niels Müller for consultation on the manuscript review (both Thünen-Institute of Forest Genetics, Grosshansdorf, Germany), and to all the staff of the Thünen-Institute of Forest Genetics, Grosshansdorf, Germany, for their support, including Katrin Groppe, Susanne Jelkmann, and Vivian Kuhlenkamp for technical assistance, and Dr. Hilke Schröder and Dr. Hans Hönicka for helpful discussions on the methylation and demethylation process, and Thanks are also due to Ana Montalvao and Avneesh Kumar (Plant Breeding Institute, University of Kiel, Germany) for assistance in R programing language. Many thanks to the staff from the Institute for Agricultural Process Engineering, Christian-Albrechts University of Kiel, Germany. Thanks are due to Nicola Wilton (Language Boutique®, Hamburg, Germany) and Dina Führmann (Thünen Institute, Braunschweig, Germany) for English language editing.

4.9 Supplementary material

The supplemental data for this section in the Microsoft word document format is available on the computer hard drive and can be distributed upon request (contact: Dr. Christian R. Moschner, cmoschner@ilv.uni-kiel.de).

4.10 References

Alvarez MV, MdR Moreira, SI Roura, JF Ayala-Zavala and GA González-Aguilar (2015) 13 - Using natural antimicrobials to enhance the safety and quality of fresh and processed fruits and vegetables: Types of antimicrobials. In: Handbook of Natural Antimicrobials for Food Safety and Quality. Taylor TM (ed) Oxford: Woodhead Publishing, pp 287-313

Ament K, V Krasikov, S Allmann, M Rep, FLW Takken and RC Schuurink (2010) Methyl salicylate production in tomato affects biotic interactions. The Plant Journal 62(1):124-134. https://dx.doi.org/10.1111/j.1365-313X.2010.04132.x

Araminienė V and I Varnagiryte-Kabasinskiene (2014) Research on birch species in Lithuania: A review study. In: Research for Rural Development 2. pp 50-56.

Ashburner K, HA McAllister, J Hague, A Brown, P Williams, M Williams and M Rix (2013) The Genus Betula: A Taxonomic Revision of Birches. Royal Botanic Gardens, London: Kew Publishing

Aspelmeier S and C Leuschner (2004) Genotypic variation in drought response of silver birch (*Betula pendula*): leaf water status and carbon gain. Tree Physiology 24(5):517-528. https://dx.doi.org/10.1093/treephys/24.5.517

Atkinson MD (1992) *Betula Pendula* Roth (*B. Verrucosa* Ehrh.) and *B. Pubescens* Ehrh. Journal of Ecology 80(4):837-870. https://dx.doi.org/10.2307/2260870

Chaiprasongsuk M, C Zhang, P Qian, X Chen, G Li, RN Trigiano, H Guo and F Chen (2018) Biochemical characterization in Norway spruce (*Picea abies*) of SABATH methyltransferases that methylate phytohormones. Phytochemistry 149:146-154. https://dx.doi.org/https://doi.org/10.1016/j.phytochem.2018.02.010

Chen F, JC D'Auria, D Tholl, JR Ross, J Gershenzon, JP Noel and E Pichersky (2003) An *Arabidopsis thaliana* gene for methylsalicylate biosynthesis, identified by a biochemical genomics approach, has a role in defense. The Plant Journal 36(5):577-588. https://dx.doi.org/10.1046/j.1365-313X.2003.01902.x

Cheong J-J and YD Choi (2003) Methyl jasmonate as a vital substance in plants. Trends in Genetics 19(7):409-413. https://dx.doi.org/https://doi.org/10.1016/S0168-9525(03)00138-0

D'Auria JC, F Chen and E Pichersky (2003) Chapter eleven The SABATH family of MTS in *Arabidopsis thaliana* and other plant species. In: Recent Advances in Phytochemistry. Romeo JT (ed): Elsevier, pp 253-283

Delker C, M Raschke A Fau - Quint and M Quint (2008) Auxin dynamics: the dazzling complexity of a small molecule's message. Planta 227, 929–941 (2008). https://doi.org/10.1007/s00425-008-0710-8.

Die JV, J Gil and T Millan (2018) Genome-wide identification of the auxin response factor gene family in Cicer arietinum. BMC Genomics 19(1):301. https://dx.doi.org/10.1186/s12864-018-4695-9

Dogru E, F Warzecha H Fau - Seibel, S Seibel F Fau - Haebel, F Haebel S Fau - Lottspeich, J Lottspeich F Fau - Stöckigt and J Stöckigt (2000) The gene encoding polyneuridine aldehyde esterase of monoterpenoid indole alkaloid biosynthesis in plants is an ortholog of the alpha/betahydrolase super family. 267(5):1397-406. doi: 10.1046/j.1432-1327.2000.01136.x. PMID: 10691977.

Dubois H, E Verkasalo and H Claessens (2020) Potential of Birch (*Betula pendula Roth* and *B. pubescens Ehrh*.) for Forestry and Forest-Based Industry Sector within the Changing Climatic and Socio-Economic Context of Western Europe. Forests 11(3):336. https://dx.doi.org/http://dx.doi.org/10.3390/f11030336

Dudareva N, LM Murfitt, CJ Mann, N Gorenstein, N Kolosova, CM Kish, C Bonham and K Wood (2000) Developmental Regulation of Methyl Benzoate Biosynthesis and Emission in Snapdragon Flowers. The Plant Cell 12(6):949. https://dx.doi.org/10.1105/tpc.12.6.949

Effmert U, S Saschenbrecker, J Ross, F Negre, CM Fraser, JP Noel, N Dudareva and B Piechulla (2005) Floral benzenoid carboxyl methyltransferases: From in vitro to in planta function. Phytochemistry 66(11):1211-1230. https://dx.doi.org/https://doi.org/10.1016/j.phytochem.2005.03.031

El-Gebali S, J Mistry, A Bateman, SR Eddy, A Luciani, SC Potter, M Qureshi et al., (2018) The Pfam protein families database in 2019. Nucleic Acids Res 47(D1):D427-D432. https://dx.doi.org/10.1093/nar/gky995

Fischer A, M Lindner, C Abs and P Lasch (2002) Vegetation dynamics in central european forest ecosystems (near-natural as well as managed) after storm events. Folia Geobotanica 37(1):17-32. https://dx.doi.org/10.1007/BF02803188

Forouhar F, IS Lee, J Vujcic, S Vujcic, J Shen, SM Vorobiev, R Xiao, TB Acton, GT Montelione, CW Porter and L Tong (2005) Structural and functional evidence for *Bacillus subtilis PaiA* as a novel N1-spermidine/spermine acetyltransferase. J Biol Chem 280(48):40328-40336. https://dx.doi.org/10.1074/jbc.M505332200

Gang H, R Li, Y Zhao, G Liu, S Chen and J Jiang (2019) Loss of GLK1 transcription factor function reveals new insights in chlorophyll biosynthesis and chloroplast development. Journal of Experimental Botany 70(12):3125-3138. https://dx.doi.org/10.1093/jxb/erz128

Gel B and E Serra (2017) karyoploteR: an R/Bioconductor package to plot customizable linear genomes displaying arbitrary data. bioRxiv:122838. https://dx.doi.org/10.1101/122838

Gilbert W (1987) The exon theory of genes. 52:901-5. doi: 10.1101/sqb.1987.052.01.098. PMID: 2456887.

Han X-M, Q Yang, Y-J Liu, Z-L Yang, X-R Wang, Q-Y Zeng and H-L Yang (2017) Evolution and Function of the Populus SABATH Family Reveal That a Single Amino Acid Change Results in a Substrate Switch. Plant and Cell Physiology 59(2):392-403. https://dx.doi.org/10.1093/pcp/pcx198

Hemery GE, JR Clark, E Aldinger, H Claessens, ME Malvolti, E O'Connor, Y Raftoyannis, PS Savill and R Brus (2010) Growing scattered broadleaved tree species in Europe in a changing climate: a review of risks and opportunities. Forestry: An International Journal of Forest Research 83(1):65-81. https://dx.doi.org/10.1093/forestry/cpp034

Holmquist M (2000) Alpha Beta-Hydrolase Fold Enzymes Structures, Functions and Mechanisms. Current Protein and Peptide Science 1(2):209-235. https://dx.doi.org/10.2174/1389203003381405

Hynynen J, P Niemistö, A Viherä-Aarnio, A Brunner, S Hein and P Velling (2009) Silviculture of birch (*Betula pendula* Roth and *Betula pubescens* Ehrh.) in northern Europe. Forestry: An International Journal of Forest Research 83(1):103-119. https://dx.doi.org/10.1093/forestry/cpp035

Joshi CP and VL Chiang (1998) Conserved sequence motifs in plant S-adenosyl-L-methionine-dependent methyltransferases. Plant Molecular Biology 37(4):663-674. https://dx.doi.org/10.1023/A:1006035210889

Kapteyn J, AV Qualley, Z Xie, E Fridman, N Dudareva and DR Gang (2007) Evolution of Cinnamate/p-coumarate carboxyl methyltransferases and their role in the biosynthesis of methylcinnamate. The Plant cell 19(10):3212-3229. https://dx.doi.org/10.1105/tpc.107.054155

Köllner TG, C Lenk, N Zhao, I Seidl-Adams, J Gershenzon, F Chen and J Degenhardt (2010) Herbivore-induced SABATH methyltransferases of maize that methylate anthranilic acid using s-adenosyl-L-methionine. Plant physiology 153(4):1795-1807. https://dx.doi.org/10.1104/pp.110.158360

Kong H, LL Landherr, MW Frohlich, J Leebens-Mack, H Ma and CW DePamphilis (2007) Patterns of gene duplication in the plant SKP1 gene family in angiosperms: evidence for multiple mechanisms of rapid gene birth. The Plant Journal 50(5):873-885. https://dx.doi.org/10.1111/j.1365-313X.2007.03097.x

Koonin EV (2006) The origin of introns and their role in eukaryogenesis: a compromise solution to the introns-early versus introns-late debate? Biol Direct 1:22-22. https://dx.doi.org/10.1186/1745-6150-1-22

Koski V and M Rousi (2005) A review of the promises and constraints of breeding silver birch (*Betula pendula* Roth) in Finland. Forestry: An International Journal of Forest Research 78(2):187-198. https://dx.doi.org/10.1093/forestry/cpi017

Kumar D and F Daniel, Klessig (2003) High-affinity salicylic acid-binding protein 2 is required for plant innate immunity and has salicylic acid-stimulated lipase activity. Proceedings of the National Academy of Sciences 100, 16101–16106. https://dx.doi.org/doi.org/10.1073/pnas.0307162100

Kumar D and DF Klessig (2003) High-affinity salicylic acid-binding protein 2 is required for plant innate immunity and has salicylic acid-stimulated lipase activity. 100 (26) 16101-16106; DOI: 10.1073/pnas.0307162100

Kumar S, G Stecher, M Li, C Knyaz and K Tamura (2018) MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. Molecular Biology and Evolution 35(6):1547-1549. https://dx.doi.org/10.1093/molbev/msy096

Lescot M, P Déhais, G Thijs, K Marchal, Y Moreau, Y Van de Peer, P Rouzé and S Rombauts (2002) PlantCARE, a database of plant *cis*-acting regulatory elements and a portal to tools for *in silico* analysis of promoter sequences. Nucleic Acids Research 30(1):325-327. https://dx.doi.org/10.1093/nar/30.1.325

Lima Silva CCd, HM Shimo, R de Felício, GF Mercaldi, SA Rocco and CE Benedetti (2019) Structure-function relationship of a citrus salicylate methylesterase and role of salicylic acid in citrus canker resistance. Scientific Reports 9(1):3901. https://dx.doi.org/10.1038/s41598-019-40552-3

Martin DM, J Gershenzon and J Bohlmann (2003) Induction of Volatile Terpene Biosynthesis and Diurnal Emission by Methyl Jasmonate in Foliage of Norway Spruce. Plant Physiology 132(3):1586. https://dx.doi.org/10.1104/pp.103.021196

Moore RC and MD Purugganan (2003) The early stages of duplicate gene evolution. Proceedings of the National Academy of Sciences 100(26):15682. https://dx.doi.org/10.1073/pnas.2535513100

Nardini M and BW Dijkstra (1999) α/β Hydrolase fold enzymes: the family keeps growing. Current Opinion in Structural Biology 9(6):732-737. https://dx.doi.org/https://doi.org/10.1016/S0959-440X(99)00037-8

Park S-W, E Kaimoyo, D Kumar, S Mosher and DF Klessig (2007) Methyl Salicylate Is a Critical Mobile Signal for Plant Systemic Acquired Resistance. Science 318(5847):113. https://dx.doi.org/10.1126/science.1147113

Patthy L (1987) Intron-dependent evolution: Preferred types of exons and introns. FEBS Letters 214(1):1-7. https://dx.doi.org/10.1016/0014-5793(87)80002-9

Prévosto B and T Curt (2004) Dimensional relationships of naturally established European beech trees beneath Scots pine and Silver birch canopy. Forest Ecology and Management 194(1):335-348. https://dx.doi.org/https://doi.org/10.1016/j.foreco.2004.02.020

Qin G, H Gu, Y Zhao, Z Ma, G Shi, Y Yang, E Pichersky, H Chen, M Liu, Z Chen and L-J Qu (2005) An indole-3-acetic acid carboxyl methyltransferase regulates Arabidopsis leaf development. The Plant cell 17(10):2693-2704. https://dx.doi.org/10.1105/tpc.105.034959

Ranta H, T Hokkanen, T Linkosalo, L Laukkanen, K Bondestam and A Oksanen (2008) Male flowering of birch: Spatial synchronization, year-to-year variation and relation of catkin numbers and airborne pollen counts. Forest Ecology and Management 255(3):643-650. https://dx.doi.org/https://doi.org/10.1016/j.foreco.2007.09.040

Rosenvald K, A Tullus, I Ostonen, V Uri, P Kupper, J Aosaar, M Varik, J Sõber, A Niglas, R Hansen, G Rohula, M Kukk, A Sõber and K Lõhmus (2014) The effect of elevated air humidity on young silver birch and hybrid aspen biomass allocation and accumulation – Acclimation mechanisms and capacity. Forest Ecology and Management 330:252-260. https://dx.doi.org/https://doi.org/10.1016/j.foreco.2014.07.016

Ross JR, KH Nam, JC D'Auria and E Pichersky (1999) S-Adenosyl-l-Methionine:Salicylic Acid Carboxyl Methyltransferase, an Enzyme Involved in Floral Scent Production and Plant Defense, Represents a New Class of Plant Methyltransferases. Archives of Biochemistry and Biophysics 367(1):9-16. https://dx.doi.org/https://doi.org/10.1006/abbi.1999.1255

Salojärvi J, O-P Smolander, K Nieminen, S Rajaraman, O Safronov, P Safdari et al., (2017) Genome sequencing and population genomic analyses provide insights into the adaptive landscape of silver birch. Nature Genetics 49(6):904-912. https://dx.doi.org/10.1038/ng.3862

Seo HS, JT Song, JJ Cheong, YH Lee, YW Lee, I Hwang, JS Lee and YD Choi (2001) Jasmonic acid carboxyl methyltransferase: a key enzyme for jasmonate-regulated plant responses. Proc Natl Acad Sci U S A 98(8):4788-4793. https://dx.doi.org/10.1073/pnas.081557298

Singewar K, CR Moschner, E Hartung and M Fladung (2020a) Identification and analysis of key genes involved in methyl salicylate biosynthesis in different birch species. PLOS ONE 15(10):e0240246. https://dx.doi.org/10.1371/journal.pone.0240246

Singewar, K., Moschner, C.R., Hartung, E., Fladung M. (2020b) 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. https://doi.org/10.1007/s00468-020-01984-x

Stuhlfelder C, MJ Mueller and H Warzecha (2004) Cloning and expression of a tomato cDNA encoding a methyl jasmonate cleaving esterase. European Journal of Biochemistry 271(14):2976-2983. https://dx.doi.org/10.1111/j.1432-1033.2004.04227.x

Teale WD, K Paponov Ia Fau - Palme and K Palme (2006) Auxin in action: signalling, transport and the control of plant growth and development. Nat Rev Mol Cell Biol 7, 847–859 (2006). https://doi.org/10.1038/nrm2020

Thompson JD, DG Higgins and TJ Gibson (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic acids research 22(22):4673-4680. https://dx.doi.org/10.1093/nar/22.22.4673

Tuskan GA, S DiFazio, S Jansson, J Bohlmann, I Grigoriev, U Hellsten, N Putnam et al., (2006) The Genome of Black Cottonwood, Populus trichocarpa (Torr. & Gray). Science 313(5793):1596. https://dx.doi.org/10.1126/science.1128691

Velasco R, A Zharkikh, J Affourtit, A Dhingra, A Cestaro, A Kalyanaraman, P Fontana et al., (2010) The genome of the domesticated apple (Malus × domestica Borkh.). Nature Genetics 42(10):833-839. https://dx.doi.org/10.1038/ng.654

Vidgren J, LA Svensson and A Liljas (1994) Crystal structure of catechol *O*-methyltransferase. Nature 368(6469):354-358. https://dx.doi.org/10.1038/368354a0

Vlot AC, P-P Liu, RK Cameron, S-W Park, Y Yang, D Kumar, F Zhou et al., (2008) Identification of likely orthologs of tobacco salicylic acid-binding protein 2 and their role in systemic acquired resistance in Arabidopsis thaliana. The Plant Journal 56(3):445-456. https://dx.doi.org/10.1111/j.1365-313X.2008.03618.x

Westfall CS, AM Muehler and JM Jez (2013) Enzyme action in the regulation of plant hormone responses. The Journal of biological chemistry 288(27):19304-19311. https://dx.doi.org/10.1074/jbc.R113.475160

Xie R, Y Li, S He, Y Zheng, S Yi, Q Lv and L Deng (2014) Genome-Wide Analysis of Citrus R2R3MYB Genes and Their Spatiotemporal Expression under Stresses and Hormone Treatments. PLOS ONE 9(12):e113971. https://dx.doi.org/10.1371/journal.pone.0113971

Yang Y, R Xu, C-j Ma, AC Vlot, DF Klessig and E Pichersky (2008) nactive Methyl Indole-3-Acetic Acid Ester Can Be Hydrolyzed and Activated by Several Esterases Belonging to the AtMES Esterase Family of Arabidopsis. Plant Physiology 147(3):1034. https://dx.doi.org/10.1104/pp.108.118224

Zhao N, B Boyle, I Duval, J-L Ferrer, H Lin, A Seguin, J Mackay and F Chen (2009) SABATH methyltransferases from white spruce (*Picea glauca*): gene cloning, functional characterization and structural analysis. Tree Physiology 29(7):947-957. https://dx.doi.org/10.1093/treephys/tpp023

Zhao N, J-L Ferrer, J Ross, J Guan, Y Yang, E Pichersky, JP Noel and F Chen (2008) Structural, Biochemical, and Phylogenetic Analyses Suggest That Indole-3-Acetic Acid Methyltransferase Is an Evolutionarily Ancient Member of the SABATH Family. Plant Physiology 146(2):455. https://dx.doi.org/10.1104/pp.107.110049

Zhao N, J Guan, F Forouhar, TJ Tschaplinski, Z-M Cheng, L Tong and F Chen (2009) Two poplar methyl salicylate esterases display comparable biochemical properties but divergent expression patterns. Phytochemistry 70(1):32-39. https://dx.doi.org/https://doi.org/10.1016/j.phytochem.2008.11.014

Zhao N, H Lin, S Lan, Q Jia, X Chen, H Guo and F Chen (2016) VvMJE1 of the grapevine (*Vitis vinifera*) VvMES methylesterase family encodes for methyl jasmonate esterase and has a role in stress response. Plant Physiology and Biochemistry 102:125-132. https://dx.doi.org/https://doi.org/10.1016/j.plaphy.2016.02.027

Zhao N, J Yao, M Chaiprasongsuk, G Li, J Guan, TJ Tschaplinski, H Guo and F Chen (2013) Molecular and biochemical characterization of the jasmonic acid methyltransferase gene from black cottonwood (*Populus trichocarpa*). Phytochemistry 94:74-81. https://dx.doi.org/https://doi.org/10.1016/j.phytochem.2013.06.014

Zubieta C, JR Ross, P Koscheski, Y Yang, E Pichersky and JP Noel (2003) Structural Basis for Substrate Recognition in the Salicylic Acid Carboxyl Methyltransferase Family. The Plant Cell 15(8):1704. https://dx.doi.org/10.1105/tpc.014548

5 General discussion

The thesis was aimed to (1) determine the phylogenetic relationship and apply multiple chloroplast and genomic regions to ascertain genetic variations among high and low MeSA producing birches; (2) identify the markers and analyze the expression of the two prominent candidate genes in different high and low MeSA producing birches, and (3) identify and facilitate our understanding of the SABATH and MES family members at the gene and enzyme levels.

5.1 Phylogenetic investigations in different birch species revealed the high MeSA producing ability of the ancestor

The first research area of the project (Chapter 2) was based on determinations of birch species and their phylogenetic relationships on their high and low MeSA producing ability. Eight chloroplast and five nuclear genomic regions were included to evaluate their potential application for species determination and phylogenetic analysis. Several studies have utilized molecular genetics approaches to investigate the phylogenetic relationship of birches. The studies include amplified fragment length polymorphism (AFLP) markers (Schenk et al., 2008), a single genomic region (Li et al., 2007), and a combination of chloroplast barcoding and nuclear genes (Järvinen et al., 2004; Bina et al., 2016) to resolve the taxonomy of *Betula*. The first (Li et al., 2005) and the most recent (Wang et al., 2016) molecular phylogenetic analyses of the genus *Betula*, based on internal transcribed spacers (ITS), also did not resolve all the ambiguities in the systematics of birches. Additionally, no investigations have been conducted to analyse the phylogenetic relationship of high and low MeSA producing birches so far, although it is one of the significantly important pharmaceuticals and medicinal substances (Clark, 1999; Demirci et al., 2000; Başer et al., 2007; Rastogi et al., 2015).

In the current thesis, eight chloroplast and five nuclear genomic regions were used to construct three different phylogenetic trees and a network to analyze the relationships between different birches. The examinations were well supported by the previous phylogenetic analysis where *B. lenta* formed an ancient clade with *B. grossa*, *B. alleghaniensis*, and *B. costata* that also carried high MeSA (Järvinen et al., 2004; Li et al., 2005; Bina et al., 2016). Remarkably, *B. lenta* formed an ancient node in the network analysis. The own investigations revealed that the diploid *B. lenta* is one of the ancestors of the genus *Betula*. The analyzed chloroplast and nuclear regions disclosed 25 and 61 nucleotide variations in chloroplast and nuclear regions respectively.

The chloroplast and nuclear DNA variations identified in the current thesis could be important for the Barcode of Life project (https://ibol.org/) and will be used to develop rapid molecular test methods to reduce the cost of sequencing for the selection of birches for breeding. Precisely, the achieved results indicate, in the evolutionary events, the ancestrally high MeSA-producing trait was transferred to some younger *Betula* species, while some species lost their MeSA-producing ability, suggesting divergence in the evolutionary process among the species (Hughes, 2012; Bina et al., 2016).

Nonetheless, comprehensive sampling, together with genome-wide nucleotide variation analysis could resolve the questionable phylogenetic position of some *Betula* species. Additionally, chemical analysis using Gas chromatography-mass spectrometry can help to make precise decisions about the high and low MeSA-producing *Betula* species that could also guide to solve taxonomical issues.

5.2 Candidate gene, marker, and tissue-specific expression analysis in different birches

The meritorious phylogenetic and network analysis assisted to select high (*B. lenta, B. alleghaniensis, B. medwediewii,* and *B. grossa*) and low (*B. pendula, B. utilis, B. alnoides,* and *B. nana*) MeSA-producing species for the candidate gene analysis (Chapter 3). In this section, a comprehensive bioinformatics and expression analysis of the two candidate genes (*SAMT* and *SABP2*) was performed. Since many studies have shown that *SAMT* and *SABP2* genes are important for MeSA biosynthesis in various plants (Shulaev et al., 1997; Ross et al., 1999; Chen et al., 2003; D'Auria et al., 2003; Forouhar et al., 2005; Park et al., 2007).

Intensive studies have been conducted to reveal the role of MeSA in plant immunity and the signaling cascades of the SAR mechanism in plants (Chen et al., 2003; Zhao et al., 2009; Ament et al., 2010). SAMT, synthesize MeSA by the methylation of SA (Ross et al., 1999), while MeSA reverts to SA by SABP2 possessing a strong esterase activity (Park et al., 2007; Kumar, 2014). Almost all plants produce MeSA as a long-distance mobile signal in the SAR mechanism (Shulaev et al., 1997), including the birch species *B. utilis, B, pendula, B. nana*, and *B. alnoides* (mentioned as “low MeSA producers”) (Singewar et al., 2020a). In addition to its role in immunity and the SAR signaling cascade, MeSA is produced constitutively in some wintergreen shrubs and birch species like *B. lenta, B. alleghaniensis, B. medwediewii,* and *B.*

grossa, (mentioned as “high MeSA producers”) and is therefore believed to be an important plant constituent (Ashburner et al., 2013; Singewar et al., 2020a).

The results of the current thesis, including multiple sequence alignments of the amino acid, revealed the occurrence of important residues for substrate specificity of *SAMT* and *SABP2* candidate genes. The analysis of *cis*-elements in different birches indicated a functional multiplicity of *SAMT* and *SABP* and provided insights into the regulation of both genes. The relative tissue-specific expression analysis of *SAMT* in leaf and bark tissue of two high and two low MeSA-synthesizing birches revealed high expression in the bark of both high MeSA-synthesizing birches. In contrast, *SABP2* expression in tissues revealed different levels of expression between species belonging to the two groups.

Furthermore, six prominent single nucleotide substitution markers were discovered on the *SABP2* candidate gene and validated with 38 additional birch individuals to differentiate high and low MeSA-producing birches.

The comparative expression and bioinformatics analyses provided vital information, which will be beneficial for the functional genomics studies where the function of the gene could be confirmed *in vivo* followed by mass production of organic MeSA from the birches. Further, including more *Betula* species from the diverse geographical region for the marker validation could be advantageous to raise confidence in the developed markers to differentiate high and low MeSA producing birches.

5.3 Evolutionary relationships and putative substrate specificities of SABATH and MES enzyme family members of silver birch

During the study of the candidate genes mediating biosynthesis of MeSA in the previous section of the current thesis, it is clear that the *SAMT* and *SABP2* genes are the members of *SABATH* and *MES* families, respectively (D'Auria et al., 2003; Yang et al., 2008). Plant SABATH and MES enzyme family members catalyze the methylation and demethylation of many hormones, floral scent metabolites, and signaling molecules like salicylic acid (SA), jasmonic acid (JA), and indol-3 acetic acid (IAA) according to physiological conditions (Zhao et al., 2008; Zhao et al., 2009; Han et al., 2017). The resulting methyl esters undergo demethylation by members of the MES enzyme family (Yang et al., 2008).

Previous analyses revealed that the members of both families share high nucleotide sequence similarities (D'Auria et al., 2003; Zhao et al., 2016; Han et al., 2017). Further, an evolutionary

and functional study of *Populus trichocarpa* SABATH family members at the gene and enzyme level disclosed that a change in a single amino acid position promotes the divergence in substrate selection (Han et al., 2017). Considering the important aspects of both families, detailed bioinformatics and phylogenetic analysis of SABATH and MES family members were performed at gene and enzyme level in *B. pendula* that revealed putative substrate specificities and evolutionary relationships of both enzyme families (Chapter 4).

In the current thesis, ten and twelve members of the SABATH (BpSABATH1-10) and MES (BpMES1-12) family in *B. pendula* were characterized at the gene and enzyme level. The *SABATH* and *MES* genes were distributed only on seven of fourteen *B. pendula* chromosomes. The phylogenetic analysis and chromosomal mapping of the gene family confirmed the moderate duplication events while conserved amino acid residues suggest their putative physiological role. The monophyletic clade has formed by the BpSABATH8 with IAA methyltransferases from different plant species suggests, it potentially methylates IAA; while BpSABATH5, BpSABATH6, and BpSABATH7 methylate JA to form methyl jasmonate (MeJA), indicating a functional divergence after gene duplication events. In the phylogenetic functional Group 1, only BpSABATH9 prefers SA as a substrate and form MeSA. The uneven distribution of BpSABATHs supports the possibilities of multiple independent evolutionary events and functional diversification. Likewise, BpMES5 and BpMES12 could be involved in the demethylation of the methyl ester of IAA, while BpMES6, BpMES7, and BpMES8 might be responsible for the demethylation of MeJA. BpMES9 and BpMES10 were clustered with all MESs that prefers MeSA as a substrate.

The essential knowledge will be helpful for further functional genomic and enzymatic substrate specificity analysis. Moreover, the valuable information of the candidate genes could be exploited for genetic modification or targeted mutagenesis to decide only the function of a gene. Further, trait-specific markers would be designed to breed birch varieties that are adapted to certain environments and to facilitate uninterrupted plant developmental processes.

6 Conclusions

The comprehensive molecular genetics, phylogenetic, and bioinformatics analysis of candidate genes mediating MeSA biosynthesis provided insights into the socio-ecological important tree species of the genus *Betula*. The phylogenetic and network analysis demonstrated that *B. lenta* is one of the ancestors and the high MeSA-producing trait has been lost several times during the evolution of younger *Betula* species. The study allowed to understand the distribution of the high and low MeSA-producing trait in the genus. Examination of different chloroplast and nuclear genomic regions suggests, considering even a small segment of a nuclear DNA sequence produces high-resolution levels and provides more parsimony information to the chloroplast DNA.

To the best of our knowledge, this is the first study that has attempted to gather information about genetic architecture and expression of the candidate genes, *SAMT,* and *SABP2* mediate the biosynthesis of MeSA in eight *Betula* species. Besides, nucleotide variations that are specific to high and low MeSA-production were identified and validated with species from different botanical gardens. The discovered nucleotides will be used to develop SNP markers associated with low and high MeSA content for molecular breeding purposes. The detailed examination of the *SAMT* and *SABP2* genes lead to the identification of their affiliated gene families, *SABATH* and *MES* respectively.

Further, extensive bioinformatics analysis of SABATH and MES family members was conducted at gene and protein levels due to their involvement in the biosynthesis of many hormones, signaling molecules, and floral scent. The vital comparative analysis revealed discrete patterns present in the SABATH and MES family members. Phylogenetic analysis revealed the evolutionary relationship and putative substrate specifications of both family members. The advantageous information of SABATH and MES genes could be exploited for genetic modification or targeted mutagenesis to decide the function of the genes.

Finally, the thesis promotes the idea of agroforestry with short-rotation copies (SRC) of high MeSA producing birch along with MeSA extraction and distillation from the bark of the trees that could give rise to the mass production technology of organic MeSA.

7 Summary

The perennial woody plant birch (*Betula* L.) is crucial for the economy and to maintain the biodiversity of the boreal forests. Numerous studies have reported the taxonomical complexities of birches due to extensive hybridization and periodic introgression events. Besides their classification arguments and conventional benefits, some birches constitutively synthesize the anti-inflammatory methyl salicylate (MeSA). It is used in many drugs as an ointment for relief against rheumatic complaints, muscles, and joint pains. Although MeSA is available naturally in many birch species, now it is mainly produced synthetically. Knowledge about the molecules involved in the biosynthesis of MeSA in birches following the evolutionary relationships of high and low- MeSA producing birch species are almost missing in the literature. Moreover, molecular markers for the breeding of birches with high MeSA content were never identified. In this thesis, intraspecific phylogenetic relationships, genetic variations between high and low- MeSA producing birches were investigated using molecular genetic approaches. Additionally, candidate genes mediating MeSA biosynthesis and their affiliated families were also studied at gene and enzyme levels.

In the first examination, eight chloroplast and five nuclear genomic regions were examined to determine the phylogenetic relationship and genetic variation present in high and low MeSA producing birch species. *Betula lenta*, always clustered within a clade with other high MeSA-producing birches and also formed an ancestral node in the network analysis. The phylogenetic and network analysis suggests, *B. lenta* is one of the ancestors and can produce high MeSA. In the process of evolution, the ancestrally high MeSA-producing trait was transmitted to some juvenile *Betula* species, whereas some species lost their MeSA-producing ability, advocating speciation in the evolutionary developments among the genus *Betula*.

Further, candidate genes, *salicylic acid methyltransferase* (*SAMT*), and *salicylic acid-binding protein 2* (*SABP2*) that mediate the MeSA biosynthesis were studied from the eight high (*B. lenta, B. alleghaniensis, B. medwediewii,* and *B. grossa*) and low (*B. pendula, B. utilis, B. alnoides,* and *B. nana*) MeSA producing birches. Additionally, six distinguishable single nucleotide substitution markers were determined and validated with 38 additional birch individuals originated from different botanical gardens to acknowledge high and low MeSA-producing birch species. In the relative tissue-specific expression analysis, the *SAMT* gene revealed increased expression in the bark of high MeSA-synthesizing birches, while, differential

expression of *SABP2* gene had been observed in the examined tissues and species. These results suggested that *SAMT* and *SABP2* have a significant effect on MeSA biosynthesis in birches. The detailed examination of the *SAMT* and *SABP2* genes lead to the identification of their affiliated gene families, *SABATH* (SAMT, benzoic acid carboxyl methyltransferase, and theobromine synthase) and *MES* (methylesterases) respectively.

Furthermore, ten SABATH and twelve MES family members were identified and investigated at the gene and enzyme level in the *Betula pendula* (BpSABATH and BpMES). Members of both families are involved in the methylation and demethylation of many hormones, signaling molecules, and floral scent metabolites like Salicylic acid (SA), jasmonic acid (JA), and indol-3 acetic acid (IAA) responsible for many plant developmental processes. The phylogenetic analysis and chromosomal mapping of the gene family confirmed the moderate duplication events while conserved amino acid residues suggest their putative physiological role. Phylogenetic analysis revealed the putative substrate specificities of BpSABATH and BpMES enzymes. The monophyletic clade has formed by the BpSABATH8 with IAA methyltransferases from different plant species suggests, it potentially methylates IAA; while methylation of JA is carried out by three members, BpSABATH5, BpSABATH6, and BpSABATH7. Synthesis of MeSA by methylation of SA is directed by BpSABATH9 which was denoted as *SAMT* in the earlier candidate gene analysis. The uneven distribution of BpSABATHs supports the multiple independent evolutionary events and functional diversification. Similarly, BpMES5 and BpMES12 possibly be involved in the demethylation of the methyl ester of IAA, and the demethylation of methyl ester of JA might be organized by three members including BpMES6, BpMES7, and BpMES8. BpMES9 prefers MeSA as a substrate to revert into SA that is represented as *SABP2* in the previous chapter.

The knowledge of high MeSA producing trait, candidate genes followed by SABATH, and MES families will be essential for further functional genomic and enzymatic substrate specificity analysis. Identified nuclear markers could be used in the birch breeding for mass production of organic MeSA. Moreover, the valuable information of the candidate genes could be exploited for genetic modification or targeted mutagenesis to decide only the function of a gene. This thesis encourages the commercial production of birches for the timber in short-rotation coppices (SRC) along with pharmacological benefits through extracting MeSA from the bark. The SRC and extracted MeSA would be highly advantageous for forest farming as well as for bio-economy.

8 Zusammenfassung

Die verschiedenen Arten der Gattung Birke (*Betula* L.) sind äußerst wichtig für die Ökologie und die Erhaltung der biologischen Vielfalt der borealen Wälder. Die hohe morphologische Variation innerhalb der Arten sowie die Hybridisierung zwischen verschiedenen Arten und der damit verbundenen genetischen Introgression stellte die Systematik der Birken vor große Herausforderungen, was manchmal auch zu Fehlklassifizierungen einiger Arten führte. Neben ihrer ökologischen Bedeutung sind Birken auch ökonomisch interessant, da sie eine breite Palette von pharmakologisch wichtigen Sekundärmetaboliten synthetisieren. Studien haben gezeigt, dass Methylsalicylat (MeSA) einer der Hauptbestandteile des ätherischen Öls einiger Birkenarten ist. MeSA wird als Arzneimittel in vielen entzündungshemmenden Produkten zur Linderung von Muskel- und Gelenkschmerzen eingesetzt. Obwohl MeSA in vielen Birkenarten natürlich vorkommt, wird es heute hauptsächlich synthetisch hergestellt. Die Verwendung von natürlichem MeSA ist aufgrund unzureichender Kenntnisse über seine Herstellung in bestimmten Birkenarten begrenzt. Daher ist es dringend erforderlich, das Wissen über den MeSA-Biosyntheseweg und die an der Reaktion beteiligten Kandidatengene sowie den molekularen und genetischen Eigenschaften von Birkenarten, die MeSA in niedrigen oder hohen Mengen produzieren, zu steigern. Das Ziel dieser Arbeit war daher, die interspezifische phylogenetische Beziehung und die genetischen Variationen zwischen Birkenarten mit hohen und niedrigen MeSA-Gehalten durch molekulargenetische Ansätze zu untersuchen sowie die Gene zu identifizieren, die im MeSA-Biosyntheseweg involviert sind. In der ersten publizierten Arbeit wurden bei Birkenarten mit hohen und niedrigen MeSA-Gehalten acht Chloroplasten und fünf genomische Kernregionen untersucht, um deren phylogenetische Beziehung und genetische Variation zu bestimmen. Die Arten *B. lenta, B. grossa* und *B. alleghaniensis* bildeten in allen konstruierten phylogenetischen Bäumen immer eine gemeinsame Gruppe, die ihren Ahnenzustand in der Analyse darstellte. Ferner bildete *B. lenta* einen Ahnenknoten in der Netzwerkanalyse. Die in diesem Kapitel erhaltenen Ergebnisse legten nahe, dass im Verlauf der Evolution das ursprüngliche MeSA-produzierende Merkmal auf einige später entwickelte Birkenarten übertragen wurde, während andere Arten die Fähigkeit zur MeSA-Produktion verloren haben. Darüber hinaus wurden Kandidatengene, die im MeSA-Biosyntheseweg involviert sind, in acht Birkenarten mit hohen (*B. lenta, B. alleghaniensis, B. medwediewii* und *B. grossa*) und mit niedrigen (*B. pendula*, *B. utilis, B. alnoides* und *B. nana*) MeSA-Gehalten auf Sequenzunterschiede hin untersucht. Sechs

Einzelnukleotidsubstitutions- (SNP-) Marker, die Birken mit hohen und niedrigen MeSA-Gehalten unterscheiden, wurden gefunden. Diese SNP-Marker wurden mit 38 zusätzlichen Birkenindividuen validiert, die aus verschiedenen geografischen Regionen stammten. In einer relativen Expressionsanalyse zweier MeSA-Kandidatengene (*SAMT* [salicylic acid methyltransferase] und SABP2 [salicylic acid-binding protein 2]) in Rinde und Blatt zeigte das SAMT-Gen eine erhöhte Expression in der Rinde von Birken mit hohen MeSA-Gehalten, während für das SABP2-Gen bei diesen Arten unterschiedliche Expressionsniveaus beobachtet wurden. Diese Ergebnisse legen nahe, dass SAMT und SABP2 einen signifikanten Einfluss auf die MeSA-Biosynthese in Birken haben. Die vergleichende Expressions- und SNP-Analyse lieferte wichtige Informationen, die für eine gezielte Selektion von Birken mit hohen MeSA-Gehalten verwendet werden können. In der zweiten publizierten Arbeit wurden weitere Mitglieder der Enzymfamilien identifiziert und auf Gen- und Enzymebene untersucht, die an der Methylierung und Demethylierung vieler Hormone, Signalmoleküle und volatilen Substanzen beteiligt sind (SABATH [SAMT benzoic acid carboxyl methyltransferase, and theobromine synthase] und MES [methylesterase]). Die chromosomale Kartierung der Gene sowie die konservierten Aminosäurereste innerhalb der Genfamilie bestätigten nicht nur die moderaten genomischen Duplikationsereignisse während der Evolution der verschiedenen Birkenarten, sondern auch ihre jeweilige physiologische Rolle. Bei den methylierenden Enzymen zeigte die Substratspezifitätsanalyse, dass BpSABATH8 möglicherweise das Hormon Indol-3-Essigsäure (IAA) methyliert, während BpSABATH5, BpSABATH6 und BpSABATH7 die Methylierung von Jasmonsäure (JA) katalysieren. Die Methylierung von Salicylsäure (SA) und die Synthese von MeSA wird von BpSABATH9 gesteuert (im Kapitel zur Kandidatenanalyse als SAMT bezeichnet). In der umgekehrten Reaktion sind BpMES5 und BpMES12 möglicherweise an der Demethylierung des Methylesters von der IAA (MeIAA) beteiligt. Die Demethylierung des Methylesters der JA (MeJA) könnte von den drei Enzymen, BpMES6, BpMES7 und BpMES8 katalysiert werden. BpMES9 (im Kapitel zur Kandidatenanalyse als SABP2 bezeichnet) bevorzugt MeSA als Substrat, um wieder SA zu bilden. Das in dieser Arbeit zusammengetragene Wissen über BpSABATH- und BpMES-Gene wird für die weitere Analyse der genomischen Organisation sowie die der funktionellen und enzymatischen Substratspezifität der Enzyme hilfreich sein.

Darüber hinaus konnten wertvolle Informationen zur genetischen Veränderung oder gezielten Mutagenese der Kandidatengene gewonnen werden, um Metabolitgehalte gezielt zu verändern. Die aktuelle Arbeit kann eine an der Bioökonomie orientierte kommerzielle Verwendung von Birken in Kurzumtriebsplantagen (KUPs) fördern. Aus den geernteten Stämmen könnte zunächst das pharmakologisch bedeutsame MeSA aus der Rinde der Birken extrahiert und anschließend das verbleibende Holz einer thermischen Verwertung zugeführt werden.

9 References

Alonso-Serra, J., Safronov, O., Lim, K.-J., Fraser-Miller, S.J., Blokhina, et al., (2019), Tissue-specific study across the stem reveals the chemistry and transcriptome dynamics of birch bark. New Phytol, 222: 1816-1831. doi:10.1111/nph.15725

Ament, K., Krasikov, V., Allmann, S., et al.,(2010) Methyl salicylate production in tomato affects biotic interactions. The Plant Journal 62(1), 124-134, 10.1111/j.1365-313X.2010.04132.x.

Anderson, A., McConville, A., Fanthorpe, L., et al.,(2017) Salicylate Poisoning Potential of Topical Pain Relief Agents: From Age Old Remedies to Engineered Smart Patches. Medicines (Basel) 4(3), 10.3390/medicines4030048.

Ashburner, K., McAllister, H.A., Hague, J., et al., (2013) The Genus *Betula*: A Taxonomic Revision of Birches. Royal Botanic Gardens, London: Kew Publishing.

Başer, K.H.C., Demirci, B.,(2007) Studies on Betula Essential Oils. ARKIVOC: archive for organic chemistry 2007(7), 335-348, http://dx.doi.org/10.3998/ark.5550190.0008.730.

Bina, H., Yousefzadeh, H., Ali, S.S., et al.,(2016) Phylogenetic relationships, molecular taxonomy, biogeography of *Betula*, with emphasis on phylogenetic position of Iranian populations. Tree Genetics & Genomes 12(5), 84, 10.1007/s11295-016-1037-4.

Cai M, Huang H, Ni F, Tong Z, Lin E, Zhu M. 2018. RNA-Seq analysis of differential gene expression in Betula luminifera xylem during the early stages of tension wood formation. PeerJ 6:e5427 https://doi.org/10.7717/peerj.5427

Chen, F., D'Auria, J.C., Tholl, D., et al.,(2003) An *Arabidopsis thaliana* gene for methyl salicylate biosynthesis, identified by a biochemical genomics approach, has a role in defense. The Plant Journal 36(5), 577-588, 10.1046/j.1365-313X.2003.01902.x.

Chen, Z.-D., Manchester, S.R., Sun, H.-Y.,(1999) Phylogeny and evolution of the *Betulaceae* as inferred from DNA sequences, morphology, and paleobotany. American Journal of Botany 86(8), 1168-1181, 10.2307/2656981.

Chen, Z.,(1994) Phylogeny and phytogeography of the *Betulaceae*. Acta Phytotaxonomica Sinica 32(1), 1-31.

Chen L, Wang W-S, Wang T, Meng X-F, Chen T-Tt, Huang X-X, Li Y-Jj, Hou B-K (2019) Methyl salicylate glucosylation regulates plant defense signaling and systemic acquired resistance. Plant Physiol. 180:2167 doi:10.1104/pp.19.00091

Chern, M., Fitzgerald, H.A., Canlas, P.E., et al.,(2005) Overexpression of a rice NPR1 homolog leads to constitutive activation of defense response and hypersensitivity to light. Mol Plant Microbe Interact 18(6), 511-520, 10.1094/MPMI-18-0511.

Christenhusz, M., Byng, J.,(2016) The number of known plant species in the world and its annual increase. Phytotaxa 261, 201-217, 10.11646/phytotaxa.261.3.1.

Cindy, E., M.,, Van Den Boom, Van Beek, T.A., et al.,(2004) Qualitative and Quantitative Variation Among Volatile Profiles Induced by Tetranychus urticae Feeding on Plants from Various Families. Journal of Chemical Ecology 30(1), 69-89, 10.1023/b:joec.0000013183.72915.99.

Clark, G.S.I., (1999) An aroma-chemical profile. Methyl salicylate, or oil of wintergreen. Perfumer and Flavorist.

Clarke JD, Sigrid M. Volko, Heidi Ledford, et al.,(2000) Roles of Salicylic Acid, Jasmonic Acid, and Ethylene in cpr-Induced Resistance in Arabidopsis. The Plant Cell 12, 2175–2190, doi.org/10.1105/tpc.12.11.2175.

Crowther, T.W., Glick, H.B., Covey, K.R., et al.,(2015.) Mapping tree density at a global scale. Nature 525(7568), 201-205, 10.1038/nature14967.

D'Auria, J., Chen, F., Pichersky, E.,(2003) Chapter eleven: The SABATH family of MTS in *Arabidopsis thaliana* and other plant species. Recent Advances in Phytochemistry 37, 253-283, 10.1016/S0079-9920(03)80026-6.

De Jong, P.C.,(1993) An introduction to *Betula*: its morphology, evolution, classification and distribution, with a survey of recent work. Proceedings of the IDS Betula symposium, 2-4 October 1992.

Demirci, F., B. Demirci, K. H. C. Baser, Guven, a.K.,(2000) The composition and antifungal bioassay of the essential oils of different *Betula* species growing in Turkey. Chemistry of Natural Compounds Vol. 36, 2.

Fischer, A., Lindner, M., Abs, C., et al.,(2002) Vegetation dynamics in central european forest ecosystems (near-natural as well as managed) after storm events. Folia Geobotanica 37(1), 17-32, 10.1007/BF02803188.

Forouhar, F., Lee, I.S., Vujcic, J., et al.,(2005) Structural and functional evidence for *Bacillus subtilis* PaiA as a novel N1-spermidine/spermine acetyltransferase. J Biol Chem 280(48), 40328-40336, 10.1074/jbc.M505332200.

Furlow, J.J.,(1990) The genera of *Betulaceae* in the southeastern United States. Journal of the Arnold Arboretum. 71, 1-67, 10.5962/bhl.part.24925.

Gadino AN., Vaughn M., Walton, J., et al.,(2012) Evaluation of methyl salicylate lures on populations of *Typhlodromus pyri* (Acari: Phytoseiidae) and other natural enemies in western Oregon vineyards,. Biological Control, 63(1), 48-55, doi.org/10.1016/j.biocontrol.2012.06.006.

Gailis, A., Kārkliņa, A., Purviņš, A., et al.,(2020) Effect of Breeding on Income at First Commercial Thinning in Silver Birch Plantations. Forests 11, 327, 10.3390/f11030327.

Gaglioti AL (2015) Sistemática, Filogenia e Biogeografia de Pourouma Aubl. (Urticaceae).

Gatehouse, J.A.,(2002) Plant resistance towards insect herbivores: a dynamic interaction. New Phytologist 156(2), 145-169, 10.1046/j.1469-8137.2002.00519.x.

Gauthier, S., Bernier, P., Kuuluvainen, T., et al.,(2015) Boreal forest health and global change. Science 349(6250), 819, 10.1126/science.aaa9092.

Han, X.-M., Yang, Q., Liu, Y.-J., et al.,(2017) Evolution and Function of the Populus SABATH Family Reveal That a Single Amino Acid Change Results in a Substrate Switch. Plant and Cell Physiology 59(2), 392-403, 10.1093/pcp/pcx198.

Hinchliff, C.E., Smith, S.A., Allman, J.F., et al.,(2015) Synthesis of phylogeny and taxonomy into a comprehensive tree of life. Proceedings of the National Academy of Sciences 112(41), 12764, 10.1073/pnas.1423041112.

Hopkins H, C. F. (1986) Parkia (Leguminosae: Mimosoideae). Flora Neotropica 43:1-123.

Huang, H., Wang, S., Li, H. et al. (2020) Effect of the suppression of BpAP1 on the expression of lignin related genes in birch. J. For. Res. https://doi.org/10.1007/s11676-020-01232-5

Hughes, A.L.,(2012) Evolution of adaptive phenotypic traits without positive Darwinian selection. Heredity 108(4), 347-353, 10.1038/hdy.2011.97.

Hynynen, J., Niemistö, P., Viherä-Aarnio, A., et al.,(2009) Silviculture of birch (*Betula pendula Roth* and *Betula pubescens Ehrh.*) in northern Europe. Forestry: An International Journal of Forest Research 83(1), 103-119, 10.1093/forestry/cpp035.

Järvinen, P., Palmé, A., Orlando Morales, L., et al.,(2004) Phylogenetic relationships of *Betula* species (*Betulaceae*) based on nuclear ADH and chloroplast matK sequences. American Journal of Botany 91(11), 1834-1845, 10.3732/ajb.91.11.1834.

Jianhua, L., Suzanne, S., Zhiduan, C.,(2005) Phylogenetics of Betula (*Betulaceae*) inferred from sequences of nuclear ribosomal DNA. Rhodora 107(929), 69-86, 10.3119/04-14.1.

Joó, É., J. Dewulf, C. Amelynck, N.S., O. Pokorska, M. Šimpraga, K. Steppe, Marc Aubinet, H. Van Langenhove,(2011) Constitutive versus heat and biotic stress induced BVOC emissions in *Pseudotsuga menziesii*. Atmospheric Environment 45(22), 3655-3662,, doi.org/10.1016/j.atmosenv.2011.04.048.

Koropachinskii, I.Y.,(2013) Natural hybridization and taxonomy of birches in North Asia. Contemporary Problems of Ecology 6(4), 350-369, 10.1134/S1995425513040045.

Kumar, D.,(2014) Salicylic acid signaling in disease resistance. Plant Sci 228, 127-134, 10.1016/j.plantsci.2014.04.014.

Kumar, D., Klessig, D.F.,(2003) High-affinity salicylic acid-binding protein 2 is required for plant innate immunity and has salicylic acid-stimulated lipase activity. PNAS. 100. 16101-6. 10.1073/pnas.0307162100..

Lapczynski, A., Jones, L., McGinty, D., et al.,(2007) Fragrance material review on methyl salicylate. Food Chem Toxicol 45 Suppl 1, S428-452, 10.1016/j.fct.2007.09.053.

Leak, W.B., (1965) Sweet birch (*Betula lenta L.*). In Silvics of forest trees of the United States. p. 99-109. H. A. Fowells, comp. U.S. Department of Agriculture, Agriculture Handbook 271. Washington, DC.

Lee, S., Nkongolo, K., Park, D. et al. (2019) Characterization of chloroplast genomes of Alnus rubra and Betula cordifolia, and their use in phylogenetic analyses in Betulaceae. Genes Genom 41, 305–316. https://doi.org/10.1007/s13258-018-0762-5

Li, J., Shoup, S., Chen, Z.,(2005) Phylogenetics of *Betula* (*Betulaceae*) inferred from sequences of nuclear ribosomal DNA. Rhodora 107(929), 69-86.

Li, J., Shoup, 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, 10.1600/036364407781179699.

Lin L, Mu H, Jiang J, Liu G. (2013) Transcriptomic analysis of purple leaf determination in birch. Gene. 10;526(2):251-8. doi: 10.1016/j.gene.2013.05.038

Liu P-P, von Dahl CC, Park S-W, Klessig DF (2011) Interconnection between methyl salicylate and lipid-based long-distance signaling during the development of systemic acquired resistance in Arabidopsis and Tobacco. Plant Physiol. 155:1762 doi:10.1104/pp.110.171694

Liu W-R WR, Qiao WL, Liu ZZ, Wang XH, Jiang R, Li SY, Shi RB, She GMet al. (2013) Gaultheria: Phytochemical and pharmacological characteristics. Molecules 18:12071-12108 doi:10.3390/molecules181012071

Lopes-Lutz D, Dettmann J, Nimalaratne C, Schieber A (2010) Characterization and quantification of polyphenols in Amazon grape (Pourouma cecropiifolia Martius) Molecules 15:8543-8552 doi:10.3390/molecules15118543.
Mallinger, R.E., Hogg, D.B., Gratton, C.,(2011) Methyl Salicylate Attracts Natural Enemies and Reduces Populations of Soybean Aphids (Hemiptera: Aphididae) in Soybean Agroecosystems. Journal of Economic Entomology 104(1), 115-124, 10.1603/ec10253.
McAllister, H.A., Rushforth, K.,(2011) 706. *BETULA ASHBURNERI*. Curtis's Botanical Magazine 28(2), 111-118, 10.1111/j.1467-8748.2011.01735.x.
Menković, N., Šavikin, K., Tasić, S., et al.,(2011) Ethnobotanical study on traditional uses of wild medicinal plants in Prokletije Mountains (Montenegro). Journal of Ethnopharmacology 133(1), 97-107, https://doi.org/10.1016/j.jep.2010.09.008.
Moose, S.P., Mumm, R.H.,(2008) Molecular plant breeding as the foundation for 21st century crop improvement. Plant physiology 147(3), 969-977, 10.1104/pp.108.118232.
Mu, H.-Z.; Liu, Z.-J.; Lin, L.; Li, H.-Y.; Jiang, J.; Liu, G.-F. (2012) Transcriptomic Analysis of Phenotypic Changes in Birch (Betula platyphylla) Autotetraploids. Int. J. Mol. Sci., 13, 13012-13029.
Nardini, M., Dijkstra, B.W.,(1999) α/β Hydrolase fold enzymes: the family keeps growing. Current Opinion in Structural Biology 9(6), 732-737, https://doi.org/10.1016/S0959-440X(99)00037-8.
NCBI, N.C.f.B.I., Methyl salicylate, CID=4133. PubChem Database.
Pal, M., Mishra, T., Kumar, A., et al.,(2015) Chemical Constituents and Antimicrobial Potential of Essential Oil from *Betula utilis* Growing in High Altitude of Himalaya (India). Journal of Essential Oil Bearing Plants 18(5), 1078-1082, 10.1080/0972060X.2015.1036569.
Park, S.W., Kaimoyo, E., Kumar, D., et al.,(2007) Methyl salicylate is a critical mobile signal for plant systemic acquired resistance. Science 318(5847), 113-116, 10.1126/science.1147113.
Pizzolatti, M.G., Mendes, B.G., Soldi, C., et al.,(2009) Analysis of Volatile Compounds Released From Flowers and Roots of *Polygala cyparissias* and *Polygala paniculata* by Headspace/SPME. Journal of Essential Oil Research 21(3), 255-258, 10.1080/10412905.2009.9700163.
Rasmussen, S.,(2020) Molecular Genetics, Genomics, and Biotechnology in Crop Plant Breeding. Agronomy 10, 439, 10.3390/agronomy10030439.
Rastogi, S., Pandey, M.M., Kumar Singh Rawat, A.,(2015) Medicinal plants of the genus *Betula* traditional uses and a phytochemical-pharmacological review. Journal of ethnopharmacology 159, 62-83, 10.1016/j.jep.2014.11.010.
Regel, E., 1866. Bemerkungen über die Gattungen *Betula* und *Alnus*: nebst Beschreibung einiger neuer Arten. In der Buchdruckerei der Kaiserlichen Universität.
Ribnicky, D.M., Poulev, A., Raskin, I.,(2003) The Determination of Salicylates in Gaultheria procumbens for Use as a Natural Aspirin Alternative. Journal of Nutraceuticals, Functional & Medical Foods 4(1), 39-52, 10.1300/J133v04n01_05.
Ross, J.R., Nam, K.H., D'Auria, J.C., et al.,(1999) S-Adenosyl-l-Methionine:Salicylic Acid Carboxyl Methyltransferase, an Enzyme Involved in Floral Scent Production and Plant Defense, Represents a New Class of Plant Methyltransferases. Archives of Biochemistry and Biophysics 367(1), 9-16, https://doi.org/10.1006/abbi.1999.1255.
Salojärvi, J., Smolander, O., Nieminen, K.e.a.,(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.
Schenk, M.F., Thienpont, C.-N., Koopman, W.J.M., et al.,(2008) Phylogenetic relationships in *Betula* (*Betulaceae*) based on AFLP markers. Tree Genetics & Genomes 4(4), 911, 10.1007/s11295-008-0162-0.
Shaw, K., Stritch, L., Rivers, M., et al., 2015. The red list of *Betulaceae*.
Shikov, A.N., Djachuk, G.I., Sergeev, D.V., et al.,(2011) Birch bark extract as therapy for chronic hepatitis CA pilot study. Phytomedicine 18(10), 807-810, https://doi.org/10.1016/j.phymed.2011.01.021.
Shine MB, Xiao X, Kachroo P, Kachroo A (2019) Signaling mechanisms underlying systemic acquired resistance to microbial pathogens. Plant Sci. 279:81-86, https://doi.org/10.1016/j.plantsci.2018.01.001
Shulaev, V., Silverman, P., Raskin, I.,(1997) Airborne signalling by methyl salicylate in plant pathogen resistance. Nature 385, 718–721, doi:10.1038/385718a0.
Singewar, K., Moschner, C.R., Hartung, E., et al.,(2020a) 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, 10.1007/s00468-020-01984-x.
Singewar K, Moschner CR, Hartung E, Fladung M (2020b) Identification and analysis of key genes involved in methyl salicylate biosynthesis in different birch species. PLOS ONE 15(10): e0240246. https://doi.org/10.1371/journal.pone.0240246

Singh, B., Sharma, R.A.,(2015.) Plant terpenes: defense responses, phylogenetic analysis, regulation and clinical applications. 3 Biotech 5(2), 129-151, 10.1007/s13205-014-0220-2.
Skvortsov, A.K., 2002. A New System of the Genus *Betula* L.—the Birch. (Accessed 28/08/2020.
Stener, L.-G., Jansson, G.,(2005) Improvement of *Betula pendula* by clonal and progeny testing of phenotypically selected trees. Scandinavian Journal of Forest Research 20(4), 292-303, 10.1080/02827580510036265.
Su, C.; Wang, C.; He, L.; Yang, C.; Wang, Y. (2014) Shotgun Bisulfite Sequencing of the *Betula platyphylla* Genome Reveals the Tree's DNA Methylation Patterning. Int. J. Mol. Sci., 15, 22874-22886.
Theriault G, Michael P, Nkongolo K (2016) Comprehensive Transcriptome Analysis of Response to Nickel Stress in White Birch (*Betula papyrifera*). PLOS ONE 11(4): e0153762.
Thórsson, Æ.T., Pálsson, S., Lascoux, M., et al.,(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(11), 2098-2110, 10.1111/j.1365-2699.2010.02353.x.
Tyler VE, B.L., Robbers JE.,(1981) Volatile oils. Pharmacognosy (Philadelphia, pp. 103-143,).
Vakkari, P.,(2009.) Technical Guidelines for genetic conservation and use of silver birch (*Betula pendula*). Bioversity International, Rome, Italy, ISBN 13: 978-92-9043-803-8.
Viherä-Aarnio, A., Velling, P.,(2008) Seed Transfers of Silver Birch (*Betula pendula*) from the Baltic to Finland Effect on Growth and Stem Quality. Silva Fennica 42, 10.14214/sf.226.
Walle, I., Camp, N., Casteele, L., et al.,(2007) Short-rotation forestry of birch, maple, poplar and willow in Flanders (Belgium) I—Biomass production after 4 years of tree growth. Biomass and Bioenergy 31, 267-275, 10.1016/j.biombioe.2007.01.019.
Wang, N., Thomson, M., Bodles, W.J.A., Crawford, R.M.M. at al., (2013) Genome sequence of dwarf birch (*Betula nana*) and cross-species RAD markers. Mol Ecol, 22: 3098-3111. doi:10.1111/mec.12131
Wang C, Zhang N, Gao C, Cui Z, Sun D, et al. (2014) Comprehensive Transcriptome Analysis of Developing Xylem Responding to Artificial Bending and Gravitational Stimuli in *Betula platyphylla*. PLOS ONE 9(2): e87566. https://doi.org/10.1371/journal.pone.0087566
Wang, N., McAllister, H.A., Bartlett, P.R., et al.,(2016) Molecular phylogeny and genome size evolution of the genus *Betula* (*Betulaceae*). Annals of Botany 117(6), 1023-1035, 10.1093/aob/mcw048.
Wang, S., Yang, C., Zhao, X. et al. (2018) Complete chloroplast genome sequence of *Betula platyphylla*: gene organization, RNA editing, and comparative and phylogenetic analyses. BMC Genomics 19, 950.
Williams, J.J.H., Arnold, M.L.,(2001) Sources of Genetic Structure in the Woody Perennial *Betula occidentalis*. International Journal of Plant Sciences 162(5), 1097-1109, 10.1086/321921.
Winkler, H., 1904. *Betulaceae*.
Wittek, F., Kanawati, B., Wenig, M., et al.,(2015) Folic acid induces salicylic acid-dependent immunity in Arabidopsis and enhances susceptibility to Alternaria brassicicola. Mol Plant Pathol 16(6), 616-622, 10.1111/mpp.12216.
Yan, S., Dong, X.,(2014.) Perception of the plant immune signal salicylic acid. Curr Opin Plant Biol 20, 64-68, 10.1016/j.pbi.2014.04.006.
Yang, Y., Xu, R., Ma, C.-J., et al.,(2008) Inactive methyl indole-3-acetic acid ester can be hydrolyzed and activated by several esterases belonging to the AtMES esterase family of Arabidopsis. Plant physiology 147(3), 1034-1045, 10.1104/pp.108.118224.
Yin, J., Li, X., Zhan, Y., et al.,(2017) Cloning and expression of *BpMYC4* and *BpbHLH9* genes and the role of *BpbHLH9* in triterpenoid synthesis in birch. BMC Plant Biology 17(1), 214, 10.1186/s12870-017-1150-z.
Zeng, J., Li, J., Chen, Z.,(2008) A new species of *Betula* section Betulaster (*Betulaceae*) from China. Botanical Journal of The Linnean Society - BOT J LINN SOC 156, 523-528, 10.1111/j.1095-8339.2007.00764.x.
Zhang, M, Wang, S., et al.,(2016) Molecular cloning and promoter analysis of squalene synthase and squalene epoxidase genes from *Betula platyphylla*. Protoplasma 253(5), 1347-1363, 10.1007/s00709-015-0893-3.
Zhao, N., Ferrer, J.-L., Ross, J., et al.,(2008) Structural, Biochemical, and Phylogenetic Analyses Suggest That Indole-3-Acetic Acid Methyltransferase Is an Evolutionarily Ancient Member of the SABATH Family. Plant Physiology 146(2), 455, 10.1104/pp.107.110049.
Zhao, N., Guan, J., Forouhar, F., et al.,(2009) Two poplar methyl salicylate esterases display comparable biochemical properties but divergent expression patterns. Phytochemistry 70(1), 32-39.
Zhao, N., Lin, H., Lan, S., et al.,(2016) VvMJE1 of the grapevine (*Vitis vinifera*) VvMES methylesterase family encodes for methyl jasmonate esterase and has a role in stress response. Plant Physiology and Biochemistry 102, 125-132, https://doi.org/10.1016/j.plaphy.2016.02.027.
Zoladeski, C.H.K., 2013. Recovery strategy for the Cherry Birch (*Betula lenta*) in Ontario, Peterborough, Ontario [Online], 28/08/2020 ed., http://files.ontario.ca/environment-and-energy/species-at-risk/stdprod_075574.pdf.

10 Declaration of co-authorship

Chapters 2, 3, and 4 were prepared as manuscripts for publication. The respective declarations of co-authorship are presented on the following pages.

C A U Christian-Albrechts-Universität zu Kiel	Agrar- und Ernährungs-wissenschaftliche Fakultät

Declaration of co-authorship

If a dissertation is based on already published or submitted co-authored articles, a declaration from each of the authors regarding the part of the work done by the doctoral candidate must be enclosed when submitting the dissertation.

1. Doctoral candidate
Name: Kiran Singewar

2. This co-author declaration applies to the following article:
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

The extent of the doctoral candidat's contribution to the article is assessed on the following scale:

A. Has contributed to the work (0-33%)
B. Has made a substantial contribution (34-66%)
C. Did the majority of the work independently (67-100%)

3. Declaration on the individual phases of the scientific work (A,B,C)	Extent
Concept: Formulation of the basic scientific problem based on theoretical questions which require clarification, including a summery of the general questions which, it is assumed, will be answerable via analyses or concrete experiments/investigations	**C**
Planning: Planning of experiments/analyses and formulation of investigative methodology, including choice of method and independent methodological development, in such a way that the scientific questions asked can be expected to be answered	**C**
Execution: Involvement in the analysis or the concrete experiments/investgation	**C**
Manuscript preparation: Presentation, interpretation and discussion of the results obtained in article form	**C**

4. Signature of all co-authors		
Date	**Name**	**Signature**
5/11/20	**Eberhard Hartung**	
5.11.20	**Christian R. Moschner**	
5.11.20	**Matthias Fladung**	

5. Signature of doctoral candidate		
Date	**Name**	**Signature**
5.11.20	**Kiran Singewar**	

C | A | U Christian-Albrechts-Universität zu Kiel	Agrar- und Ernährungs-wissenschaftliche Fakultät

Declaration of co-authorship

If a dissertation is based on already published or submitted co-authored articles, a declaration from each of the authors regarding the part of the work done by the doctoral candidate must be enclosed when submitting the dissertation.

1. Doctoral candidate
Name: Kiran Singewar

2. This co-author declaration applies to the following article:
Identification and analysis of key genes involved in methyl salicylate biosynthesis in different birch species

The extent of the doctoral candidat's contribution to the article is assessed on the following scale:

A. Has contributed to the work (0-33%)
B. Has made a substantial contribution (34-66%)
C. Did the majority of the work independently (67-100%)

3. Declaration on the individual phases of the scientific work (A,B,C)	Extent
Concept: Formulation of the basic scientific problem based on theoretical questions which require clarification, including a summery of the general questions which, it is assumed, will be answerable via analyses or concrete experiments/investigations	**C**
Planning: Planning of experiments/analyses and formulation of investigative methodology, including choice of method and independent methodological development, in such a way that the scientific questions asked can be expected to be answered	**C**
Execution: Involvement in the analysis or the concrete experiments/investgation	**C**
Manuscript preparation: Presentation, interpretation and discussion of the results obtained in article form	**C**

4. Signature of all co-authors		
Date	**Name**	**Signature**
5/11/20	**Eberhard Hartung**	
5/11/20	**Christian R. Moschner**	
5.11.20	**Matthias Fladung**	

5. Signature of doctoral candidate		
Date	**Name**	**Signature**
5.11.20	**Kiran Singewar**	

C | A | U Christian-Albrechts-Universität zu Kiel | Agrar- und Ernährungswissenschaftliche Fakultät

Declaration of co-authorship

If a dissertation is based on already published or submitted co-authored articles, a declaration from each of the authors regarding the part of the work done by the doctoral candidate must be enclosed when submitting the dissertation.

1. Doctoral candidate
Name: **Kiran Singewar**

2. This co-author declaration applies to the following article:
Genome-wide bioinformatics and phylogenetic analysis revealed putative substrate specificities of SABATH and MES family members in silver birch (Betula pendula)

The extent of the doctoral candidat's contribution to the article is assessed on the following scale:

A. Has contributed to the work (0-33%)
B. Has made a substantial contribution (34-66%)
C. Did the majority of the work independently (67-100%)

3. Declaration on the individual phases of the scientific work (A,B,C)	Extent
Concept: Formulation of the basic scientific problem based on theoretical questions which require clarification, including a summery of the general questions which, it is assumed, will be answerable via analyses or concrete experiments/investigations	**C**
Planning: Planning of experiments/analyses and formulation of investigative methodology, including choice of method and independent methodological development, in such a way that the scientific questions asked can be expected to be answered	**C**
Execution: Involvement in the analysis or the concrete experiments/investgation	**C**
Manuscript preparation: Presentation, interpretation and discussion of the results obtained in article form	**C**

4. Signature of all co-authors		
Date	**Name**	**Signature**
5/11/20	**Eberhard Hartung**	
5.11.20	**Christian R. Moschner**	
5.11.20	**Matthias Fladung**	

5. Signature of doctoral candidate		
Date	**Name**	**Signature**
5.11.20	**Kiran Singewar**	

11 Acknowledgement

I would like to thank my supervisors, Prof. Dr. Eberhard Hartung and PD Dr. Matthias Fladung for their continuous guidance. I cannot thank enough Prof. Hartung for providing me the opportunity to do my doctoral research at the Institut für Landwirtschaftliche Verfahrenstechnik, Kiel. He always helped me to shape my thesis and motivated me to be a better researcher by conducting scientific discussions and project meetings.

Matthias has been the greatest help throughout my Ph.D. work at the Thünen Institute of Forest Genetics. From him, I learned that research might show ups and downs but there is no alternative to hard work and it always pays off in the end. I learned a lot from him. Matthias, in my professional life, I idealize you and I wish one day I become a good researcher as you are.

Dr. Christian R. Moschner, being a principal investigator has been very supportive throughout my Ph.D. work. He was always available for me to solve all my small or big problems. Thank you for writing such a splendid project and collecting vital plant samples for our experiments.

I am grateful to Birgit, Niels, Ana, Malte, and Avneesh for their help in bioinformatics. Thanks to Hilke, Hans, Tobias, and Khira for scientific discussions on the project. Many thanks to Katrin, Susanne, Vivian, and Stefan for their technical assistance. I want to thank Mr. Graf and Maike for their assistance in the greenhouse. Many thanks to all the colleagues from the Thünen Institute of Forest Genetics for support and help in my good and bad times.

Many thanks to all my colleagues from the Institut für Landwirtschaftliche Verfahrenstechnik, Kiel for your valuable support and discussions through meetings including Angelika, Susanne, Eiko, Torsten, Stefan, and Anna. Special thanks to Renate for helping with difficult documentation. I appreciate the financial support by the Ministry for Energy Transition, Agriculture, Environment, Nature, and Digitalization of Schleswig-Holstein, Germany (Project number: LPW-L/1.2/24).

I was very lucky to have great friends like Smit-Riddhi, Gaurav, Ankush, Amit-Priyanka, Jay-Riddhi, Dinesh, Kanishk, Abhijeet, Komathy, Soumya, Shailja, and Stefan Girlich. The time we all spent together was priceless, and no matter where we go we will always cherish those wonderful moments. You guys are family to me.

Most importantly, I would like to thank my parents, brother, sisters-in-law, my wife Divya, and all the family members for their constant support, and love. I would especially like to mention my late brother for his blessings.